



**Langues & Langage  
à la croisée des Disciplines**

**1ère Rencontre annuelle LLcD**

Colloque : 9-11 septembre, 2024

Ecole d'été : 12-14 septembre, 2024

Sorbonne Université - 28 Rue Serpente, 75006 Paris, France



**Book of Abstracts**

**Livret de résumés**

## Workshops

<b>WS</b>	<b>Métrique, musication</b>	<b>10</b>
<b>1</b>	<b>Metrics, Textsetting</b>	
	Julien Meyer and Denny Moore	11
	Jean-Louis Aroui	13
	Cathal Twomey	14
	Timothée Premat	16
	Romain Benini	18
	Eliane Delente	19
	Richard Renault and Stéphane Ferrari	21
<b>WS</b>	<b>Applicative uses of spatial markers: Typological and diachronic perspectives</b>	<b>23</b>
<b>2</b>	<b>Les marqueurs spatiaux ayant une valeur applicative : Perspectives typologiques et diachroniques</b>	
	Timofey Mukhin, An Van linden & Dana Louagie	24
	Cristian Juarez	27
	George Aaron Broadwell	30
	Julie Marsault	32
	Ann-Sophie Vrielynck	34
	Johan Pedersen	36
	Peter Arkadiev	38
	Nimet Mezane Konuk	41
	Aurore Montebran	43
	Deokhyun Nam	46
	Nicholas Evans	49
<b>WS</b>	<b>Les images dans le discours numérique : enjeux, formes, intégration, fonction</b>	<b>50</b>
<b>3</b>	<b>Images in Digital Discourse: Issues, Forms, Integration, Function</b>	
	Michel Marcoccia	51
	Agathe Cormier	53
	Charlotte Rep	56

	européenne franco-allemande	
Barbara Dancygier and Lieven Vandelanotte	Towards a grammar of Internet memes combining image and text	59
Justine Simon	Le même en contexte socionumérique : une mosaïque interdiscursive	61
Soufyane Chafik	Embrayage et relation texte-image dans les mêmes-Internet type « moi quand »	63
Marie Boscaro	Assertions and Presuppositions with hyperlinks in X	65

**WS The demonstrative in the evolutionary process: Typological variation and change 68**

**4 Le démonstratif en évolution: variation typologique et changement**

Guillaume Kurz	Les démonstratifs du louvite hiéroglyphique à la lumière du hittite : une contribution fonctionnaliste à la linguistique historique anatolienne	69
Benjamin Dufour	This my lettere - l'émergence d'un démonstratif de révérence en anglais moderne	74
Fanguang Kong	La pragmatization des marqueurs discursifs alors en français et nà en chinois mandarin	77
Jean Cruchet	L'évolution des démonstratifs français : du qualitatif au quantitatif	79
Sabine Lehmann	Configurations discursives et syntaxiques des pronoms ceci et cela en diachronie (du moyen français au français classique)	82
Christine da Silva-Genest, Marine Le Mené Guigourès and Anne Salazar-Orvig	"Ça c'est à moi" : usages et spécificités du pronom démonstratif dans les dialogues adulte-enfant	84
Ulrike Albers and Patricia Cabredo Hofherr	Contextes d'usage de la forme la en français et dans les créoles français	86
Tatyana Slobodchikoff	Definiteness, Demonstrative Pronouns, and the Absence of the Definite Article in Old Russian	88
Alexander Rostovtsev-Popiel and Lela Tsikhelashvili	From Demonstratives to Nominal Markers in the History of Georgian: Greenberg's Predictions Proving to Be Correct	90
Ion Tudor Giurgea	From demonstrative to definite article: the case of Gothic	92
Alexandra-Corina Stavinschi	The role of ECCE in the emergence of the demonstratives in Romance	94
Julen Manterola	Reinforced and non-reinforced plural distal demonstratives in Eastern Basque texts	96
Fabio Gasparini	From determiners to nominalizers: the case of Modern South Arabian	98

**WS Subordination in Creole languages 100**

**5 Subordination dans les langues créoles**

Stefano Manfredi and Nicolas Quint	SCroll: the database of subordination in Creole languages	101
Muhsina Alleesaib, Shrita Hassamal, Fon Sing Guillaume, Guilhem Florigny and Anne Abeillé	Des subordonnées sans subordonnant en créole mauricien	102

Minella Duzerol	Martinican complement clauses in ke: a statistical approach	103
Patricia Cabredo Hofherr	Grammaticalisation of final complementisers from GIVE in Haitian	105
Nicolas Quint	Comparing the syntax of subordination in insular and continental Upper Guinea Creoles	107
<b>Communication invitée:</b> Sibylle Kriegel	Recent evolutions of say-complementizers in Creole languages	109
Bettina Migge and Paula Prescod	Subordinate constructions in English-lexifier Creoles: Adverbial Complementation	110
Susanne Maria Michaelis	Against simplicity in subject relative clause marking in Creole languages	112
<b>Communication invitée:</b> Bern Kortmann	Subordination and subordinators in Caribbean English-based Creoles: A (dialect) typological view	114

## **WS Grammatical gender in diachrony** **115**

### **6 Le genre grammatical en diachronie**

Gerd Carling	The evolutionary dynamics of lexical concepts: the connection between semantic property and gender	116
John Mansfield, Ruth Singer and Frank Seifart	Areal and phylogenetic distribution of lexical partitioning imposed by noun classes	117
Francesca Di Garbo and Annemarie Verkerk	Preliminary steps towards an integrated evolutionary model of Bantu gender systems	120
Jayden L. Macklin-Cordes	Towards a phylogenetic typology of Australian nominal classification	122
Tom Gueldemann and Ines Fielder	"Noun classes" in Niger-Congo: from classifier to gender system	125
Michael Franjeh, Alexandra Grandison and Greville Corbett	From Classifier to Grammatical Gender: Evidence from Oceanic	127
Jurriaan Wiegertjes	Tukanoan highly grammaticalized classifiers are an expanded gender system	129
Denis Creissels	Non-contextual uses of gender-number agreement markers and the emergence of orphan agreement markers in Atlantic languages	132
Gauthier Dubois and Sylvie Voisin	Changements dans les systèmes de classes des langues atlantiques Nord à travers la dissociation du genre et du nombre	134
Tom Güldemann	From asymmetric grammatical treatment of nouns to gender: the case of Ubangi	137
Kristin Haug	Gender and declension at an early stage of gender loss	138
Fabian Zuk	Humanity, Animacy and "the Rest": diatopic variation in Gender-Marking and its consequences for Proto-Dogon	140

## **WS Morph2Read: Des connaissances morphologiques à la compréhension de la lecture chez des apprenants en L1 et L2** **142**

### **7 Morph2Read: From Morphological Knowledge to Reading Comprehension in L1 and L2 Learners**

Despoina Stefanou	Explorer la connaissance morphologique dans les mots préfixés et suffixés : le cas des locuteurs francophones plurilingues	143
-------------------	--	-----

Michela Redolfi, Chiara Melloni and Maria Vender	Processing morphology while reading: An eye-tracking study with typical and atypical adult readers	146
Madeleine Voga	Organisation et fonctionnement du lexique bilingue et plurilingue : un modèle fondé sur le transfert des connaissances morphologiques.	148
Anne-Sophie Besse, Felipe Valentini and Márcia Maria Peruzzi Elia da Mota	Conscience morphologique et compréhension en lecture chez les étudiants de français L1 : Lien direct, lien indirect et modération par le niveau scolaire	150
Marta Tagliani, Maria Vender and Chiara Melloni	Enhancing morphological awareness in a-typical learners: Effects of a primary school intervention on morphological and reading skills	152
Hélène Giraud	Morph-to-Read : Un modèle cognitif holistique de la compréhension du langage écrit	155

## **WS Future directions in Word and Paradigm morphology** **158**

### **8 Nouvelles pistes pour la morphologie "Mot et Paradigme"**

Farrell Ackerman and Robert Malouf	The usefulness of the developmental sciences for WP models of morphology	159
Borja Herce and Clayton Marr	Diachronizing WP morphology: quantifying the complexity effects of sound change vs analogy	161
Alexandru Craevschi, Sarah Babinski and Chundra Cathcart	The evolution of principal parts and paradigmatic leveling in Germanic	164
Baerman Matthew, Jérémy Pasquereau and Helen Sims-Williams	Paradigmatic complexity as diachronic layering	166
Chundra Cathcart and Borja Herce	Pattern naturalness and stability in Romance stem alternations	168
Matteo Pellegrini, Federica Iurescia, Eleonora Litta and Marco Passarotti	An integrated approach to inflectional and derivational predictability. The case of Latin deverbal agent and action nouns	170
Olivier Bonami, Louise McNally and Denis Paperno	It takes a family. How morphological families shape the interpretation of derived words	172
Peter Nyhuis	Towards an abstractive theory of morphophonological alternations: Empty morphs, neutralisation and paradigm predictability in Wubuy	174
Nazanin Shafiabadi, Nina Nusbaumer and Olivier Bonami	Assessing the import of morphology in gender assignment: A computational study of French	176
Yulia Zinova, Ruben van de Vijver, Anna Sophia Stein and Melanie Schmidt	Addressing the Problem of Syncretism in Russian Nominal Paradigms for the Task of Unsupervised Learning of Morphology	178
Gabrielle Aguila-Multner and Berthold Crysman	Multiple periphrastic relations in the morphology of French	182

	Maria Copot and Andrea Sims	Community structure in inflectional networks: Exploring inflectional class structure	184
<b>WS</b>	<b>9</b>	<b>Cooccurrences et marquage discursif</b>	<b>187</b>
	Mathilde Dargnat and Agnès Tutin	L'expression de la surprise. Phrases préfabriquées et marqueurs discursifs	188
	Olivier Kraif	Modèles de cooccurrences collostructionnels pour l'études des expressions préfabriquées de l'oral	190
	Jacques Jayez	L'extraction des marqueurs de discours dans les corpus	192
	Ludivine Crible	Ordre, force et variation en genre des co-occurrences de marqueurs discursifs	194
	Mathilde Dargnat, Paola Herreño Castañeda and Maeva Sillaire	S'associer pour mieux contrer. Combinaison de marqueurs discursifs dans le processus argumentatif de la concession	196
	Gaétane Dostie	Phrasèmes interactifs intégrant le morphème interro-exclamatif "-tu" en français québécois parlé (1970-2020). "On s'entend-tu", "ça se peut-tu", "tu penses-tu", "tu veux-tu ben"	198
	Bujar Rushiti	The discourse markers vallë and a thua in Albanian	201
<b>WS</b>	<b>10</b>	<b>Phonologie empirique Empirical phonology</b>	<b>203</b>
	Marie-Hélène Côté and Isabelle Racine	Les oppositions de durée vocalique en syllabe finale ouverte en français de Suisse romande	204
	Marc Chalier, Linda Bäumlér, Elisabeth Heisenberger	Variation et codages inter-juges : le rôle de la variété native du codeur sur la qualité du codage à l'épreuve des données	206
	Fabian Santiago	Interactions prosodie-phonétique en français L2	208
	Jun Wang	Analyse exploratoire des corrélats rythmiques du français lu chez les apprenants chinois de FLE	210
	Svenja Krieger	Prosodic aspects of discourse particles in L2 learners of French	212
<b>WS</b>	<b>11</b>	<b>Expressing extreme quantity: a myriad of variants: Contrastive, synchronic, diachronic, empirical and theoretical approaches Exprimer la quantité extrême, une kyrielle de variantes : Approches contrastives (langues germaniques, slaves, romanes) synchroniques, diachroniques, empiriques et théoriques</b>	<b>214</b>
	<b>Communication invitée:</b> Daniel Petit	L'origine du lituanien daug "beaucoup"	215
	Delphine Pasques, Adrien Morvan, Natalia Bernitskaïa	Expression de la grande quantité en moyen-haut-allemand et en vieux-russe : approche sémantique et pragmatique	217
	Olivier Azam	Une syntaxe saisie en pleine transition : la syntaxe des numéraux de 1 à 5 en russe littéraire moderne entre diachronie et synchronie	219
	Krisztina Hever-Joly	Expression de la quantité extrême dans les langues finno-ougrienne (hongrois, mordve erzya)	222

Martina Verdelli	How to express extreme quantity through Italian multi-word expression: a proposal study	223
Katharina Fezer	"Il y en avait tant, tant, tant !" - Comment les enfants acquièrent-ils l'expression de la quantité extrême ?	226
Zhuo Jing-Schmidt	Spatial metaphors of exhaustive quantification in Chinese	228
Giacomo Bucci	Downward entailing properties of the Partitive of Negation from a typological perspective	229
<b>Communication invitée:</b> Laure Gardelle et Wiltrud Mihatsch	Les degrés de grammaticalisation dans les binomiaux anglais et français désignant les grandes quantités	231

## **WS Natural language processing and linguistic variation analysis** **233**

### **12 Traitement automatique du langage et analyse de la variation**

Natasha Romanova and Rayan Ziane	(Vers) l'adaptation des modèles à des données hors du domaine d'entraînement pour l'analyse syntaxique du français en diachronie : Exemple du français des îles Anglo-Normandes du 16ème siècle	234
Timo Korhonen	Latin treebanker's pain: suboptimal recall of key syntactic constructions	237
Santiago Herrera, Ioana-Madalina Silai, Caio Corro, Bruno Guillaume and Sylvain Kahane	Building quantitative contrastive grammars from syntactic treebanks	239
Luigi Talamo, Julius Steuer and Annemarie Verkerk	Measuring the accessibility of personal pronouns through surprisal in a UD-parsed parallel corpus	242
Marina Seghier, Alice Millour and Jean-Yves Antoine	Étude de la variation textuelle par le prisme des entités nommées	245
Dana Roemling, Yves Scherrer, Aleksandra Miletić and Janine Siewert	Model Interpretability as Error Analysis for Variation Research	247

## **Workshop & Tutorial**

### **Research infrastructures and FAIR principles in Linguistics: Interdisciplinary perspectives, theoretical implications and practical applications** **250**

Efstathia Soroli & Francesca Frontini	CLARIN The European research infrastructure for linguistic data resources and technology	252
Olivier Baude & Nicolas Larousse	The role of the Huma-Num infrastructure in supporting language sciences	252
Chiara Chelini & Cécile Fabre	The French national networks of humanities and social sciences centres, their platforms and the role of linguistics.	253
Edward Gray	DARIAH: Helping humanities researchers - including linguists - confront the digital turn	253

Christophe Parisse	CORLI and ORTOLANG: Providing tools and guidance for sharing and reusing open science language data	253
Annemarie Verkerk, Luigi Talamo & Andrew Dyer	Compiling and annotating mini-CIEP+: a sharable parallel corpus of prose	254
Matthias Urban	Using cross-linguistic data formats (CLDF) databases for historical linguistics and comparative anthropological research: the example of CINWA, a database of names of cultivated plants in South American Indigenous communities	254
Cyriel Mallart, Andrew Simpkin, Rémi Venant, Nicolas Bailler, Bernardo Stearns, Jen-Yu Li & Thomas Gaillat	Linguistic interoperability within a unified architecture: the case of Analytics for Language Learning	255
Antonio Balvet	Teaching linguistics with CLARIN resources: preliminary results for syntax	256
<b>Tutorial session.</b> Christophe Parisse, Alina Tsikulina, Adélie Buclier & Eva Soroli	Basic tools for spoken discourse analysis with the CLARIN Knowledge Centre TalkBank and the CLAN (Computerized Language Analysis) software	256
	<b>Session générale - General session</b>	258
Johan van der Auwera	Cycles?	259
Paola Marongiu	Modal combination patterns in Latin: A corpus analysis by genre	261
Zinaida Geylikman	General human nouns in Medieval French: borderline grammaticalization?	264
Inka Wissner	Variabilité en grammaire et vernacularité : les adverbes du type en neuf et à blanc en français	267
Axelle Domingues	Corrélation entre l'émergence des prépositions fonctionnelles et la perte de la flexion casuelle en italien et en français : perspectives diachronique et comparée	269
Michail Marinis	Language contact and paradigmatic levelling in Italiot nominal inflection	273
Kelig Bertin	Micro-diachronie en langue des signes française (LSF) : quels changements lexicaux ?	275
Agnès Lenepveu-Hotz	Differential Object Marking in New Persian: Diachronic and Dialectal Variations	277
Alexander Rostovtsev-Popiel	Spatial Applicatives in Georgian in Diachronic Perspective	279
Juliette Thuilier, Gilles Corminboeuf and Christophe Benzitoun	Une alternance en question : avoir été et être allé en français oral	281
José Deulofeu	When corpus linguistics helps solving a theoretical puzzle: the case of French relative qui	282
Anne Dister	Corpus oraux et variation diatopique. Documenter le français parlé à Bruxelles	284
Mathieu Avanzi and Laure Anne Johnsen	OFROM+ : documenter la variation du français en Suisse romande et dans la francophonie	285



Suzanne Lesage	Marqueurs de haut degré et variation régionale : État des lieux des contraintes sémantico-syntaxiques sur les MDH en usage en Suisse romande	286
Wim Remysen	La recherche sur le français québécois et la mutualisation des ressources	288
Izold Guégan	À la recherche de la syntaxe du breton contemporain : enquête auprès des locuteurs natifs	290
Anne Dagnac, Juliette Thuilier and Iona Brédif	Differential Object Marking in Southern French: Effects of Local factors	291
Saghie Sharifzadeh and Thomas Sähn	Représentations picturales et langues des signes : étude diasystémique de processus d'iconisation et de variations transculturelles	294
Juliette Thuilier, Pegah Faghiri, Léa Pineau and Yanis da Cunha	Prototypicalité des arguments et animéité dans le domaine postverbal du français	296
Giulio Massari, Fanny Meunier and Raphaël Fargier	Sharp as a tack: assessing the role of iconicity and non-arbitrariness in word production processes	299
Fanny Meunier, Anaïs Tran Ngoc and Julien Meyer	Le traitement de la parole sifflée par des auditeurs non entraînés	301
Fabio Montermini	L'émergence d'un nouveau suffixe : -(i)ssimo vs. -issime en français contemporain	303
Léa Pineau, Fabio Montermini and Juliette Thuilier	Les marqueurs typographiques et linguistiques associés aux néologismes : études en production et en réception.	305
Juliette Thuilier, Pegah Faghiri, Léa Pineau and Yanis da Cunha	Alternance de voix des verbes psychologiques en français : une étude expérimentale	308
Nina Dobrushina	Who fulfills a wish? On supernatural ergatives in Nakh-Daghestanian languages	310
Natalia Logvinova	Nominative-Genitive variation in Russian close appositional constructions	312
Michail Marinis and Brian Joseph	Comparing Headline grammars: Article Omission in Greek and English	313
Fayssal Tayalati	Les réalisations du na't sababī en français. De l'arabe standard au français, il n'y a qu'un pas	315
Amazigh Bedar	L'allomorphie à l'interface phonologie-syntaxe	318
Abdulrahman Burour Ibrahim	French as the Second Official Language in Nigeria: Symbolic or Reality	320
Ludovic Mompelat	«Vers l'Infini(tif) et Au-delà », Identifying Finiteness in Creoles: A Contrastive Study of the Complementation Systems of Martinican Creole and Haitian Creole.	321
Noémie François-Haugrin	Insécurité linguistique chez les jeunes créolophones de la Martinique : le jeu de rôle comme outil d'investigation de la compétence langagière dans la langue minorée.	323
Lina Inčiuraitė-Noreikienė	Competition between personal and agent nouns with borrowed roots in contemporary Lithuanian	325
Liudmila Arcimavičienė and Lina Marčiulionytė	Discursive representation of wildfires in the news media narrative: the case of anthropocentrism	328

Eleonora Gualdoni and Gemma Boleda	Why do objects have many names? A study on word informativeness in language use and lexical systems.	330
Christophe Combe and Dejan Stosic	Le traitement linguistique et cognitif de la composante de manière : Le cas de l'interprétation simultanée français-anglais et anglais-français	332
Natia Botkoveili	Resilience Predictors in International Students – the Role of Multilingualism	335
Zygmunt Frajzyngier	'Event' as a linguistic category	337
Denys Teptiuk, Tatiana Nikitina and Stef Spronck	Visual and auditory perception in narrative discourse: a contrastive approach	338
Alice Vittrant and Karl Seifen	(A)symétrie dans l'expression de la Source et du But dans quelques langues de Birmanie	340
Shogo Mizuno	Axial forms exhibit stronger cohesion with the ground nominal than directional markers	342
Martin Haspelmath	FAIR database publication in typology: Language parameters and construction parameters in the CrossGram data registry	345
Daniele Sanacore, Nabil Hathout and Fiammetta Namer	Estimer la prédictibilité dans les paradigmes dérivationnels à partir de scénarios générés par des LLM	347
Hing Yuet Fung	An oblique correspondence between dominant word orders and corpus variations	348
Marc Evrard, Philippe Boula de Mareüil, Alexandre François and Antonio Romano	Classification phylogénétique des dialectes romans : à la croisée entre linguistique historique et apprentissage automatique	351
Christophe Benzitoun	Entre normes et variations : réflexions sur les usages contemporains des corpus en linguistique française	353
<b>Programme</b>		<b>354</b>
<b>Index des auteurs</b>		<b>380</b>
<b>Author index</b>		<b>380</b>



## **Langues & Langage à la croisée des Disciplines**

### **1ère Rencontre annuelle LLcD**

Colloque : 9-11 septembre, 2024

Ecole d'été : 12-14 septembre, 2024

Sorbonne Université - 28 Rue Serpente, 75006 Paris, France



## **Workshop 1**

### **Métrique, musication**

### **Metrics, Textsetting**

## Comment les Gavião du Rondônia (Brésil) imitent la parole grâce à trois instruments de musique traditionnels d'Amazonie

Meyer Julien <sup>(1)</sup>, Moore Denny <sup>(2)</sup>

(1) Université Grenoble Alpes, CNRS, GIPSA-Lab, Grenoble, France

(2) Secteur de linguistique, Museu Paraense Emílio Goeldi (MPEG), Ministério de Ciência, Tecnologia e Inovação, Belém, Brésil

De nombreuses populations de la planète utilisent diverses formes de parole auxiliaire non vocale qui s'appuient sur les productions sonores d'instruments de musique pour imiter la parole usuelle, principalement dans des situations de performance artistique et de communication de messages publics (Meyer et Manfredi 2024). Ce type de "langage de substitution" (Nketia 1971) emploie des techniques de "réduction de la parole" (Stern 1957) basées sur des "icônes acoustiques" (Sebeok et Umiker-Sebeok 1976). Ainsi, les Gavião, un groupe amazonien originaire du Rondônia, au Brésil, utilisent trois instruments de musique traditionnels différents qu'ils identifient comme "parlants" ou "chantants" et qui reposent sur une relation de similarité musique-parole étroite (au niveau des profils mélodiques (F0) et rythmiques). Il s'agit d'une flûte (appelée *kotiráp*), d'une paire d'arcs à bouche (*iridináp*) et d'un ensemble de trois grandes 'clarinettes' à une note en bambou (*totoráp*) (Moore and Meyer 2014). Ce dernier instrument est très caractéristique des cultures d'Amazonie (Beudet 1997) et est joué par plusieurs musiciens ayant chacun un instrument à hanche à une seule note.

Notre recherche s'appuie sur 7 mois d'enquête de terrain sur ce sujet - répartis en 5 périodes sur 5 ans - et fournit une approche comparative entre ces différentes modalités de parole encore en pratique chez les Gavião. Ces instruments parlants représentent un patrimoine culturel peu étudié du substrat traditionnel amazonien qui est menacé de disparition par un faible taux de transmission intergénérationnelle. Leur rôle est à la fois artistique et pratique car ils contribuent à entretenir le patrimoine oral d'un large corpus de chants et de poésies.

Ici, nous explorons la nature phonétique et phonologique de l'iconicité acoustique qui existe entre et la musique jouée sur ces instruments et les paroles de chansons en gavião du rondônia (une langue à ton également dotée de quantités vocaliques distinctives). Celle ci repose sur l'association cognitive entre, d'une part, le ton et la longueur de chacune des syllabes de cette langue et, d'autre part, les notes jouées avec ces instruments de musique. De plus, on remarque une différence entre la voix qui permet de moduler la courbe mélodique fréquentielle et le jeu musical l'imitant qui se limite à des notes fixes. Cet aplatissement mélodique est une adaptation aux contraintes acoustiques des instruments de musique qui offre un regard alternatif sur le système tonal de la langue. En effet, les interprètes continuent de restituer la hauteur relative de chaque syllabe de la version parlée par rapport aux syllabes environnantes, y compris quand des relations phonétiques/phonologiques complexes s'appliquent (tons montants, ou effets descendants tels que le 'downstep') (Meyer et Moore 2021). Pour enseigner ces mélodies 'aplanies' imitant la parole, les musiciens d'Amazonie utilisent souvent un 'double chanté' (Beudet 1997), véritable intermédiaire qui fait office de modèle de la mélodie instrumentale. Nous montrons également de manière plus détaillée qu'auparavant comment la trame rythmique de la phrase parlée est reflétée par la musique instrumentale, conservant les quantités phonologiquement distinctives de la langue tout en s'adaptant aussi à d'autres effets comme la versification ou les fusions syllabiques de la parole rapide lorsqu'elle est chantée ('parole allegro').

**Mots clefs:** relation musique-langage, instruments parleurs, métrique, parole chantée, iconicité mélodique

### **Bibliographie**

Beudet, J. M. (1997). *Souffles d'Amazonie, les Orchestres tute des Wayãpi*. Nanterre: Société d'Ethnologie.

Meyer, J., and Manfredi, V. (2024). Auxiliary speech by vocal tract modulation and musical surrogacy. In Sammler, D. (ed.), *Oxford Handbook of Language and Music*. Oxford: Oxford University Press.

Meyer, J., and Moore, D. (2021). A Flute, Musical Bows and Bamboo Clarinets that “Speak” in the Amazon Rainforest; Speech and Music in the Gavião Language of Rondônia. *Front. Psychol.* 12:674289.

Moore, D., and Meyer, J. (2014). The study of tone and related phenomena in an Amazonian tone language, Gavião of Rondônia. *Lang. Document. Conserv.* 8, 613–636.

Nketia, J.H. K. (1971) *Surrogate Languages of Africa*, in Thomas A Sebeok (ed.), *Current Trends in Linguistics vol.7: Linguistics in Sub-Sahara Africa*, 669-732. The Hague: Mouton.

Stern, T. (1957). Drum and whistle languages: an analysis of speech surrogates. *Am. Anthropol.* 59, 487–506.

# **Singing Moras: a Musical Textsetting Problem in Japanese Lullabies and Nursery Rhymes**

Jean-Louis Aroui  
*Université Paris 8*  
U.M.R. 7023 Structures Formelles du Langage  
FRANCE

Japanese native speakers hardly have an intuition for the *syllable* unit in their language; but they easily chunk a text into moras. Some of these moras are produced by gemination and can be voiceless plosives (Labrune 2012). Words like *sakka* (“author”), *koppu* (“goblet”) or *motte* (“to hold and...”) contain three moras (*sa.k.ka*, *ko.p.pu*, *mo.t.te*).

Phonetically, such moras are unsingable. But Japanese songs often include such moras. How do Japanese singers deal with them?

In this paper, building from Hayes & Kaun’s (1996: 260) “syllable duration” constraint and Dell & Elmedlaoui’s (2008: 204) “rightward shift” operation, I suggest that these moras are always mentally associated to a musical beat, *via* a silence or a melisma. The corpora consists of lullabies and nursery rhymes included in Grosléziat (2007) and Sakura (2004). When a same song appears in both corpora, the treatments of geminate voiceless plosive are compared.

## **References**

- Dell, F. (2011). Singing in Tashlhiyt Berber, a Language that Allows Vowel-less Syllables. In C.E. Cairns & E. Raimy (Eds.), *Handbook of the Syllable* (pp. 173–193). Leiden/Boston: Brill.
- Dell, François & Elmedlaoui, Mohamed. 2008. *Poetic Meter and Musical Form in Tashlhiyt Berber Songs*. Köln: Rüdiger Köppe Verlag.
- Grosléziat, Chantal. 2007. *Comptines et berceuses des rizières. 29 comptines de Chine et d’Asie* [1 CD + 1 book]. Paris: Didier Jeunesse.
- Hayes, B. & Abigail Kaun, A. (1996). The Role of Phonological Phrasing in Sung and Chanted Verse. *The Linguistic Review*, 13: 3-4, 243–303.
- Jackendoff, R. (1989). A comparison of Rhythmic Structures in Music and Language. In P. Kiparsky & G. Youmans (Eds.), *Rhythm and Meter* (pp. 15–44). San Diego, Academic press (Phonetics and Phonology, 1).
- Labrune, Laurence. 2012. *The Phonology of Japanese*. Oxford: Oxford university press.
- Sakura. 2004. *Japon. Rondes, comptines et berceuses* [1 CD + 1 book], s.l.: ArB Music.

## **Keywords:**

musical textsetting, song, Japanese, mora.

In the 1990s, Katherine Rohrer noted the prevalence of dance forms in Henry Purcell's vocal music, arguing that the dance chosen depended on the poetic metre of the text. Rohrer also observed that Purcell and his contemporaries employed local rhythmic effects for certain types of poetic line, but she described these as transcriptions of verbal stress into music, rather than theorising a set of stereotyped musical rhythms suited to specific types of poetry. In short, Rohrer offered a theory of large-scale English Baroque musico-poetics, but saw smaller-scale word-setting patterns as text-derived rather than musically or musico-linguistically systematic.

This paper expands upon Rohrer's analysis, filtered through Robert O. Gjerdingen's theory of galant musical schemata, and informed by Megan Kaes-Long's and Ruth I. DeFord's concept of schematic textsetting in music of the late Renaissance and by Yonatan Malin's concept of declamatory schemas in nineteenth-century *Lieder*. Like Rohrer's, its poetic methodology is essentially generative, seeing poetic metre as a sequence of strong and weak positions and textsetting as exhibiting a number of normative matching rules for well-formedness. However, it also draws upon the textsetting research of Nigel Fabb and on traditional rhetorical concepts (a scheme or *skhêma* as a standardised pattern that distorts the formal norms of language) to posit a new, synthetic theory of textsetting: the existence of musico-linguistic 'skemes', musical stock gestures that match specific poetic-metrical patterns to produce mismatches between poetic and musical metrical strength in specific, stereotypical ways. This analytical method, developed for my PhD thesis on Handel, is also the subject of my forthcoming monograph derived in part from that thesis.

The paper first examines three common, small-scale (two-to-four-beat) rhythmic devices, arguing that each was used exclusively to set a different type of short (two-to-four-syllable) poetic metrical pattern:

- The Chanson Opening matches *weak-STRONG-weak-STRONG* syllable strings to *STRONG-STRONG-weak-STRONG* strings of musical beats in duple time.
- The Compressed Triple Upbeat matches *weak-STRONG-weak-STRONG* syllable strings to *STRONG-weak-weak-STRONG* strings of musical beats in triple time.
- The Strong Delay matches *STRONG-weak* syllable pairs to *weak-weak* pairs of musical beats in triple time, or a weak beat and then the weaker latter half of that beat in duple.

Some of the resulting poetic-to-musical-strength-mismatches are theorised as 'ideal fits' for a mismatch already present in the poetry, a standard means of reflecting common surface-prosodic deviations from the metrical template such as initial inversion. Other, however, seem unmotivated by surface prosody, instead serving to provide composers with a way to produce musical variety while remaining recognisably faithful to poetic form. Following Graeme M. Boone's and Lawrence Earp's theories of French medieval musico-poetics, I describe this as 'recitational dissonance', and expand upon this concept to suggest that, much like dissonance in harmony, recitational dissonance must later 'resolve', which is to say, be followed by recitational consonance, a point in the setting completely devoid of mismatch between poetic and musical strength.

The second part of the paper examines two frequently-used, large-scale formal types also employed exclusively for specific poetic metres and stanza-forms. Like the smaller-scale skemes, these encompass musical rhythmic patterns, but unlike their small-scale counterparts, also encompass specific approaches to word repetition, melisma, and dance-derived structure, suggesting that poetic and musical form were far more closely connected than musicologists often suppose:

- The Lullian Minuet sets duple-metre texts in triple time, with every strong syllable on a strong beat, little to no repetition of words, no long melismas, and a predominant SHORT-long rhythm (a one-beat note followed by a two-beat note, beginning on the strong beat)
- The Ballad Song sets common-measure texts in lyrical duple time, with every strong syllable on a strong beat, little to no repetition of words, no long melismas, and often strophically (if the text is strophic) and for a solo voice

The paper concludes with a case study of a specific skeme in action throughout a specific song ('Let Monarchs Fight' from Purcell's dramatick opera *Dioclesian*), demonstrating how a skematic analysis can aid not only in our understanding of the setting's formal operations, but also in delivering a more sensitive performance of the song.



## Un élément de parophonologie médiévale : l'élision

Timothée PREMAT

SFL (CNRS & Univ. Paris 8 : UMR 7023)

La parophonologie désigne une grammaire phonologique spécifique au vers (Kiparsky 1977). C'est elle qui, dans le vers français, décide du maintien des schwas finaux devant initiale consonantique : l'élision est ainsi prohibée au sein d'un hémistiche, indépendamment de la structure syntaxique de la phrase. Comme illustré en (1) et (2), l'isosyllabisme ne peut être maintenu qu'au prix d'un maintien de la finale atone de *fébriles* et *aise*, alors même que ces mots sont en fin de syntagme intonatif et que le schwa ne devrait pas s'y maintenir.

- (1) *Calme un peu ces transports fébriles, ma charmante* (Verlaine, 1866, « Lassitude », v. 2)  
\*[febɛʁil]
- (2) *Laisse-la trompeter à son aise, la gueuse !* (Idem, v. 10)  
\*[ɛz]

Dans les vers (1) et (2) de Verlaine, c'est donc l'hémistiche qui règle, seul, le traitement de l'élision, et toute finale atone devant initiale consonantique au sein d'un hémistiche est préservée de l'élision. Il en va de même dans les vers médiévaux : tous les mots lexicaux paroxytons perdent leur finale si, et seulement si, ils sont suivis d'un mot à initiale vocalique au sein du même hémistiche. Ainsi, en (3), l'isosyllabisme impose l'élision sur *nostre* mais l'interdit sur *emperere*.

- (3) *Carles li reis, nostre emperere magnes,* (Roland, éd. Moignet, 1972, v. 1)  
[ˈnɔstr ɛmpəˈrɛrə ˈmaɲəs]  
« Le roi Charles, notre grand empereur »

La situation est parfois moins claire pour les mots fonctionnels, avec notamment des élisions non transcrites par la graphie. Ainsi, en (4), l'isosyllabisme impose l'élision de *ki* dans *ki est* ; mais l'élision n'est pas permise sur *une* (ni sur *Sarraguce* ou *muntaigne*, dont la finale est extramétrique).

- (4) *Fors Sarraguce, ki est en une muntaigne.* (Idem, v. 6)  
[saraˈgutsə ˈkɛst] [ynə mɔ̃nˈtɛɲə]  
« À part Saragosse, qui est situé sur une montagne »

Si l'on admet que l'élision métrique a pour domaine l'hémistiche ou le vers (pour les vers simples), cela impose de considérer que cette définition n'est pas le résultat de la seule phonologie. En effet, avant l'élargissement du domaine computationnel (à partir de la fin du XII<sup>e</sup> s.), l'élision ne concerne probablement que certains monosyllabes fonctionnels (Premat 2023b). Ainsi, en (3), la prononciation isosyllabique [ˈnɔstr ɛmpəˈrɛrə] pratique une élision qui n'était probablement pas généralisée dans la langue spontanée. De même, en (5), la prononciation isosyllabique (a), impliquant l'élision, n'est probablement pas conforme à la phonologie médiévale, qui aurait préservé le schwa final de *Carlemagne* (b).

- (5) *Seignurs baruns, a Carlemagne irez.* (Idem, v. 70)  
a. [karləˈmaɲ iˈrɛts]  
b. [karləˈmaɲə iˈrɛts]  
« Seigneurs barons, vous irez rejoindre Charlemagne »

Dans cette présentation, nous interrogerons donc l'écart entre la phonologie de l'ancien français et celle du vers médiéval, et proposerons une formalisation de la parophonologie de l'élision. Les données sont issues de deux grands corpus de textes littéraires (NCA : Stein et al., 2006 ;

BFM : Guillot-Barbance et al., 2017), analysés *via* deux programmes informatiques (le PAM et Comparalem : Poggio et Premat 2019, Premat et Poggio [2024], Premat 2023a, 2024). L'analyse formelle proposée se joue à l'intersection entre formalisme métrique et phonologie CVCV (Lowenstamm 1996, Scheer 2004).

## Références

- GUILLOT-BARBANCE, Céline, HEIDEN, Serge et LAVRENTIEV, Alexei (2017) « Base de français médiéval : une base de référence de sources médiévales ouverte et libre au service de la communauté scientifique », *Diachroniques*, n° 7, p. 168-184.
- KIPARKSY, Paul (1977) « The Rhythmic Structure of English Verse », *Linguistic Inquiry*, vol. 8, n° 2, p. 189-247.
- LOWENSTAMM, Jean (1996) « CV as the only syllable type », In J. Durand et B. Laks (éd.), *Current trends in Phonology. Models and Methods*, Salford, Manchester, ESRI, p. 419-441.
- MOIGNET, Gérard éd. (1972) *Chanson de Roland*, Paris, Bordas. Publiée en ligne par l'ENS de Lyon dans la Base de français médiéval, dernière révision le 30-6-2006.
- POGGIO, Enzo et PREMAT, Timothée (2019) « Le PAM, un Programme d'Analyse Métrique pour le français médiéval », In *Actes des Rencontres lyonnaises des jeunes chercheurs en linguistique historique*, Lyon, Diachronies contemporaines, p. 59-70.
- PREMAT, Timothée (2024) « Comparalem, Outil pour une phonétique historique de corpus ». Présentation au congrès *DIACHRO XI*, Madrid, 24/05/2024.
- PREMAT, Timothée (2023a) « Comparalem : Outil(s) pour l'étude de la variation grapho-phonologie médiévale ». Présentation à la journée d'étude *Le traitement numérique des matériaux galloromans : sources anciennes, nouveaux problèmes*, Neuchâtel, 16/02/1993.
- PREMAT, Timothée (2023b) *La genèse de l'élision*, Thèse de doctorat, Nice, Université Côte d'Azur.
- PREMAT, Timothée et POGGIO, Enzo [2024] « Restituer des fragments de phonologie avec le PAM (*Programme d'Analyse Métrique*) : l'élision de schwa et des voyelles pleines dans les monosyllabes fonctionnels en très ancien français », In R. Hesselbach et T. Prohl (éd.), *Approches numériques des corpus historiques des langues de France*, Romanistische Dossiers I, Munich, AVM Verlag.
- SCHEER, Tobias (2004) *A Lateral Theory of Phonology: What is CVCV, and why should it be?*, Berlin/New York, Mouton de Gruyter.
- STEIN, Achim, KUNSTAMM, Pierre et GLEBGEN, Martin-D. (2006) *Nouveau Corpus d'Amsterdam. Corpus informatique de textes littéraires d'ancien français (ca 1150-1350)*, Établi par Anthonij Dees (Amsterdam : 1987), remanié par Achim Stein, Pierre Kunstmann et Martin-D. Gleßgen. Stuttgart, Institut für Linguistik/Romanistic.
- VERLAINE, Paul (1866) *Poèmes saturniens*, Paris, Alphonse Lemerre.

**Mots-clefs** : Métrique ; phonologie ; ancien français ; élision ; sandhi externe.

## Congruence, alignement et naturel dans les vers

Dell & Benini (2021 et à paraître) ont étudié l'aspect syntaxique, qu'ils appellent *congruence*, de l'association entre les expressions linguistiques et la forme métrique. La congruence n'associe pas strictement une catégorie de constituants syntaxiques avec une catégorie de constituants métriques. Contrairement à ce que proposait Hanson 2009, il ne s'agit donc pas d'une contrainte d'alignement au sens défini par McCarthy & Prince (1993), dans lequel le bord d'un constituant d'une structure donnée doit systématiquement coïncider avec celui d'un constituant d'une autre structure (« Align(Cat1, Edgel, Cat2, Edge2) =<sub>def</sub> », 1993 : 80).

La congruence repose bien sur la prise en compte des bords de constituants mis en jeu dans les deux structurations métrique et syntaxique, mais elle associe ces deux structurations de façon asymétrique. L'asymétrie repose entre autres sur le fait suivant : tout bord de constituant métrique doit être associé à un bord de constituant syntaxique, alors que l'inverse n'est pas vrai (dans un vers donné, même en situation de congruence, les bords d'un constituant syntaxique peuvent ne coïncider ni à gauche ni à droite avec ceux d'un constituant métrique). Ainsi dans

[1] *Dans ce chaos du siècle | où [votre cœur] se serre*  
[DV<sub>1</sub> X X X X X X ] [DV<sub>2</sub> X X X X X X ]  
(V. Hugo – la barre verticale indique la césure)

*Votre cœur* est un constituant syntaxique dont aucun des bords n'est aligné avec ceux d'un constituant métrique, puisque le constituant métrique pertinent pour l'analyse de la congruence est DV<sub>2</sub>.

Par ailleurs, la définition de la congruence suppose un alignement sur le bord droit de constituants qui sont en situation de chevauchement. Il s'agit en quelque sorte d'une contrainte d'alignement secondaire, visant à réguler l'empan des configurations non strictement alignées, comme en [2], où la congruence contraint le bord droit du SA et celui du distique à coïncider :

[2] *Il voit plus que jamais | ses campagnes [SA couvertes*  
*De Romains que la guerre | enrichit de nos pertes SA].* (Racine)

Malgré sa dimension asymétrique et secondaire, la congruence a pour conséquence de produire, à l'intérieur des configurations chevauchantes, des constituants strictement alignés à gauche et à droite (ainsi du syntagme prépositionnel *de Romains que la guerre enrichit de nos pertes* en [2])

La présente communication aura pour objectif l'étude de ces constituants doublement alignés et de leurs conséquences sur la cognition des vers, en lien avec l'idée traditionnelle selon laquelle la concordance favorise le *naturel* dans l'écriture métrique.

## RÉFÉRENCES

- HANSON, Kristin, 2009, "Metrical alignment", dans Jean-Louis Aroui et Andy Arleo, *Towards a Typology of Poetic Forms*, Amsterdam, Benjamins, p. 268-286.  
DELL, François et Romain BENINI 2021, *La Concordance chez Racine*, Paris, Classiques Garnier.  
—, à paraître, "Incongruent enjambments", *Linguistic Inquiry*.  
MCCARTHY, John & Alan PRINCE, 1993, "Generalized Alignment", dans Geert Booij & Jaap van Marle (dir.), *Yearbook of Morphology*, Dordrecht, Kluwer, p. 79-153.

**Éliane Delente**  
**Crisco (EA 4255)**  
**Université de Caen Normandie**

## **Une approche temporelle et dynamique des décalages rythme / sens dans la poésie française versifiée du XVIIe à la fin du XIXe siècle.**

Les décalages rythme / sens (enjambement, rejet, contre-rejet) appartiennent à une longue tradition poétique et n'ont cessé d'être étudiés au sein de poèmes particuliers mais peinent à se constituer en véritable objet théorique. Pourtant, de nouvelles études (Dell & Benini 2020) et de récents programmes automatiques de leur repérage (Ruiz Fabo, P., Martínez Cantón, C., Poibeau, T. et González-Blanco, E. 2017) témoignent d'un regain d'intérêt pour ces questions.

L'exposé, s'appuyant sur un corpus heuristique (Scheer 2004) qui a permis d'observer les décalages rythme / sens chez Boileau (traitement rythmique peu discordant), et Chénier (décalages assez brutaux) fait un certain nombre de propositions générales visant à contribuer à la construction d'un objet théorique :

1- je montrerai que la question de la discordance entre constituants grammaticaux et constituants rythmiques n'est pas spécifique à la poésie versifiée. Les décalages prosodie / syntaxe dans les discours oraux, par exemple, sont l'objet d'études de plus en plus nombreuses (Degand, L. & Simon, A. C. (2009) ; Simon, A. C. & Christodoulides, G. 2016 ; Martin, P. (2018) ; Lacheret-Dujour, A. & Kahane, S. (2020)). En conséquence, je propose d'adopter un traitement rythmique unifié.

2- méthodologiquement, le rythme métrique organisant intégralement le poème, nous prenons comme unités d'étude les expressions métriques elles-mêmes (EMs) -hémistiches, vers, modules de strophe, strophes, poème-, plutôt que la phrase et ses constituants. Sera défendue l'hypothèse que les processus phonologiques et rythmiques à l'œuvre dans le vers permettent de le constituer comme unité de planification et de traitement du discours. Les EMs correspondent ainsi aux étapes essentielles -parce que rythmiques-, de la progression du discours. Une des implications de cette hypothèse est que les relations syntaxiques à l'entre-vers, en cas de décalages rythme / sens, n'ont pas la préséance sur les relations rythmiques de vers à vers. Formellement, le vers peut ainsi correspondre à un constituant grammatical ou non (achevé, complet provisoirement, incomplet, interrompu, etc.). Mais rythmiquement, sa fin est un moment d'interprétation.

3- vue comme un défaut d'alignement entre deux structures achevées -mètre / syntaxe (et/ou prosodie)- la discordance est souvent traitée comme une propriété objective du texte, vu comme un produit fini. Pourtant, le discours poétique est une action en cours ('a process in time'), se déroulant dans le temps de la lecture. Les unités morpho-syntaxiques et phonologiques sont traitées au fur et à mesure qu'elles se déploient dans le temps. De même, le rythme régulier -métrique- se construit mentalement dans le temps. Nous intégrons donc à l'analyse une perspective temporelle, en envisageant le discours EM après EM et une perspective dynamique, qui implique chez le lecteur une phase prospective de prédictions et une phase rétrospective de satisfaction ou frustration, qui sont des parties intégrantes de l'interprétation, entraînant des réajustements continuels. Ce processus interprétatif implique les composantes de tout discours -morpho-syntaxe, phonologie, rythme, sémantique, pragmatique, interaction, d'où la formulation 'rythme / sens' de l'interface examinée.

### **Bibliographie**

- Attridge, D. (1982) *The rhythms of English poetry*, London : Longman.  
(1995) *Poetic Rhythm: An Introduction*, Cambridge University Press.
- Auer, P., Couper-Khulén, E., Müller, F. (1999) *Language-in-time*, Oxford University Press, 236 pages.
- Auer, P. (2009) : "On-line syntax: Thoughts on the temporality of spoken language", *Language Sciences* 31 (2009) 1–13.

- Bakker, E. (1997) *Poetry in Speech. Orality and Homeric Discourse*, Cornell University Press.
- Bateux, l'Abbé (1810) *Traité de la construction oratoire*, Demonville, Paris.
- Blanche-Benveniste, C. (2003) : « La naissance des syntagmes dans les hésitations et répétitions du parler », en J.-L. Aroui (ed.), *Le sens et la mesure. Hommages à Benoît de Cornulier*, Paris, Honoré Champion, 2003, pp. 40-55.
- Blanche-Benveniste, C. et Martin, P. (2011) « Structuration prosodique, dernière réorganisation avant énonciation », *Langue Française*, 170, 127-142.
- Brazil, D. (1995) *A Grammar of Speech*, Oxford: OUP.
- Chafe, W. (1994) *Discourse, Consciousness, and Time. The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago & London: The University of Chicago Press.
- Cornulier, B. de (2000) « La place de l'accent, ou l'accent à sa place. Position, longueur, concordance », dans *Le vers français. Histoire, théorie, esthétique*. Textes réunis par Michel Murat, Paris, Champion.
- Cornulier, B. de (2003) : « Problèmes d'analyse rythmique du non-métrique », *Semen* n° 16, Rythme de la prose, éd. Eric Bordas, Presses universitaires Franc-Comtoises, 107-118.
- Cornulier, B. de (2009) *De la métrique à l'interprétation. Essais sur Rimbaud*, Paris, Editions Classiques Garnier.
- Cornulier Benoît de (2012) : « Si le mètre m'était compté... Sur la notion fallacieuse de mesure de vers », dans Louis de Saussure, Andrée Borillo et Marcel Vuillaume (dir.), *Lexique, Référence, Regards sur le sens, Mélanges offerts à Georges Kleiber pour ses quarante ans de carrière*, Berne, Peter Lang, coll. « Sciences pour la communication », p. 355-376.
- Delente, E. (2018) « La dimension textuelle du rythme. Étude chez Verlaine », col. «Textes & Langue», Presses universitaires de Lyon.
- Delente, E. (2002) "Can Relationships between Rhythm and Meaning be automated?", Anne-Sophie Bories; Petr Plecháč; Pablo Ruiz Fabo (éds), *Computational Stylistics in Poetry, Prose, and Drama*, De Gruyter, pp.99-116
- Dell, F. & Benini, R. (2020) *La Concordance chez Racine. Rapports entre structure grammaticale et forme métrique dans le théâtre de Racine*, Paris, Classiques Garnier, coll. Versification, métrique et formes de la poésie, 229 p.
- Du Gardin, L. (1620) *Les premières addresses du chemin de Parnasse, pour monstrier la prosodie françoise par les minutez des vers françois, minutees en cent reigles*, Douay, B. Bellere.
- Fraisse, P. (1974) *Psychologie du rythme*, PUF.
- Grasserie, R. de la (1900) *Des principes scientifiques de la versification française*, Maisonneuve, Paris.
- Golomb, Harai (1979) *Enjambment in Poetry : Language and verse in interaction*. Porter Institut for Poetics and Semiotics, Tel Aviv University, 1979
- Ruiz Fabo, P., Martínez Cantón, C., Poibeau, T. & González-Blanco, E. (2017). Enjambment detection in a large diachronic corpus of Spanish sonnets. In LaTeCH-CLFL 2017, Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature. Association for Computational Linguistics, Aug 2017, Vancouver, Canada. pp.27 - 32.
- Scheer, T. (2004) « Le corpus heuristique : un outil qui montre mais ne démontre pas », *Corpus* 3, Corpus 3, 153-192.

## Analyse automatique d'un corpus de textes versifiés

Stéphane Ferrari et Richard Renault  
(CRISCO, Université de Caen - Normandie)

Le traitement automatique de textes versifiés appliqué à un large corpus de poèmes et de pièces de théâtre permet de mettre en évidence des régularités et des particularités à l'échelle d'un recueil, d'un auteur, ou d'une période. Le corpus analysé est relativement conséquent ; il comporte plus d'un million de vers pour 280 auteurs (25 000 poèmes et 141 pièces de théâtre) ; corpus réparti sur une période qui s'étale du XVIIe au début du XXe siècle. L'analyse automatique porte sur 1) l'identification des noyaux syllabiques, 2) le traitement des "e" instables, 3) le traitement des diérèses, 4) le calcul de la longueur métrique des vers, 5) la détermination du profil métrique des poèmes et du mètre des vers, 6) l'appariement des vers en rimes, l'identification des schémas strophiques et la reconnaissance de la forme globale du poème, 7) l'évaluation et le calcul de l'extension des rimes ainsi que l'évaluation de la qualité des rimes. Le projet étant relativement bien avancé, ces différents points ne nécessitent plus que des ajustements et améliorations.

Le projet de recherche en cours porte sur l'évaluation de la concordance syntaxe/mètre. L'analyse syntaxique automatique du corpus fait l'objet d'un travail, portant sur l'évaluation de la catégorisation des mots. Cette expérimentation est actuellement mise en place sur un échantillon du corpus, constitué de 10 sonnets issus de périodes, d'auteurs et de genres de poésie variés. Quatre outils distincts ont été utilisés pour cette annotation : treetagger (Schmid, 1994), Hops (Grobol et Crabbé, 2021), Arborator (Gerdes, 2013) et UDPipe (Straka et al., 2016). Quatre experts ont alors corrigé ces annotations manuellement et un standard de référence (*gold*) a été mis en place par consensus entre ces experts.

Notre communication portera sur une présentation globale du projet et mettra l'accent sur : 1) la mise en œuvre de l'analyse statistique appliquée à la caractérisation d'un auteur et à l'évolution du traitement des diérèses et de la corrélation entre la richesse des rimes et qualité des rimes ; 2) les nécessaires améliorations du traitement des rimes et de la reconnaissance des formes périodiques indépendamment de la disposition typographique ; 3) la mise en place d'une typologie des enjambements par pondération afin de préciser le degré de concordance entre unités métriques et structuration syntaxique du texte ; et 4) concernant l'évaluation de l'analyse syntaxique, nous présenterons les différentes mesures relatives aux performances des différents outils et à l'accord entre les annotateurs.

Une analyse syntaxique contrôlée du corpus (lemmes, catégories et fonctions), combinée à une analyse métrique des vers et strophes permettra à terme de dégager des observations utiles aux métriciens pour l'étude de la concordance ; évaluation notamment des différents types d'enjambements dans le cadre d'une approche telle que celle proposée par Dell et Benini (2021) pour les distiques du théâtre classique.

### Sites internet :

<https://crisco4.unicaen.fr/verlaine> (métrique en ligne)  
<https://git.unicaen.fr/malherbe/corpus> (corpus Malherbe)  
<https://www.ortolang.fr/market/corpora/malherbe> (corpus Malherbe)

<https://versologie.cz/poetree/index> (Poetree)  
<https://zenodo.org/records/10008459> (corpus Malherbe catégorisé)

<https://universaldependencies.org> (Universal Dependencies)  
<https://ufal.mff.cuni.cz/udpipe/2> (UDPipe 2)

### Quelques Références :

- Cohen, Jean, 1966, Structure du langage poétique. Paris, Flammarion.
- Cornulier, Benoît de, 1995, Art poétique. Notions et problèmes de métrique, PU Lyon
- Delente, Éliane et Richard Renault (dir.), 2015, Traitement automatique des textes versifiés : problématiques et pratiques, Langages n°199.
- Dell, François et Romain Benini, 2021, La Concordance chez Racine. Rapports entre structure grammaticale et forme métrique dans le théâtre de Racine, Paris, Classiques Garnier.
- Gerdes, Kim 2013. Collaborative Dependency Annotation, *Proceedings of Depling 2013*, 88–97, Prague.
- Grobol, Loïc et Crabbé, Benoît 2021. Analyse en dépendances du français avec des plongements contextualisés. *TALN/RECITAL 2021*, Jun 2021, Lille (virtuel), France. [〈hal-03223424〉](https://hal.archives-ouvertes.fr/hal-03223424)
- Monget, Eulalie et Ruiz, Pablo. "Opportunités et limites en stylistique computationnelle de la poésie : Détection automatique de l'enjambement en anglais". *Humanistica 2020*, May 2020, Bordeaux, France. 2020, Humanistica 2020.
- Renault, Richard (2022), « Traitement automatique de la métrique appliqué au corpus de Germain Nouveau ; quelques observations », in Germain Nouveau, l'insaisissable , textes rassemblés par Michèle Monte, Var et Poésies 10 , Laboratoire Babel, Université de Toulon.
- Ruiz, Pablo, et al., 2017, “Enjambment Detection in a Large Diachronic Corpus of Spanish Sonnets”, *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, p. 27-32.
- Schmid, Helmut, 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK.
- Straka, Milan, Jan Hajič, and Jana Straková. 2016. UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4290–4297, Portorož, Slovenia. European Language Resources Association (ELRA).



**Langues & Langage  
à la croisée des Disciplines**

**1ère Rencontre annuelle LLcD**

Colloque : 9-11 septembre, 2024

Ecole d'été : 12-14 septembre, 2024

Sorbonne Université - 28 Rue Serpente, 75006 Paris, France



## **Workshop 2**

**Applicative uses of spatial markers:  
typological and diachronic perspectives**

**Les marqueurs spatiaux ayant une valeur  
applicative : Perspectives typologiques et  
diachroniques**



## A typological study of applicative uses of spatial markers

Timofey Mukhin, An Van Linden & Dana Louagie

(University of Liège)

Keywords: applicatives, spatial verb morphology, typology, valency, grammaticalization

This paper investigates cross-linguistic evidence for the functional extension of spatial verb morphology into applicative uses. Spatial markers (SM) have only recently been established as a source for applicatives; see Van Linden (2022) on Harakmbut and Payne (2021) on Nilotic languages. It is still unknown how widespread this pathway is, and what the main types of variation are. This paper presents the results of a pilot study investigating these issues from a typological perspective in a 75-language sample, compiled using Miestamo's (2005) Genus-Macroarea method.

First, applicative uses are observed for different functional types of SM. A first type concerns *locationals*, as in (1a) from Jarawara (Arauan), where the prefix *ka-* 'inside' changes the semantics of the predicate from 'grasp with hand' (1b) to 'cup in hand' without affecting its valency. Contrariwise, in (1d), the same prefix introduces the applied phrase *otara* 'us' as a core argument; in the non-applicative counterpart (1c), the same participant is coded as an oblique adjunct.

- (1) Jarawara (Dixon 2004: 255, 259)
- a. *sina tama o-ka-na o-ke*  
snuff(F) hold 1.SG.A-APPL-AUX 1.SG-DCL.F  
'I hold the snuff in my hand.'
- b. *jimawa0 tama o-ne o-ke*  
knife(F) hold 1.SG.A-CONT. F 1.SG-DCL.F  
'I am holding the knife with my hand.'
- c. [*otaa ni-jaa*] *mee bosa na-maki-hete-ke tasa*  
1.EXC PERI 3.NSG.S get.up.early AUX-following-RP.NE.F-DCL.F again  
'Then they got up early on us again.'
- d. [*otara*] *mee bosa ka-na-hani*  
1.EXC.O 3.NSG.A get.up.early APPL-AUX-IP.NE.F  
'They got-up-early-on us.'

In addition to *locationals*, we also find *directionals*, see (2), and *associated motion* markers (e.g. in Tungusic languages (Pakendorf & Stoyanova 2021)); some of these functions may be coded by the same SM (Guillaume & Koch 2021: 7).

Secondly, our dataset shows that SM are not only found in prototypical "P-applicative constructions", where a core argument is added, see (1d), but also in so-called "X-applicatives", where a non-core argument is introduced to the clause, see (3b) (Zuniga & Creissels 2024: 19). Our dataset also evidences *redirecting* effects of SM. This is illustrated in (2) from Agar Dinka (Nilotic), where the itive verb form triggers a rearrangement of semantic roles but no valency-change; it causes the object to shift from goal ('bird') (2a) to moving theme ('stone') (2b) (Payne 2021: 719).

- (2) Agar Dinka (Andersen 1992-1994: 10)
- a. *dɔɔk à-bòk dít*  
 boy DCL-throw bird  
 ‘The boy is throwing at the bird.’
- b. *dɔɔk à-bóok doòot*  
 boy DCL-**throw:ITV** stone  
 ‘The boy is throwing a stone thither.’

Thirdly, in terms of the semantic role of the applied phrase, we find that it is often a locative participant, as is expected from the (original) spatial meaning of the marker, but not necessarily so. In (3b) from Dagik (Kordofanian), the suffix *-tɛ* ‘towards’ introduces the landmark ‘room’, while in (1d), the prefix *ka-* introduces a maleficiery. More generally, we will investigate whether Peterson’s (2007: 229) hierarchy of semantic roles of applied phrases can be upheld.

- (3) Dagik (Vanderelst 2016: 96)
- a. *a-ŋɪ b-ɔ-bɛk:-ɔ*  
 REF-1.SG CL-PM-bath\_o.s.-FV(-PFV)  
 ‘I bathed.’
- b. *a-ŋɪ b-ɔ-bɛk:-ɔ-tɛ [ri=g-əðu]*  
 REF-1.SG CL-PM-bath\_o.s.-FV-(**PFV**.)APPL LOC=CL-room  
 ‘I bathed in the room.’

### Harmonized abbreviations

A	transitive subject	FV	final vowel	PFV	perfective
APPL	applicative	IP	immediate mood	PM	predicative marker
AUX	auxiliary	ITV	itive	REF	referential marker
CL	class marker	LOC	locative	RP	recent past
CONT	continuous	NE	non-eyewitness	S	intransitive subject
DCL	declarative	NSG	non-singular	SG	singular
EXC	exclusive	O	transitive object	1, 3	first, third person
F	feminine	PERI	peripheral		

### Acknowledgments

Work on this article was funded by grant ARC 23/27-14 – SPACEGRAM of the Research Council of the University of Liège.

### References

- Andersen, Torben (1992-1994), Morphological stratification in Dinka: On the alternations of voice quality, vowel length and tone in the morphology of transitive verbal roots in a monosyllabic language, *Studies in African Linguistics* 23 (1), 1–63.
- Dixon, Robert M. W. (2004), *The Jarawara Language of Southern Amazonia*, Oxford University Press.

- Guillaume, Antoine & Harold Koch (eds.) (2021), *Associated Motion*, Berlin: Mouton de Gruyter.
- Miestamo, Matti (2005), *Standard negation*, Berlin: Mouton de Gruyter.
- Pakendorf, Brigitte & Natalia Stoynova (2021), Associated motion in Tungusic languages: a case of mixed argument structure, in A. Guillaume & H. Koch (eds.), (2021), *Associated Motion*, 855–898, Berlin: Mouton de Gruyter.
- Payne, Doris L. (2021), The extension of associated motion to direction, aspect and argument structure in Nilotic languages, in A. Guillaume & H. Koch (eds.), (2021), *Associated Motion*, 695–746, Berlin: Mouton de Gruyter.
- Peterson, David A. (2007), *Applicative constructions*, Oxford University Press.
- Vanderelst, John (2016), *A Grammar of Dagik*, Köln: Rüdiger Köppe.
- Van linden, An (2022), Spatial prefixes as applicatives in Harakmbut, in S. Pacchiarotti & F. Zúñiga (eds.), (2022), *Applicative morphology: Neglected syntactic 18 and non-syntactic functions*, 129–159, Berlin: Mouton de Gruyter.
- Zuniga, Fernando & Denis Creissels (2024), *Applicative Constructions in the World's Languages*, Berlin, Boston: Mouton de Gruyter.

# Spatial markers and their effects on valency: A view from Lowland South American languages

Cristian Juárez

*Max Planck Institute for Evolutionary Anthropology*

The South American Guaycuruan languages, Mocoví, Toba, Pilagá, Kadiweo, possess a rather elaborate system of verbal suffixes encoding spatial information. This includes, but is not limited to, different dimensions of the figure-ground relation and directionality. Compare, for example, the use of cognate forms *-ge* : *-ge* : *-ge'*, indicating a distant location in (1a)-(1c).<sup>1</sup> With the growing body of individual studies on these languages over the last twenty years or more (see, among others, Carpio 2012; Carrió 2009; Censabella 2002; Gualdieri 1998; Grondona 1998; Messineo 2003; Sandalo 1997; Vidal 2001), it has become clear that Guaycuruan languages present relevant data for the typologically salient link between spatial markers and valency increasing morphology. That is that, in addition to their spatial function, it has been argued that Guaycuruan spatial markers increase verb valency of monovalent predicates, but only modify the verb meaning of bivalent predicates without necessarily increasing verb valency (see, e.g., Carrió 2013; Censabella 2024; González 2009, 2010). Note in (2), for example, how *-ge* adds a locative participant to the intransitivized verb stem.

Despite the past descriptive progress, comparative work exploring parameters of similarities and variation within the Guaycuruan group is still missing. This presentation aims at filling that gap, while also bringing new data for the cross-linguistic study of spatial morphology in relation to applicativization processes (see, for instance, Van Linden 2022; Kohlberger 2022).

- (1) a. Mocoví (fieldwork notes, mocCA180709)  
*t-aja?-ge Villa Angela Pastoril*  
3.II-go-LOC<sub>2</sub> Villa Angela Pastoril  
'She (Cacica Dominga) goes to Villa Angela and Pastoril.'
- b. Toba (González 2009)  
*ta-ta-ge ze pueblo*  
3-go-AWAY DLYING town  
'He/She is going to town.'
- c. Pilagá (Vidal 2001: 237)  
*w-eta-ge' ñi' Formosa*  
3-go-DIR:THIT CL Formosa.city  
'He/She is in Formosa city (a distant place from where the speaker is located).'

---

<sup>1</sup>ABBREVIATIONS: 3 = third person, AWAY = directional: away, CAUS = causative, CL = classifier, DET<sub>2</sub> = determiner 2: going, far, DIR = directional, DLYING = determiner: lying, F = feminine, II = set II bound person form, INTR = intransitive, LOC<sub>2</sub> = locative type 2, THIT = thither, VM = valency modifier.

(2) Mocoví (fieldwork notes, mocCA191108)

*a-so ?a:lo r-afiwi-cat-can-ge*

F-DET<sub>2</sub> woman 3.II.INTR-be/get.dry-CAUS-VM:INTR-LOC<sub>2</sub>

‘The woman made someone else to dry (something) somewhere else.’

In order to carry out consistent and transparent comparative work, I follow methodological principles of AUTOTYP (Witzlack-Makarevich et al. 2022). Additionally, I only consider a subset of spatial markers, i.e., cognate forms of *-lek*, *-ge*, *-git* and *-gi* in Mocoví, as such a subset allows us a balanced coverage of datapoints in the dataset. This paper pursues two specific goals. First, it seeks to capture the distributional pattern of spatial markers according to verb types, beyond the typical monovalent vs. bivalent distinction. Second, it aims to capture the variation in terms of the semantic roles that each of the studied spatial markers can introduce to the verbal predication. These two tasks will also contribute to understanding the extent to which spatial markers might have been subject to a process of change across the Guaycuruan languages.

**Keywords:** locatives, directionals, valence extension, Guaycuruan, South American languages

## References

- Carpio, M. B. (2012). *Fonología y morfosintaxis de la lengua hablada por grupos tobas en el oeste de Formosa (Argentina)*. München: LINCOM.
- Carrió, C. (2009). *Mirada generativa a la lengua mocoví (familia guaycurú)*. Ph.D. Dissertation, Universidad Nacional de Córdoba.
- Carrió, C. (2013). Clasificación y análisis de las construcciones direccionales y locativas en mocoví. In M. Censabella and C. Messineo (Eds.), *Lenguas Indígenas de América Del Sur II. Morfosintaxis y Contacto de Lenguas*, pp. 39–57. Mendoza: FFyL-UNCuyo y SAL.
- Censabella, M. (2002). *Descripción funcional de un corpus en lengua toba (familia Guaycurú, Argentina). Sistema fonológico, clases sintácticas y derivación. Aspectos de sincronía dinámica*. Ph.D. Dissertation, Universidad Nacional de Córdoba.
- Censabella, M. (2024). Applicatives in toba/qom (guaycuruan). In F. Zúñiga and D. Creissels (Eds.), *Applicative Constructions in the World's Languages*, pp. 143–178. Berlin, Boston: De Gruyter Mouton.
- González, R. (2009). Verbos seriales en toba (familia guaycurú, Argentina). Conference presentation at II Encuentro de lenguas indígenas Americanas y II Simposion de Lingüística Amerindia. (Asociación de Lingüística y Filología de América Latina). Resistencia, Chaco.
- González, R. (2010). Análisis sintáctico y semántico de dos aplicativos locativos en toba (familia guaycurú). *Lingüística* 24, 123–140.
- Grondona, V. (1998). *A grammar of Mocovi*. Ph.D. Dissertation, Department of Linguistics, University of Pittsburgh.
- Gualdieri, B. (1998). *Mocovi (Guaicuru): Fonología e Morfosintaxe*. Ph.D. Dissertation, Universidade Estadual de Campinas.
- Kohlberger, M. (2022). The function of applicative morphology in shiwiari. In S. Pacchiarotti and F. Zúñiga (Eds.), *Applicative morphology: neglected syntactic and non-syntactic functions*, pp. 51–72. Berlin/Boston: De Gruyter Mouton.
- Messineo, C. (2003). *Lengua toba (Guaycurú). Aspectos gramaticales y discursivos*. München: LINCOM EUROPA.

- Sandalo, F. (1997). *A Grammar of Kadiwéu*. Ph.D. Dissertation, Department of Linguistics, University of Pittsburgh.
- Van Linden, A. (2022). Spatial prefixes as applicatives in harakmbut. In S. Pacchiarotti and F. Zúñiga (Eds.), *Applicative morphology: neglected syntactic and non-syntactic functions*, pp. 129–159. Berlin/Boston: De Gruyter Mouton.
- Vidal, A. (2001). *Pilagá grammar (Guaykuruan family, Argentina)*. Ph.D. Dissertation, Department of Linguistics, University of Oregon.
- Witzlack-Makarevich, A., J. Nichols, K. A. Hildebrandt, T. Zakharko, and B. Bickel (2022). Managing AUTOTYP Data: Design Principles and Implementation. In A. L. Berez-Kroeker, B. McDonnell, E. Koller, and L. B. Collister (Eds.), *The Open Handbook of Linguistic Data Management*, pp. 631–642. The MIT Press.

## Applicatives in Timucua, a language isolate of Florida

Timucua is a language isolate, formerly spoken in North Florida. Though no longer spoken, there is an extensive text corpus of the language from the Spanish Colonial period. The description here is based on a corpus of about 134,000 words of Timucua. Timucua is a polysynthetic head-marking verb-final language with two applicative morphemes. This paper provides the first description of the Timucua applicative system.<sup>1</sup>

**The dative applicative** The *-si* affix is an applicative with dative (goal, benefactive, malefactive, and source) arguments, as in (1, 2). Objects of *-si* are always human.

- (1) Chitaco hebuano uqua-si-bi-cho ?  
 Who word learn-APP:SRC-PST-2:ABS-Q  
 ‘Who did you learn (these) words from?’ [1612 Cat f49r]
- (2) Hachaqueniqe Iesu Christo hebuano n-eca-si-bo-bi ?  
 Why Jesus Christ word 1:ABS-teach-APP:GOAL-PL:ABS-PST  
 ‘Why did Jesus Christ teach us (these) words?’ [1612 Cat f34v]

**The spatial/instrumental applicative** The *na-* applicative is used with non-human sources and goals, as well as locations and instruments. The examples below illustrate instrumental (3), source (4), goal (5) and locative (6) readings of *na-*.

- (3) ...numaabo pesolo-leqe atichicolo na-balu-su-te-le.  
 heaven heaven-FOC soul APP:INS-live-CAUS-TNS-DEC  
 ‘it is the bread of heaven with which (one) causes the soul to live’ [1635 Mov f41r]
- (4) Mine Iesu Christo-ma nalimoabo beta na-pita-ta...  
 RESP Jesus Christ-DEF heaven from APP:SRC-descend-SS  
 ‘The honored Jesus Christ descends from heaven and...’ [1628 Cat f34r]
- (5) caqi altari-ma-qua, mine Diosi-mano, na-pitata  
 this altar-DEF-OBL RESP God-TOP APP:GOAL-descend-SS  
 mine-qua, na-hono+le-ta, ano na-cumeca-ta  
 RESP-OBL APP:INS-food+BE-SS HON APP:INS-sustain-SS  
 ‘God descends to this altar and honorably sustains by becoming food and...’ [1628 Cat f13v]
- (6) hiti hica-ma infierno mono-ma na-calubo-so-ta...  
 evil city-DEF hell call-DEF APP:LOC-be:punished-CAUS-SS  
 ‘(Lucifer) punishes him in the evil city called Hell and...’ [1614 Arte f129]

<sup>1</sup> Glosses use the following abbreviations: ABS = absolutive, APP = applicative, CAUS = causative, DEC = declarative, DEF = definite, FOC = focus, INS = instrumental, LOC = locative, PL = plural, PST = past, Q = question, SRC = source, SS = same-subject, TNS = default tense.

Some Timucua applicatives associate with NP bare objects (1,2,3,6), some with NP objects with the oblique -qua (5) and some with PP objects (4).<sup>2</sup> Example (6) shows incorporation of the noun ‘food’ into the copula with a preceding applicative.

The combined use of a single applicative for both spatial and instrumental senses (to the exclusion of benefactive/dative senses) is an underdocumented pattern in the typology of applicatives, and may point to a diachronic connection between these applicatives in Timucua. In addition, the preverbal position of the spatial/instrumental applicative bears a similarity to the position of incorporated nouns in Timucua, suggesting a nominal origin for this affix.

## References

- Marsault, Julie. 2023. Unusual properties of applicative constructions in Umó<sup>n</sup>ho<sup>n</sup>. *International Journal of American Linguistics* 89(3):357-390.
- Movilla, Gregorio de. 1635. *Explicacion de la Doctrina que compuso el cardenal Belarmino, por mandado del Señor Papa Clemente 8. Traducida en Lengua Floridana: por el Padre Fr. Gregorio de Movilla*. Mexico: Imprenta de Iuan Ruyz. [=1635 Mov]
- Pareja, Francisco. 1612. *Catechismo y breve exposición de la doctrina christiana*. Mexico City: Casa de la viuda de Pedro Balli. [= 1612 Cat]
- Pareja, Francisco. 1614. *Arte y Pronunciacion de la Lengua Timucvana y Castellana*. Mexico City: Empronta de Iuan Ruyz. [= 1614 Arte]
- Pareja, Francisco. 1628. *III Parte del Catechismo, en lengua timuquana y castellana*. Mexico City: Imprenta de Iuan Ruyz. [= 1628 Cat]

---

<sup>2</sup> A similar phenomenon is noted for Umó<sup>n</sup>ho<sup>n</sup> by Marsault (2023).



## Spatial applicative prefixes in Umó<sup>h</sup>o<sup>n</sup> and Pá<sup>n</sup>ka (Siouan)

Native languages of North America are known for a number of features, including complex verbal morphology (Mithun 2015, 1999). The Umó<sup>h</sup>o<sup>n</sup> and Pá<sup>n</sup>ka language (Omaha and Ponca, ISO 639-3; Siouan family) is a typical example, for it displays a complex verbal system including a number of valency-changing affixes that combine between themselves. Among these, five applicative prefixes introduce distinct thematic roles.

In particular, three applicative prefixes introduce three distinct spatial relations, as can be seen in examples (1) through (4), where applied phrases are in curly brackets.

(1) **gthi<sup>n</sup>** “to sit”

**shkó<sup>n</sup>=azhi gthi<sup>n</sup>=ga, á=biamá Ishúnik<sup>h</sup>e ak<sup>h</sup>á.**  
 move=NEG sit=IMP.M say=HEARSAY I. DEF:ANIM.PROX  
 “‘Sit still”, said Ictinike.’ (Dorsey 1890: 97.7 / speaker: Frank La Flesche)

(2) **á-gthi<sup>n</sup>** “to sit on {x}”

**“wí-shi<sup>e</sup> {umí<sup>n</sup>zhe tho<sup>n</sup> hébe} á-gthi<sup>n</sup> te,**  
 POSS:1SG-sister’s\_husband rug DEF:INAN.RND part APPL-sit IRR  
**ehé” á=biamá.**  
 A1SG.say say=HEARSAY  
 “‘Let my sister’s husband sit on part of the rug” [said she, they say]’ (Dorsey 1890: 594.4 / speaker: George Miller)

(3) **u-gthi<sup>n</sup>** “to sit in {x}”

**{xáde búta unáshte} o<sup>n</sup>g-ú-gthi<sup>n</sup>=i.**  
 grass round left\_from\_fire A1PL-APPL-sit=PL  
 ‘We sat on a round track of grass which has not been burnt by the prairie fire’ (Dorsey 1890: 456.6 / speaker: Kaxé-tho<sup>n</sup>ba)

(4) **í-gthi<sup>n</sup>** “to sit by {x}”

**Ní<sup>n</sup>kashi<sup>n</sup>ga {áwat<sup>h</sup>e} í-gthi<sup>n</sup> thi<sup>n</sup>k<sup>h</sup>é, á=biamá.**  
 person which\_thing APPL-sit DEF:ANIM.OBV say=HEARSAY  
 “‘Where is the person sitting?’” said the Pawnees.’ [Literally: next to which thing was the person sitting? ] (Dorsey, 1890:409.4 / speaker: Ó<sup>n</sup>p<sup>h</sup>o<sup>n</sup>-to<sup>n</sup>ga)

The prefixes **á-** ‘on’, **u-** ‘in’, and **í-** ‘by’ are attested throughout Siouan languages and are inherited from Proto-Siouan (Helmbrecht 2006, 2008). By contrast, I have been unable to find a similar semantic distinction in applicative systems outside the Siouan family.

In this talk, I will present the three applicative prefixes of Umó<sup>h</sup>o<sup>n</sup> and Pá<sup>n</sup>ka:

- their semantic distinctions, especially the ‘in’ versus ‘on’ distinction, and the other meanings associated with the prefix **í-**;
- the syntactic constructions they are attested in, especially the possibility for the applied argument to be a NP or a PP;

- alternative constructions with no applicative and a peripheral PP complement (scarcely attested);
- cases of semantic demotivation.

This description is based on various textual and lexicographic corpora (esp. Dorsey 1890, 1891 available in Rankin 2008; Dorsey n.d.; OLIT-UNL 2018).

## Abbreviations

1 = 1 <sup>st</sup> person	M = masculine
A = agentive person marker	NEG = negative
ANIM = animate	OBV = obviative
APPL = applicative	PL = plural
DEF = definite	POSS = possessive
HEARSAY = hearsay (evidential marker)	PROX = proximate
IMP = imperative	RND = round
INAN = inanimate	SG = singular
IRR = irrealis	

## References

- Dorsey, James Owen. 1890. *The Čegiha Language* (Contributions to North American Ethnology 6). Washington, D. C.: U.S. government.
- Dorsey, James Owen. 1891. *Omaha and Ponka letters* (Bulletin 11). Bureau of American Ethnology.
- Dorsey, James Owen. n.d. Omaha dictionary. (Unpublished manuscript on note cards.)
- Helmbrecht, Johannes. 2006. Applicatives in Siouan languages: A study in comparative Siouan grammar. (Paper presented at the Workshop on Comparative Siouan Grammar at the SCLC in Billings, MT. Unpublished manuscript)
- Helmbrecht, Johannes. 2008. Decay and loss of applicatives in Siouan languages: A grammaticalization perspective. In Verhoeven, Elisabeth & Skopeteas, Stavros & Shin, Yong-Min & Nishina, Yoko & Helmbrecht, Johannes (eds.), *Studies on Grammaticalization*, 135–155. Berlin: De Gruyter Mouton.
- Mithun, Marianne. 1999. *The languages of Native North America*. Cambridge: Cambridge University Press.
- Mithun, Marianne. 2015. Morphological complexity and language contact in languages indigenous to North America. *Linguistic Discovery* 13(2). 37–55.
- OLIT-UNL. 2018. Part II: Lessons from the Omaha language class at the University of Nebraska-Lincoln. In ULCC & OLIT (eds.), *Umó<sup>n</sup>ho<sup>n</sup> Íye t<sup>h</sup>e, Umó<sup>n</sup>ho<sup>n</sup> Úshko<sup>n</sup>-t<sup>h</sup>e: The Omaha Language and the Omaha Way. An Introduction to Omaha Language and Culture*, 243–636. Lincoln: University of Nebraska Press.
- Rankin, Robert L. 2008. James Owen Dorsey’s The Čegiha language. (Unpublished manuscript)

## Diachronic corpus study on Dutch compound verbs: a pilot study

Ann-Sophie Vrielynck (Universities of Liège and Antwerp)

Dutch compound verbs are traditionally classified as separable (1) and inseparable (2) verbs. The non-verbal element typically has spatial meaning, e.g. *onder* ‘under’ in (1) and (2), and may serve an applicative function, as in (2).

(1) *De zon gaat onder.* (‘The sun goes down.’)

(2) *Hij ondergaat zijn lijden.* (‘He undergoes his suffering.’)

Blom (2004) argues that these verbs are the result of a grammaticalization process in which inseparable verbs have developed from separable ones. This study aims to test Blom’s (2004, 2005) hypotheses, which posit two distinct developments. On the one hand, non-resultative elements could grammaticalize into separable particles, progressing into inseparable prefixes. On the other hand, resultative elements might only evolve into separable particles (Blom 2004: 45). This hypothesis predicts that a separable verb with a resultative meaning must lose its resultative meaning before becoming inseparable, with resultativity being defined as the preverbal element expressing a result that affects the direct object (for transitive verbs) or the subject (for intransitive verbs). This study tests this prediction in two ways.

First, we examine the verb *ondergaan* ‘go down; undergo’ with 706 occurrences in the EDGeS-corpus of Bible translations from the 14th to 20th century (195,948 tokens) (Bouma et al. 2020) and in a subset of the DBNL, comprising literary texts from the 16th to 20th century (9,116,714 tokens) (Van Olmen 2019). In contemporary Dutch, the verb *ondergaan* appears both separable and inseparable, with the former having a compositional resultative meaning and the latter a figurative meaning (cf. (1)-(2)). Historical material is expected to reveal the - likely older - resultative meaning primarily in separable use, while the possibly younger figurative meaning can occur with both usages.

Second, in the same corpora, we scrutinize five verbs that only appear as inseparable in contemporary Dutch, but, contrary to Blom’s hypothesis, express a resultative meaning: *onderdrukken* ‘suppress’ (236 instances), *onderwerpen* ‘subject’ (668 cases), *overtuigen* ‘convince’ (828 attestations), *overreden* ‘persuade’ (145 occurrences), and *overhandigen* ‘hand over’ (101 instances). For the first three verbs, it can be argued that this meaning is indeed resultative but no longer literally compositional, although this is more challenging for the last two. Therefore, we suspect that these verbs also underwent a semantic shift before becoming inseparable, their original meaning being more literally compositional. Like *ondergaan*, a compositional meaning is not expected with inseparable usage.

Our quantitative results show that the verb *ondergaan* seems to exhibit an (almost) one-to-one relationship between form and meaning since Middle Dutch. The separable variants always have a compositional resultative meaning, whereas the inseparable variants usually convey a more figurative, holistic meaning. Additionally, verbs with the prefix *over-* were only found as inseparable verbs in our corpora. The verbs *onderwerpen* and *onderdrukken* appear in both variants. Here, these variants almost always express the same holistic meaning. Nevertheless, separable usage decreases, suggesting that the assumed semantic change has already taken place in an earlier period, and the observed formal change, namely from separable to inseparable, followed. Overall, our findings support Blom’s (2004, 2005) hypothesis.

## References

- Blom, C. (2004). On the Diachrony of Complex Predicates in Dutch: Predicative and Nonpredicative Preverbs. *Journal of Germanic Linguistics* 16(1), 1–75.
- Blom, C. (2005). *Complex Predicates in Dutch: Synchrony and Diachrony*. Utrecht: Netherlands Graduate School of Linguistics.
- Bouma, G., Coussé, E., Dijkstra, T., & van der Sijs, N. (2020). The EDGeS Diachronic Bible Corpus. In Calzolari, N., Béchet, F., Blache, P., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., & Piperidis, S. (Eds.), *Proceedings of the 12<sup>th</sup> Conference on Language Resources and Evaluation*, (pp. 5232-5239). Marseille: European Language Resources Association.
- Van Olmen, D. (2019). A diachronic corpus study of prenominal zo'n 'so a' in Dutch: Pathways, analogy and (inter)subjectification. *Functions of Language* 26(2), 217-248.

**Deciphering the Variability of Danish *be/for* Applicative Constructions:  
Insights from Corpus Analysis**

Johan Pedersen

University of Copenhagen

This study investigates the use of Danish *be/for*-applicative constructions, aiming to shed light on their variability and type. We adopt a *constructionist approach* (cf. Michaelis & Ruppenhofer 2000; Petré & Cuyckens 2008), which challenges lexical rule explanations. We hypothesize, though, that the internal structure of the constructions is crucial to determine the different types and their variability (cf. Pijpops et al. 2021).

Based on corpus data from *KorpusDK* and employing collocation methodology, we identify three meaning variants, exemplified in (A), (B), and (C) (Pedersen 2023):

- |  |                                      |
|--|--------------------------------------|
| (A) <i>Peter be-snakker den ældre dame</i><br>Peter be-talks the old lady<br>'Peter manipulates the old lady with his talk.' | Core meaning:<br><br>≈ 'X affects Y' |
| (B) <i>Kunstneren be-sjæler sit værk</i><br>The artist be-souls his work<br>'The artist gives soul to his work of art.'      | ≈ 'X "furnishes" Y'                  |
| (C) <i>Moderen be-roliger datteren</i><br>The mother be-calms the daughter<br>'The mother calms the daughter'                | ≈ 'X causes Y to be Z'               |

The A-type may or may not involve argument promotion in the classical sense of the applicative. From our constructionist perspective, the core meaning of the A-type is directly conveyed by the schematic transitive prefix-construction, interacting productively with a variable base verb. Conversely, the core meanings of the B- and C-types – sometimes referred to as *applicative lookalikes* (e.g., Zúñiga & Creissels 2024) – arise from the constructional derivation of the transitive verb. The B-type involves *valence creation* and interaction with a variable base noun ('soul → transferring soul') and the C-type entails *process creation* and interaction with a variable base adjective/noun ('calm → making calm') (cf. Booij 2010).

These structural differences manifest in distinct core features of their creative usage, exemplified by (A<sub>1</sub>)-(C<sub>1</sub>) (Pedersen 2023):

- |  |
|--|
| (A <sub>1</sub> ) <i>Min far på 78 med de dårlige knæ be-dansede alle damerne.</i><br>'My 78 years old father with the bad knees be-danced all the women.'                     |
| (B <sub>1</sub> ) <i>Første bil til at be-hjule dansk grund var den store Buick</i><br>'First car to be-wheele Danish ground was the big Buick.'                               |
| (C <sub>1</sub> ) [...] <i>tilbage i studiet, hvor de har hevet folk ind som kan be-kloge os</i><br>'(we are) back in the studio, where they have people that can be-wise us.' |

While the creative usage in (A<sub>1</sub>) involves a semantic conflict typical of other clause constructions (cf. Goldberg 1995; Michaelis 2003; Pedersen 2019), in which the schematic transitive construction overrides the activity meaning of the base verb, no such conflict occurs in (B<sub>1</sub>) and (C<sub>1</sub>), where unexpected creative effects stem from the transitive verb formation. Consequently, we argue that the A-types are *schema-based constructions*, while the B- and C-types are *verb-based constructions*.

We conclude that most Danish *be-/for-* constructions are schema-based (the A-type), although the verb-based types are also frequently used. While both *be-* and *for-* constructions are attested with all three variants, we observe a distinct semantic profile for each prefix. We will argue that these findings have language typological implications since similar constructions in Spanish and other Romance languages appear to be verb-based types, for instance, parasynthetic verb constructions or specific types of preverb constructions (cf. B, C; Zúñiga et al. 2024: 420).

**Keywords:** Danish applicatives, morphological constructions, variability

## References

- Booij, G. 2010. *Construction Morphology*. Oxford: Oxford University Press.
- Goldberg, A. 1995. *Constructions: A Construction Grammar approach to argument structure*. Chicago: The University of Chicago Press.
- KorpusDK. <https://ordnet.dk/korpusdk>.
- Michaelis, L. A. 2003. Headless constructions and coercion by construction. In E. J. Francis and L. A. Michaelis (eds.), *Mismatch: Form-function incongruity and the architecture of grammar*, pp. 259–310. Stanford, CA: CSLI Publications.
- Michaelis, L., and Ruppenhofer, J. 2000. Valence creation and the German applicative: The inherent semantics of linking patterns. *Journal of Semantics* 17, 335–395.
- Pedersen, J. 2019. Verb-based vs. schema-based constructions and their variability: On the Spanish transitive directed motion construction in a contrastive perspective. *Linguistics*, 57(3), 473–530.
- Pedersen, J. 2023. Danish verb prefixes and the schematizing transitive prefix construction. In Evie Coussé, Steffen Höder, Benjamin Lyngfelt, and Julia Prentice (eds.), *Constructional Approaches to Nordic Languages*, CAL 37, pp. 212–246. Amsterdam: John Benjamins.
- Petré, P., and Cuyckens, H. 2008. Bedusted, yet not beheaded: The role of *be-*'s constructional properties in its conservation. In A. Bergs and G. Diewald (Eds.), *Constructions and Language Change* (pp. 133–169). Berlin: Mouton de Gruyter.
- Pijpops, D., D. Speelman, F. Van de Velde, and S. Grondelaers. 2021. Incorporating the multi-level nature of the constructicon into hypothesis testing. *Cognitive Linguistics*, 32(3): 487–528.
- Zúñiga, F., and D. Creissels. 2024. Applicative constructions: An introductory overview. In F. Zúñiga and D. Creissels (eds.), *Applicative Constructions in the World's Languages*, 3–56. Berlin, Boston: De Gruyter Mouton.
- Zúñiga, F., P. Arkadiev, and V. Hegedűs. 2024. Applicativizing preverbs in selected European languages. In F. Zúñiga and D. Creissels (eds.), *Applicative Constructions in the World's Languages*, pp. 419–471. Berlin, Boston: De Gruyter Mouton.

## From incorporated nouns to (locative) applicatives in Northwest Caucasian

Peter Arkadiev

Freiburg Institute for Advanced Studies

The polysynthetic Northwest Caucasian languages possess rich systems of applicative prefixes (traditionally called “preverbs”) introducing indirect objects indexed by pronominal markers immediately preceding the respective preverb (Arkadiev et al., 2024). These include a large set of locative applicatives that semantically specify the spatial configuration of the situation expressed by the verb (1).

- (1) Abaza (O’Herin, 2001: 481)  
*d-ʕa-hə-ʒqa-jə-r-gəl-t*  
3SG.H.ABS-CSL-**1PL.IO-LOC:beside**-3SG.ERG-CAUS-stand-DCL  
‘He caused him/her to stand **near to us**.’

Locative applicatives vary as to the degree of abstractness of their meanings and their formal and semantic relations to lexical elements. While some preverbs express very broad spatial relations and have no obvious counterparts among lexical roots, others have clear nominal cognates, mostly among body-part nouns (Lomtadze, 1983; Klychev, 1994; Avidzba, 2017; Arkadiev & Maisak, 2018: 121–125). Synchronically, it is possible to distinguish between the following putative stages of development from incorporated nouns to applicatives in Northwest Caucasian:

(i) The incorporated noun always retains its original meaning, and the accompanying pronominal prefix refers to the possessor of the body-part (2); the whole complex expresses a spatial relation.

- (2) Abaza (AbTa 2015: 142)  
*arqan-g’əj ʕ-a-qʷdə-j-χ-χə-n*  
rope-ADD CSL-**3SG.N.IO-LOC:neck**-3SG.M.ERG-take-RE-PST  
‘He took the rope off its (the stallion’s) neck’

(ii) The incorporated noun expresses both the original meaning (3a) and an abstract spatial meaning (e.g. ‘on one’s head’ > ‘above’), in the latter case the accompanying pronominal prefix indexes the spatial landmark (3b).

- (3) Abaza (Klychev, 1995: 260, 258)  
a *j-qəlpa j-qa-š-əw-n*  
3SG.M.PR-hat **3SG.M.IO-LOC:head**-fall-IPF-PST  
‘The hat fit his head well (lit. fell **on his head**).’  
b. *a-č’kʷən a-ʒəχ’ d-a-qa-č-t*  
DEF-youth DEF-spring 3SG.H.ABS-**3SG.N.IO-LOC:above**-sleep-DCL  
‘The guy fell asleep **over the spring of water**.’

(iii) The preverb expresses only an abstract spatial meaning and retains only an etymological connection to the noun (4).

- (4) West Circassian (Kumakhov, 1964: 177–179)  
*kʷeçə* ‘intestines’: *kʷeçə-λhe-n* ‘put inside smth’  
*ʔwə* ‘mouth’: *ʔwə-cʷe-n* ‘stand near smth’

Stages (i) and (ii) are particularly characteristic of incorporated body-part nouns in Abaza and Abkhaz, but a few preverbs in Circassian also belong here (5).

- (5) West Circassian (Rogava & Kerasheva, 1966: 131)  
*gʷəʃʷe-kʷəwe-maqe-r qə-že-de-zə-κ*  
 joy-cry-voice-ABS CSL-LOC:mouth-LOC:in-exit-PST  
 Lit. ‘A joyful scream went out of his mouth.’

Most preverbs of Circassian languages exemplify stage (iii). Further developments, starting off at stages (ii) or (iii), lead to non-spatial functions such as e.g. the Circassian ‘spike’ > malefactive (6) or Circassian ‘under someone’s hand’ > involuntary agent (7).

West Circassian

- (6) a. *čʻale-r pχeʔebzane-m ʃʷe-na-κ*  
 boy-ABS rake-OBL LOC:spike-remain-PST  
 ‘The boy ran into a rake.’ (Mazurova, 2009: 438)
- b. *nepsə-r qə-s-ʃʷa-kʷe*  
 tear-ABS CSL-1SG.IO-MAL-go  
 ‘Tears appear against my will.’ (Arkadiev et al., 2024: 886)
- (7) a. *senefə-bzə-r qə-ʔe-čʻ-e-zə*  
 wine-horn-ABS CSL-hand-LOC:under-DYN-fall  
 ‘The horn with wine falls from his hand.’ (Arkadiev et al., 2024: 899)
- b. *s-jə-šʻewekʷə-r s-ʔečʻe-wəʔa-κ*  
 1SG.PR-POSS-friend-ABS 1SG.IO-INADV-injure-PST  
 ‘I unintentionally injured my friend.’ (Arkadiev et al., 2024: 898)

The Northwest Caucasian languages thus present a clear case of development of applicatives from nouns incorporated as spatial modifiers of the verb and provide evidence of several stages of this historical scenario, complementing the findings of Nordlinger (2019) and Van linden (2022).

## References

- AbTa 2015: *Абза турыхква* [Abaza Tales]. Mineral'nye vody: Alashara.
- Arkadiev, P., Lander, Y., & Bagirokova, I. (2024). Applicative constructions in the Northwest Caucasian languages. In: F. Zúñiga, & D. Creissels (Eds.), *Applicative Constructions in the World's Languages* (pp. 869–912). Berlin, Boston: De Gruyter Mouton.
- Arkadiev, P., & Maisak, T. (2018). Grammaticalization in the North Caucasian languages. In: H. Narrog, & B. Heine (Eds.), *Grammaticalization from a typological perspective* (pp. 116–145). Oxford: Oxford University Press.
- Avidzba, A. V. (2017). *Lokal'nye preverby v abxazskom i abazinskom jazykax* [Locative preverbs in Abkhaz and Abaza]. PhD Dissertation, The D. I. Gulia Abkhazian Institute for Research in the Humanities, Sukhum.
- Klychev, R. N. (1994). *Lokal'no-preverbnoe obrazovanie glagolov abazinskogo jazyka*. [The locative preverbal derivation of verbs in Abaza]. Cherkessk: Adž'pa.



- Klychev, R. N. (1995). *Slovar' sočetaemosti lokal'nyx preverbov s suffiksoidami i glagol'nymi kornjami v abazinskom jazyke* [The collocational dictionary of locative preverbs with suffixoids and verbal roots in Abaza]. Cherkessk: Karačaevo-čerkeksskoe knižnoe izdatel'stvo.
- Kumakhov, M. A. (1964). *Morfologija adygskix jazykov. Sinxronno-diaxronnaja xarakteristika. I. Vvedenie, struktura slova, slovoobrazovanie častej reči* [Morphology of Circassian languages. A synchronic and diachronic characteristic. I. Introduction, word structure, derivation of parts of speech]. Nal'čik: Kabardino-balkarskoe knižnoe izdatel'stvo.
- Lomtadze, K. V. 1983. Osnovnye typy lokal'nyx preverbov v abxazskom i abazinskom jazykax [The basic types of locative preverbs in Abkhaz and Abaza]. In N. T. Tabulova, & R. X. Temirova (Eds.), *Sistema preverbov i poslelogov v iberijsko-kavkazskix jazykax* [The system of preverbs and postpositions in Ibero-Caucasian languages] (pp. 10–13). Cherkessk: Karachaevo-Cherkesskij NII istorii, filologii i èkonomiki.
- Mazurova, J. V. (2009). Semantika lokativnyx preverbov *pə-* i *š<sup>w</sup>e-* [Semantics of the locative preverbs *pə-* i *š<sup>w</sup>e-*]. In: J. G. Testelec (Ed.), *Aspekty polisintetizma: Očerki po grammatike adygejskogo jazyka* [Aspects of polysynthesis: Essays in the grammar of West Circassian] (pp. 429–453). Moscow: Russian State University for the Humanities.
- Nordlinger, R. (2019). From body part to applicative: Encoding 'source' in Murrinhpatha. *Linguistic Typology*, 23(3), 401–433.
- O'Herin, B. (2001). Abaza applicatives. *Language*, 77(3), 477–493.
- Rogava, G. V., & Kerasheva, Z. I. (1966). *Grammatika adygejskogo literaturnogo jazyka* [A Grammar of Standard Adyghe]. Krasnodar, Majkop: Adygejskoe knižnoje izdatel'stvo.
- Van linden, A. (2022). Spatial prefixes as applicatives in Harakmbut. In: S. Pacchiarotti, & F. Zúñiga (eds.), *Applicative morphology: Neglected syntactic and non-syntactic functions* (pp. 129–159). Berlin, Boston: Mouton de Gruyter.

Keywords: applicatives, noun incorporation, grammaticalisation, spatial morphology, Northwest Caucasian languages

N. Mezane KONUK  
 ATER - Université Sorbonne Nouvelle  
 LACITO – CNRS  
[mezanekonuk@gmail.com](mailto:mezanekonuk@gmail.com)

### Locative applicatives in Abzakh: a clear diachronic origin, a wealth of divergent synchronic patterns

Abzakh (North-West Caucasian) has a verbal structure with seventeen applicative markers, no less than thirteen of which are locatives. These constructions are defined as non-direct applicatives (Zúñiga & Kittilä 2019, 58). Even though they do not promote an argument to direct object but to indirect object position, the applied argument is obligatorily indexed in the verb (Letuchiy 2012, 337), promoting a non-core argument to the core argument status (Konuk 2022; 2023).

Applicatives are highly productive in North-West Caucasian languages. They are used with both transitive and intransitive roots and are highly grammaticalized. However, they assume a lexical derivational function as well: adding a locative applicative marker to a root yields a static verb, i.e. *jəʔən* “to have” (*jə-* “inside” + *ʔən* “hand” + *-n* “masdar”).

Example (1) (elicited):

zəj fəpx<sup>w</sup> s-jə-ʔən-p  
 none sister 1SG.OBL.A-LOC<sub>INSIDE</sub>-hand-NEG  
 “I don’t have any sisters.” (litt: *in my hand*)

A contribution of the present work, based on a corpus collected on Abzakh dialect spoken in the diaspora in Turkey (the corpus is gradually being made available online at Pangloss Collection, open archive of LACITO-CNRS), is the observation that this lexical derivational function is attested in some dynamic verbs as well, i.e. *to go/come out* -> *to publish*:

Example (2):

ḏʒə fəʔən txət-xən-r qə-ʔən-d-ʔən-tʃə-ʔən  
 DEM like book-PL-ABS/DEF CIS-LOC<sub>UNDER</sub>-1PL.OBL.C<sub>EUR</sub>-CAUS-come\_out-PERF  
 “That is how we published (litt: *made-come-out-beneath*) the books.”

In example (2), the root *come out* is an intransitive verb. It is marked by the causative marker -*ʔən* and thus, the object of the verb, *books*, is marked by the absolutive case. However, the locative applicative marker *ʔən* does not refer to any argument in the sentence, which clearly shows that this locative applicative assumes a lexical derivational function.

Though locative applicative markers are not obligatory with dynamic verbs, the use of at least one such marker is mandatory for static verbs, i.e. *to be standing*, *to be sitting*, etc. From a diachronic point of view, it is clear that some of them have been grammaticalized from a nominal root, especially from body parts: e.g. *hand* > *under*; *nose* > *at the edge (of sth)*. However, not all locative applicatives allow for such a straightforward diachronic analysis. Arkadiev et al. (2024) propose that the grammaticalized applicatives (comitative, benefactive, malefactive and instrumental) originate from locative applicatives. This insightful generalization highlights the general observation that locative applicative markers in Abzakh

(and in Circassian in general) are highly dynamic. They constitute a promising research area for typologists studying the expansion of the applicatives' functions.

In a nutshell, the proposed account brings out the remarkable expansion of the grammatical derivational function of body part nouns, reaching a lexical-derivational function after and, in a sense, *through* grammaticalization as applicative markers.

**Key words:** Locative applicatives, non-direct applicatives, Abzakh, Circassian, Adyghe.

### References

- Arkadiev, Peter, Yury Lander, et Irina Bagirokova. 2024. « Applicative constructions in the Northwest Caucasian languages ». In *Applicative Constructions in the World's Languages*, édité par Denis Creissels et Fernando Zúñiga, 869-912.
- Konuk, Mezane. 2022. « Description grammaticale de l'abzakh - tcherkesse occidentale ». Université Sorbonne Nouvelle-Paris 3.
- . 2023. « Applicative Constructions in Abzakh - a West Circassian Dialect [Bati Çerkesçesinin Abzah Lehçesinde Aplikatif Yapılar] ». *Journal of Caucasian Studies* 8 (14): 83-108. <https://doi.org/10.21488/jocas.1280744>.
- Letuchiy, Alexander. 2012. « Ergativity in the Adyghe system of valencychanging derivations ». In *Ergativity, Valency and Voice*, édité par Gilles Authier et Katharina Haude. Berlin, Boston: DE GRUYTER. <https://doi.org/10.1515/9783110227734.323>.
- Zúñiga, Fernando, et Seppo Kittilä. 2019. *Grammatical voice*. Cambridge University Press.

## Applicative uses of directional suffixes: a case study

MONTÉBRAN Aurore

(LLACAN - CNRS, INALCO, EPHE)

Mankanya (knf, mank1251) is an Atlantic language (Niger-Congo phylum), originally spoken in Guinea-Bissau, nowadays also spread in Senegal, Gambia; and diaspora. There are two main varieties of Mankanya, we will focus on the main one: the Hula from Bula. Atlantic languages are agglutinative, is a non-tonal language, with a quite strict SVO word order.

This talk aims to provide a description of the Mankanyan verbal polyfunctional and polysemic suffixes *-ar* ~ *-ir* ~ *-ad* (the allomorph is selected following the form of the verbal stem) and *-na* ~ *-:na* (the second allomorph occurring after a vowel).

Synchronically, the suffix *-ar* ~ *-ir* ~ *-ad* occurs with motion verbs whose stems express motion of the subject; the derived verb expresses motion of the subject with regards to the object (goal), with a nuance of purpose, the suffix has a directional meaning.

- (1) *Aposar du katoh*  
a-    poş    **-ar**    d-    -u    ka-    toh  
3SG-    walk    **-DIR**    EXT-    -DIST    CL3SG-    house  
“He is walking towards the house.”

(Gaved 2020: 146)

In addition, this suffix is also used for benefactive (only with transitive verbs, without subject restrictions) and for reciprocal (only with plural subjects). Both of these uses change the valency of the verb, increasing it for the benefactive, reducing it for the reciprocal.

- (2a) *Şompi alemp*  
Şompi    a-    lemp  
PN        3SG-    work  
“Shompi works.”
- (2b) *Şompi alempar Naala*  
Şompi    a-    lemp    **-ar**    Naala  
PN        3SG-    work    **-BEN**    PN  
“Shompi works for Naala.”

(Gaved 2020: 71)

- (3a) *p-ηal* “to love”  
(3b) *p-ηal-ad* “love one another”  
(Gaved 2020: 71)

The study of this first suffix gives an example of typical grammaticalization path of spatial morphemes: directional (allative, pure spatial use with motion verbs) > benefactive (metaphoric use with non-motion verbs) (Kuteva et al. 2019). A third step has been added to the chain: benefactive > reciprocal. Through this case study, we will analyse the evolution of the suffix, including its historical development, i.e its grammaticalization pathway, and its semantic changes.

The suffix *-na ~ -:na* is not so easily analyzable. Synchronically, *-na ~ -:na* is an applicative marker with an instrumental meaning (it is worth adding that *na* also means “with”), but with motion verbs it also has the meaning of path (“through”).

(4) *p-de* “to eat” > *a-de-ena kataam* “he eats with a spoon”  
(Gaved 2020: 74)

(5) *p-neej* “to enter” > *a-neej-na p-lëman* “he entered through the door”  
(Gaved 2020: 75)

Diachronically, we assume that the suffix *-na ~ -:na* is an ablative marker, which nowadays only occurs in lexical frozen motion verbs.

(6a) *p-pën* “to go out” > *p-pën-na* “to go out of”  
(6b) *p-wo* “to be” > *p-wo-na* “to be from”

Considering the unidirectionality principle of grammaticalization, we will analyse the language-internal processes leading to the development of these suffixes, using two data collection strategies: gathering data from grammars (Trifkovič 1969, Gaved 2020) and collecting data with dedicated tools such as the Questionnaire of the Berlin-Utrecht Reciprocals Survey (Dimitriadis 2008).

## References

- Dimitriadis A., 2008. *Berlin – Utrecht Reciprocals Survey: Marker Inventory & Overview*. (Version:10 January 2008).
- Gaved T., 2020. *A grammar of Mankanya: An Atlantic language of Guinea-Bissau, Senegal and the Gambia*. PhD Thesis, University of Leiden, The Netherlands.
- Kari E. E., 2016. “Restrictions in the attachment of verbal extensions in Degema”, in *Studies in Linguistics*, Vol. 39.
- Kuteva T., Heine B., Hong B., Long H., Narrog H., & Rhee S., 2019. *World Lexicon of Grammaticalization* (2nd ed.). Cambridge University Press, Cambridge.
- Mithun M., 2000. “Valency-changing derivation in Central Alaskan Yup'ik”. In R. Dixon & A. Aikhenvald (Eds.), *Changing Valency: Case Studies in Transitivity*. Cambridge University Press, Cambridge.
- Trifkovič M., 1969. *Le mancagne. Etude phonologique et morphologique*. Initiations et études africaines XXVI. IFAN, Dakar, Senegal.

## Abbreviations

BEN: benefactive; CL: nominal classifier; DIST: distance DIR: directional; EXT: external; PN: proper noun; SG: singular.

## Spatial origin of the Ainu applicative prefix *e-* and the grammaticalization mechanism

As known since Kindaichi (1993 [1931]: 264-276), Ainu (moribund, isolate; Japan) has three applicative markers, today mainly found in oral literature. The prefix *e-* is one of them and introduces semantic roles including location, goal, instrument and theme. Kindaichi (1993 [1931]: 264) postulated that the prefix originates in the element *e-* ‘there’ (noun-adverb), as in *e-un hosari* (there-at look\_back) ‘He looked back there.’ (p. 181) and *e-un a!* (there-at sit) ‘Sit there!’ (p. 264). Another proposed origin of it is the bound noun *e-* (< *he-*) ‘head’, which, according to Bugaeva (2010: 762), is attributed to Hiroshi Nakagawa. The latter scenario seems somewhat radical and I consider *e-* ‘there’ the immediate origin of the applicative *e-*. Meanwhile, Kindaichi did not provide any specific support for his own hypothesis. Thus, the present study aims to elaborate its validity in light of syntactic properties of *e-* applicatives, through analysis of oral literature (epics and old tales), in a functional-typological framework.

I suppose that the original construction of *e-* applicatives with an explicit applied argument was like the following (as connoted in Kindaichi 1993 [1931]: 264):

- (1) 

spatial element	<i>e(-)</i> ‘there’	verb
-----------------	---------------------	------

and that in (1) the spatial element (typically, a nominal or postpositional phrase) and *e(-)* are apposed, the latter an anaphor referring to the former. Thus, for example, the literal meaning of (2) is ‘at the side of the treasure, there, I sit.’:

- (2) *ikit        tukari a-e-horari*  
 treasure side I-APPL-sit  
 ‘I sit at the side of the treasure.’

(Kannari & Kindaichi 1964: 305)

As a syntactic property, *e-* can realize “nondirect applicatives” (Zúñiga & Kittilä 2019: 58), as manifested in epic texts. The explanation I suppose is: As long as its referent is shared with *e-*, the spatial element may have any syntactic status, like in:

(3) *shiso-sam ne e-horari*

right-side at APPL-sit

‘He sits down at the side of the right seat.’

(Kannari & Kindaichi 1963: 137)

Also, as Bugaeva (2010: 755) and oral literature texts suggest, implicit applied arguments (zero anaphora) are common for *e-*. The explanation I suppose is: As the fundamental meaning of *e-* itself is ‘there’, its referent can be identified in a previous sentence or through spatial deixis, instead of through apposition within the same clause, like in:

(4) *kapar-pe kasa kamuiranke-tam kani-uwokkut kani*

thin-NMLZ hat godgiven-sword golden-belt cotton

*kosonte a-ko-ebittekka atusa numi a-e-tursere*

cloak I-APPL-tear\_off naked stature I-APPL-fall

‘I tear off his thin hat, god-given sword, golden belt and cotton cloak, and knocked his naked body down (there).’

(Kannari & Kindaichi 1963: 65)

In such cases, it may be ambiguous whether *e-* is an applicative marker with an implicit applied argument or a spatial prefix (‘there’).

Implicit applied arguments are even common with non-locative semantic roles. It thus seems that *e-* retained well its original syntactic category, adverb, during its semantic extension:

(5) *e-atusi-kara-chi-hi neampe taa*

APPL-tie-make-they-NMLZ when at

‘After they tied up (the crow) (with the rope),’

(Asai & Murasaki 2001)



## References

- Asai, Take (teller) and Kyōko Murasaki (edition and Japanese translation) (2001), *Kaitei Asai Take mukashibanashi zenshū* [Complete tales by Take Asai revised] (“Kantaiheiyō-no gengo” chōsa hōkokusho [Research report about “languages of pacific rim”], 1346-082X A2-010). Suita: Osaka Gakuin University.
- Bugaeva, Anna (2010), “Ainu applicatives in typological perspective”. *Studies in Language* 34 (4): 749-801.
- Kannari Matsu (original texts) & Kyōsuke Kindaichi (Japanese translation and notes) (1963), *Ainu jojishi yūkara shū* [Collection of Ainu epic Yukar] Vol. 3. Tōkyō: Sanseidō.
- Kannari Matsu (original texts) & Kyōsuke Kindaichi (Japanese translation and notes) (1964), *Ainu jojishi yūkara shū* [Collection of Ainu epic Yukar] Vol. 4. Tōkyō: Sanseidō.
- Kindaichi, Kyōsuke (1993 [1931]), *Ainugo 1* [Ainu 1]. (Kindaichi Kyōsuke zenshū [Complete works of Kyōsuke Kindaichi]; 5). Tōkyō: Sanseidō.
- Zúñiga, Fernando & Seppo Kittilä (2019), *Grammatical voice* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

## Sleeping in the cave: an incipient locative applicative in Dalabon

Abstract submitted for workshop

‘Applicative uses of spatial markers: typological and diachronic perspectives’ (Dana Louangie & Dirk Pijpops, organisers), Sorbonne, 9-11 Sept 2024

In this talk I will discuss the emergence of a locative applicative in Dalabon (Gunwinyguan; Australian) to take its place among three other applicatives (benefactive, comitative and instrumental) in a suite of four, each with subtly different morphosyntactic effects on valency change (or not), interaction with reflexive/reciprocal formation, and the relation of the applicative to its etymological sources. I will also contextualise the evolution and morphosyntactic properties of the Dalabon applicatives against that of applicative morphemes in other Gunwinyguan languages with related applicative sets, including Bininj Kunwok, Kunbarlang, Rembarrnga and Ngalakgan.

To understand how applicatives evolve, a crucial step is finding bridging cases where they are on the cusp of reanalysis, still at a stage where they could either be analysed as applicatives or as some other construction from which they are derived. In this talk I will examine one such ‘smoking gun’, by looking at the diachronic origins of the incipient locative applicative *kun* in Dalabon, exemplified in:

*Bala-h-djarrk-dun-kun-ni,*                      *bala-h-djarrk-dun-kun-daddi.*  
3pl.S-ASSERT-together-cave-LOC-sit:PR    3pl.S-ASSERT-together-cave-LOC-be.inside:PR  
‘They are all together in the cave, they are all together inside the cave.’

The diachronic origins of this incipient locative applicative are as a genitive/purposive suffix on NPs, though one encoding rather a wide range of functions both in Dalabon and in the closely-related Bininj Kunwok, including genitive, purposive, gentilic, quantity and manner uses. As with the proprietive/instrumental applicative in Bininj Kunwok and Dalabon, its applicative use appears to have developed through a reanalysis by which an incorporated verb-internal sequence [N-CASE.SUFFIX]<sub>incorp</sub>-V was reanalysed as [N<sub>incorp</sub>-[APPLIC.PREFIX-V]]. With the comitative/instrumental, a further stage allowed the omission of the incorporated noun, to give just APPLIC.PREFIX-V, but with the incipient locative applicative this last stage has not been attested and it only occurs when it follows a noun.

Other differences indicate that the incipient locative applicative is not yet on a par with the others. The benefactive and comitative applicatives both increase valency, as indicated both by the pronominal prefix series (which index subject, and object in a broad sense) and by the possibility of feeding the outcome of applicativisation into the valency-reducing operations of reciprocalisation and reflexivisation. This does not happen with the incipient locative applicative: the relevant verbs remain intransitive. This is one of the reasons for considering it incipient rather than as a true applicative, since in contrast to all the other applicatives there is no morphosyntactic evidence that the entity it introduces into the verb is a core argument. On the other hand, constructional conventionalisation has already occurred: the locative meaning conveyed in this construction is not available when the noun is not incorporated, so that *dun-kun* can only mean ‘of the cave’, not ‘in the cave’, this latter requiring the locative case to express it, *dun-kah* [cave-LOC]. I will explore possible routes by which this development occurred, drawing also on Bininj Kunwok data which shows some marginal but related uses of the cognate suffix, but without anything comparable to the Dalabon incipient locative applicative.



**Langues & Langage  
à la croisée des Disciplines**

**1ère Rencontre annuelle LLcD**

Colloque : 9-11 septembre, 2024

Ecole d'été : 12-14 septembre, 2024

Sorbonne Université - 28 Rue Serpente, 75006 Paris, France



## **Workshop 3**

**Les images dans le discours numérique :  
enjeux, formes, intégration, fonctions**

**Images in computer-mediated discourse:  
issues, forms, functions, integration**

## **La fabrication des émojis : un langage universel, entre technique et politique**

**Michel Marcoccia**

*Université de technologie de Troyes / LIST3N (Lab. Informatique & Société Numérique)*

[michel.marcoccia@utt.fr](mailto:michel.marcoccia@utt.fr)

De nombreux travaux sont consacrés aux émojis : à leur fonction (proche du non verbal dans la communication en face à face ; Evans, 2017), à leur signification et interprétation (par exemple Danesi 2016, Herring et al., 2018, Zappavigna & Logi, 2024 ) ou à leur standardisation (Seargeant, 2019).

Cette présentation sera consacrée à un aspect particulier : le processus de fabrication des émojis et les enjeux techniques et sociolinguistiques qu'il soulève.

En adoptant une approche « techno-socio-linguistique », nous décrivons le processus technique et institutionnel de création et de validation des émojis par le consortium Unicode (Berard, 2018) et les critères de sélection qu'il a choisis.

Puis, à travers l'étude de divers exemples de propositions de nouveaux émojis – validées ou rejetées par le consortium Unicode - nous analyserons les enjeux sémiolinguistiques et politiques qu'implique ce processus.

A la différence des émoticônes que les utilisateurs composaient eux-même en combinant divers signes de ponctuation, par exemple, les émojis utilisés de nos jours sont « préfabriqués ». Ils sont produits en deux étapes. Des nouveaux émojis sont d'abord proposés par des individus ou des groupes d'intérêt au consortium Unicode, qui les valident ou les rejettent. Le consortium Unicode a pour objectif de standardiser les caractères d'écriture du monde entier pour les rendre lisibles sur n'importe quel terminal (ordinateur ou smartphone), de n'importe quelle marque et avec n'importe quel système d'exploitation. Cette nécessaire standardisation technique des émojis soulève de nombreux enjeux sémiolinguistiques et politiques.

D'un point de vue sémiolinguistique, il est intéressant de noter que les émojis Unicode sont supposés correspondre à un langage universel et être interprétables par tout individu (c'est un critère important dans la procédure de validation), mais que de nombreux exemples montrent la difficulté d'atteindre cet objectif. Ainsi, de nombreux émojis ont donné lieu à des débats qui peuvent porter sur le signifiant, le symbole choisi. Par exemple, en août 2016, la modification de l'emoji revolver (remplacé par un pistolet à eau) a provoqué la colère de la communauté pro-armes aux Etats-Unis. Des débats peuvent aussi porter sur la signification des émojis. Par exemple, l'utilisation de l'emoji « OK » est problématique depuis que ce symbole est aussi utilisé pour signifier son appartenance au mouvement White Power.

Les enjeux sont aussi politiques. De nombreux combats, moyens d'identification ou d'appartenance passent par les émojis. Divers exemples seront abordés : la reconnaissance du genre neutre, la visibilité des femmes voilées, de la culture afro ou des cheveux roux, les enjeux identitaires et territoriaux (avec les débats sur les émojis-drapeaux corse ou québécois, par exemple), la promotion du véganisme (avec la disparition de l'œuf dans l'emoji « salade »), le débat sur la standardisation culturelle (avec l'emoji « pizza », trop américain aux yeux des Italiens), etc.

Ainsi, on voit que la fabrication des émojis recouvre de nombreuses problématiques sociolinguistiques et politiques : la question de la standardisation culturelle dissimulée en universalité, la représentativité des communautés, le pouvoir performatif du langage (l'emoji rend visible la communauté qu'il symbolise) et, corrélativement, le pouvoir de l'institution qui contrôle ce langage.

- Berard, B. (2018). I second that emoji: The standards, structures, and social production of emoji. *First Monday*, 23(9). <https://doi.org/10.5210/fm.v23i9.9381>
- Danesi, M. (2016), *The Semiotics of Emoji*, London, Bloomsbury Academics.
- Evans, V. (2017). *The Emoji Code*, New York, Picador.
- Herring, S.C. & Dainas, A. (2018). Receiver Interpretation of Emoji Functions: A Gender Perspective, *Proceedings of the 1<sup>st</sup> international Workshop on Emoji Understanding and Applications in Social Media (Emoji2018)*, Stanford, California, USA, 2018.
- Seargeant, P. (2019). *The emoji Revolution*, Cambridge, Cambridge University Press.
- Zappavigna, M. & Logi, L. (2024). *Emoji and Social Media Paralanguage*, Cambridge, Cambridge University Press.

## Intégration des émojis à la syntaxe : des logogrammes dans l'écriture du français ?

Agathe Cormier

*UPEC-Céditec*

Alors que les emplois des émojis comme émoticônes (Wilson, 1993 ; Mourlhon-Dallies & Colin, 1995 ; Halté, 2016a, 2016b, 2018, 2019) et leur rôle dans la segmentation et la ponctuation des énoncés (Halté, 2017 ; Magué et al., 2020) sont désormais bien décrits, les cas d'intégration de ces signes à la chaîne syntaxique, où ils se substituent à un syntagme, ont été relevés (Magué et al., 2020) mais les études systématiques sur corpus d'énoncés attestés restent encore peu nombreuses (voir par exemple Dürscheid & Siever, 2017 ; Schneebeli, 2018 ; Escouflaire, 2021).

C'est précisément sur les cas d'émojis occupant la position syntaxique d'un nom, d'un verbe voire d'une proposition que portera notre communication. Il s'agira ainsi d'interroger les ressources de la langue écrite dans des textes qui mêlent verbal et non-verbal, et de se demander quel statut accorder à ces signes dans le système d'écriture du français.

Notre approche sémiolinguistique (voir Cormier, 2023) articule l'autonomisme d'Anis (Anis, 1988), qui propose d'aborder l'écrit indépendamment de l'oral, et la perspective intégrationnelle de Harris (Harris, 1994), dont les principes d'analyse syntagmatique permettent d'isoler les unités minimales de la langue écrite.

Cette étude exploratoire porte sur un échantillon d'énoncés extrait de la base de données du projet ANR Amulex (« Analyse multiplateforme de l'expression politique en ligne »), projet lancé en janvier 2024 qui vise notamment à saisir la critique du politique dans les commentaires suscités par des interviews de candidats à l'élection présidentielle française de 2022 sur Instagram, Twitch, Twitter et YouTube.

Nous analysons d'une part la distribution syntaxique des émojis dans les énoncés sélectionnés, et d'autre part leur signification au regard des substitutions possibles de séquences de lettres. L'analyse sémantique est par ailleurs complétée par la comparaison avec des énoncés où les émojis considérés intensifient de manière redondante l'expression verbale. Ainsi, l'énoncé *Les médias sont à 🤮* peut être glosé par « les médias sont à vomir » ou « les médias sont à gerber », ce que suggèrent d'autres énoncés du corpus comme *Quel crevard ce Jadot à gerber 🤮* ou *Cette France est à vomir 🤮*.

Pour rendre compte de ces emplois des émojis, nous partons des deux hypothèses complémentaires suivantes :

- que les émojis représentent l'incursion de logogrammes dans l'écriture alphabétique du français et qu'ils intègrent à ce titre le lexique du français écrit,
- qu'un emoji redondant par rapport à un syntagme tend à hériter de son sens et à pouvoir se substituer à ce syntagme dans une phrase.

Les premiers résultats de l'analyse montrent que ce phénomène émerge de substitution d'émojis à des syntagmes s'applique à une typologie réduite d'émojis dont la fréquence semble stabiliser une signification conventionnelle. La question se pose dès lors du statut linguistique de ces signes écrits : faut-il les considérer comme de simples pictogrammes à valeur iconique ou comme de véritables logogrammes possédant un sens conventionnel en langue ?

Mots clés : émoji, logogramme, système d'écriture, intégration syntaxique

### Références bibliographiques

Anis, J. (1988). *L'écriture : Théories et descriptions*. De Boeck Université.

Cormier, A. (2023). Essai de typologie des paramètres signifiants du support. *Linguistique de l'écrit*, 4, 67-108. <https://linguistique-ecrit.org/pub-265977>

Dürscheid, C., & Siever, C. M. (2017). Jenseits des Alphabets – Kommunikation mit Emojis. *Zeitschrift für germanistische Linguistik*, 45(2), 256-285. <https://doi.org/10.1515/zgl-2017-0013>

Escoufflaire, L. (2021). Signaling irony, displaying politeness, replacing words : The eight linguistic functions of emoji in computer-mediated discourse. *Linguisticae Investigationes*, 44(2), 204-235. <https://doi.org/10.1075/li.00062.esc>

Halté, P. (2018). *Les émoticônes et les interjections dans le tchat*. <http://www.lambert-lucas.com/livre/les-émoticônes-et-les-interjections-dans-le-tchat/>

Halté, P. (2016a). Émoticônes et modalisation dans un corpus d'enseignement par t'chat. *Éla. Études de linguistique appliquée*, 184(4), 441-453. <https://doi.org/10.3917/ela.184.0441>

Halté, P. (2016b). Enjeux pragmatiques et sémiotiques de l'étude des émoticônes. *Réseaux*, 197-198(3), 227-252. <https://doi.org/10.3917/res.197.0227>

Halté, P. (2017). Positionnement syntaxique des interjections et des émoticônes : Modalisation, portée, visée. *Cahiers de praxématique*, 69. <https://doi.org/10.4000/praxématique.4680>

Halté, P. (2019). Emojis, émoticônes, smileys ? Proposition de classement terminologique selon des critères sémiotiques et énonciatifs. *Interfaces numériques*, 8(2), 365-385. <https://doi.org/10.25965/interfaces-numeriques.3956>

Harris, R. (1994). *La sémiologie de l'écriture*. CNRS Éditions.

Magué, J.-P., Rossi-Gensane, N., & Halté, P. (2020). De la segmentation dans les tweets : Signes de ponctuation, connecteurs, émoticônes et émojis. *Corpus*, 20. <https://doi.org/10.4000/corpus.4619>

Mourlhon-Dallies, F., & Colin, J.-Y. (1995). Les rituels énonciatifs des réseaux informatiques entre scientifiques. *Les carnets du Cediscor*, 3, 161-172. <https://doi.org/10.4000/cediscor.530>

Schneebeli, C. (2018). Les modalités iconiques dans le discours médié par ordinateur : Du neuf dans l'interaction ? *Études de stylistique anglaise*, 13, 45-66. <https://doi.org/10.4000/esa.3126>

Wilson, A. (1993). A pragmatic device in electronic communication. *Journal of Pragmatics*, 19(4), 389-392. [https://doi.org/10.1016/0378-2166\(93\)90098-A](https://doi.org/10.1016/0378-2166(93)90098-A)



## **Émoticônes et stratégies conversationnelles en contexte de mobilité franco-allemande**

Charlotte Rep

*UMR-EFTS, Université Toulouse Jean Jaurès, INSPE TOP*

**Mots clés :** émoticônes - plurilinguisme - multimodalité

### **Abstract :**

Notre recherche porte sur l'emploi d'émoticônes en contexte de mobilité scolaire franco-allemande pour des élèves de 14-15 ans, en fin de cycle 4, de niveau A2 à B1.

L'évolution des pratiques de communication numérique vers un usage des littératies médiatiques multimodales (Lacelle, 2017) constitue un changement de paradigme de la situation de communication qui « comporte désormais l'utilisation de dispositifs dont la fonction est de transmettre des signes ou des images, et qui créent une nouvelle dynamique sociale des échanges » (Fischer, 2020). La multimodalité permet ainsi de pallier « l'absence d'un espace commun de perception, la référence à la situation de l'émetteur et/ou à la situation de communication au moyen de pratiques nouvelles » (Pappert, 2017). Les distances et les frontières se trouvent modifiées par les espaces multimodaux à travers lesquels la présence de l'autre est incarnée par le biais de plusieurs modes sémiotiques qui font « co-exister des éléments graphiques, iconiques et picturaux, statiques ou mobiles, et des fonctions de communication synchrones et asynchrones » (Develotte & Paveau, 2017). A ce contexte numérique s'ajoute un contexte européen de politiques linguistiques qui entendent renforcer la mobilité des élèves afin de développer leurs compétences linguistiques mais aussi interculturelles, comme le préconise le Cadre européen commun de référence pour les langues (2018), les dernières consistant en des « aptitudes qui facilitent les interactions efficaces et appropriées » (Dollander, 2022). Nous considérons ainsi un système dans lequel, à la complexité de l'écosystème numérique tel que l'analyse Paveau (2017), s'ajoute la complexité du contexte plurilingue de la mobilité scolaire.

Nous inscrivons cette recherche dans notre travail doctoral en cours qui ambitionne de considérer l'ensemble des interactions dans lesquels les usages des émoticônes dépassent une visée esthétique (Navé, 2021) et mobilisent les différents modes et dimensions du langage, selon une approche interdisciplinaire. Dans la mesure où des échanges hors du cadre scolaire s'établissent entre les correspondants, où l'écriture s'est muée en écriture (Liénard, 2012), nous

interrogeons à travers les émoticônes le rapport entre les langues et cultures, et les pratiques langagières.

Halté (2020) définit comme stratégie conversationnelle « la visée d'un locuteur quant au résultat d'une de ses interventions au sein d'un échange ». La question que nous posons est alors la suivante : Dans quelle mesure les emplois d'émoticônes en contexte de communication extrascolaire plurilingue relèvent de stratégies conversationnelles de la part des élèves et s'inscrivent dans le champ des compétences interculturelles ? Nous avons ainsi exploité des données qualitatives issues d'une étude exploratoire, constituées - dans le cadre de notre communication- de chats entre correspondants français et allemands. Nous avançons l'hypothèse que les émoticônes constituent un outil des stratégies conversationnelles des élèves observés dans ce contexte de communication plurilingue franco-allemand, dans la mesure où ils facilitent la compréhension de l'interlocuteur, neutralisent les possibilités d'interprétations (Durr, 2016 ; Halté, 2018 ; Weise, 2021), affirment ou créent une identité numérique (Amossy, 2002) , et témoignent également de la prise en compte de l'altérité, dans le sens d'Abdallah-Preteille (1997).

## **Bibliographie**

Abdallah-Preteille, M. (1997). Pour une éducation à l'altérité. *Revue des sciences de l'éducation*, 23(1), 123–132. <https://doi.org/10.7202/031907ar>

Amossy R. (2002), « Ethos », In Chteau P. & Maingueneau D. (Dir), *Dictionnaire d'Analyse du Discours*, Paris, Seuil, p.238-240

Develotte, C. & Paveau, M. (2017). Pratiques discursives et interactionnelles en contexte numérique. Questionnements linguistiques. *Langage et société*, 160-161, 199215. <https://doi.org/10.3917/ls.160.0199>

Dollander, M. (2022). *Formation à l'éducation interculturelle au regard du CECRL : Le cas de 6 universités suédoises et d'une haute école pédagogique suisse*. 27-51. <https://doi.org/10.57086/dfles.332>

Durr, A. (2016). Interagir à distance pour nouer un rapport vivant à la langue-culture. *Recherches en didactique des langues et des cultures*, 13(2). <https://doi.org/10.4000/rdlc.859>

Fischer, G. (2020). Chapitre 6. La communication sociale. In G. Fischer, *Les concepts fondamentaux de la psychologie sociale* (pp. 193-232). Paris: Dunod. <https://doi.org/10.3917/dunod.fisch.2020.01.0193>

Halté, P. (2018). *Les émoticônes et les interjections dans le tchat*. Lambert-Lucas.

Halté, P. (2020). Les émoticônes : De la signification des affects aux stratégies conversationnelles. *Communiquer. Revue de communication sociale et publique*, 28, Article 28. <https://doi.org/10.4000/communiquer.5921>

Lacelle, N., Boutin, J.-F. et Lebrun, M. (2017). *La littératie médiatique multimodale appliquée en contexte numérique, LMM@ : Outils conceptuels et didactiques*. Presses de l'Université du Québec.

Navé, E. & Farini, C. (2021). Quelles multimodalités pour quelles correspondances dans le cadre d'une didactique du français fondée sur l'approche interculturelle ? 68, en ligne. (hal-03065916).

Paveau, M.-A. (2017). *L'analyse du discours numérique : Dictionnaire des formes et des pratiques*. Hermann.

Weise, H. & Labrenz, A. (2021). Chapitre 10. Emoji as graphic discourse markers : Functional and positional associations in German WhatsApp® messages. In *Pragmatic Markers and Peripheries*. 277-300. John Benjamin Publishing company.

## Towards a grammar of internet memes combining image and text

Barbara Dancygier & Lieven Vandelanotte

*University of British Columbia & Université de Namur/KU Leuven*

In recent years, linguistic analyses of internet memes have started to address the constructional properties of image-text combinations in these kinds of artefacts (e.g. Dancygier & Vandelanotte 2017, Lou 2017, Zenner & Geeraerts 2018, Bülow et al. 2018, Vandelanotte 2021). Based on detailed analysis of an extensive data collection (Dancygier & Vandelanotte forthcoming), we expand on existing work in proposing to distinguish three major types of image-text relations in memes.

A first type concerns *when*-memes (Lou 2017, Piata 2020) and related cases (e.g. *if*-memes), where the *when*- or *if*-clause is above the image. Here the text does not explain or label the content of the image; rather, the image fills a slot in the structure of the meme as a whole. Specifically, it functions as an equivalent of the main clause in a predictive (*when/if X then Y*) construction. In **Figure 1**, for instance, the image represents not a person who is bored because their phone has no charge left, but rather the *mood* of such a person, represented by the mood of the meme character of a bored monkey.

A second prominent pattern is where the space of the meme is completely filled by the image, and the text is superimposed on it in a structured way, distinguishing a line of Top Text and one of Bottom Text. In **Figure 2**, for example, repeated use of the pattern has so entrenched the image used (turning it into a so-called ‘image macro’) that it calls up a frame of behaviour (inconsiderate behaviour associated with a meme character called ‘Scumbag Steve’). Among the striking linguistic features is the way in which here, the conjunction *when* is suppressed, as is any textual mention of a grammatical subject, which instead is provided visually, in a case of genre-based argument suppression (Ruppenhofer & Michaelis 2010).

The third major type we distinguish is that of so-called labelling memes, where the superimposed text shows no Top Text / Bottom Text division of labour, but instead ‘labels’ various components of the image which, as a whole, represents a salient situation. **Figure 3**, for instance, shows the famous situation of a ‘distracted boyfriend’, turning appreciatively to a passing girl, and so prompting feelings of betrayal in his girlfriend. Labelling such an image is a constructional process of re-filling the slots available in the meme – in this case, the slots of ‘person who is attracted to a new option’, ‘new option’ and ‘old option’. For the labelling to make sense, the situation to which it is applied (in our example, changing political leanings of young people) first has to be recognized as an instance of the constructional pattern the original image represents.

In our analysis, we discuss the role of images, including the difference between non-entrenched images and (entrenched) image macros, across these different types. We focus on the emergent grammar of memes, and consider the role of frame evocation (Fillmore 1985), depiction of embodied behaviour, and constructional forms in the meaning emergence of these multimodal artefacts.

### *Keywords*

internet memes, multimodality, construction grammar, embodiment, frames

## Figures



Figure 1



Figure 2



Figure 3

## References

- Bülow, Lars, Marie-Luis Merten & Michael Johann. 2018. Internet-Memes als Zugang zu multimodalen Konstruktionen. *Zeitschrift für Angewandte Linguistik* 69: 1–32.
- Dancygier, Barbara & Lieven Vandelanotte. 2017. Internet memes as multimodal constructions. *Cognitive Linguistics*, 28(3), 565–598.
- Dancygier, Barbara & Lieven Vandelanotte. Forthcoming. *The Language of Memes: Patterns of Meaning across Image and Text*. Cambridge: Cambridge University Press.
- Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di semantica*, 6(2), 222–253.
- Lou, Adrian. 2017. Multimodal simile: The “when” meme in social media discourse. *English Text Construction*, 10(1), 106–131.
- Piata, Anna. 2020. Stylistic humor across modalities: The case of Classical Art Memes. *Internet Pragmatics*, 3(2), 174–201.
- Ruppenhofer, Josef & Laura A. Michaelis. 2010. A constructional account of genre-based argument omissions. *Constructions and Frames*, 2(2), 158–184.
- Vandelanotte, Lieven. 2021. Creative constructs, constructions, and frames in Internet discourse. *Constructions and Frames*, 13(1), 160–191.
- Zenner, Eline & Dirk Geeraerts. 2018. One does not simply process memes: Image macros as multimodal constructions. In Esme Winter-Froemel & Verena Thaler (eds.), *Cultures and Traditions of Wordplay and Wordplay Research* (pp. 167–193). Berlin and Boston: De Gruyter Mouton.

---

# Le mème en contexte socionumérique : une mosaïque interdiscursive

Justine Simon\*<sup>1,2</sup>

<sup>1</sup>Edition, Littératures, Langages, Informatique, Arts, Didactique, Discours - UFC (EA 4661) –  
Université de Franche-Comté – 30 rue Mégevand, 25030 Besançon cedex, France

<sup>2</sup>Centre de Recherche sur les Médiations – Université de Lorraine, Université de Lorraine : UR3476,  
Université de Lorraine : EA3476 – France

## Résumé

Les mèmes sont des réalisations sémiotiques difficiles à appréhender du point de vue interdiscursif du fait de leur apparition dans un contexte participatif numérique. Les mèmes sont considérés comme des **unités technodiscursives plurisémiotiques** (Paveau, 2017 ; 2021) traversés par l'**interdiscursivité** (nous focalisons ici l'attention sur le dialogisme interdiscursif). Ce rapport au déjà-dit/vu/entendu se manifeste à différents niveaux :

- lien avec des discours antérieurs du même genre, comme pour la construction des mèmes à l'aide d'images macro ou de *templates* de vidéos sur fond vert ;

- lien avec des références interdiscursives citées dans le texte, l'hypertexte, l'image, le son et/ou la musique ; qui peuvent être de nature variée :

- références issues de la culture populaire de masse,
- références à l'actualité sociale et politique.

Du point de vue de la " variation " de mèmes ensuite (détournements ou imitations, et non simple " réplique " au sens de Bonenfant, 2014), la **dynamique créative** se réalise à différents degrés de transformation du discours d'origine. Dans la lignée des travaux de Genette (1982) et Lugrin (2006), nous distinguons deux types de réappropriations par variation : l'" intertextualité " et l'" hypertextualité ". Ces deux procédés relèvent du dialogisme interdiscursif montré. On parle d'" intertextualité " lorsque le discours d'origine est explicitement " visible ". Il est modifié (rogné, augmenté) mais reconnaissable directement. Certaines réalisations mémétiques sont de la sorte préformatées et imposent une transformation restreinte (adjonction ou substitution de textes et d'images). L'" hypertextualité " correspond à une forte transformation du discours d'origine, qui n'est plus reconnaissable littéralement. Ces variations créatives de mèmes exploitent d'autres modalités sémiotiques : formes d'imitations, réalisations allusives " faites maison ", transpositions sémiotiques (déclinaison d'un mème dessiné en photographie par exemple). Ces catégories ne sont pas totalement étanches. Parfois, nous sommes face à des transformations qui fusionnent ces deux types de transformations.

---

\*Intervenant

Au sein des réseaux socionumériques, les mèmes se construisent enfin différemment en fonction du rôle accordé à l'image partagée (au sens large de contenu inséré mêlant plusieurs matérialités sémiotiques) et au **discours d'escorte** (discours original qui accompagne l'image partagée et qui donne des indications de lecture et d'interprétation de celle-ci). La réalisation d'un mème est ainsi dépendante des fonctionnalités propres au " dispositif " socionumérique (Appel, Boulanger & Massou, dirs, 2010) dans lequel il s'inscrit. Cette réflexion portant sur la tension entre contraintes et opportunités rejoint la notion d'affordance technodiscursive (Ghliss, Perea & Ruchon, Catherine, dirs, 2019), imposée différemment selon les réseaux socionumériques (on se focalise dans cette étude sur Twitter/X, Instagram et TikTok). Les réalisations mémétiques dépendent de ces affordances. Le discours d'escorte peut faire partie intégrante du mème (comme sur Twitter/X par exemple). Ou, dans le cas où le mème partagé est simplement (re)mis en circulation, le discours d'escorte relève d'un nouvel acte d'énonciation. Analyser ces configurations permet de mettre au jour l'entremêlement des points de vue énonciatifs (Rabatel, 2010) qui peuvent se croiser dans l'interdiscours. Et dans la continuité de l'analyse de Pierre Halté (2022) portant sur les relations entre texte et image dans les mèmes, il s'agit d'être attentif à la fois aux contraintes et opportunités offertes par les différents dispositifs numériques (composante technologique) ainsi qu'à la combinaison entre discours d'escorte (texte, *emoji*, technodiscours, hyperliens) et image (pouvant combiner : textes, *emoji*, technodiscours, hyperliens, GIFs, images fixes, images animées, sons, musiques, etc.).

Cette proposition se situe dans la continuité du travail mené sur les mèmes de #ChatonsMignons (Simon, 2024). À partir d'un large corpus de mèmes partagés sur Twitter/X, Instagram et TikTok, l'objectif est d'analyser le plus finement possible ces relations multiformes avec d'autres discours en mettant au cœur de l'étude la dynamique créative et le rôle du discours d'escorte.

### Références bibliographiques

Appel, Violaine, Boulanger, Hélène et Massou, Luc (dirs), 2010, *Les dispositifs d'information et de communication : concept, usages et objets*, Bruxelles, De Boeck.

Bonenfant, Maude, 2014, " Le mème numérique : étude sémiotique des réseaux à partir des concepts de trace et d'indice ", *RISCP*, 12, pp. 27-42. <https://doi.org/10.4000/communiquer.1295>. (consulté le 12/07/2022)

Genette, Gérard, 1982, *Palimpsestes, La littérature au second degré*, Paris, Seuil.

Ghliss, Yosra, Perea, François et Ruchon, Catherine (dirs), 2019, " Les affordances langagières : textualité numérique, matérialité discursive ", *Corela* (en ligne), HS-28. <https://doi.org/10.4000/corela.82> (consulté le 15/02/2022)

Halté, Pierre, 2022, " Modalité et jugement dans les relations texte / image ", *Signata* (En ligne), 13. <https://doi.org/10.4000/signata.3673>. (consulté le 02/02/2023)

Lugrin, Gilles, 2006, *Généricité et intertextualité dans le discours publicitaire de presse écrite*, Berne, Peter Lang.

Paveau, Marie-Anne, 2021, " Le gif, outil d'iconisation du discours sur Twitter ", *Forum Linguístico* (en ligne), 18.

**Mots-Clés:** Réseaux socionumériques, Discours plurisémiotiques, Mèmes, Interdiscursivité, Discours d'escorte, #ChatonsMignons

## **Embrayage et relation texte-image dans les mèmes-Internet type « moi quand »**

Emblématiques de la culture Internet, les mèmes fournissent un exemple multimodal privilégié pour analyser l'imbrication de l'iconique et du linguistique (oral ou écrit). Un genre particulier y consiste à illustrer une situation (« moi quand je vais enfin à la plage ») par une image ou une vidéo. L'interprétation de ces mèmes déclenche un jeu complexe d'inférences : bien que formulées à la première personne du singulier, ces situations invitent régulièrement l'énonciataire à s'y identifier et produire une lecture non-référentielle du pronom personnel ; ces situations mettent donc également en place une situation d'énonciation complexe où la frontière énonciateur/énonciataire est brouillée ; d'autre part, les situations décrites peuvent être plus ou moins réalistes, invitant à une lecture non référentielle de la valeur temporelle de l'énoncé. Ces mèmes se présentent donc comme des exemples d'embrayage, c'est-à-dire « l'effet de retour à l'énonciation, produit par la suspension de l'opposition entre certains termes des catégories de la personne et/ou de l'espace et/ou du temps, » (Fiorin, 2016). Ce sont aussi des embrayeurs qui sont proposés dans la grammaire texte-image par Klinkenberg (2020) comme connecteurs, appliqués comme dans l'analyse de la modalité en bande-dessinée ou dans les mèmes, dans (Halté, 2022).

A partir d'un corpus de 150 mèmes vidéos francophones et anglophones recueillis entre novembre 2023 et avril 2024 sur la fonction Reels d'Instagram, issu d'un corpus plus large de thèse, nous proposons d'analyser l'embrayage comme clef de voute de lectures permettant entre autres l'identification ou la compréhension ironique. Notre approche sera d'analyser la compréhension pragmatique des pronoms personnels comme embrayeurs, et les éléments écrits, visuels et contextuels qui permettent de leur attribuer un ou des sens. Nous espérons ainsi fournir une étude de cas permettant d'éclairer plus spécifiquement les modalités d'interprétation de vidéo avec composante écrite, dont la production semble s'imposer comme le nouveau standard.



## Bibliographie indicative

Barb eris, J.-M. (2010). « Quand t’es super bobo »... La deuxi me personne g n rique dans le fran ais parisien des jeunes. *2 me Congr s Mondial de Linguistique Fran aise*, 124. <https://doi.org/10.1051/cmlf/2010258>

Fiorin, J. L. (2016).   propos des concepts de d brayage et d’embrayage : Les rapports entre la s miotique et la linguistique. *Actes S miotiques*, 119, Article 119 <https://doi.org/10.25965/as.5605>

Halt , P. (2022). Modalit  et jugement dans les relations texte / image. *Signata. Annales des s miotiques / Annals of Semiotics*, 13, Article 13. <https://doi.org/10.4000/signata.3673>

Klinkenberg, J.-M. (2020). Pour une grammaire g n rale de la relation texte-image. *Pratiques. Linguistique, litt rature, didactique*, 185-186, Article 185-186. <https://doi.org/10.4000/pratiques.8436>

Odin, R. (2000). *De la fiction*. De Boeck Sup rieur. <https://doi.org/10.3917/dbu.odin.2000.01>

Odin, R. (2011). *Les espaces de communication : Introduction   la s mio-pragmatique*. Presses universitaires de Grenoble.

Rabatel, A. (1998). Les embrayeurs du point de vue du narrateur. In *La construction textuelle du point de vue* (p. 101-138). Delachaux et Niestl . <https://www.cairn.info/la-construction-textuelle-du-point-de-vue--9782603011300-p-101.htm>

Yus, F. (2019). *Multimodality in Memes : A Cyberpragmatic Approach*. 105-131. [https://doi.org/10.1007/978-3-319-92663-6\\_4](https://doi.org/10.1007/978-3-319-92663-6_4)

## Assertions and Presuppositions with hyperlinks in X

**Main Claim** We address the semantico-pragmatic contribution of hyperlinks in X, focusing on what role they have in the formation of a veridicality judgement, *i.e.*, the judgement about the truthfulness of content. Our study is cast within the framework of the semantics of gestures and pictures [17, 13, 21, 22, 23, 5, 6, 19, 1, 10, 15, 16, 11, 9]. We propose two families of hyperlink functions (cf Figure 1) : hyperlinks that contribute to the *presupposition* of the text, and others that contribute to the actual *assertion*. Given that the *at-issue* [20] hyperlinks are less frequent, we conclude that hyperlinks, like evidentials [18], contribute mostly to presupposition while they fulfill different goals.

**Methodology** We rely on a French ecological crises corpus of 19, 595 tweets collected by [12]. We have randomly selected 800 tweets which have already been annotated for speech act categories following [14] who distinguished : *Assertives*, *Subjectives*, *Interrogatives*, *Jussives*. We focus on *Assertives* and *Subjectives* (containing perspectival elements, evaluatives, epistemic attitudes and modals), see Figure 1, Figure 2 for the numerical distribution).

Category	Function	Description
<b>Presupposition</b>	<b>Source</b>	Hyperlink marks source of information, evidential [3, 2, 24].
	<b>Illustration</b>	Hyperlink is a picture illustrating the text with narrow or wide scope
	<b>Exemplification</b>	We use the term of [6]. Hyperlink is a "generic" picture, its content generalizes the accompanying text.
	<b>Setting up an indexical referent</b>	Hyperlink sets up referent signaled by an indexical.
	<b>Explicitation of evidence</b>	Hyperlink clarifies one's evidence for the inference.
<b>Assertion</b>	<b>Rhetorical</b>	Hyperlink and associated text may be contradictory or unrelated, creating an ironic or humorous effect.
	<b>Prospeech</b>	Hyperlink completes the accompanying text, <i>at issue</i> [20].

Table 1: The relationships types between hyperlinks and their accompanying text

The numerical analysis (see Table 2) highlights that :

- The majority of annotated tweets have at least one hyperlink (65, 37%).
- The most important functions of hyperlinks are information *source* (53%) and *illustration* (26%) and are mostly present in *Assertives*.
- Most of the annotated tweets which do not have an hyperlink are *Subjectives* (56 %).

Speech acts	No hy- perlinks	Source	Illustration	Exempl.	Index.	Expl. ev- idence	Rhetorical	Prospeech
<b>Assertives</b>	28,88%	76, 71 %	70,00%	83,33%	46,15 %	50, 00%	100, 00%	66,67 %
<b>Interrogatives</b>	6,50%	0,47%	2,38 %	0%	0%	50,00%	0 %	0%
<b>Jussives</b>	7,94 %	9,65%	9,52%	0%	3,85%	0 %	0%	0 %
<b>Subjectives</b>	56,32%	13,18%	18,10%	16,67%	50,00%	0 %	0%	33,33 %
<b>Total</b>	<b>41,13 %</b>	<b>53,13 %</b>	<b>26,25 %</b>	<b>0,75%</b>	<b>4,88 %</b>	<b>1,00 %</b>	<b>0,13 %</b>	<b>0,38 %</b>

Table 2: Annotation results

**Discussion** Per Figure 1, hyperlinks contribute to *Assertion* or to *Presupposition*. We saw that mainly they contribute to *Presupposition*. Additionally, the strong correlation between *Assertives* and *Source* or *Illustration* indicated that hyperlink may fulfill different goals:

- **Goal 1** : as *evidence for a veridical judgement* (following [8]).
- **Goal 2** : as *norms of citizen journalism*, [7, 4],
- **Goal 3** : as a way to *highlight one's affiliation to a political, ideological or social group*. Selecting one specific source instead of another is a way to emphasize the belonging to a community.
- **Goal 4** : we propose that all hyperlinks ( regardless of function or content), aim to *sustain or establish relevance (be popular)*. Integrating one hyperlink is a way to make a tweet more visible and more likely to be shared or retweeted.

## References

- [1] D. Abusch and M. Rooth. Parallel and differential contributions from language and image in the discourse representation of picturebooks. In Proceedings of Sinn und Bedeutung, volume 27, pages 1–18, 2023.
- [2] A. Aikhenvald. The grammar of knowledge: a cross-linguistic view of evidentials and the expression of information source. The grammar of knowledge: A cross-linguistic typology, 1:51, 2014.
- [3] A. Y. Aikhenvald. Evidentiality. Oxford University Press, 2004.
- [4] D. Crystal. Language and the Internet. 2nd édition, Cambridge: Cambridge University Press, 2006.
- [5] C. Ebert. Semantics of gesture. Annual Review of Linguistics, 10:169–189, 2024.
- [6] C. Ebert and M. Steinbach. The semantics of visual communication : Theoretical approaches to visual meaning aspects in co-speech gestures and sign languages. The European Summer School in Logic, Language and Information (ESSLLI), 2023.
- [7] D. Gillmor. We the media: The rise of citizen journalist. Nat'l Civic Rev., 93:58, 2004.
- [8] P. Grice, H. Logic and conversation. Syntax and semantics 3 : Speech acts, Cole et al (eds), 3:pp41–58, 1975.
- [9] P. G. Grosz, G. Greenberg, C. De Leon, and E. Kaiser. A semantics of face emoji in discourse. Linguistics and Philosophy, 46(4):905–957, 2023.
- [10] P. Halté. Les gestes à l'écrit dans les interactions numériques: Description et fonctions. Pratiques. Linguistique, littérature, didactique, (183-184), 2019.
- [11] J.-M. Klinkenberg. Pour une grammaire générale de la relation texte-image. Pratiques. Linguistique, littérature, didactique, (185-186), 2020.
- [12] D. Kozłowski, E. Lannelongue, F. Saudemont, F. Benamara, A. Mari, V. Moriceau, and B. A. A three-level classification of french tweets in ecological crises. Information Processing & Management., 57(5):1–46, 2020.

- [13] A. Lascarides and M. Stone. A formal semantic analysis of gesture. Journal of Semantics, 26(4):393–449, 2009.
- [14] E. Laurenti, N. Bourgon, F. Benamara, A. Mari, V. Moriceau, and C. Courgeon. Give me your intentions, i’ ll predict our actions: A two-level classification of speech acts for crisis management in social media. In 13th Conference on Language Resources and Evaluation (LREC 2022), pages 4333–4343, 2022.
- [15] E. Maier. Picturing words: The semantics of speech balloons. In Proceedings of the 22nd Amsterdam Colloquium, pages 584–592. 2019.
- [16] E. Maier. Emojis as pictures. Ergo: An Open Access Journal of Philosophy, 10(n/a), 2023.
- [17] D. McNeill. Gesture and Thought. University of Chicago Press; Illustrated edition, (November 8, 2005).
- [18] S. E. Murray. The semantics of evidentials, volume 9. Oxford University Press, 2017.
- [19] M. Nikolajeva and C. Scott. How picturebooks work. Routledge, 2013.
- [20] C. Potts. Conversational implicatures via general pragmatic pressures. In New Frontiers in Artificial Intelligence: JSAI 2006 Conference and Workshops, Tokyo, Japan, June 5-9 2006, Revised Selected Papers, pages 205–218. Springer, 2007.
- [21] P. Schlenker. Gesture projection and cosuppositions. Linguistics and Philosophy, 41(3):295–365, 2018.
- [22] P. Schlenker. Iconic pragmatics. Natural Language & Linguistic Theory, 36:877–936, 2018.
- [23] P. Schlenker. Gestural grammar. Natural Language & Linguistic Theory, 38(3):887–936, 2020.
- [24] T. Willett. A cross-linguistic survey of the grammaticization of evidentiality. Studies in Language, 1988.



**Langues & Langage  
à la croisée des Disciplines**

**1ère Rencontre annuelle LLcD**

Colloque : 9-11 septembre, 2024

Ecole d'été : 12-14 septembre, 2024

Sorbonne Université - 28 Rue Serpente, 75006 Paris, France



## **Workshop 4**

**The demonstrative in the evolutionary process.  
Typological variation and change**

**Le démonstratif en évolution. Variation  
typologique et changement**

**LES DÉMONSTRATIFS DU LOUVITE HIÉROGLYPHIQUE  
À LA LUMIÈRE DU HITTITE :  
UNE CONTRIBUTION FONCTIONNALISTE  
À LA LINGUISTIQUE HISTORIQUE ANATOLIENNE**

Guillaume KURZ  
*Université Libre de Bruxelles & École Normale Supérieure (Paris)*

Les langues anatoliennes, attestées dès la seconde moitié du deuxième millénaire avant J.-C., et jusqu'aux premiers siècles de notre ère (MELCHERT, 2017 : 172), sont d'un intérêt majeur pour les études indo-européennes. Les données offertes par cette branche figurent à la fois parmi les plus anciennes à l'échelle de la famille indo-européenne, mais également parmi les plus archaïques (MELCHERT, à paraître). La langue la mieux attestée, le hittite, constitue un champ d'investigation exceptionnel pour l'étude des démonstratifs ; c'est ce qu'a notamment démontré GOEDEGEBUURE (2014) en développant un cadre théorique original, d'inspiration typologique et fonctionnelle.

En combinant des approches pragmatiques, discursives et cognitives, cette dernière aboutit à un modèle où chaque fonction pragmatique (DIESSEL, 1999 : 93–114) est étudiée à la lumière des effets textuels – introduction d'une nouvelle référence ou bien procédé de cohésion – et informationnels – *Hearer Old - Discourse New* (DIESSEL, 1999 : 106). Son approche (GOEDEGEBUURE, 2003 : 57–59) est modélisée grâce à des tableaux à plusieurs entrées (figures 1 et 2) :

<i>geographical parameter</i>	<i>cognitive parameter</i>				
	continuity procedure			centering procedure	
	E-Topic	U-Topic	A-Focus	U-Topic	Focus
Deictic	<i>s/he</i>	S/HE	S/HE	—	<i>dem N</i>
Recognitional	—	—	—	—	<i>that N</i>
Anaphora	<i>s/he</i>	S/HE	S/HE	—	—
in-node	<i>dem N</i>	<i>dem N</i>	—	<i>that</i>	<i>that N</i>
across-node	—	—	—	—	—
Discourse Deictic	<i>it</i>	?	THAT	—	—
in-node	—	—	—	—	—
across-node	—	—	—	—	<i>dem N</i>

Figure 1 : matrice initiale

<i>geographical parameter</i>	<i>discourse parameter</i>			
	reference inside discourse unit		reference across discourse node (including first mention)	
	E-Top	Other	creates cohesion	no cohesion
world	∅, <i>s/he</i>	SI/HE	<i>demNP</i>	<i>defNP</i>
text	∅, <i>s/he</i>	SI/HE	<i>demNP</i>	<i>defNP</i>
memory	—	—	<i>demNP</i>	<i>defNP</i>

Figure 2 : proposition d'une variante de la matrice initiale

L'attention portée aux démonstratifs des autres langues anatoliennes a en revanche été moins soutenue, en témoigne seulement l'identification très récente de certaines formes casuelles en louvite hiéroglyphique (GOEDEGEBUURE, 2007 ; KLOEKHORST, 2012). Par ailleurs, les démonstratifs *za-* (proximal) et *aba-* (distal) de cette dernière langue n'ont fait l'objet d'aucune étude approfondie. Tout au plus trouve-t-on une étude contrastive entre grec ancien et

louvite sur la position relative du démonstratif et du nom (GIUSFREDI, 2017) et quelques pages par BAUER (2014 : 42–61) traitant de l’interaction entre démonstratif et définitude. Cette communication entend donc contribuer à la linguistique des langues anatoliennes anciennes : si l’on compte jusqu’à cinq formes démonstratives différentes en hittite (HOFFNER & MELCHERT, 2008 : 142–148), le louvite semble n’en comporter que deux. Il sera donc question de dresser une image du système de démonstratif d’une langue en regard de l’autre : on étudiera la manière dont sont distribuées les fonctions pragmatiques, informationnelles – en termes de “peripheral importance” ou de “low topicality” (HIMMELMANN, 1996 : 230) – et textuelles en louvite, avant de confronter brièvement les résultats avec ceux de GOEDEGEBUURE (2014).

L’étude des démonstratifs d’une langue attestée uniquement à travers un corpus limité, tant dans la taille que dans la variété de genres textuels, n’est pas une entreprise vaine ; au contraire, celle-ci a toutes les chances d’être prometteuse, comme le montre notamment l’examen des données sabelliennes, pourtant très fragmentaires, réalisé par DUPRAZ (2012b). Cela est d’autant plus vrai dans le cas du louvite puisque cette langue dispose, si ce n’est d’une grammaire ou d’un dictionnaire, d’un excellent corpus comportant une édition des textes, de riches indices ainsi qu’une bibliographie à jour (HAWKINS, 2024). Parmi les premiers résultats obtenus se trouve notamment la reconnaissance, inédite, de plusieurs occurrences de démonstratifs témoignant d’un emploi dit recognitionnel (HIMMELMANN, 1996 : 230–239 ; DIESSEL, 1999 : 105–109) ou anamnétique (DUPRAZ, 2012a). On montrera que le démonstratif proximal louvite semble tout aussi compatible avec un emploi de ce type que son analogue distal, phénomène certes aussi observé en hittite mais surprenant d’un point de vue typologique.

## Bibliographie

- Bauer, A. H. (2014). *Morphosyntax of the Noun Phrase in Hieroglyphic Luwian*. Leiden: Brill. <https://doi.org/10.1163/9789004260030>
- Diessel, H. (1999). *Demonstratives: Form, Function and Grammaticalization*. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.42>
- Dixon, R. (2003). Demonstratives: A cross-linguistic typology. *Studies in Language*, 27(1), 61–112. <https://doi.org/10.1075/sl.27.1.04dix>
- Dupraz, E. (2012a). *Ille* comme démonstratif anamnétique en latin classique. In C. Denizot & E. Dupraz (Éds.), *Anaphore et anaphoriques : variété des langues, variété des emplois* (pp. 77–96). PURH.
- Dupraz, E. (2012b). *Sabellian Demonstratives: Forms and Functions*. Leiden: Brill. <https://doi.org/10.1163/9789004216990>
- Giusfredi, F. (2017). The position of determiners in Ancient Greek and Luwian: Between syntax and information flow. *Folia Linguistica*, 51(3), 611–637. <https://doi.org/10.1515/flin-2017-0023>
- Goedegebuure, P. (2003). *Reference, Deixis and Focus in Hittite. The Demonstrative ka- “this”, apa- “that” and asi “yon”*. Thèse de l’Université d’Amsterdam.
- Goedegebuure, P. (2007). The Hieroglyphic Luwian demonstrative ablative-instrumentals *zin* and *apin*. In A. Archi & R. Francia (Éds.), *VI Congresso Internazionale di Ittitologia, Roma, 5-9 settembre 2005* (pp. 319–334). CNR.
- Goedegebuure, P. (2014). *The Hittite Demonstratives: Studies in Deixis, Topics and Focus*. Wiesbaden: Harrassowitz.
- Hawkins, J. D. (Éd.). (2024). *Corpus of Hieroglyphic Luwian Inscriptions*. Vol. 3: *Inscriptions of the Hittite Empire and New Inscriptions of the Iron Age*. Berlin: De Gruyter. <https://doi.org/10.1515/9783110778854>

- Himmelmann, N. P. (1996). Demonstratives in Narrative Discourse: A Taxonomy of Universal Uses. In B. A. Fox (Éd.), *Studies in Anaphora* (pp. 205–254). John Benjamins. <https://doi.org/10.1075/tsl.33.08him>
- Hoffner, A. H., Jr. & Melchert, C. (2008). *A Grammar of the Hittite Language*. Vol. 1: *Reference Grammar*. Winona Lake: Eisenbrauns. <https://doi.org/10.1515/9781575065731>
- Kloekhorst, A. (2012). Pronominal Morphology in the Anatolian Language Family. *Altorientalische Forschungen*, 39(2), 254–264. <https://doi.org/10.1524/aofo.2012.0017>
- Melchert, C. (2017). Anatolian. In M. Kapović (Éd.), *The Indo-European Languages* (2<sup>e</sup> éd., pp.171–201). Routledge.
- Melchert, C. (à paraître). The position of Anatolian. In A. Garrett & M. T. Weiss (Éds.), *Handbook of Indo-European Studies*. Cambridge University Press.

### **Keywords**

Linguistique historique ; indo-européanistique ; langues anatoliennes ; typologie fonctionnelle ; démonstratif



*Ille et ipse dans l'Itinerarium Ætheriae, articloïdes ou démonstratifs ?*

La grammaticalisation usuelle des démonstratifs en articles définis (Greenberg, 1978) a incité à étudier le latin tardif au prisme de ce changement et à rechercher des signes avant-coureurs dès cet état de langue de l'évolution à venir dans les langues romanes (Aebischer, 1948, Selig 1992, Hertenberg, 2014...). En nous appuyant sur la distinction théorique et empirique entre définitude pragmatique et définitude sémantique (Löbner, 2011, Ortman, 2014), nous analysons l'intégralité des contextes d'emploi des démonstratifs dans un texte bas-latin, l'*Itinerarium Ætheriae* (4<sup>ème</sup> siècle, 17 592 mots) sous l'angle de cette division. Il a en effet été observé que les démonstratifs étaient restreints à la définitude pragmatique (déixis, anaphore directe, relative déterminative) quand les articles définis pouvaient également être employés dans des contextes de définitude sémantique (anaphore associative, substantifs avec un sélecteur, noms uniques, généricité...). Nous avons ainsi annoté manuellement chaque occurrence selon ces sept contextes d'emploi. L'analyse des premiers stades du développement des articles définis dans ce cadre théorique a récemment été employée de façon euristique pour les langues scandinaves médiévales (Skrzypek *et alii*, 2021) ou dans les langues slaves contemporaines (Czardybon, 2017, Dvorak, 2020).

Si l'emploi adnominal des démonstratifs est plus fréquent dans ce texte (355 occurrences) que dans un corpus cicéronien de même longueur (210 occurrences), et si *ipse* a bel et bien développé des emplois démonstratifs qui peuvent préfigurer son emploi comme source de l'article défini dans certaines langues romanes (sarde *su*, catalan des Baléares *es...*), l'analyse qualitative des occurrences révèle que ces démonstratifs ne sont jamais indubitablement employés en contexte de définitude sémantique :

	<i>hic</i>	<i>iste</i>	<i>ille</i>	<i>ipse</i>
<i>Déixis</i>	17	13	/	4
<i>Anaphore directe</i>	21	3	42	151
<i>Relative déterminative</i>	13	1	64	21
<i>Anaphore associative</i>	/	/	1 (?)	2 (?)
<i>Nom + sélecteur</i>	/	/	/	/
<i>Nom unique</i>	/	/	1 (?)	/
<i>Générique</i>	/	/	/	/

Fig. 1 – *Emploi des démonstratifs dans l'Itinerarium Ætheriae.*

Quatre exemples peuvent être lus comme des cas de définitude sémantique, mais une explication concurrence et compatible avec la définitude pragmatique peut toujours être formulée, par exemple comme démonstratif anamnétique (Dupraz, 2012). L'exemple suivant pourrait être un cas d'anaphore associative, mais une analyse d'*ipse* comme un marqueur d'identité – un rôle par ailleurs bien attesté en latin tardif – est également possible :

(1) <i>Nam</i>	<i>duxit</i>	<i>nos</i>	<i>statim</i>	
Car.SUB	mena.V_3PS	nous.PRO_PERS_A	immédiatement.ADV	
<i>ad</i>	<i>ecclesiam</i>	<i>quae</i>	<i>est</i>	<i>foras</i>
vers.PREP	église.N_A	qui.REL_N	est.V_3PS	hors_PREP
<i>ciuitatem</i>	<i>in</i>	<i>eo</i>	<i>loco</i>	<i>ubi</i>
ville.N_A	dans.PREP	ce.DEM_ABL	lieu.N_ABL	où.REL
<i>fuit</i>	<i>domus</i>	<i>sancti</i>	<i>Abrahamae</i>	
était.V_3PS	maison.N_N	saint.ADJ_G	Abraham.N_G	

<i>id</i>	<i>est</i>	<i>in</i>	<i>ipsis</i>
cela.PRO_ANAP_N	est.V_3PS	dans.PREP	ces.DEM_ABL
<i>fundamentis</i>	<i>et</i>	<i>de</i>	<i>ipso</i>
fondements.N_ABL	et.CONJ	de.PREP	cette.DEM_ABL

*lapide* (I.Æ. §20.3)  
pierre.N\_ABL

« Il nous mena alors aussitôt vers une église qui est hors de la ville, dans ce lieu où se dressait la maison de Saint Abraham, c'est-à-dire dans les (mêmes) fondements et à partir de la (même) pierre ».

Notre analyse suggère ainsi que les démonstratifs, dans ce texte bas-latin, ne présentent aucun signe indubitable de début de grammaticalisation en article défini.

### Sources

- Éthérie (1948): *Journal de Voyage*. (Ed.) Hélène Pétré Paris: Éditions du Cerf.  
Cicero (1959): *Discours, tome XIX*. (Ed.) André Boulanger & Pierre Wuilleumier Paris: Les Belles Lettres.

### Références

- Aebischer, Paul (1948). Contribution à la protohistoire des articles *ille* et *ipse* dans les langues romanes. *Cultura neolatina* (8). 181-203.  
Czardybon, Adrian (2017): *Definiteness in a Language without Articles - A Study on Polish*. Düsseldorf: Düsseldorf University Press.  
Dupraz, Emmanuel (2012): *Ille* comme démonstratif anamnestique en latin classique. In Camille Denizot & Emmanuel Dupraz (Hrsg.), *Anaphore et anaphoriques : variété des langues, variétés des emplois (2012)*, 73-96. Mont-Saint-Aignan: Publications des universités de Rouen et du Havre.  
Dvorak, Jan (2020): The emerging definite article ten in Czech. *Naše řeč* (103). 297–319.  
Greenberg, Joseph, Harold (1978): How does a language acquire gender markers? In Joseph Greenberg Harold, Charles Ferguson & Edith Moravcsik (Hrsg.), *Universals of Human Language, vol.3, Word Structure*, 47–81. Stanford: Stanford University Press.  
Hertenberg, Mari (2014): „The valley“ or „that valley“? *Ille* and *ipse* in the *Itinerarium Egeriae*. *Latin Vulgaire, Latin Tardif X*, vol. 3, 873–892. Bergame: Sestante Edizioni.  
Löbner, Sebastian (2011): Concept types and determination. *Journal of Semantics* 3(28). 279–333.  
Ortmann, Albert (2014): Definite Article Asymmetries and Concept Types : Semantic and Pragmatic Uniqueness. *Frames and Concept Types, Applications in Language and Philosophy* (Studies in Linguistics and Philosophy 94), 293–321. Heidelberg, New York, Dordrecht, London: Springer.  
Selig, Maria (1992): *Die Entwicklung der Nominaldeterminanten im Spätlatein*. Tübingen: Gunter Narr Verlag.  
Skrzypek, Dominika, Alicja Piotrowska & Rafał Jaworski (2021): *The Diachrony of Definiteness in North Germanic*. Leiden ; Boston: Brill.

*This my lettere* – l'émergence d'un démonstratif de révérence en anglais moderne

Benjamin Dufour (École Normale Supérieure – AOROC)

Si l'emploi conjoint du possessif et du démonstratif est bien attesté en vieil anglais, (Wood 2007), cette construction disparaît de nos sources en moyen anglais (Allen, 2004), avant de réapparaître au début de l'anglais moderne (16<sup>e</sup> siècle), mettant en péril l'unidirectionnalité de la constitution du système des déterminants anglais : en suivant la distinction établie par Lyons (1999), l'anglais passerait d'un type AG – où le possessif se comporte comme un adjectif et est compatible avec un déterminant – à un type DG – où il se comporte comme un déterminant et exclut donc l'emploi d'un autre déterminant – puis AG de nouveau. Dans cette étude, nous montrons que le démonstratif est employé dans ce tour comme un marqueur pragmatique de révérence ou de politesse, dépassant ainsi l'illusion d'un retour à la situation du vieil anglais.

Une analyse quantitative du *Penn-Helsinki Parsed Corpus of Early Modern English* (212 occurrences du phénomène) révèle plusieurs tendances : plus grande fréquence du proximal *this* que du distal *that*, surreprésentation des première et deuxième personnes, concentration de ce tour dans les lettres non-privées (22,64% des occurrences, quand ces lettres constituent 5,35% du corpus). Afin de concilier cette approche quantitative avec une approche qualitative, nous avons annoté la correspondance de Thomas Cromwell (1485?-1540), qui regroupe 351 lettres (140 000 mots), dont la connaissance des contextes de rédaction et des destinataires permet une étude de pragmatique historique (Palander-Collin, 2010). Le relevé des occurrences dans la correspondance de Cromwell s'est fait manuellement à partir d'une édition moderne des lettres. L'analyse des occurrences révèle que le tour présente une part indéniable de lexicalisation (41,89% des attestations se font avec *letteres*, y compris dans les lettres rédigées en latin : *nostris his litteris*, Lettre 74):

(01) At	the	sight	of	<b>these</b>
<i>à.PREP</i>	<i>la.DEF</i>	<i>vue.N</i>	<i>de.PREP</i>	<i>ces.DEM</i>
<b>my</b>	<b>Letters</b> (Lettre 237)			
<i>mes.POSS</i>	<i>lettres.N</i>			

« À la lecture de mes lettres ».

Outre ces cas de lexicalisation, plusieurs critères pragmatiques favorisent l'apparition du démonstratif avec le possessif : la structure est surreprésentée dans les lettres adressées à Henry VIII et au clergé (40,54% quand ces lettres représentent 20,33% du corpus), et lorsque le possessif désigne le roi ou une personne de haut rang:

(02) A	byll	that	no	person	within	<b>this</b>
<i>une.INDEF</i>	<i>loi.N</i>	<i>que.REL</i>	<i>pas.NEG</i>	<i>personne.N</i>	<i>dans.PREP</i>	<i>ce.DEM</i>
<b>your</b>	<b>Realme</b>	shall		herafter	kepe (Lettre 67)	
<i>votre.POSS</i>	<i>royaume.N</i>	<i>va.MOD</i>		<i>ensuite.ADV</i>	<i>suivre.INF</i>	

« Une loi que nul homme dans votre royaume ne suivra dans le futur ».

Le démonstratif se rencontre dans les mêmes conditions avec un article défini, et ce dans des contextes de définitude sémantique où le démonstratif est normalement exclu (Löbner, 1985):

(03) Understanding	that	the	body	of	<b>that</b>
<i>comprenant.GERUND</i>	<i>que.COMP</i>	<i>le.DEF</i>	<i>corps.N</i>	<i>de.PREP</i>	<i>cette.DEM</i>

the Université of Cambridge hath elected... (Lettre 116)  
 la.DEF université.N de.PREP Cambridge.N a.V\_3PS élu.PART  
 « Comprenant que le corps de l'Université de Cambridge m'a élu... ».

Notons par ailleurs que, dans la correspondance de Thomas Cromwell, le démonstratif proximal *this* est de loin plus fréquent dans cet emploi (65 occurrences) que le démonstratif distal *that* (9 occurrences).

Nos données invitent à voir dans ce démonstratif un marqueur de révérence ou de politesse, dont le développement ne serait pas surprenant dans cette période caractérisée par l'émergence et la codification de nouvelles normes de politesse et de comportements de cour (Jucker, 2020, Jucker & Taavitsainen, 2013). Diachroniquement, ce « démonstratif de révérence » proviendrait de l'emploi *affectif* du démonstratif (Kleiber, 1991, Dvorak, 2021). Cette hypothèse permet de résoudre l'interrogation diachronique sur la cooccurrence de deux déterminants dans l'histoire de la langue anglaise, mais ouvre également de nouvelles perspectives sur l'analyse pragmatique de cet état de langue.

### Sources

Cromwell, Thomas. *Life and Letters of Thomas Cromwell, Vol. 1: Life, Letters to 1535*. Édité par Roger Bigelow Merriman. Oxford : Oxford University Press, 1902.

———. *Life and Letters of Thomas Cromwell, Vol. 2: Letters from 1536*. Édité par Roger Bigelow Merriman. Oxford : Oxford University Press, 1902.

Kroch, Anthony, Beatrice Santorini, et Lauren Delfs. « The Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME) », 2004.

### Bibliographie

Allen, Cynthia. « These our letters: The dem poss construction from old to Early Modern English ». Édité par Gunnar Bergh, Jennifer Herriman, et Mats Mobärg. *An international master of syntax and semantics. Papers presented to Aimo Seppänen on the occasion of his 75th birthday*, 2004, 11-19.

Dvorak, Jan. « Les emplois adnominaux des démonstratifs tchèque et français dans la langue orale informelle ». École Normale Supérieure de Lyon, 2021.

Jucker, Andreas H., et Irma Taavitsainen. *English historical pragmatics*. Edinburgh textbooks on the English language. Advanced. Edinburgh: Edinburgh University Press, 2013.

Jucker, Andreas H. *Politeness in the History of English*. Cambridge: Cambridge University Press, 2020.

Kleiber, Georges. « Du nom propre non modifié au nom propre modifié: le cas de la détermination des noms propres par l'adjectif démonstratif ». *Langue française*, n° 92 (1991): 82-103.

Lyons, Christopher. *Definiteness*. Cambridge : Cambridge University Press.

Löbner, Sebastian. « Definites ». In *Journal of Semantics*, n°4 (1985): 279-326.

Palander-Collin, Minna. « Correspondence ». In *Historical pragmatics*, édité par Andreas Jucker et Irma Taavitsainen, 651-678. Berlin ; New York : De Gruyter Mouton, 2010.

Wood, Johanna. « Demonstratives and possessives from Old English to present-day English ». In *Nominal Determination : Typology, context constraints, and historical emergence*, édité par Elisabeth Stark, Elisabeth Leiss, et Werner Abraham, 339-61. Amsterdam ; Philadelphia: John Benjamins Publishing Company, 2007.

## La pragmaticalisation des marqueurs discursifs *alors* en français et *nà* en chinois mandarin

S'appuyant sur un corpus oral informel, cette communication vise à comparer la « pragmaticalisation » (Dostie 2004) de deux marqueurs discursifs formés sur les démonstratifs : *alors* en français et *nà* en chinois mandarin. L'approche adoptée pour cette étude est principalement comparative et typologique, avec un apport de la linguistique diachronique.

Cette recherche montrera un processus de pragmaticalisation similaire des deux marqueurs discursifs : 1) connecteur discursif anaphorique temporel ou propositionnel > 2) connecteur discursif anaphorique consécutif et hypothétique > 3) marqueur de structuration du discours (anaphorique, mémoriel, phatique).

Côté français, on ne voit plus dans *alors* la trace du démonstratif. Étymologiquement, *alors* tire sa source en français médiéval de la préposition *à* combinée à l'adverbe *lors*, qui marque « une référence dans le passé à distance du moment de l'énonciation » (Buridant 2019 : 463). En français moderne, *alors* garde toujours sa valeur temporelle (*Arrive alors le temps de l'éméritat*) et développe une valeur consécutif (*Ils ont le même thème, alors j'ai dit non*) ou hypothétique (*Si tu changes d'avis, alors le cauchemar deviendra la réalité*) comme connecteur discursif. En outre, étant « marqueur de structuration du discours » (Schlamberger Brezar 2012) en début d'énoncé, *alors* sert d'introducteur de topique (*Alors cette séance de kiné c'était comment ?*), d'introducteur de sous topique par rapport au topique central (*Il y a des draconiens les pixies et les morganes, alors les draconiens sont...*) ou de « planificateur du discours » (*Qu'est-ce que t'en penses ? Alors... je ne sais pas*). Mais, il peut aussi se situer en fin d'énoncé comme marqueur de conclusion (*Il n'y a pas de bouchons pour rentrer à Paris alors*).

Côté chinois, *nà* est un morphème de démonstratif distal. Il fonctionne comme pronom déictique discursif (ou anaphorique) reprenant le contenu propositionnel de l'énoncé précédent, d'où dérive son emploi de connecteur discursif consécutif ou hypothétique (Zhang & Fang 2014 [1996]), comme *alors* en français. En tant que marqueur de structuration du discours, *nà* sert à changer le topique discursif par rapport à un ancien topique, ou à planifier des idées discursives de la part du locuteur. Dans ce cas, il se trouve toujours en début d'énoncé de même que *alors* en français.

À travers cette étude comparative, nous voudrions montrer que *alors* en français est plus pragmaticalisé que sa contrepartie *nà* en chinois mandarin. Cela réside notamment dans leurs différentes valeurs sémantiques et distributions syntaxiques :

- 1). Sémantiquement, *alors* peut déjà introduire un nouveau topique discursif provenant des connaissances partagées des interlocuteurs (avec fonction phatique), tandis que *nà* doit toujours recourir au contexte discursif.
- 2). Syntaxiquement, la position de *alors* est relativement libre dans un énoncé. Comme connecteur discursif, il se trouve soit autour du verbe, soit au début de la proposition principale ; en tant que marqueur discursif, il est déplaçable au début ou à la fin de l'énoncé. En revanche, *nà* dans ces deux cas se restreint toujours à la position initiale.

## Bibliographie

BURIDANT, C. (2020). *Grammaire du français médiéval (XIe-XIVe siècles)*. Strasbourg : ELIPHI.

DOSTIE, G. (2004). *Pragmaticalisation et marqueurs discursifs : Analyse sémantique et traitement lexicographique*. De Boeck Supérieur. DOI : <https://doi.org/10.3917/dbu.dosti.2004.01>

SCHLAMBERGER BREZAR, M. (2012). « Les marqueurs discursifs “mais” et “alors” en tant qu’indicateurs du degré de l’oralité dans les discours officiels, les débats télévisés et les dialogues littéraires ». *Linguistica* 52/1, p. 225-237.

ZHANG, B.-J. & FANG, M. (2014 [1996]). « Zhidai fanchou de yufahua zhiyi : xiandai hanyu li de jieduanxing fazhan » [« L’une des grammaticalisations des démonstratifs : l’évolution périodique en chinois moderne »]. *Hanyu gongneng yufa yanjiu [La Grammaire fonctionnelle du chinois]*. Pékin : Presses Commerciales, p. 216-243.

## L'évolution des démonstratifs en français : du qualitatif au quantitatif

L'évolution du système des démonstratifs français a jusqu'ici été décrite, à bon droit, comme le passage d'une opposition dite « sémantique » (Marchello-Nizia 1995), entre un paradigme proximal et un paradigme distal, à une opposition dite « morpho-syntaxique », entre un paradigme de pronoms et un paradigme de déterminants. Si le constat est indéniable, la formulation pose le problème de situer sur deux plans différents le point de départ et le point d'arrivée de l'évolution. Or, il est bien évident que le système de départ comme le système d'arrivée sont aussi bien descriptibles en terme de sémantique qu'en terme de morphosyntaxe. Afin de remédier à cette difficulté, nous proposons une description permettant d'unifier les deux plans, en ayant recours à l'opposition entre *signifiant* et *signifié*. Sur le plan du signifié, les démonstratifs français sont d'abord passés de la notion de *sphère personnelle du locuteur* à celle d'*appariement contigu* (Kleiber 1987, Marchello-Nizia 2006a, Guillot-Barbance 2017), pour aboutir, en français moderne, à une opposition entre *prédicativité* et *non prédicativité* (Soutet 1992). La première est qualitative (le référent appartient ou non à la sphère personnelle) et relève pour une bonne part de l'évaluatif et du modal, la seconde est quantitative (conceptualisation en amont faible VS forte) et relève de la sémantique référentielle (Soutet 1997), et la troisième est également quantitative (support seul VS support + apport). Sur le plan du signifiant, on passe d'une opposition qualitative, l'alternance consonantique, à une opposition quantitative entre *formes courtes* et *formes longues* (Combettes 2006).

Cette reformulation nous permet de poser l'hypothèse d'une évolution des démonstratifs en trois étapes. Jusqu'au 13<sup>ème</sup> siècle l'opposition sémantique se fonde sur la notion de sphère personnelle (Céline Guillot-Barbance 2017), et elle est signifiée à l'aide d'une opposition consonantique /-st / VS /-l/. Parallèlement, émerge un paradigme de formes courtes exclusivement déterminants (Marchello-Nizia 1995, 2006b) et, sur le plan sémantique, non prédicatifs (Soutet 1992). La seconde étape, jusqu'au quinzième siècle, consiste en un changement exclusivement sémantique, de l'évaluatif au référentiel, sous maintien de l'alternance consonantique. Le paradigme de déterminants se maintient en marge du système. Du 15<sup>ème</sup> siècle au 17<sup>ème</sup> siècle, les formes de cas sujet disparaissent, et une sélection de formes pronominales longues s'opère sur la base d'une analogie avec les pronoms personnels *lui*, *elle*, *eux*, désormais prédicatifs (Zink 1997). Les déterminants démonstratifs non prédicatifs peuvent



désormais former un système central par opposition à un paradigme de pronoms démonstratifs prédicatifs nouvellement constitués. La première étape peut donc se définir comme un alignement sémiologique qualitatif, la seconde comme un décalage entre signifié quantitatif et signifiant qualitatif, et la troisième comme un alignement sémiologique quantitatif. Nous espérons également, à travers cette proposition, pouvoir situer l'évolution des démonstratifs dans une dynamique de changement typologique plus général (Carlier-Combette 2015). Les modifications qu'il connaît se produisent en relation avec des changements touchant d'autres secteurs : nous avons déjà évoqué une coïncidence chronologique avec les pronoms personnels, nous pouvons ajouter que le paradigme de déterminants démonstratifs se construit par analogie avec le système de l'article alors que de ce dernier vient de s'instituer (Carlier-Lamiroy 2018). Le passage d'une sémiologie qualitative, l'alternance phonémique /-st/ VS /-l/, mais aussi /-i/ VS /-e/ à plus grande échelle (Soutet 1992), à une sémiologie quantitative qui oppose forme courte et forme longue doit par ailleurs retenir l'attention dans le contexte d'une transition de l'oral vers l'écrit (Zumthor 1993).

## **Bibliographie**

CARLIER, Anne, COMBETTES, Bernard, *Typologie et catégorisation morphosyntaxique : du latin au français moderne*, Langue Française n°187, Larousse, Paris, 2015.

CARLIER, Anne, LAMIROY, Béatrice, *The emergence of the grammatical paradigm of nominal determiners in French and in Romance: Comparative and diachronic perspectives*, in Canadian journal of linguistics, 2018.

COMBETTES, Bernard, *Grammaticalisation et parties du discours : la différenciation des pronoms et des déterminants en français*, in *A la quête du sens*, ENS édition, Lyon, 2006.

GUILLOT-BARBANCE, Céline : *Le démonstratif en français : étude de sémantique grammaticale diachronique (9<sup>ème</sup> – 15<sup>ème</sup> siècles)*, Peeters, Leuven, Paris, 2017.

KLEIBER, Georges, *Sur la spécialisation grammaticale des démonstratifs en français ancien*, Mélanges H. Nais, De la plume d'oie à l'ordinateur, pp.99-113, 1987

MACHELLO-NIZIA, Christiane, *L'évolution du français – ordre des mots, démonstratifs, accent tonique*, Armand Colin, Paris, 1995.

MARCHELLO-NIZIA, Christiane, *Du subjectif au spatial: l'évolution sémantique des démonstratifs en français*, Langue Française 152, 2006a.

MARCHELLO-NIZIA, Christiane, *Grammaticalisation et changement linguistique*, De Boeck, Bruxelles, 2006b.

SOUTET, Olivier, *Etudes d'ancien et de moyen français*, PUF, Paris, 1992.

SOUTET, Olivier, *Epistémologie et linguistique diachronique : l'exemple des démonstratifs dans l'histoire du français*, in *Les formes du sens*, De Boeck supérieur, Bruxelles, 1997.

ZINK, Gaston, *Morphosyntaxe du pronom personnel (non réfléchi) en moyen français*, Librairie Droz, Genève, 1997

ZUMTHOR, Paul, *La mesure du Monde – Représentation de l'espace au moyen âge*, Seuil, Paris, 1993.

## Configurations discursives et syntaxiques des pronoms *ceci* et *cela* en diachronie (du moyen français au français classique)

Dans le cadre de l'étude proposée, nous nous intéresserons au fonctionnement discursif des démonstratifs *ceci* et *cela* : c'est à partir de l'identification de leur position dans des chaînes de référence construites dans le discours que nous distinguerons, dans un premier temps, les emplois anaphoriques et cataphoriques. Cette première opposition nous permettra de localiser et de délimiter les espaces discursifs avec lesquels les pronoms sont liés et qui rendent possible l'accès au sens de l'entité pronominale. La prise en considération de la perspective diachronique nous permettra également d'observer et de relever des particularités liées aux types de textes et aux séquences qui les structurent. L'exploitation des données du *DMF* fait apparaître des emplois anaphoriques de *ceci* qui, dans le rôle du sujet d'une proposition participe (*ceci estre fait ; ceci estre dit ; ceci voyant ...*), condense le contenu d'un espace discursif plus ou moins étendu du cotexte de gauche (anaphore résomptive) et fonctionne en même temps comme un relais entre deux macro-propositions (déclencheur – action) de la séquence narrative. Nous montrerons que cette fonction de relais – impliquant un pontage vers le cotexte de gauche (fonction anaphorique) et une ouverture vers l'espace discursif qui suit (par anticipation et implication) – est également attestée dans des séquences explicatives et argumentatives où un complément circonstanciel de but axé sur la compréhension d'un fait (*pour mieux cecy entendre*) déclenche souvent une démarche illustrative (*il est bon de donner aucuns exemples*<sup>1</sup>).

Concernant la relation cataphorique, nous partirons des résultats de l'analyse de M. Achard (2010), réalisée à partir d'un corpus des 91 textes en français contemporain (rédigés après 1950) de la base de données FRANTEXT. Achard a montré que les emplois cataphoriques des pronoms démonstratifs *ceci*, *cela* et *ça* recouvrent des organisations discursives très différentes, surtout en fonction de l'intégration sémantique de l'élément qui suit le pronom dans la thématique générale du discours. Tandis que *ceci* possède un « sens cataphorique pur » (Achard), l'entité qui suit les démonstratifs *cela* et *ça* entretient un lien sémantique fort avec le contexte qui précède. Nous retiendrons cette idée d'un continuum cataphorique mise en avant par Achard à partir des réflexions de Smith (2000) concernant la présence d'un espace mental (Fauconnier, 1985) indiquée par les pronoms cataphoriques. L'espace mental désigné par les pronoms annonce donc l'entité à venir mais peut aussi s'appuyer sur différentes informations contextuelles déjà fournies. Nous nous proposons d'observer le parcours diachronique de ce continuum cataphorique tout en questionnant le rôle du genre discursif et donc l'inscription des emplois cataphoriques dans des genres discursifs et des structures séquentielles et syntaxiques spécifiques. Afin de mettre en valeur le facteur séquentiel (et plus particulièrement la comparaison de séquences narratives et explicatives / argumentatives), nous compléterons notre corpus de travail (*DMF*, *Frantext*) par des ouvrages à dominante explicative (traités de médecine, d'agriculture, d'art militaire...).

---

<sup>1</sup> LEGRAND (Jacques) Archiloge Sophie, c.1400, 258 : et se il ne souffist pour soustraire trois cens, tu pues deviser une chascune poitevine en soixante parties, et a ce faire il ne fault nemois multiplier le nombre des poitevines par soixante, et ainsi tousjours tu pues proceder sanz fin. Et pour cecy mieulx entendre il est bon de donner aucuns exemples, et pour ce je figureray quatre nombres. (source *DMF*)

## Eléments de bibliographie

Achard, M. (2010), « Le sens cataphorique de *ceci, cela, et ça* », 2ème Congrès Mondial de Linguistique Française, (2010) 108 , <http://www.linguistiquefrancaise.org/>

Achard, M. (1998), *Representation of Cognitive Structures: Syntax and Semantics of French Sentential Complements*. Berlin and New-York : Mouton de Gruyter.

Fauconnier, G. (1985), *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge, Mass.: MIT Press, and London: Bradford.

Guillot, C. (2017), *Le démonstratif en français : étude de sémantique grammaticale diachronique (9ème-15ème siècles)*, Leuven-Paris, Peeters.

Guillot, C. (2006), « Démonstratif et déixis discursive : analyse comparée d'un corpus écrit de français médiéval et d'un corpus oral de français contemporain », *Langue française*, vol. 152, no. 4, 2006, pp. 56-69.

Guillot C. (2004), « Ceste parole et ceste aventure dans la Queste del Saint Graal , marques de structuration discursive et transitions narratives », *L'Information grammaticale*, 103, p. 29-36.

Kleiber, G (2005), « Démonstratifs et pratique des textes littéraires ». *De la langue au style*, édité par Jean-Michel Gouvard, Presses universitaires de Lyon.

Marchello-Nizia, Ch. (2012), « Savoir lire ce qui est écrit : le rôle des démonstratifs et des possessifs dans le français médiéval ». *Langue et histoire*, édité par Jean-Marie Bertrand et al., Éditions de la Sorbonne.

## “Ça c’est à moi” : usages et spécificités du pronom démonstratif dans les dialogues adulte-enfant

Christine da Silva-Genest<sup>1&4</sup>, Marine Le Mené Guigourès<sup>2&5</sup> & Anne Salazar Orvig<sup>3&6</sup>  
<sup>1</sup>Université de Lorraine, <sup>2</sup>Université du Québec à Montréal, <sup>3</sup>Université Sorbonne Nouvelle  
<sup>4</sup>UR 3450 DevAH & UMR 7118 ATILF, <sup>5</sup>CRBLM & ISC, <sup>6</sup>EA 7345 CLESTHIA

En français, l’acquisition des pronoms démonstratifs est peu décrite, comparativement au développement d’autres formes pronominales. Pourtant, la plupart des études sur corpus soulignent leur usage précoce (entre 1 an et 1 an 6 mois) et ceci dans des proportions importantes (Diessel & Monakhov, 2023 ; Clark, 2003) : une forme sur dix produite par les tout jeunes enfants est un démonstratif, fréquemment associée à des gestes (pointage, offrande, monstration), accentuant sa nature déictique. Selon Diessel & Monakhov (2023), ces proportions tendent à diminuer avec l’âge, au profit des pronoms de troisième personne. Cette évolution correspondrait au développement de la capacité des enfants à poursuivre un fil de continuité de type (proto)-anaphorique. Toutefois, l’étude des contextes pragmatique-discursifs dans lesquels apparaissent les démonstratifs révèle des usages en reprise (da Silva-Genest et al., 2021) de type anaphorique, voire anadéictique (Cornish, 2017).

Il convient donc de s’interroger sur l’éventail et le poids respectif des contextes de production des démonstratifs par les jeunes enfants en dialogue. En quoi leurs usages se distinguent-ils des usages des pronoms de 3<sup>e</sup> personne ? Et, à quels usages les enfants sont-ils exposés ?

Nous avons constitué un corpus de 38 sessions d’interactions adulte-enfant (24 enfants âgés entre 1 an 10 mois et 2 ans 6 mois) à partir duquel nous avons relevé toutes les formes de démonstratifs (*ça, c’est* et les formes disloquées *ça c’est/c’est ...ça*) et comparé leurs usages à ceux des pronoms de 3<sup>e</sup> personne. Pour chaque occurrence, nous avons analysé le statut discursif et attentionnel du référent (nouveau, cognitivement activé dans le contexte, déjà mentionné dans le discours ou réintroduit à distance), ainsi que les caractéristiques dialogiques de leur production (répétition, réponse, continuité et contraste).

Les résultats confirment la fréquence des démonstratifs dans le discours des jeunes enfants comme dans celui des adultes avec lesquels ils interagissent. Les démonstratifs présentent des contextes de production qui leur sont propres et qui se complètent. Ainsi, chez les enfants, les différentes formes de démonstratifs introduisent préférentiellement des référents non préalablement mentionnés mais cognitivement activés dans la situation, alors que les pronoms clitiques de 3<sup>e</sup> personne sont spécialisés dans l’usage anaphorique. Les adultes présentent une configuration similaire pour les démonstratifs forts mais chez eux le démonstratif clitique est moins souvent mobilisé pour les référents activés et plus souvent pour les référents préalablement mentionnés. Par ailleurs, en contexte de reprise, quel que soit le locuteur, les démonstratifs forts marquent souvent un contraste dans le dialogue, alors que le pronom de 3<sup>e</sup> personne maintient le focus d’attention discursif. Le démonstratif clitique se situe entre ces deux valeurs, présentant un plus fort pourcentage de continuité référentielle que les formes fortes.

Ainsi, nos résultats permettent de saisir la diversité de valeurs des démonstratifs auxquelles sont exposés les enfants, et qu’ils reprennent dès le plus jeune âge. Ces spécificités de fonctionnement seront illustrées par une analyse qualitative des données. Nous discuterons enfin la manière dont elles nous éclairent sur la façon dont se construisent les valeurs déictiques, anadéictiques et anaphoriques dans les premières années.

*Mots-clés* : pronoms démonstratifs – valeur déictique – valeur anaphorique - acquisition du langage – dialogue

*Références bibliographiques*

- Clark, Eve Vivienne. (2003). *First language acquisition*. Cambridge: Cambridge University Press.
- Cornish, F. (2017). SN démonstratifs et anadeixis : sens ‘spatial’, ou valeurs tributaires d'une stratégie pragmatique potentielle? *Journal of French Language Studies*, 27(2), 215-239. <https://doi:10.1017/S0959269515000514>
- da Silva-Genest, C., Salazar Orvig, A., Marcos, H., Caët, S., & Heurdier, J. (2021). Chapter 3. Young children’s uses of referring expressions. In A. Salazar Orvig, G. de Weck, R. Hassan, & A. Rialland (Eds.), *The Acquisition of Referring Expressions : A dialogical approach* (p. 82-113). <https://doi.org/10.1075/tilar.28.03das>
- Diessel, H., & Monakhov, S. (2023). Acquisition of demonstratives in cross-linguistic perspective. *Journal of child language*, 50(4), 922–953. <https://doi.org/10.1017/S030500092200023X>

## Contextes d'usage de la forme *la* en français et dans les créoles français

Ulrike Albers, Université de La Réunion / UMR 7023 *Structures Formelles du Langage*  
Patricia Cabredo Hofherr, UMR 7023 *Structures Formelles du Langage* (CNRS & Paris 8)

La présente étude examine un élément de forme /la/ cognat avec l'adverbe locatif *là* du français dans plusieurs créoles à base lexicale française (l'haïtien, le mauricien et le réunionnais). Tandis que *-là* en français a évolué en "intensifieur" post-nominal dans les SN démonstratifs comme *ce sac-là* (*demonstrative renforcer*, Diessel 1999), dans les créoles à base française le *-là* s'est grammaticalisé en affixe phrasal en fin de SN, qui remplit des fonctions d'article défini (voir Wespel 2008 pour les articles définis dans des créoles à base lexicale française, voir Heine et al. 2020 pour la grammaticalisation d'adverbes démonstratifs en démonstratifs adnominaux). Or, on observe que la distribution de *la* varie à travers les créoles examinés. En mauricien, par exemple, le LA est devenu un défini pragmatique. i.e. qui reprend des référents déjà introduits dans le discours, (cf. Wespel 2008) et la forme a perdu sa fonction de démonstratif : elle ne peut plus apparaître dans un contexte comme en (1).

- (1) #Donn mwa [liv la] ek [liv la]. (mauricien)  
donner 1SG livre DEF et livre DEF  
#'Donne-moi le livre et le livre.' (Déprez 2007:270)

En revanche, en réunionnais, l'exemple parallèle à (1) permet un emploi démonstratif.

- (2) Donne amwin [liv -la] èk [liv -la]. (réunionnais)  
donner 1SG.OBJ livre DEM et livre DEM  
'Donne-moi ce livre et ce livre-là'.

Il est bien connu (cf. Greenberg 1978 ; Hawkins 2004, De Mulder & Carlier 2011) que d'un point de vue diachronique, il est commun que les démonstratifs d'une langue s'affaiblissent pour couvrir une portion de plus en plus large du champ sémantique de la définitude. En général c'est le démonstratif distal qui évolue pour couvrir ensuite différents usages déictiques et anaphoriques, et qui apparaît plus tard dans les anaphores associatives, et finalement dans des contextes où il n'est plus fait appel au contexte comme les entités uniques (*le soleil*) et les génériques (*les pandas*).

**DÉMONSTRATIF DISTAL > DÉMONSTRATIF > ARTICLE DÉFINI (définitude pragmatique) > ARTICLE DÉFINI (définitude sémantique) > ARTICLE SPÉCIFIQUE > MARQUEUR NOMINAL<sup>1</sup>**

Au cours du changement linguistique le déterminant peut aussi perdre certains usages, par exemple la faculté de contraster deux référents. Une hypothèse est que l'évolution se reflète de façon typologique : les expressions référentielles des langues du monde occuperaient ces espaces contigus sur l'échelle de définitude (voir e.g. Egedi 2014), ce qui revient à une interprétation de cette échelle comme une carte sémantique (semantic map).

Dans notre étude nous examinons les contextes contrastifs typiques des démonstratifs pour la forme /la/ de l'haïtien, du mauricien et du réunionnais ; nos premiers résultats indiquent que une partie de l'échelle de définitude fonctionne comme une carte sémantique (*semantic map*) cependant, certains emplois d'anaphore associative ne semblent pas être ordonnées entre elles (voir Czardybon 2021 sur le polonais silésien). Nous évaluons cette hypothèse, et nous

<sup>1</sup> Schéma basé sur Greenberg (1978) et De Mulder & Carlier (2011).

examinons quel est le rapport entre démonstratif et défini pragmatique (cf. Löbner 1985, 2011 pour une discussion des types de définitude et des types de concepts).

Nous examinons les contextes suivants afin de déterminer la distribution de la forme LA et les emplois démonstratifs qui lui sont associés :

Démonstratifs – définis anaphoriques (*un garçon et une fille ... la fille*) – différents types d’anaphores associées (*un ville - la mairie, une voiture – le moteur, un enfant – la mère, un mariage – la mariée, une symphonie – le compositeur*) – définis d’unicité (*le soleil, le président*) – génériques (*les pandas*). Nous nous appuyons pour cela sur des exemples extraits de la littérature, et de divers corpus écrits et oraux, ainsi que sur des jugements obtenus auprès de consultants natifs.

Mots-clés : Démonstratifs ; définis ; langues créoles à base lexicale française

### Références

- Czardybon, A. 2021. *Definiteness in a Language without Articles – A Study on Polish*, Berlin, Boston: düsseldorf university press. Open access <https://doi.org/10.1515/9783110720426>
- Diessel, H. 1999. The morphosyntax of demonstratives in synchrony and diachrony. *Linguistic Typology* 3 (pp. 1–49). <https://doi.org/10.1515/lity.1999.3.1.1>
- Egedi, B. 2014. The DP-cycle in Hungarian and the functional extension of the noun phrase. In K. É. Kiss (eds.) *The Evolution of Functional Left Peripheries in Hungarian Syntax* (pp. 56–82). Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198709855.003.0003>
- Greenberg, J. H. 1978. How does a language acquire gender markers? In C. A. Ferguson, J. H. Greenberg & E. A. Moravcsik (eds.) *Universals of Human Language 3: Word Structure* (pp. 47–82). Stanford: Stanford University Press.
- Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199252695.001.0001>
- Heine, B., Kuteva, T., Long, H., Narrog, H., & Wu, F. 2020. Where do demonstratives come from? *STUF - Sprachtypologie und Universalienforschung*, 73(3) (pp. 403–434). <https://doi.org/10.1515/flin-2021-2027>
- Löbner, S. 1985. Definites. *Journal of semantics*, 4-4 (pp. 279–326). <https://doi.org/10.1093/jos/4.4.279>
- Löbner S. 2011. Concept types and determination. *Journal of Semantics*, 28 (pp. 279–333). <https://doi.org/10.1093/jos/ffq022>
- De Mulder, W. & Carlier, A. 2011. The grammaticalization of definite articles. In Narrog, H. & Heine, B. (eds.), *The Oxford handbook of grammaticalization* (pp. 522–535). Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199586783.013.0042>
- Wespel, Johannes. 2008. *Descriptions and their domains: the patterns of definiteness marking in French-related creoles*. Stuttgart: Universität Stuttgart dissertation.



### *Definiteness, Demonstrative Pronouns, and the Absence of the Definite Article in Old Russian*

It is well known cross-linguistically that a definite article can originate from the demonstrative pronoun. Among the Slavic languages, Bulgarian and Macedonian are a testament to this process of grammaticalization, in that both languages have developed the definite article. The rest of the Slavic languages including Russian did not develop the definite article. Although the problem of the diachronic development of the demonstrative pronouns and the emergence of the definite article has been addressed in Old Bulgarian (Dimitrova-Vulchanova and Vulchanov 2012) and Old Church Slavonic (Kuznetsov 2020), it has not received much attention in Old Russian. This article aims to analyze the evolution of the demonstrative pronouns in Old Russian and shed light on why the definite article did not emerge from the demonstratives. I argue that over time, the Old Russian demonstrative pronouns *сѣ* (*s'*) ('this') and *тѣ* (*tŭ*) ('that') have lost their deictic meaning and began to function as anaphoric definites. However, they failed to grammaticalize as the definite article. There are two main claims that I make in the article. First, I suggest that the Old Russian noun phrase should be analyzed as an NP without a DP layer. This claim is supported by an obvious absence of the definite article in Russian. I show that the absence of the definite article correlates with a number of generalizations about article-less languages, such as left-branch extraction, adjunct-extraction, scrambling, etc. Following research (Corver 1992; Bošković 2005, 2009, 2013; Willim 2000; Despić 2013; Zlatić 2013) on the absence of the DP projection in article-less languages, I suggest that the demonstratives should be analyzed as structural adjuncts within NP. Second, regarding diachronic change, I claim that the syntactic position occupied by the demonstrative pronouns, namely not the 2<sup>nd</sup> Wackernagel slot within a noun phrase, prevented them from grammaticalizing into the definite article. To substantiate these two claims, I have created a diachronic sub-corpus of the syntactic distribution and discourse use of the demonstrative pronouns *сѣ* ('this') and *тѣ* ('that') based on the corpus of Old Russian texts dating from the 11<sup>th</sup> to the 15<sup>th</sup> centuries.

In Old Russian, the demonstratives were used as bare arguments (1-2) or with a noun in noun phrases (3). The bare demonstrative *s'* ('this') appears in the subject position (1). The bare demonstrative *tŭ* ('that') appears in the object position (2). The demonstrative *tŭ* ('that') appears in a noun phrase (3). In (3), the word order of the Dem+NP is scrambled; in that the demonstrative follows the adjective not precedes it. In Contemporary Standard Russian, bare demonstratives are disallowed in subject and object positions (4). These diachronic data raise the questions: 1). Why were bare demonstratives allowed in subject and object positions in Old Russian? 2). Why was scrambling allowed in Dem+NP in Old Russian? 3). Why did the demonstratives not grammaticalize as the definite article in Old Russian similarly to Bulgarian and Macedonian?

- (1) и сѣ оубо любѣвь многоу имѣаше къ сѣѣма·  
i s' ubo ljubŭv' mnogu imējaše kŭ styima  
and this emph love much had to saint  
'And this one had much love for the saints.'

(*Boris and Gleb's Miracles according to the Uspenskij Collection, 2<sup>nd</sup> half of the 11<sup>th</sup> century*)

- (2) и тѣ покажнень будеть по закону грѣцкому  
i tŭ pokaznenŭ budet' po zakonu gr'ckomu  
and that punish be.fut on law Greek  
'And that one will be punished according to the Greek law.'

(*The Primary Chronicle according to the Ipat'ev Collection, 1110s*)

- (3) бѣ же немилосердыи тѣ.  
 bū že nemiloserdyj tŭ  
 God emph unmerciful that  
 ‘the unmerciful God’

(*Russian Truth*, middle of the 11<sup>th</sup> century)

- (4) \*Sej/etot/tot prišel domoj.  
 this came home  
 ‘This came home.’

In my analysis I assume Minimalism and Distributed Morphology as my theoretical frameworks. Regarding my 1<sup>st</sup> claim, I argue that bare demonstratives in the subject and object positions in Old Russian should be analyzed as definites having their own syntactic projection DemP. The demonstrative itself is the syntactic head Dem<sup>o</sup> encoding the [±definiteness] feature. It takes the Num P as its complement. The Num<sup>o</sup> head carries [±atomic] feature which encodes the singular/plural number distinction (5). Old Russian examples (1-2) yield a definite interpretation because the demonstratives ‘this’ and ‘that’ are overtly realized and filled by a Vocabulary Item specified with the [+definiteness] feature. Consider example (6) in which the demonstrative is realized by the plural exponent *eti* (‘these’). Even though the demonstrative is overtly realized, (6) yields both a definite and an indefinite interpretation: ‘these three definite letters (mentioned previously)’ or ‘three indefinite letters.’ I argue that the reason why (6) yields both a definite and an indefinite interpretation is that a bare plural noun *plat’ja* (‘dresses’) is underspecified for definiteness and can be interpreted as both definite and indefinite at Logical Form. The sentence in (6) has a simple syntactic structure given in (7). I further argue that the syntactic structure in (7) is a simple NP crucially without a DP layer modified by an adjunct DemP. The evidence for treating Dem+NP as NP not DP comes from generalizations correlating with article-less languages; these generalizations include left-branch extraction, adjunct-extraction, and scrambling (8b). As is well-known, languages with the definite article do not allow left-branch extraction. Thus, my analysis supports DP/NP parametrization and restricting linguistic differences to the lexicon.

- (5) [DemP [Dem<sup>o</sup>] [Num P [Num<sup>o</sup>]]]  
 (6) Eti tri plat’ja *Contemporary Standard Russian*  
 these three dress-.gen.pl  
 ‘these three letters’  
 (7) [Dem P [NP [N<sup>o</sup>]]]  
 (8) a. eti tri krasivyx plat’ja  
 these three beautiful dress-gen.pl  
 b. krasivyx etix trex plat’ev Left-Branch Extraction  
 c. trex etix plat’ev krasivyx Scrambling

Regarding my 2<sup>nd</sup> claim, I argue that demonstratives within Dem+NP phrases should be analyzed as structural adjuncts. Furthermore, unlike Bulgarian and Macedonian counterparts, the demonstratives never assumed the 2<sup>nd</sup> Wackernagel position within NP which prevented them from grammaticalizing as the definite article. Old Russian NPs exhibit scrambling of the constituent parts which would not be possible if the demonstrative occupied the 2<sup>nd</sup> Wackernagel position within NP (cf. 3 and 9).

- (9) v’se to selo Old Bulgarian  
 all.acc.neut art.acc.neut place (*Codex Suprasliensis*, 207, 4)

## From Demonstratives to Nominal Markers in the History of Georgian: Greenberg's Predictions Proving to Be Correct

This talk aims at reviewing syntactic factors that conditioned the grammaticalization of demonstratives into definite and specific articles in Old Georgian and addresses a similar process in Old Georgian's unwritten predecessor (Early Georgian) in which it had gone further, up to the development of nominal markers. This evidence supports so far unsubstantiated claims that Greenberg made in his 1978 milestone paper on the emergence of nominal markers from demonstratives.

Unlike Modern Georgian, Old Georgian (5<sup>th</sup>–11<sup>th</sup> centuries AD) had both definite and indefinite inflected articles (Šaniže 1953, 1976; Imnaišvili 1956; Fähnrich 1994). The former rested upon the demonstratives — *ese* 'this', *ege* 'that', and *igi* 'yonder.' These could either precede their heads and thus remain deictically marked demonstratives or follow those, thus becoming articles:

- (1) *ese*                      *k'ac-i*              *ara*      *ar-s*                      *γmrt-isa=gan.*              (Jo. 9:16)  
 DEM.PROX:NOM    man-NOM      NEG      be.PRS-S3SG      God-GEN=EL  
 'This man is not from God'.
- (2) *ara*      *v-i-c-i*                      *k'ac-i*              *ese.*                      (Mk.14:71)  
 NEG      S1-VAL-know-PERM      man-NOM      DEM.PROX//DEF.ART:NOM  
 'I do not know **this//the** man'.

Once attached to the right of the noun, their deictic semantics could bleach thus contributing to their grammaticalization into definite articles.

Should the NP contain an attributive constituent, the article appeared between the head and the attribute (Imnaišvili 1956):

- (3) *kar-i*              *igi*                      *zlier-i*                      (Mt. 14:3)  
 wind-NOM      DEF.ART:NOM      strong-NOM  
 'the strong wind'

The definite article also developed the so-called "specific" function (cf. Greenberg 1978; Harris 1985:77):

- (4) *da*      *ay-qd-a*                      *nav-s=a*                      *ma-s...*                      (Mt. 9:1)  
 and      PRV-step-S3SG.AOR      boat-DAT=EMPH      SPEC.ART:OBL-DAT  
 'And (Christ) stepped into **a (certain)** boat...'      cited after (Harris 1985:78)

This function turned out to be the endpoint of this grammaticalization pathway in Old Georgian. However, a similar process must have already taken place in Early Georgian, before the first extant texts were written, and gone one step further, as specific articles grammaticalized into the markers of NOM (Šaniže 1925, 1953) and ERG (Čikobava 1939; Šaniže 1953:620), thus splitting the function of indefinite ABS into the separate marking of NOM and ERG:

- (5) *kal*                      →      *kal*                      *igi*                      →      *kal-i*  
 woman:ABS                      woman:ABS      DEF.ART:NOM                      woman:NOM  
 'woman'                      'the woman'                      'woman'
- (6) *kal*                      →      *kal*                      *ma-n*                      →      *kal-man*  
 woman:ABS                      woman:ABS      DEF.ART:OBL-ERG                      woman-ERG  
 'woman'                      'the woman'                      'woman'

Particular oblique cases, such as GEN, DAT, and INSTR, attached the emphatic clitic =*a* (Šaniže 1953:48; Čikobava 1956:265–268; Harris 1985:77):

(7)	<i>kal-s</i>	→	<i>kal-s</i>	<i>a-ma-s</i>	→	<i>kal-s=a</i>
	woman-DAT		woman-DAT	PROX-DEF.ART:OBL-DAT		woman-DAT=EMPH
	‘to woman’		‘to <b>the</b> woman’			‘to <b>the</b> woman’

This was a remainder of the proximal deictic vowel *a* originally prefixed to the oblique stem of the demonstrative (cf. *a-ma-* vs. distal *i-ma-* and unmarked *ma-*).

Typologically, the factual evidence reconstructed for Early Georgian vindicates the grammaticalization pathway of demonstratives into case markers (the third sub-type of “nominal markers” according to Greenberg), via definite and specific articles, which Greenberg could pre-estimate but not substantiate. Furthermore, the grammaticalization process in issue turned out to be replicated anew, although without reaching the case marker endpoint milestone, as in Early Georgian, at a later stage of the history of the language, in the Old Georgian period, which fact supports the existence of recurrent grammaticalization pathways.

#### Abbreviations

ABS – absolutive; AOR – aorist; ART – article; DAT – dative; DEF – definite; DEM – demonstrative; EL – elative; EMPH – emphatic marker; ERG – ergative; GEN – genitive; INSTR – instrumental; NEG – negation marker; NOM – nominative; NP – noun phrase; OBL – oblique stem; PERM – permansive; PROX – proximal; PRS – present; PRV – preverb; S – subject person marker; SG – singular; SPEC – specific; VAL – valency operator.

#### References

- Čikobava, A. 1939. Motxrobiti brunvis genezisisatvis kartvelur enebši [On the Emergence of the Ergative Case in the Kartvelian Languages]. (*Works of the Tbilisi State University*. Vol. X). Tbilisi.
- Čikobava, A. 1956. Ist’oriulad gansxvavebuli ori morpologiuri ṭ’ip’isatvis kartul brunvata šoris [On Two Historically Distinct Morphological Types of Georgian Cases], in: Topuria, V. (ed.). *Saxelis brunebis ist’oriisatvis kartvelur enebši* [On the History of Nominal Cases in the Kartvelian Languages]. Tbilisi: Universit’et’i: 265–268.
- Fähnrich, H. 1994. *Grammatik der Altgeorgischen Sprache*. Hamburg: Helmut Buske Verlag.
- Greenberg, J. 1978. How Does a Language Acquire Gender Markers? in: Greenberg, J., Ferguson, C. and E. Moravcsik (eds.). *Universals of Human Language, I: Method and Theory*. Stanford: Stanford University Press: 47–82.
- Harris, A. 1985. *Diachronic Syntax: The Kartvelian Case*. (Syntax and Semantics. Vol. 18.) Academic Press.
- Heine, B. and T. Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge University Press.
- Imnaišvili, I. 1956. Nac’evari zvel kartulši [Article in Old Georgian]. (*Works of the Tbilisi State University*. Vol. 61.). Tbilisi: 249–276.
- Šaniže, A. 1953. *Kartuli enis gramat’ik’is sapuzvlebi* [Foundations of Georgian Grammar]. Tbilisi.
- Šaniže, A. 1976. *zveli kartuli enis gramat’ik’a* [A Grammar of the Old Georgian Language]. Tbilisi.
- Vogt, H. 1947. Le système des cas en géorgien ancien. *Norsk Tidsskrift for Sprogvidenskap* XIV: 98–140.

## From demonstrative to definite article: the case of Gothic

Ion Giurgea

The “Iorgu Iordan – Al. Rosetti” Institute of Linguistics of the Romanian Academy

It has been argued that a first stage in the development demonstrative > definite article is that of a determiner generalized for anaphoric and exophoric uses (Hawkins 2004), which would correspond to the “strong articles” documented for various West Germanic varieties (De Mulder & Carlier 2011). The Gothic determiner *sa* is a candidate for this stage, being traditionally described as an anaphoric article (Bernhardt 1874, Streitberg 1920, Pimenova 2017, Miller 2019). The Gothic corpus provides a good testing ground for this article system, because it consists, for the most part, in translations from (New Testament) Greek, a language that had a fully developed definite article. I studied this system by examining all the correspondences of the Greek definite article in the (extant parts of) the gospels of Matthew and Luke, using the electronic corpus available at [www.wulfila.be](http://www.wulfila.be).

**Results:** (i) the article is obligatory in definite phrases without an overt noun (but with an overt modifier, e.g. *þaim af hleidumein ferai* ‘ART.PL.DAT from left part’ = Gr. *τοῖς ἐξ ἐβωνύμων*); elsewhere: (ii) the article is almost obligatory for anaphoric definites, including the so-called ‘discourse deixis’ (reference to propositions or events described in the previous utterance): 218 ex. +*sa* vs. 12 without *sa* (plus 36 situations in which a non-anaphoric analysis is possible); (iii) the article is also regularly used for the exophoric use (8 ex.): although situation-based uniqueness is satisfied, the article appears if the referent is perceptually accessible, cf. the distinct rendering of the definite article in the following example:

(1) *urraisands nim þana ligr þeinana jah gagg in gard þeinana* (Mk.9.6)  
*ἐγερθεῖς ἄρον σου τὴν κλίνην καὶ ὑπαγε εἰς τὸν οἶκόν σου*  
‘[He then said to the paralytic:] Rise, take up **your bed** and go to **your house**.’

(iv) the article can be used, but is more often dropped, in cases of non-anaphoric definites unique in a restricted situation, based on shared knowledge, and associative anaphora; (v) for other types of definites, the article is overwhelmingly not translated (96%). Some exceptions can be accounted for by a preference for using the article with pronominal modifiers and by the licensing effect of relative clauses (also attested for modern Germanic strong articles and for demonstratives, see Wolter 2006, Schwarz 2009, Simonenko 2014; but the Gothic article is not obligatory with relative clauses); (vi) the article systematically appears for the second member of Greek double- or poly-definites, e.g. *hlaif unsarana þana sinteinan* ‘bread our the daily’ = *τὸν ἄρτον ἡμῶν τὸν ἐπιούσιον* (21 +*sa*, 1 without), which under a close apposition analysis is covered by the generalization (i).

The data indicate two paths that act conjointly in the emergence of the definite article: the generalization of a demonstrative for all cases of referents available in the discourse situation and the use of demonstratives in noun phrases without an overt noun.

### References

- Bernhardt, Ernst. (1874). *Der Artikel im Gotischen*. Erfurt: Gerhardt & Schreiber.
- De Mulder, Walter & Anne Carlier. 2011. Definite Articles. In Heine Bernd & Narrog Heiko (eds.) *The Oxford Handbook of Grammaticalization*. Oxford: Oxford University Press.
- Hawkins, J. A. 2004. *Efficiency and complexity in grammars*. Oxford: OUP.
- Miller, D. Gary. 2019. *The Oxford Gothic Grammar*. Oxford: Oxford University Press.
- Pimenova, Natalia. 2017. „Figur“, „Grund“ und narrative Kontinuität in der Grammatikalisierung des gotischen Protoartikels. *Zeitschrift für Germanistische Linguistik* 45(3): 359-391.
- Schwarz, Florian. 2009. *Two types of definites in natural language*. PhD diss., University of Massachusetts Amherst.

- Simonenko, Alexandra. 2014. *Grammatical ingredients of definiteness*. PhD diss., McGill University, Montreal.
- Streitberg, Wilhelm August. 1920. *Gotisches Elementarbuch*. Heidelberg: Winter.
- Wolter, Lynsey Kay. 2006. *That's that: The Semantics and Pragmatics of Demonstrative Noun Phrases*. PhD dissertation, University of California – Santa Cruz.

## **The role of ECCE in the emergence of the new Romance demonstrative systems**

The aim of this paper is to revise previous theories concerning the reorganisation of the demonstrative system in the transition from Late Latin to Romance. In particular, it will highlight the pivotal role of the presentative ECCE. The similar but different configurations of the three-term systems in Central and Southern Italo-Romance, on the one hand, and of Ibero-Romance, on the other, with the Hearer as an unstable anchoring point, will provide crucial evidence in this respect. Conflicting accounts in the literature have linked the emergence of the second person to either the use of ECCE (Vincent 1999) or to the subjectivisation process (Stavinschi 2012). Although they capture important pieces of the puzzle, neither of these theories can fully explain what happened; a more comprehensive and convincing framework is needed.

The data has been gathered from different sources, including previous research on the topic (such as Ledgeway & Smith 2016), combined with data from the Corpus *OVI* dell'italiano antico, Corpus Informatizat del Català Antic (CICA), Corpus del castellano medieval (OLDES), Corpus Histórico do Português Tycho Brahe. It has also benefited from the new insights on Late Latin (Płocharz 2021), in particular on the distribution of the terms inside the dyad of conversation.

It can be shown that the linguistic change followed a two-step process; the first one can be further broken down in two: the grammaticalisation stage appears quite distinct from the intersubjectification process (Hopper and Traugott 2003, Boye and Harder 2016), while the second stage is due to systemic pressure. Last but not least, demonstratives behave differently depending on their syntactic function. This doesn't concern only adverbs, as previously shown, but also adnominals and pronominals. Despite often being treated together in the literature, they appear to behave quite differently and play different roles in the grammaticalisation process, with adnominals far more sensitive to change.

Apart from shedding light on some aspects of the evolution of the Romance demonstratives, this paper is hoped to contribute to a more general understanding of the internal structure of the demonstratives (see Terenghi 2023 and Dayal, forthcoming), on the one hand, as well as to the ongoing debate on grammaticalisation and subjectification, on the other (Van Olmen & Cuyckens 2016).

## References

- Bonilla Carvajal, Camilo Andrés, 2020, "The syntax of the Latin presentative adverb *ecce*: Relation to focus phrase", *Journal of Latin Linguistics*, vol. 19, no. 1, 27-57.
- Boye, Kasper & Harder, Peter, 2014, "(Inter)subjectification in a functional theory of Grammaticalization", *Acta Linguistica Hafniensia*, Vol. 46, No. 1, 7–24.
- Dayal, Veneeta (forthcoming) "Bare Nominals, Definites, Demonstratives: What competes with What".
- Ledgeway Adam, Smith John Charles, 2016. "Deixis", in Adam Ledgeway and Martin Maiden (eds), *The Oxford Guide to the Romance Languages*. Oxford: Oxford University Press, 879-896
- Płocharz, Piotr, 2021, *Le démonstratif en latin tardif et altimédiéval : Étude de la documentation notariale privée rédigée en Gaule et en Italie (7e-10e siècles)*. Linguistique. École Normale Supérieure de Lyon, Université de Lyon, 2021.
- Stavinschi, Alexandra, 2012, "On the development of the Romance demonstrative systems: Historical remarks and theoretical conclusions", *Diachronica*, Volume 29, Issue 1, 72 - 97
- Terenghi, Silvia, 2023. *Missing Person: Structure and change in Romance demonstratives*. [Doctoral thesis 1 (Research UU / Graduation UU), Universiteit Utrecht].
- Van Olmen Daniël, Cuyckens Hubert, 2014, "Grammaticalization and (inter)subjectification", *Acta linguistica hafniensia*, Vol.46 (1), 1-6.
- Vincent, Nigel, 1999, *The synchrony and diachrony of deixis in Old Neapolitan (with special reference to the Cronache e ricordi of Loise De Rosa)*. MS, University of Manchester



## Reinforced and non-reinforced plural distal demonstratives in Eastern Basque texts

### 1. Introduction

According to Schulze (2003: 292), “demonstrative pronouns often constitute quite short-lived paradigms”. The morphological rearrangement of demonstrative forms is illustrated by well-known diachronic evolutions like that of Lat *ille* ‘that’ → \**accu ille* → Sp *aquel* ‘that’ (Piel 2005: 109-125), where a new series of demonstratives reinforced with the particle \**accu* emerged, while the original demonstrative *ille* ‘that’ specialized into the personal pronoun *él* ‘he’. This paper addresses a similar evolution in Basque, and explores the possible motivations of the use of reinforced vs non-reinforced demonstratives when both types are present in texts by the same author.

### 2. The grammaticalization of the definite inflection and the remorphologization of the plural distal demonstratives in Basque

As in other European languages (for German, see now Flick 2020: v), the Basque definite inflection emerged in the Middle Ages through the grammaticalization of distal demonstratives (Mitxelena 1979: 208). This is illustrated in (1) for the plural inflection:

- |   |   |   |
|---|---|---|
| (1) <i>etxe hek</i> house that.PL ‘those houses’    | > | <i>etxe-ek</i> house-the.PL ‘the houses (ERG)’    |
| <i>etxe hen</i> house that.PL.GEN ‘of those houses’ | > | <i>etxe-en</i> house-the.PL.GEN ‘of the houses’   |
| <i>etxe hetan</i> house that.PL.LOC ‘in the houses’ | > | <i>etxe-etan</i> house-the.PL.LOC ‘in the houses’ |

In historical records, however, two types of plural distal demonstratives are present: simple-stemmed demonstratives (*hek*, *hen*, *hetan*...) and demonstratives with a bisyllabic stem (*haek*, *haen*, *haetan*...). Demonstratives of bisyllabic stems are considered as morphologically complex (Azkarate & Altuna 2001; Manterola 2015): their initial reinforcing particle *ha-* may be related to a particle of deictic force (in some dialects, *ha* is indeed the singular demonstrative ‘that’), also present in exclamatory sentences such as (2):

- |                          |                       |
|--------------------------|-----------------------|
| (2) <i>ha gau luzea!</i> | (Axular, Guero, 1643) |
| that night long          |                       |
| ‘What a long night!’     |                       |

This paper assumes that the remorphologization of plural distal demonstratives by means of the reinforcing particle *ha-* occurred somewhat concurrently with the grammaticalization process depicted in (1). However, contrary to languages like Spanish, where *aquel* ‘that’ (< *accu*-\* + *ille*) has completely taken over the principal function as a demonstrative (displacing *él* ‘he’ to a pronominal function), in Eastern Basque *hek*-type demonstratives and innovative *haek*-type demonstratives coexist.

### 3. The hypothesis and the corpus

By examining demonstrative pairs like *aqueste* ~ *este* ‘this’ in Old Spanish, Piel (2005: 110-125) defines the function of *aqu-* in Old Spanish as some sort of “filtre contrastif”. This paper tests the hypothesis of whether a similar discourse-based criterion may be described in order to account for the alternative use of *hek*-type vs *haek*-type demonstratives. The texts browsed in this corpus-based study include 16<sup>th</sup> and 17<sup>th</sup> century Eastern authors: Etxepare (1545), Leizarraga (1571), Oihenart (1657).

### 4. Preliminary and expected results

It has been observed that the demonstratives in grammatical cases (absolute, ergative, dative) tend to present the reinforced stem *hae-* more often than other cases (local cases, instrumental, etc.). At the moment, no conclusive results have been attained in what regards the possibility of a contrastive use of the two types of demonstrative, and no clear discourse-based pattern has been described.

## 5. References

- Axular, Pierre. 1643. *Guero*. Bordeaux: Milanges.
- Azkarate, Miren & Altuna, Patxi. 2001. *Euskal morfologiaren historia*. Donostia: Elkarlanean.
- Flick, Johanna. 2020. *Die Entwicklung des Definitartikels im Althochdeutschen. Eine kognitiv-linguistische Korpusuntersuchung (Empirically Oriented Theoretical Morphology and Syntax 6)*. Berlin: Language Science Press.
- Manterola, Julen. 2015. *Euskararen morfologia historikorako: artikulak eta erakusleak = Towards a history of Basque morphology: articles and demonstratives*. Doctoral dissertation: University of the Basque Country.
- Mitxelena, Luis. 1979. Miscelánea filológica vasca I: *vasc. on da / gauza ona da*. *Fontes Linguae Vasconum* 11, 208-218.
- Piel, Amélie Anne. 2005. *Les déictiques déclinables et indéclinables de l'espagnol médiéval: étude synchronique*. Lille: ANRT.
- Schulze, Wolfgang. 2003. "The diachrony of demonstrative pronouns in east Caucasian. In D. A. Holisky & Kevin Tuite (eds.): *Current Trends in Caucasian, East European, and Inner Asian Linguistics: Papers in Honor of Howard I. Aronson (Current Issues in Linguistic Theory 246)*, 291-348. Amsterdam: John Benjamins Publishing Company.

## From determiners to nominalizers: the case of Modern South Arabian

This talk will provide an analysis of the phenomenon of ‘stand-alone nominalization’ (Yap et al. 2011) – i. e. nominalization in a syntactically-independent context – in three Modern South Arabian languages (henceforth MSAL – Semitic, Afroasiatic), namely Mehri, Baṭṭari and Šheret. Data is taken from Rubin (2018), Morris & Gasparini (Forth.) and original field data. Stand-alone nominalization can be understood as a broad case of insubordination (Evans & Watanabe 2016: 1). This innovative structure likely developed because of the diachronic loss of participial forms, which in turn are widely attested elsewhere in Semitic. Traditionally, stand-alone nominalized clauses in MSAL are called ‘pseudorelative’ clauses (Wagner 1953), since these are main clauses headed by an associative particle – *d(a)*- in the case of Mehri, *(ə)l*- for Baṭṭari and *d*- for Šheret – more commonly found as a genitive/relative/attributive marker.

I describe the associative particle as a nominalizer (Shibatani & Makhashen 2019), according to its substantial/determinative nature from a comparative Semitic perspective (Huehnergard & Pat-El 2018).

Roughly speaking, a stand-alone nominalized verb encodes a durative (1), progressive (2) or anterior (3) aspect depending on verbal Aktionsart and TMA properties (with micro-variation between the languages):

(1) Baṭṭari

*el fēzʕak*

NMLZ be\_scared\PFV:1S

‘I am terrified’

(2) Šheret

*d=ifteréren*

NMLZ = 3S.M:yawn\IND

‘he is yawning’

(3) Mehri

*də = xtáwn*

*ḥə = bān = ihəm*

NMLZ = circumcize\PFV.3PL.M DET = son\PL = 3PL.M

‘they circumcized their sons’

Nominalization is significant in relation to discourse structure and textual organization since such clauses are mainly used to convey background information (Gasparini 2021).

This talk will deal with the MSAL nominaliser within its wider Semitic context by investigating this interesting grammaticalisation pattern. According to Yap et al. (2011: 18), nominalizers and demonstratives (in their pronominal function) both possess a ‘reifying’ function, since they create an ontological entity from an event or state, allowing speakers to comment on it by virtue of their definite nature: this would explain why demonstratives are often grammaticalized in discursive elements that indicate the speaker's attitude in relation to the narrative structure and the pragmatic and textual function.

### References

- Comrie, Bernard & Sandra A. Thompson. 2007. Lexical Nominalization. In Shopen, Timothy (eds.), *Language Typology and Syntactic Description* 3. 334-381. Cambridge: Cambridge University Press.
- Evans, Nicholas & Honoré Watanabe. 2016. The dynamics of insubordination: An overview. In Evans, Nicholas & Honoré Watanabe (eds.), *Insubordination*, 1-38. Amsterdam: John Benjamins.
- Gasparini, Fabio. 2021. Nominalizzazione indipendente in Sudarabico Moderno. In: Ballarè, Silvia & Guglielmo Inglese (eds.), *Tipologia e sociolinguistica: verso un approccio integrato allo studio della variazione. Atti del workshop della Società di Linguistica Italiana, 10 settembre 2020*, 177–206. Milan: Officina ventuno.
- Huehnergard, John, & Na‘ama Pat-El. 2018. The origin of the Semitic relative marker. *Bulletin of the School of Oriental and African Studies* 81(2). 191-204.
- Morris, Miranda J. & Gasparini, Fabio. Forthcoming. *A descriptive grammar of Baṭāḥrēt (Baṭḥari)*. Wiesbaden: Harrassowitz.
- Rubin, Aaron D. 2018. *Omani Mehri: a new grammar with texts*. Leiden: Brill.
- Shibatani, Masayoshi & Khaled A. Bin Makhshen. 2019. Nominalization in Soqotri, a South Arabian language of Yemen. In Wetzels, Leo W. (eds.), *Endangered languages: Contributions to Morphology and Morpho-syntax*, 9-31. Leiden: Brill.
- Wagner, Ewald. 1953. *Syntax der Mehri-Sprache unter Berücksichtigung auch der anderen Neusüdarabischen Sprachen*. Berlin: Akademie-Verlag
- Yap, Foong Ha, Karen Grunow-Hårsta & Janick Wrona. 2011. Introduction: Nominalization strategies in Asian languages. In Yap, Foong Ha, Karen Grunow-Hårsta & Janick Wrona (eds.), *Nominalization in Asian languages: diachronic and typological perspectives*, 1-57. Amsterdam: John Benjamins.



**Langues & Langage  
à la croisée des Disciplines**

**1ère Rencontre annuelle LLcD**

Colloque : 9-11 septembre, 2024

Ecole d'été : 12-14 septembre, 2024

Sorbonne Université - 28 Rue Serpente, 75006 Paris, France



## **Workshop 5**

**Subordination in Creole languages**

**La subordination dans les langues créoles**

### **SCroll: the database of subordination in Creole languages**

This paper introduces SCroll, a database tailored for the typological study of clausal subordination in Creole languages. SCroll (Manfredi & Quint, forth.) offers researchers a comprehensive tool to investigate both formal and semantic aspects of the encoding of subordinate relations in ten Creole languages with distinct lexifiers (English, French, Portuguese, Spanish, Arabic) and different substrate languages (Niger-Congo, Nilo-Saharan, Austronesian).

SCroll aims to contribute to renewing the semantic and morphosyntactic parameters for cross-linguistic comparison of subordinate clauses (Cristofaro 2003). The database stands out from other databases for typological comparison of Creole languages (Michaelis et al. 2013) in that it provides two integrated query interfaces, allowing users for fine-grained research on adverbial, complement, and relative subordinate clauses in Creole languages. The first interface enables a multilevel comparison of the formal encoding of these relations, offering empirical examples and qualitative insights into morphosyntactic variation. The second interface focuses on the semantic features of subordinating devices across the languages of the sample and it helps to trace back the etymological and functional origins of these devices, gaining insights into their historical development, semantic properties and domains of use.

The paper outlines SCroll's primary query functions, emphasizing their relevance in identifying formal parameters crucial for defining subordinate clauses. Examples of these parameters include the (non-)sharing of one and the same subject between main and subordinate clauses, the presence/absence of subordinating devices, TAM marking, and the presence/absence of pronominal arguments. The paper also underscores SCroll's significance regarding historical linguistics, particularly in examining the grammaticalization processes of adverbial, complement, and relative subordinating devices (see Kortmann 1996 for European languages). Positioned as an open-access resource for both creolists and typologists, SCroll emerges therefore as an innovative tool meeting the general demand for cross-linguistic comparison databases while addressing the unique challenges posed by complex syntax in Creole languages (Veenstra 2015; Diessel 2004).

#### **References:**

- Cristofaro, S. 2003. *Subordination*. Oxford: Oxford University Press.
- Diessel, H. 2004. *The Acquisition of Complex Sentences*. Cambridge: Cambridge University Press.
- Kortmann, B. 1996. *Adverbial Subordination: A Typology and History of Adverbial Subordinators Based on European Languages*. Berlin: Mouton de Gruyter.
- Manfredi, Stefano & Nicolas Quint (eds.), forth. *SCroll: the database of subordination in Creole languages*. HumaNum.
- Michaelis, S. M., P. Maurer, M. Haspelmath and M. Huber (eds.) 2013. *The Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Veenstra, T. 2015. The development of subordination. In A. Trotzke and J. Bayer (eds.), *Syntactic Complexity across Interfaces*. Boston: De Gruyter Mouton, pp. 137-162.

303 words

**Des subordonnées sans subordonnant en créole mauricien**  
**Muhsina Alleesaib<sup>1</sup>, Hassamal Shrita<sup>3</sup>, Fon Sing Guillaume<sup>2</sup>, Florigny Guilhem<sup>1</sup>, Abeillé Anne<sup>2</sup>**  
<sup>1</sup>University of Mauritius, <sup>2</sup> Université Paris Cité (LLF), <sup>3</sup> Mauritius Institute of Education

La subordination a été peu étudiée en créole mauricien (Police-Michel et al 2011). Nous avons annoté les subordonnées complétives (1a) et circonstancielles (2a) dans trois corpus oraux (Alleesaib 1999, Chady 2018, Florigny 2021) et un corpus écrit (Virahsawmy 2007, 2017). Les subordonnées montrent un taux important d'omission de *ki* ('que') (Alleesaib et al. 2024, Véronique 2007). Un indice de subordination est la forme longue (FL) du verbe principal, à la différence de la forme courte (FC) qui précède les compléments (1b) et certains ajouts (2b) non phrastiques (Hassamal et al. 2018):

- (1) a. Zot kone [(ki) li pe dir laverite.]  
 3PL savoir.FL (SUB) 3SG PROG dire vérité 'Ils savent qu'il dit la vérité.'  
 b. Zot konn laverite.  
 3PL savoir.FC vérité 'Ils savent la vérité.'
- (2) a. Mo reste [pou (ki) li pa santi li tousel].  
 1sg rester.FL pour (SUB) 3SG NEG sentir 3SG seul 'Je reste pour qu'il ne se sente pas seul.'  
 b. Mo res pou lafin.  
 1SG rester.FC pour fin 'Je reste pour la fin.'

Les subordonnées circonstancielles présentent une variété d'introducteurs (voir Tableau): subordonnant, souvent issu du français (*kouma* 'comme', *kan* 'quand'), préposition, adverbe, groupe nominal, ces trois dernières catégories pouvant être suivies de *ki*, le plus souvent omis, et pouvant en être séparées (*avan mem ki*, *alor zisteman ki* 'avant même que', 'alors justement que'). A la différence des adverbes, les prépositions peuvent être suivies d'un complément nominal (*pou twa* 'pour toi') avec le même sens. La préposition *dan*, à valeur temporelle ou concessive, n'est jamais suivie de *ki*, et semble une innovation créole, comme en seychellois (Michaelis 1994) :

- (3) a. **Malgre (ki)** mo pa pe kapav lev lor mo lili [...]  
 Malgre (SUB) 1SG NEG PROG pouvoir lever sur 1SG lit [...] (Corpus Alleesaib, 1999)  
 'Même si je n'arrivais pas à sortir de mon lit, [...]'  
 b. **Dan** siklone pankor vini, pena kouran isi. (instagram, 13 janvier 2024)  
 Dans cyclone NEG.encore venir.FL, NEG.avoir courant ici  
 'Alors que le cyclone n'est pas encore là, on n'a pas de courant ici.'

En revanche, *pangar* ('bien que') (Chaudenson 2003, Fon Sing 2013), *ler*, *letan* ('quand') (4) ont été grammaticalisés comme subordonnants.

- (4) a. **Letan** mo ti ankor piti-piti, mo ti pe may pwason dan dilo (Corpus Alleesaib 1999)  
 temps 1SG PST encore petit-petit, 1SG PST PROG attraper poisson dans eau  
 'Lorsque j'étais encore tout petit, j'attrapais des poissons dans l'eau.'  
 b. **Ler** li pe pase, li nek signal enn gard [...] (Corpus Florigny 2017)  
 Heure 3SG PROG passer, 3SG juste signaler un policier [...]  
 'Quand il passe, il fait même un signe à un policier'

Introduceur	Subordonnant	Préposition (+ki)	Adverbe (+ki)	Nom (+ ki)
causale	<i>kouma, parsi, tanki</i>	<i>akoz, etan done (ki)</i>	<i>telman, sirtou ki</i>	
concessive	<i>kannem, pangar</i>	<i>dan, malgre (ki), san (ki), sof (ki)</i>	<i>alor ki, mem ki</i>	
conditionnelle	<i>si, sipa</i>	<i>anka, amwin (ki), pourvi (ki), sipoze (ki), sof si</i>		
finale et consécutive	<i>afin ki</i>	<i>pou (ki), desort (ki)</i>	<i>samem ki</i>	<i>tansion (ki)</i>
temporelle	<i>apenn, kan, kouma, ler, letan</i>	<i>amezir (ki), apre (ki), avan (ki), dan, depi (ki), pandan (ki), ziska (ki)</i>	<i>aster ki</i>	<i>letan ki, le ki,, enn fwa (ki), sak fwa (ki), tou kou (ki)</i>

**Mot-clés**

Créole Mauricien, Subordonnée complétive, Subordonnée circonstancielle, Subordonnant

**Références**

Alleesaib A., Chady S., Florigny G., Fon Sing G., Hassamal S., Abeillé, A. 2024, L'omission de *que* en français mauricien, CMLF, Lausanne.  
 Chady, Sh. 2018. *Des marqueurs aux mouvements discursifs dans des interactions entre jeunes plurilingues*, Thèse de doctorat, Université Paris Descartes.  
 Fon Sing G. 2013. Émergence et évolution de certains marqueurs en créole mauricien. *Actes du XXVII<sup>e</sup> Congrès international de linguistique et de philologie romanes*, Nancy, 25-38.  
 Hassamal S., Abeillé A. et Henri F. 2018. Les adverbes en mauricien, *Faits de Langues* No. 49.  
 Michaelis, S. 1994. *Komplexe Syntax im Seychellen-Kreol*. Tübingen: Narr.  
 Police-Michel D., Carpooran A., Florigny G. 2011. *Gramer Kreol Morisien*. Vol. 1. Akademi Kreol Morisien. Mauritius.  
 Véronique, D. 2007. *Ki* en haïtien, en mauricien et dans quelques autres créoles français. In Charolles M. et al. (éds.). *Parcours de la phrase*, Gap: Ophrys, 1-12.

## Martinican complement clauses in *ke*: how to explain variation?

Duzerol Minella, Dynamique du Langage, CNRS/Lyon 2

Keywords: Martinican, creoles, subordination, complementation, variation, sociogrammar

Martinican is a French-based creole language spoken in the French Caribbean island of Martinique by about 400.000 persons, bilingual in Martinican and French at different degrees. The literature on complement clauses in Martinican (Pinalie and Bernabé 1999; Jeannot-Fourcaud 2003; Damoiseau 2012; Colot and Ludwig 2013; Syea 2017) points to a construction alternation for semantically equivalent complement clauses. The literature states that most complement clauses are formed without any subordinator, as in ( 1 ).

(1)

[...] *zot di [zot ka monté]*  
2PL say 2PL IPFV go.up  
[...] 'you said you are going up'  
Descri OTA 1 031

An alternate construction is sometimes used, that differs from the former in that it is introduced by a subordinator, *ki*, *ké* or *ke*, as in ( 2 ).

(2)

*man pé di [ke man té ka joué]*  
1SG can say SUB 1SG PST IPFV play  
'I can say that I used to play'  
FLU Narr 006

The use of a subordinator is explained by an influence of French. Besides, following the linguistic representations, speakers that are the most exposed to French would be those producing it most frequently. Yet, few papers tackle this variation with linguistic and non-linguistic variables, on the basis of a corpus study. In this paper, I present the results of a descriptive corpus-based study of the variation between absence and presence of a subordinator in the declarative complement clauses, that benefits from the statistical approach. The database at hand (Duzerol 2023) gathers 112 complement clauses extracted from 110 minutes of narrations and descriptions collected by the author with the collaboration of 18 Martinican speakers. Twelve linguistic (morphosyntactic, semantic) and non-linguistic (information processing, sociolinguistics) variables are under scrutiny. Generalized mixed models were implemented by Marc Allasonnière-Tang. Firstly, the statistical results align with Jeannot-Fourcaud's (2003) suggestions: for the database, no social variable is statistically significant. These results come in opposition with the explanation of so-called French-influenced changes triggered by those who are the most exposed to French. Second, for the database, this



alternation can be statistically explained by information processing needs, in a similar fashion to what has been observed for languages such as the Santiago variety of Cap Verdean (Araújo Vieira Semedo 2021) or English (Casasanto and Stag 2008; Jaeger 2010). For the database, the density of information within and around a given complement clause has a statistical effect on the presence of a subordinator at the beginning of this complement clause.

## References

- Araújo Vieira Semedo, Eliane Cristina. 2021. 'Frase Complexa Em Cabo-Verdiano (Variedade de Santiago): Um Estudo Da Integração Entre Cláusulas'. Theses, Institut National des Langues et Civilisations Orientales- INALCO PARIS - LANGUES O' ; Universidade federal do Ceará. <https://theses.hal.science/tel-03469603>.
- Casasanto, Laura Staum, and Ivan A. Stag. 2008. 'The Advantage of the Ungrammatical'. *Proceedings of the Annual Meeting of the Cognitive Science Society*, no. 30. <https://escholarship.org/uc/item/6tm38849>.
- Colot, Serge, and Ralph Ludwig. 2013. 'Guadeloupean and Martinican Creole'. In *The Survey of Pidgin and Creole Languages. Volume 2: Portuguese-Based, Spanish-Based, and French-Based Languages*, edited by Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath, and Magnus Huber. Oxford: Oxford University Press. <https://apics-online.info/surveys/50>.
- Damoiseau, Robert. 2012. *Syntaxe Créole Comparée: Martinique, Guadeloupe, Guyane, Haïti*. Dictionnaires et Langues. Fort-de-France : Paris: CRDP Martinique ; Karthala.
- Duzerol, Minella. 2023. 'Complétives et Relatives En Martiniquais (Créole, Martinique)'. [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage) - [www.ortolang.fr](http://www.ortolang.fr). <https://hdl.handle.net/11403/basededonnees-these>.
- Jaeger, Florian T. 2010. 'Redundancy and Reduction: Speakers Manage Syntactic Information Density'. *Cognitive Psychology* 61 (1): 23–62. <https://doi.org/10.1016/j.cogpsych.2010.02.002>.
- Jeannot-Fourcaud, Béatrice. 2003. 'De l'opposition Verbo-Nominale En Créole Martiniquais : Approche Fonctionnaliste'. PhD Thesis. <http://www.theses.fr/2003PA05H046>.
- Pinalie, Pierre, and Jean Bernabé. 1999. *Grammaire Du Créole Martiniquais En 50 Leçons*. Editions L'Harmattan.
- Syea, Anand. 2017. *French Creoles: A Comprehensive and Comparative Grammar*. <http://site.ebrary.com/id/11326154>.

**LLcD 2024 – WS Complementisers in Creole languages**  
**Grammaticalisation of final complementisers from GIVE in Haitian**  
**Patricia Cabredo Hofherr (UMR 7023 – SFL, CNRS & U. Paris 8)**

Cognates of GIVE give rise to a range of grammaticalization patterns. Here I examine a cognate of *bay* ‘give’ (1) in Haitian (cf. Maurer&al 2013 for GIVE-SVCs) and show that it can give rise to a final complementiser.

Non-initial verbs in serial verb constructions like (2) lack verbal properties (no negation, no TAM marking allowed): it is often not clear whether these constructions are synchronically verbal or not (Jansen&al 1978, Sebba 1987).

- (1) Jan **bay** Mari yon liv. (Haitian)  
 J give M a book  
 ‘Jean gave Marie a book.’

BA(Y)+DP resembles a preposition (2a) BA(Y)+final clause resembles a complementiser (2b).

- (2) a. M a pote liv la **ba** wou. (Haitian)  
 1SG FUT bring book DEF GIVE 2SG  
 ‘I will bring you the book.’ (Fattier 2013: ex 49-174)
- b. Li achte savon **bay** Mari lave men li.  
 3SG buy soap GIVE M. wash hands 3SG  
 ‘S/he bought soap and gave (it) to Mary to wash her hands (with it).’

**HC *ba(y)* +NP** combines with Verbs of transfer (*send, return, transmit*) and introduces the recipient (2a) however, unlike e.g. Martinican *ba* ‘give’, **HC *ba(y)* +NP** does not allow introduction of a benefactive argument (Cabredo 2022).

In contrast with **HC *ba(y)*+NP**, the **HC *ba(y)*+ final clause** use is possible with verbs like *fe gato* ‘make cake’ that are not transfer verbs, showing that **HC *ba(y)*+ final clause** is not a subtype of the recipient use of **HC *ba(y)***, but a distinct development.

- (3) Li fe gato a **bay** Mari vann (Haitian)  
 3SG make cake DET GIVE M. sell  
 ‘S/he made the cake and give it to M to sell/ for M. to sell.’

In (3) the object of *fe* is interpreted as coreferential with the object gap of the embedded predicate *vann*. The object gap is not elliptical. **HC *vann*** does not allow ellipsis of the object: in the coordination in (4) only the lexical object pronoun can be interpreted as coreferential with the object of the first verb (4a), *vann* without an object is ungrammatical (4b).

- (4) a. Jan fe [gato a]<sub>k</sub> epi Mari vann <sup>ok</sup>i<sub>k</sub>. (Haitian)  
 b. Jan fe [gato a]<sub>k</sub> epi Mari vann \*[]<sub>k</sub>.  
 J make cake DET and M sell 3SG  
 ‘Jan made the cake and Mari sold it.’

This pattern is parallel to English *for*-final clauses: with objects a gap with co-referential interpretation is possible (*John made a cake for Mary to sell* \_) but in a coordination the pronoun is obligatory (*John made a cake and Mary sold \*(it)*).

Ex (2b) shows that when the shared object is in an instrumental adjunct (*wash hands **with the soap***) a syntactic gap is not necessary.

The final construction shows diachronic persistence of the transfer semantics of lexical *ba(y)* ‘give’: while abstract transfer is allowed (5), denial (6) or lack (7) of transfer are excluded. In ex like (6) the preposition *pou* ‘for’ has to be used instead.

- (5) **Abstract transfer ok**  
 Li chante chanson **bay** Mari koute (Haitian)  
 3SG sing song GIVE M. hear  
 'S/he sang a song for M. to hear.'
- (6) **Denial of transfer excluded**  
 Jan<sub>i</sub> fe gato a **bay** Mari manje. (Haitian)  
 J. make cake DET GIVE M. eat  
 # men finalman li<sub>i</sub> manje l li menm  
 but in-the-end 3SG ate 3SG 3SG-SELF  
 'Jan made the cake for Mary to eat, **#but in the end he ate it himself.**'
- (7) **No transfer excluded**  
 \*Li cache yon liv bay Mari cache/ jwenn. (Haitian)  
 3SG hide a book GIVE M. look-for/ find  
**Not** 'S/he hid a book for M to look for/ find.'

The data show that HC *bay* has developed uses introducing final clauses, broadly similar to FOR in English. GIVE > BENEFACTIVE > PURPOSE is a well-attested as a grammaticalization path (cf Newman 1996: 171-81, Heine & Kuteva 2002). However, the HC pattern provides a case in which final clauses do not depend on the availability of benefactive uses of *bay* (in contrast with Engl. FOR that has benefactive uses). Unlike final clauses with FOR in English, final-BA(Y) has retained transfer meaning from its lexical source and does not allow denial of transfer (6). This is in keeping with a source of final BAY as a serial verb construction as SVCs are characterized by an actuality entailment of the verbs in a SVC in the perfective past.

**Cabredo Hohferr, Patricia 2022.** GIVE Constructions in Haitian. *Mélanges en ligne pour Léa Nash*. [60ln.wordpress.com/happy-birthday-lea/](https://60ln.wordpress.com/happy-birthday-lea/)

**Cabredo Hohferr, P. & Téosier, S. 2023.** Give-constructions with final clauses – comparing Haitian and Martinican. In: Cabredo Hohferr, Glaude & Soare (eds). *La Grammaire est une fête / Grammar is a moveable feast. Mélanges offerts à / A Webschrift for Anne Zribi-Hertz*, 93-102. Zenodo. Open access. <https://doi.org/10.5281/zenodo.8055461>

**Fattier, Dominique 2013.** Haitian Creole structure dataset. In Michaelis, Susanne Maria, Philippe Maurer, Martin Haspelmath, and Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

**Heine, Bernd & Kuteva, Tania 2002.** *World lexicon of grammaticalization*. CUP.

**Jansen, B. Koopman, H, Muysken, P. 1978.** Serial verbs in Creole languages. *Amsterdam Creole studies* 2: 125-159.

**Lord, C., Foong Ha Yap, and S. Iwasaki 2002.** Grammaticalization of 'give'. African and Asian perspectives. In Wischerand Gabriele Diewald (eds.), *New Reflections on Grammaticalization*, 217–235. John Benjamins.

**Maurer, Ph. and the APiCS Consortium 2013.** 'Give' serial verb constructions. In Michaelis, Susanne Maria, Philippe Maurer, Martin Haspelmath, and Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

**Newman, J. 1996.** *Give: A Cognitive Linguistic Study*. Berlin & New York: Mouton de Gruyter.

**Sebba, Mark 1987.** *The Syntax of Serial Verbs*. Amsterdam: Benjamins.

## Comparing the syntax of subordination in insular and continental Upper Guinea Creoles

Today most linguists admit (e.g. Biagui, Nunez & Quint in press and references therein) that Capeverdean on the one hand and Guinea-Bissau and Casamance Creole on the other hand form two branches (henceforward called ‘insular’ and ‘continental’) of one of the same phylogenetic unit, namely Upper Guinea Creoles, a group of Portuguese-based languages, which most plausibly developed from a common ancestor (Quint & Moreira 2019), namely Upper Guinea proto-Creole (UGPC), that got into use along West-African shores in the second half of the 15<sup>th</sup> century. Historically, Upper Guinea Creoles are the result of the encounter of two different types of languages: Portuguese (an Indo-European Romance language) and Wolof, Temne and Mandinka (Niger-Congo languages belonging respectively to the Atlantic branch (Wolof and Temne) and the Mande branch (Mandinka) of Niger-Congo).

Several comparisons have already been published about insular (or Capeverdean) Creoles and continental Creoles (Marques Barros 1887-1908; Quint 2000: 99-117; Baptista & al. 2007: 53-82, Hagemeyer & Alexandre 2012). However, to-date, no systematic comparison of the syntax of subordination in the two branches has ever been undertaken. In this communication, we propose to make a first attempt at developing such a comparative study, basing ourselves on the data available in the ScroLL database (Manfredi & Quint forthcoming). In this database, the main characteristics of adverbial, complement and relative subordinates have been described and illustrated for Santiaguense (a Capeverdean insular variety) and Casamance Creole (a continental, variety). Drawing on the various criteria taken into account in ScroLL (such as TAM and subject marking or word order of the main constituents of each type of subordinate clause), we will show how continental and insular Upper Guinea Creoles, albeit closely related from a historical point of view, can also display striking contrasts in the way they build complex sentences.

Baptista, Marlyse, Heliana Mello & Miki Suzuki. 2007. Kabuverdianu, or Cape Verdean, and Kriyol, or Guinea-Bissau (Creole Portuguese). In John Holm & Peter L. Patrick (eds.), *Comparative Creole Syntax: Parallel Outlines of 18 Creole Grammars*, 53-82. London: Battlebridge.

Biagui, Noël B., Joseph J.-F. Nunez & Nicolas Quint. in press. *Casamance Creole*. In Friederike Lüpke (ed.), *Oxford Guide to the World's Languages: Atlantic*. Oxford: Oxford University Press.

Hagemeyer, Tjerk & Nélia Alexandre. 2012. Os crioulos da Alta Guiné e do Golfo da Guiné: uma comparação sintáctica. *Papia* 22(2). 233-251.

Manfredi, Stefano & Nicolas Quint (eds.), forthcoming *SCROLL: the database of subordination in Creole languages*. HumaNum.

Marques Barros, Marcelino. 1887-1908. *O Guineense*. Revista Lusitana 5, 6, 7 & 10.

Quint, Nicolas & Moreira Karin. 2019. The common African lexical core of Upper Guinea Creoles and its historical significance, *Journal of Ibero-Romance Creoles* 9, pp. 115-161.

Quint, Nicolas. 2000. *Le Cap-verdien: origines et devenir d'une langue métisse*. Paris: L'Harmattan.



## Recent evolutions of *say*-complementizers in Creole languages

Sibylle Kriegel, Aix Marseille Univ, CNRS, LPL

Based on the scroll resource (Manfredi & al. 2023) and the data in APiCS (Michaelis & al. 2013), this contribution will give an overview of *say*-complementizers (also see Parkvall 2000) in several creole languages with different lexifiers, like the English-based Nenge(e) and Naijá (scroll resource), the Arabic-lexifier Juba-Arabic (scroll-resource), the Portuguese-based Guinea-Bissau Kriyol (APiCS) and one French-lexified Creole, Seychelles Creole. I then will zoom in on Seychelles Creole, the only French-lexified Creole that uses a *say*-complementizer. Ø-complementation and the French-derived *ki* (< ‘que’) complementizer are also used. In Kriegel (2004), I stated that *pourdir*, a lexicalization of the preposition *pour* (< Fr. ‘pour’) followed by the verb *dir* (< Fr. ‘dire’) was almost exclusively used by elderly people and limited to contexts where the speaker does not assert the content of the complement-clause. Twenty years after this first study, the situation in present-day Seychelles Creole obviously has changed: we observe a much wider use of the *pourdir*-complementation in spoken and written media as well as in spontaneous spoken language. A new study based on a written corpus (*Verbatim Reports of the National Assembly of Seychelles*, data from 2024) and spoken data collected recently will analyze if there still exist any semantic restrictions on the use of *pourdir*. I claim that the increase in frequency of the *pourdir*-complementizer is linked to what I have called the Distinction-during-codification hypothesis in Kriegel (2021). Seychelles Creole is the first official language of the Seychelles since 1981 and its codification is very advanced. I claim that in this special sociolinguistic context, constructions not found in the lexifier have more chances to become stabilized than those that continue patterns of the lexifier.

Kriegel, Sibylle 2004. « Juste pour dire que *pourdir* existe toujours », *CREOLICA*, jeudi 16 décembre 2004, <http://www.creolica.net>

Kriegel, Sibylle 2021. « Grammaticalization in Seychelles Creole: the coding of reciprocity by *kanmarad* » *Isogloss. Open Journal of Romance Linguistics* 7, 11:1-27. DOI:<https://doi.org/10.5565/rev/isogloss.149>

Manfredi, Stefano, Nicolas Quint and the SCroll Consortium (eds.), 2023. *SCroll - The Database of Subordination in Creole Languages*.

Michaelis, Susanne Maria and the APiCS Consortium. 2013. « Complementizer with verbs of speaking ». in: Michaelis, Susanne Maria & Maurer, Philippe & Haspelmath, Martin & Huber, Magnus (eds.) *The atlas of pidgin and creole language structures*. Oxford: Oxford University Press.

Michaelis, Susanne Maria and the APiCS Consortium. 2013. Complementizer with verbs of knowing. in: Michaelis, Susanne Maria & Maurer, Philippe & Haspelmath, Martin & Huber, Magnus (eds.) *The atlas of pidgin and creole language structures*. Oxford: Oxford University Press.

Parkvall, Mikael. 2000 *Out of Africa. African influences in Atlantic Creoles* London: Battlebridge Publications.

## Subordinate constructions in English-lexifier Creoles: Adverbial Complementation

Bettina Migge, University College Dublin, Ireland  
Paula Prescod, Université de Picardie Jules Verne, France

Cross-clausal phenomena have received comparatively little attention in research on Creoles. There is work on purpose clauses, relative clauses and fact-type complements focusing on their general description (e.g. Winford 1985) and topics such as finiteness (Byrne 1987; van der Vate & Veenstra 2016) and substrate influence (Migge & Winford 2013; Huttar et al. 2013; Lefebvre & Loranger 2008), for example, and research has focused on the Creoles of Suriname, particularly Saamaka. It shows, for example, that there are interesting similarities but also differences with some of the main input languages, English and Gbe. Interestingly, adverbial subordination, except for purpose clauses, has received little attention. The aim of this presentation is thus to begin to close this gap through a descriptive analysis.

After an overview of subordinate constructions in three English-lexifier Creoles, Nenge(e) or the Eastern Maroon Creoles, Vincentian Creole and Naijá based on the scroll resource (Manfredi et al 2023), we zoom in on a selection of subordinate clauses: temporal clauses, reason clauses, as well as clauses expressing reality conditions. We also examine the order of main-subordinate clauses, the devices and internal structure of the subordinate clause, comparing Creole and English constructions.

- (1) *{because} e get stones I have to wash it well using calabash*  
Because it has stones, I have to wash it well, using a calabash. **Naijá**  
(pcm.26, NaijaSynCor: JOS\_12\_How-To-Prepare-Gote-Soup\_MG.conllu)
- (2) *I gaan sinama {kaaz} i lov big skriin*  
3SG gone cinema REAS.SUB 3SG love big screen **Vincentian Creole**  
'He's gone to the cinema because he loves big screens.' (svc.87)
- (3) *U an gi i pasi fu i gwe*  
1PLNEG give 2SG.OBJ path PURP.COMP 2SG leave  
*ma na {fu.di} i fuka*  
but FOC REAS.COMP 2SG overwhelm **Eastern Maroon Creoles**  
'We did not give you permission to leave but since you are under pressure.' (djK 117)

As examples (1-3) show, all three languages behave more alike with respect to their overall syntactic structure. However, Vincentian Creole and Naijá show greater similarities to each other and to English in terms of subordinating devices. There is also a fair amount of variation across constructions and languages, which we bring to the fore in this presentation.

### References

- Byrne, Francis. 1987. Grammatical Relations in a Radical Creole. Verb Complementation in Saramaccan [Creole Language Library 3]. Amsterdam: John Benjamins. doi: 10.1075/cll.3
- Huttar, George, Aboh, Enoch & Ameka, Felix. 2013. Relative clauses in Suriname Creoles and Gbe languages. *Lingua* 129 (2013), 96-123.

Lefebvre, Claire & Loranger, Virginie. 2008A synchronic and diachronic account of the multifunctionality of Saramaccan ta.. *Linguistics* 46: 1167–1228. doi: 10.1515/LING.2008.037

Manfredi, Stefano, Nicolas Quint and the SCroll Consortium (eds.), 2023. SCroll - The Database of Subordination in Creole Languages.

Migge, Bettina & Winford, Donald. 2013. Fact-type complements in Gbe and the Surinamese Creoles. *Lingua* 129 (7): 9-31.

van de Vate, Marlene & Veenstra, Tonjes. 2016. Non-finiteness in Saamáka. In *Finiteness matters: on finiteness-related phenomena in natural languages*, E. M. Eide, K.(ed.), 189-209.

Winford, Donald. 1985. The syntax of fi complements in Caribbean English Creole. *Language* 61(3). 588–624.



## Against simplicity in subject relative clause marking in Creole languages

Kuteva & Comrie (2012) argue that creole languages, such as Krio in example (1), tend to have at most one single relative marker in subject relative clauses, whereas non-Creoles often have multiple relative markers in the same construction. An extreme case cited by Kuteva & Comrie (2012) is example (2) from Ngemba, a Bantoid language, with five different relative clause markers: (i) relative determiner *wá*, (ii) complementizer *bah* (optional), (iii) pronoun-retention, (iv) multipurpose marker for topicalization, nominalization and relativization *-ne*, and (v) sentential definitizer *-la*.

(1) Krio (English-based, Finney 2013)

*Di man we kam yestade na mi padi*  
ART man REL come yesterday COP POSS friend  
'The man that came yesterday is my friend.'

(2) Ngemba (Bantoid, Chumbow 1977:290, cited after Kuteva & Comrie 2012:30)

*Nyung wá bah a-keshung-ne mung wa la a kung atsang*  
man REL COMP he-TNS.beat-ne child DET la he enter into.prison  
'The man who beat the child went to prison.'

Kuteva & Comrie (2012) explain the lack of elaborateness in Creoles with their language ecologies: Due to pidginization processes and their young age, Creoles are said not to have had enough time to develop more complex relative markers.

In this paper, I take issue with such an interpretation, and I will argue for a different explanation based on Creole data in APiCS (2013) and SCRoIL (2024). The argument goes as follows: Most languages world-wide have only one marker in subject relative clause constructions (Comrie & Kuteva 2005). Furthermore, most contributing languages of creoles, too, have only one marker in subject relative clauses, such as the (dialectal non-standard) European lexifier languages and relevant substrate languages. Therefore, we do not expect any elaboration of relative markers in Creoles compared to their ancestry languages in the first place. Instead, detailed studies, such as Huttar et al. (2015) and Corne (1994), have clearly identified the role of substrate languages in the formation of Creole relative clause patterns for Ndyuka/Saramaccan and Tayo. But cases where both the substrate and the dialectal nonstandard European lexifier patterns (for English: Herrmann 2005, for Romance: Schafroth 1993) converge are certainly not excluded.

Thus, the argument for simple-before-complex from Creole relative clause structures is very weak, and the contribution of the ancestry languages can account for many or most of the attested patterns.

## References

- Comrie, Bernard. & Kuteva, Tania. 2005. The evolution of grammatical structures and "functional need" explanations. In: M. Tallerman (ed.), *Language Origins: Perspectives on Evolution*, 185–207. Oxford: Oxford University Press.
- Corne, Chris. 1994. Relativization and thematization in Tayo and the implications for creole genesis. *Journal of Pidgin and Creole Languages* 9(2): 283-304.
- Finney, Malcolm Awadajin. 2013. Krio structure dataset. In: Michaelis, Susanne Maria & Maurer, Philippe & Haspelmath, Martin & Huber, Magnus (eds.) *Atlas of Pidgin and*

- Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Huttar, George L. & Enoch O. Aboh & Felix K. Ameka. 2013. Relative clauses in Suriname creoles and Gbe languages. *Lingua* 129, 96—123.
- Herrmann, Tanja. 2005. Relative clauses in English dialects of the British Isles. *Agreement, Gender, Relative Clauses* 21-123.
- Kuteva, Tania, & Comrie, Bernard. 2012. The evolution of language and elaborateness of grammar. In: Comrie, Bernard & Estrada, Zarina (eds.), *Relative Clauses in Languages of the Americas: A Typological Overview*, 27-46. Amsterdam: Benjamins.
- Manfredi, Stefano, Nicolas Quint and the SCroll Consortium (eds.). 2024. SCroll: *The Database of Subordination in Creole Languages*. <https://scroll.huma-num.fr>
- Michaelis, Susanne Maria, Maurer, Philippe, Haspelmath, Martin & Huber, Magnus (eds.). 2013. *The atlas of pidgin and creole language structures*. Oxford: Oxford University Press.
- Schafroth, Elmar. 1993. *Zur Entstehung und vergleichenden Typologie der Relativpronomina in den romanischen Sprachen: Mit besonderer Berücksichtigung des Substandards*, Tübingen: Niemeyer.

**Workshop “Subordination in Creole languages”**

Subordination and subordinators in Caribbean English-based Creoles: A (dialect) typological view

(Bernd Kortmann, Freiburg, Germany)

The focus of this talk will be on subordinators in Caribbean English-based Creoles (CECs), i.e. on relativizers, complementizers and, especially, adverbial subordinators. These will be looked at from both a broadly typological perspective (mapping cross-linguistic variation) and, in the first place, from a dialect typological perspective, which is concerned with mapping language-internal (in this case: English-internal) variation. Both perspectives will be informed by what is known about grammaticalization patterns and paths of subordinators across languages (as e.g. identified in Kortmann 1997 or Ohori 2011) and across different varieties and variety types of English (notably high-contact L1 and L2 Englishes; cf. e.g. Kortmann / Schneider 2011). Questions which will be addressed include the following: Is it possible to identify features in the subordination domain which are shared among all, or the vast majority of, CECs, some of which may even be diagnostic of CECs vis-à-vis English-based Pidgins and Creoles spoken in other parts of the Anglophone world? Are there subordination features in the Caribbean which reflect the distinction between Western and Eastern CECs, or the commonly made threefold distinction into ‘radical creoles’ (i.e. the Surinamese Creoles), ‘conservative’ CECs and ‘intermediate’ CECs (cf. e.g. Kortmann *ibid.*)? Most interesting from both a broadly typological and a dialect typological perspective will be the grammaticalization patterns and paths which gave rise to the inventories of subordinators in the Caribbean creoles. Do, for example, CEC subordinators resulting from grammaticalization processes markedly differ from subordinators in non-creole languages, and if so, how? Of special interest from a dialect typological perspective will be to determine the extent to which the grammaticalization processes resulting in the current inventories of subordinators in CECs are contact-/substrate-induced or rather, in a broad sense, instances of superstrate- (i.e. English-) internal developments, be it that they are known from earlier stages of English or that they represent continuations (or ‘re-enactments’) of internal grammaticalization processes known from the history of English or currently spoken (non-standard) non-creole varieties of English. An important data source for the dialect typological take on subordinators in CECs will be eWAVE, the *Electronic World Atlas of Varieties of English* (Kortmann / Lunkenheimer / Ehret 2020).

Hackert, Stephanie (2012). ‘Regional profile: The Caribbean’, in Bernd Kortmann, and Kerstin Lunkenheimer (eds.), *The Mouton World Atlas of Variation in English*. Berlin, Boston: De Gruyter Mouton, 704–733.

Kortmann, B. forthcoming. “21 CECs and typology”. In: Stephanie Hackert, ed. *The Oxford Handbook of Caribbean English*. Oxford: Oxford University Press.

Kortmann, B. 1997. *Adverbial Subordination. A Typology and History of Adverbial Subordinators Based on European Languages*. Berlin/New York: Mouton de Gruyter.

Kortmann, B./K. Lunkenheimer/K. Ehret, eds. 2020. *The Electronic World Atlas of Varieties of English*. Zenodo. DOI: 10.5281/zenodo.3712132 (Available online at <http://ewave-atlas.org>)

Kortmann, B./K. Lunkenheimer, eds. 2012. *The Mouton World Atlas of Variation in English*. Berlin/New York: De Gruyter Mouton.

Kortmann, B./A. Schneider. 2011. “Grammaticalization in Non-standard Varieties of English.” In: H. Narrog/B. Heine, eds. *The Oxford Handbook of Grammaticalization*. Oxford/New York: Oxford University Press. 263-278.

Ohori, Toshio. 2011. “The Grammaticalization of subordination.” In: H. Narrog/B. Heine, eds. *The Oxford Handbook of Grammaticalization*. Oxford/New York: Oxford University Press. 636-645.



**Langues & Langage  
à la croisée des Disciplines**

**1ère Rencontre annuelle LLcD**

Colloque : 9-11 septembre, 2024

Ecole d'été : 12-14 septembre, 2024

Sorbonne Université - 28 Rue Serpente, 75006 Paris, France



## **Workshop 6**

**Grammatical gender in diachrony**

**Le genre grammatical en diachronie**

# The evolutionary dynamics of lexical concepts: the connection between semantic property and gender

Gerd Carling

In collaboration with: Noor Efrat-Kowalsky, Marc Allasonnière-Tang, Lev Michael, Niklas Erben Johansson, Filip Larsson

Lexical concepts have differentiating behaviour in terms of their evolutionary profile. Some lexemes have lower substitution rates than others (Pagel, Atkinson, and Meade 2007), and lexemes differ when it comes to their propensity for colexification (Carling et al. 2023; Haspelmath and Tadmor 2009; Jackson et al. 2019). Further, evolutionary rates vary depending on the semantic class (humans, animates, objects of various types), but the reasons for this variation is not clear (Carling 2019). In the current paper, we will look at the evolutionary dynamics of a list of concepts (basic vocabulary and culture terms) with respect to their semantic property (animacy, humanness, sexus, and other selected semantic properties) as well as their gender. Linguistic gender is found in about 20% of the world's languages (Allasonnière-Tang et al. 2021) and gender languages typically have rules or conditions for how objects are assigned to a specific gender (Corbett 1991; Corbett Greville 2013) We hypothesize that the semantic core of gender assignment, in particular the connection between gender and animacy, have an impact on the evolutionary rates of gender change in the lexicon. We have compiled two dataset of lexemes from Indo-European and Arawak languages. These datasets (basic and culture vocabulary), are coded for different semantic properties and gender. On this data, we are running a phylogenetic reconstructing model (MCMC, RJHP, in BayesTraits, based on an informed tree), aiming at extracting the transition rates of different concepts together with their gender. Preliminary results indicate not just a connection between core semantic properties (such as animacy) and gender, but also a differentiation in transition rate between different genders.

Keywords: gender, Indo-European, Arawak, lexical evolution, phylogenetics

- Allasonnière-Tang, Marc, Olof Lundgren, Maja Robbers, Sandra Cronhamn, Filip Larsson, One-Soon Her, Harald Hammarström, and Gerd Carling. 2021. "Expansion by migration and diffusion by contact is a source to the global diversity of linguistic nominal categorization systems." *Nature Humanities & Social Science - Communications* 8:331.
- Carling, Gerd. 2019. *Mouton Atlas of Languages and Cultures. Vol. 1: Europe and West, Central and South Asia*. Berlin - Boston: Mouton de Gruyter.
- Carling, Gerd, Sandra Cronhamn, Olof Lundgren, Victor Bogren Svensson, and Johan Frid. 2023. "The evolution of lexical semantics: dynamics, directionality, and drift." *Frontiers in Communication* 8. doi: <https://doi.org/10.3389/fcomm.2023.1126249>.
- Corbett, Greville G. 1991. *Gender, Cambridge textbooks in linguistics, 99-0104661-0*. Cambridge: Cambridge Univ. Press.
- Corbett Greville, G. 2013. Gender typology. In *The Expression of Gender*.
- Haspelmath, Martin, and Uri Tadmor. 2009. *Loanwords in the world's languages: a comparative handbook*. Berlin: Mouton de Gruyter.
- Jackson, Joshua Conrad, Joseph Watts, Teague R. Henry, Johann-Mattis List, Robert Forkel, Peter J. Mucha, Simon J. Greenhill, Russell D. Gray, and Kristen A. Lindquist. 2019. "Emotion semantics show both cultural variation and universal structure." *Science* 366 (6472):1517-1522. doi: doi:10.1126/science.aaw8160.

## Areal and phylogenetic distribution of lexical partitioning imposed by noun classes

Noun classes and classifiers partition the world into semantic classes, with varying degrees of arbitrariness, especially for grammaticalised noun classes (Dixon 1986; Mel'cuk 2006). Following recent work on areal semantics (Koptjevskaja-Tamm & Liljegren 2017; Schapper & Koptjevskaja-Tamm 2022; François 2022) we ask whether such ways of partitioning show more areal or phylogenetic patterning (Allasonnière-Tang et al. 2021). In Indo-European, where noun classes (genders) are semantically bleached, class partitions have a strongly phylogenetic distribution (McCarthy et al. 2020). But for more semantically substantive classifier systems, we might expect that more areal spread could occur (Seifart 2018).

We target three clusters of languages, from northern Australia, the northwest Amazon and northern Papua New Guinea, for which we coded grammatical classification of a standard list of 395 nominal concepts (see Table 1 for examples). We then apply McCarthy et al.'s statistical methods, using adjusted mutual information (AMI) as the basis for hierarchical clustering of the conceptual classifications. Figure 1 illustrates the AMI clustering for our Australian and Amazonian languages (see Table 2 for details). As expected, the Amazonian languages (Bora and Muinane) separate from the Australian languages at the highest level. Below this, the Australian clustering broadly matches the presumed genealogy in some respects, for example placing Bininj Gun-wok and Kunbarlang in a low-level cluster (Kapitonov 2021). But in other respects it does not match presumed genealogy, for example Murrinhpatha is more tightly clustered with the neighbouring language Marri Tjevin, despite unclear genealogical relationship, while being more distant from the demonstrably related Ngan'gi (Green 2003). The continuation of our study will further investigate these mismatches. We will also apply explicit phylogenetic methods to the etymologically reconstructible classifiers, complementing the method based on alignment of partitions.

**Table 1. Examples of noun classification in Australian and Amazonian languages**

	Daly		Arnhem		NW Amazon	
	Murrinhpatha	M. Tjevin	Wubuy	Mawng	Bora	Muinane
BIRD	<i>ku</i> (ANIM)	<i>awu</i> (ANIM)	<i>ngarra-</i> (IV)	<i>na-</i> (I)	<i>-co</i> (POINT)	<i>-i</i> (ROUND)
FISH	<i>ku</i> (ANIM)	<i>awu</i> (ANIM)	<i>ngarra-</i> (IV)	<i>na-</i> (I)	<i>-be</i> (MASC)	<i>-bo</i> (MASC)
WASP	<i>ku</i> (ANIM)	<i>awu</i> (ANIM)	<i>ana-</i> (II)	<i>niny-</i> (II)	<i>-co</i> (POINT)	<i>-cu</i> (ROUND)
FIREWOOD	<i>thungku</i> (FIRE)	<i>tjendji</i> (FIRE)	<i>ngarra-</i> (IV)	<i>ma-</i> (III)	<i>-wa</i> (PLANK)	<i>-gai</i> (PLANK)
FIRESTICK	<i>thungku</i> (FIRE)	<i>tjendji</i> (FIRE)	<i>ngarra-</i> (IV)	<i>niny-</i> (II)	<i>-ho</i> (OBLONG)	<i>-hu</i> (OBLONG)
ALCOHOL	<i>kura</i> (LIQ)	<i>wudi</i> (LIQ)	<i>ngarra-</i> (IV)	<i>nung-</i> (IV)	<i>-pacyo</i> (LIQ)	<i>faihu</i> (LIQ)

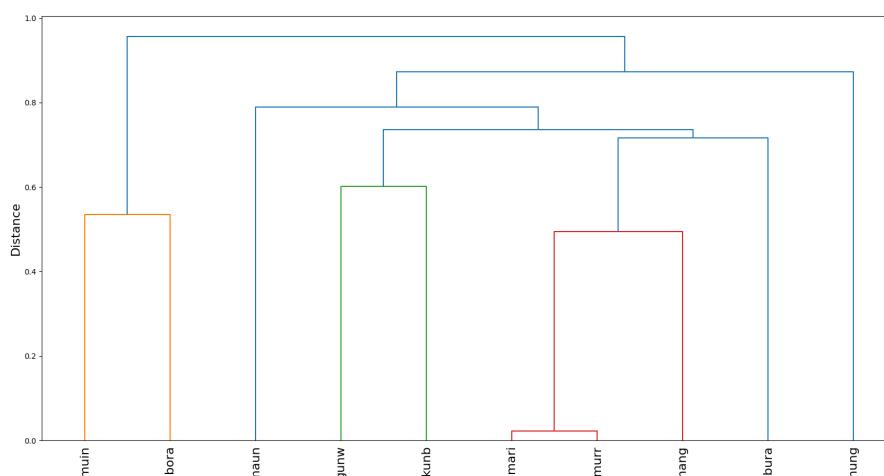


Figure 1. Hierarchical clustering of noun classifier systems, based on AMI scores.

Table 2. Languages analysed in Figure 1

	Language	Area	Glottocode	Family
1.	Burarra	Arnhem	bura1267	Maningrida
2.	Wubuy	Arnhem	nung1290	Gunwinyguan
3.	Bininj Gun-wok	Arnhem	gunw1252	Gunwinyguan
4.	Kunbarlang	Arnhem	kunb1251	Gunwinyguan
5.	Mawng	Arnhem	maun1240	Iwaidjan
6.	Murrinhpatha	Daly	murr1258	Southern Daly
7.	Ngan'gi	Daly	nang1259	Southern Daly
8.	Marri Tjevin	Daly	mari1420	Western Daly
9.	Bora	NW Amazon	bora1263	Boran
10.	Muinane	NW Amazon	muin1242	Boran

## References

- Allasonnière-Tang, Marc & Lundgren, Olof & Robbers, Maja & Cronhamn, Sandra & Larsson, Filip & Her, One-Soon & Hammarström, Harald & Carling, Gerd. 2021. Expansion by migration and diffusion by contact is a source to the global diversity of linguistic nominal categorization systems. *Humanities and Social Sciences Communications*. Palgrave 8(1). 1–6. (doi:10.1057/s41599-021-01003-5)
- Dixon, R. M. W. 1986. Noun classes and noun classification in typological perspective. In Craig, Colette G. (ed.), *Noun Classes and Categorization: Proceedings of a symposium on categorization and noun classification, Eugene, Oregon, October 1983* (Typological Studies in Language), 105–112. John Benjamins Publishing Company. (doi:10.1075/tsl.7.09dix)
- François, Alexandre. 2022. Lexical tectonics: Mapping structural change in patterns of lexification. *Zeitschrift für Sprachwissenschaft*. De Gruyter 41(1). 89–123. (doi:10.1515/zfs-2021-2041)
- Green, Ian. 2003. The genetic status of Murrinh-patha. In Evans, Nicholas (ed.), *The non-Pama-Nyungan Languages of Northern Australia*, 125–158. Canberra: Pacific Linguistics.
- Kapitonov, Ivan. 2021. *A Grammar of Kunbarlang* (Mouton Grammar Library). Vol. 89. Berlin: De Gruyter Mouton. (https://doi.org/10.1515/9783110747058)
- Koptjevskaja-Tamm, Maria & Liljegren, Henrik. 2017. Semantic patterns from an areal perspective. In Hickey, Raymond (ed.), *The Cambridge Handbook of Areal Linguistics* (Cambridge Handbooks in Language and Linguistics), 204–236. Cambridge: Cambridge University Press. (doi:10.1017/9781107279872.009)
- List, Johann-Mattis & Rzymiski, Christoph & Tresoldi, Tiago & Greenhill, Simon & Forkel, Robert (eds.). 2019. *CLICS<sup>3</sup>*. Jena: Max Planck Institute for the Science of Human History. (https://clics.clld.org/) (Accessed August 28, 2023.)
- McCarthy, Arya D. & Williams, Adina & Liu, Shijia & Yarowsky, David & Cotterell, Ryan. 2020. Measuring the similarity of grammatical gender systems by comparing partitions. *Proceedings of the 2020*

- Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 5664–5675. Online: Association for Computational Linguistics. (doi:10.18653/v1/2020.emnlp-main.456)
- Mel'cuk, Igor. 2006. Gender and noun class. *Aspects of the theory of morphology*, 322–383. De Gruyter Mouton. (doi:10.1515/9783110199864.2.322)
- Schapper, Antoinette & Koptjevskaja-Tamm, Maria. 2022. Introduction to special issue on areal typology of lexico-semantics. *Linguistic Typology*. De Gruyter Mouton 26(2). 199–209.
- Seifart, Frank. 2018. The semantic reduction of the noun universe and the diachrony of nominal classification. In McGregor, William B. & Wichmann, Søren (eds.), *The Diachrony of Classification Systems* (Current Issues in Linguistic Theory), 9–32. John Benjamins Publishing Company. (doi:10.1075/cilt.342.02sei)



### **Preliminary steps towards an integrated evolutionary model of Bantu gender systems**

Bantu languages are usually described as having pervasive systems of grammatical gender marking (also known as noun class systems), both in terms of number of gender distinctions and patterns of agreement. Yet, the diversity attested among the languages of the family in this domain of grammar can be conspicuous (Di Garbo and Verkerk 2022 for northwestern Bantu). Multiple trajectories of variation and change are attested and are likely due to contact dynamics with both related Bantu and unrelated non-Bantu languages (see Verkerk and Di Garbo 2022). In this paper, we showcase the initial steps of a newly funded project on variation and change in Bantu gender systems. This project investigates the diversity of Bantu gender systems with the aim of producing an integrated evolutionary model of how these systems are distributed in the context of contact with neighboring languages. At each of the borders of the Bantu spread zone, speakers of Bantu languages interact with communities who speak languages belonging to distantly related branches of the Niger-Congo family, such as Ubangi, as well as to genealogically unrelated groupings such as Cushitic, Nilotic, Central Sudanic, Khoe-Kwadi, Kx'a, and Tuu. Within the Bantu spread zone, inter-Bantu contact is also intense (Mous 2019). By comparing changes attested in Bantu languages in contact with non-Bantu languages with those attested in situations of Bantu-Bantu contact, and using a combination of typological, diachronic and field-based studies, we aim to show how processes of divergence and convergence differ depending on whether languages in contact share ancestry. The focus of the talk will be twofold.

We will first present the internal structure of the dataset. Given that the goal of the project is to investigate trajectories of change in Bantu gender systems using Bayesian statistics to infer diachrony from synchrony, our dataset will consist of several hundreds of languages. The sample will be organized in three sub-sets, each corresponding to the three major borders of the Bantu spread zone: northwestern, eastern, and southern. For the northwestern region we rely on an existing dataset of 179 languages; the eastern and southern datasets will be generated anew (ideally between 250 and 300 languages, to be added to the 179 languages already sampled).

Secondly, we will tackle methodological issues pertaining to identifying and dating contact events between Bantu and non-Bantu populations in the present and the past, which may have (had) an impact on Bantu gender systems. More specifically, we will present the initial results of a survey aiming at characterizing the contact profile of Bantu-non-Bantu interactions at each of the three major border zones (northwestern, eastern, and southern). The survey will be based on published literature and will serve as a baseline for generating hypotheses about possible scenarios of gender changes in contact situations, which we will tackle during the subsequent stages of the project.

We will conclude with some general considerations on the role of contact in diachronic change in gender systems, focusing on the assessment of different types of contact in terms of function, intensity, and duration.

**Word count: 499**

#### **References**

- Di Garbo, F., & Verkerk, A. (2022). A typology of northwestern Bantu gender systems. *Linguistics*, 60(4), 1169–1239. <https://doi.org/10.1515/ling-2020-0217>
- Mous, M. (2019). Language contact. In M. Van De Velde, K. Bostoen, D. Nurse, & G. Philippson (Eds.), *The Bantu languages. Second edition* (pp. 355–381). Routledge.

Verkerk, A., & Di Garbo, F. (2022). Sociogeographic correlates of typological variation in northwestern Bantu gender systems. *Language Dynamics and Change*, 12(2), 155–223.

## Towards a phylogenetic typology of Australian nominal classification

Jayden Macklin-Cordes

*University of Newcastle, Australia*

Nominal classification reflects human cognition and the way we categorise the world around us (Shcherbakova & Allasonnière-Tang 2024). In Australia, noun classes are common in languages of the north in particular, however, current surveys are relatively small in sample size (e.g. Harvey & Reid 1997; Skilton 2023) and the diachronic origins of these systems remain underexplored. This study presents a typological survey of the presence/absence of noun class systems in 94 Australian language varieties. I quantitatively model the fit of phylogeny and areality as explanatory factors for the present distribution of these systems, and perform an ancestral state reconstruction to infer the presence/absence of noun classes in the ancestral proto-Pama-Nyungan and putative proto-Australian languages.

**Methodology.** Language sample: 44 Pama-Nyungan languages, a large family spanning nearly 90% of the Australian mainland. 50 languages represent 18 genealogically distinct non-Pama-Nyungan families and 6 isolates, densely distributed in the north. Sources: Grambank (Skirgård, Haynie, Blasi, et al. 2023; Skirgård, Haynie, Hammarström, et al. 2023), Autotyp (Bickel et al. 2015), WALS (Dryer & Haspelmath 2022). A phylogenetic reference tree and location data were extracted from Glottolog (Hammarström et al. 2024).

Three Bayesian regression models, consisting of: i) a single phylogenetic predictor; ii) a single geographic predictor (modelled as a gaussian process, following Naranjo & Becker 2022); iii) a combined model with phylogenetic and geographic predictors together.

A Bayesian MCMC ancestral state reconstruction was performed using BayesTraits (Meade & Pagel 2022).

**Results.** Noun classes are present in 43/94 languages (46%) (Figure 1). There is a striking split between Pama-Nyungan (6/44, 14% of languages), and non-Pama-Nyungan families (37/50, 74%).

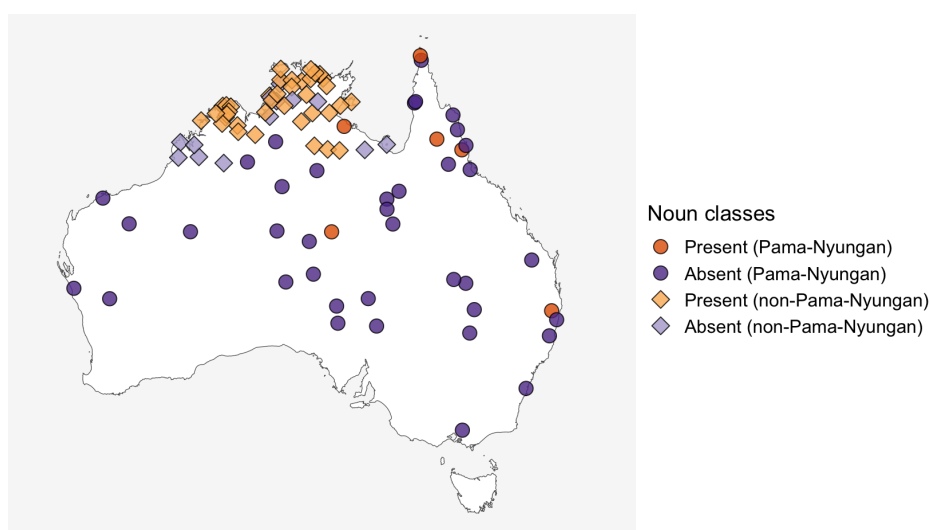
The phylogenetic model intercept (expressed as probability) is 0.64, equivalent to a phylogenetically balanced mean. The geographic model intercept, equivalent to a geographically balanced mean, is 0.17. This reflects, in the former case, downweighting of the relatively large Pama-Nyungan family. In the latter case, it is downweighting the relatively numerous northern languages. The intercept of the combined phylogenetic + geographic model is 0.38.

The geographic model ( $R^2 = 0.45$ , 95% CI [0.28, 0.60]) fits the data slightly better than the phylogenetic model ( $R^2 = 0.39$ , 95% CI [0.26, 0.50]), and the combined model fits slightly better than both ( $R^2 = 0.49$  (95% CI [0.34, 0.62])), although the differences are small.

Ancestral state reconstruction predicts that noun classes were present ( $Pr > 0.99$ ) at the root of the reference tree, which would be equivalent to a putative proto-Australian. Conversely, noun classes were likely absent from proto-Pama-Nyungan ( $Pr > 0.99$ ), perhaps unsurprising given the present scarcity of noun classes in the family.

**Discussion.** This study presents a phylogenetically-informed typological survey of the presence of noun class systems in Australian Indigenous languages. The slight improvement in fit when geography is included in the model is perhaps attributable to language contact. This explanation seems plausible given the distribution of non-Pama-Nyungan languages without noun classes, which cluster near Pama-Nyungan neighbours in the east and west. However, the more precise dynamics of noun class evolution in Australia require more investigation.

Keywords: Pama-Nyungan, phylogenetic signal, grammatical gender, linguistic phylogenetics



**Figure 1.** Presence/absence of noun classes in 94 Australian languages.

## References

- Bickel, Balthasar, Alena Witzlack-Makarevich, Kamal K. Choudhary, Matthias Schlesewsky & Ina Bornkessel-Schlesewsky. 2015. The Neurophysiology of Language Processing Shapes the Evolution of Grammar: Evidence from Case Marking. *PLOS ONE* 10(8). e0132819. <https://doi.org/10.1371/journal.pone.0132819>.
- Bürkner, Paul-Christian. 2021. Bayesian Item Response Modeling in R with brms and Stan. *Journal of Statistical Software* 100. 1–54. <https://doi.org/10.18637/jss.v100.i05>.
- Dryer, Matthew & Martin Haspelmath. 2022. The World Atlas of Language Structures Online. Zenodo. <https://doi.org/10.5281/zenodo.7385533>.
- Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2024. glottolog/glottolog: Glottolog database 5.0. Zenodo. <https://doi.org/10.5281/zenodo.10804357>.
- Harvey, Mark & Nicholas Reid. 1997. *Nominal Classification in Aboriginal Australia*. Amsterdam: John Benjamins Publishing Company. <http://ebookcentral.proquest.com/lib/uql/detail.action?docID=784229>. (5 March, 2021).
- Meade, Andrew & Mark Pagel. 2022. Ancestral State Reconstruction Using BayesTraits. In Haiwei Luo (ed.), *Environmental Microbial Evolution: Methods and Protocols*, 255–266. New York, NY: Springer. [https://doi.org/10.1007/978-1-0716-2691-7\\_12](https://doi.org/10.1007/978-1-0716-2691-7_12).

- Naranjo, Matías Guzmán & Laura Becker. 2022. Statistical bias control in typology. *Linguistic Typology*. De Gruyter Mouton 26(3). 605–670. <https://doi.org/10.1515/lingty-2021-0002>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Shcherbakova, Olena & Marc Allasonnière-Tang. 2024. Evolutionary pathways of complexity in gender systems. *Journal of Language Evolution* lzae001. <https://doi.org/10.1093/jole/lzae001>.
- Skilton, Amalia. 2023. Noun classes. In Claire Bower (ed.), *The Oxford Guide to Australian Languages*. Oxford University Press. <https://doi.org/10.1093/oso/9780198824978.003.0019>.
- Skirgård, Hedvig, Hannah J. Haynie, Damián E. Blasi, Harald Hammarström, Jeremy Collins, Jay J. Latache, Jakob Lesage, et al. 2023. Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Science Advances*. American Association for the Advancement of Science 9(16). eadg6175. <https://doi.org/10.1126/sciadv.adg6175>.
- Skirgård, Hedvig, Hannah J. Haynie, Harald Hammarström, Damián E. Blasi, Jeremy Collins, Jay Latache, Jakob Lesage, et al. 2023. Grambank v1.0. Zenodo. <https://doi.org/10.5281/zenodo.7740140>.

## "Noun classes" in Niger-Congo: from classifier to gender system

Tom Güldemann and Ines Fiedler (Humboldt-Universität zu Berlin)

Niger-Congo, the largest language family on the globe, displays a well-known system of so-called "noun classes", which conflate the marking of gender and number and involve both agreement and adnominal morphological classes. This has been analyzed for a long time within the domain of gender, both in African linguistics (Heine 1982) and globally (Corbett 1991). At the same time, it has been noticed that these Niger-Congo "noun classes" display striking similarities to classifiers (e.g., Nichols 1992: 136, Grinevald and Seifart 2004). Accordingly, the idea that "noun classes" can be traced back to classifiers of some sort is not new and has more recently been entertained prominently by Kießling (2013, 2018).

The aim of this paper is twofold. For one thing, we address the early history of the relevant gender-number system of Niger-Congo by substantiating the hypothesis of its origin in a late classifier system. In so doing, we also crucially inform the general linguistic question of the emergence and subsequent evolution of gender.

Our analysis is based on extensive and historically deep data from a wide range of Niger-Congo subgroups. This comparative material goes far beyond the examination of the best-known system of Bantu, a large but young subfamily, which heretofore has been viewed as the yardstick for the synchronic and diachronic evaluation of the larger lineage. Our analysis focusses on various parameters that inform the morpho-syntactic history of the gender-number (aka "noun class") markers in Niger-Congo, notably a) inventory, b) semantics, c) number-sensitivity, and d) morpho-syntax of both adnominal and agreement exponents. Regarding the first two parameters, the Niger-Congo system had already been recognized to be classifier-like (cf., e.g., Grinevald and Seifart 2004). We argue in addition that the two latter parameters, namely the number-sensitivity and the morphologically bound nature of gender-number exponents, are later features of the system, which makes the earlier state typologically yet more similar to a system of classifiers. Hence, the central conclusion of our investigation is that the Niger-Congo case is a likely example for the scenario in which gender, as the combination of noun classification and morpho-syntactic agreement, arises from adding agreement to a classifier system, similar to cases attested in the Amazon (cf. Seifart 2005).

Heine, Bernd. 1982. African noun class systems. In Hansjakob Seiler and Christian Lehmann (eds.), *Apprehension: Das sprachliche Erfassen von Gegenständen, Teil I: Bereich und Ordnung der Phänomene*. Language Universals Series 1, 189-216. Tübingen: Gunter Narr.

Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.

Grinevald, Colette, and Frank Seifart. 2004. Noun classes in African and Amazonian languages: Towards a comparison. *Linguistic Typology* 8: 243-285.

- Kießling, Roland. 2013. On the origin of Niger-Congo nominal classification. In: Ritsuko Kikusawa and Lawrence A. Reid (eds.), *Historical linguistics 2011: Selected papers from the 20th international conference on historical linguistics*, Osaka, 25-30 July 2011, 43-65. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Kießling, Roland, 2018. Niger-Congo numeral classifiers in a diachronic perspective. In: William B. McGregor and Søren Wichmann (eds.), *The diachrony of classification systems*, *Current Issues in Linguistic Theory* 342, 33–76. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- Seifart, Frank. 2005. The structure and use of shape-based noun classes in Miraña (North West Amazon). *MPI Series in Psycholinguistics* 32. Wageningen: Ponsen and Looijen.

## From Classifier to Grammatical Gender: Evidence from Oceanic

Michael Franjeh, Alexandra Grandison, Greville G. Corbett  
*University of Surrey, University of Surrey, University of Surrey*

Possessive classifier systems are known to be ‘relational’: they signal the interaction between the possessor and the referent:

(1) Lewo (Vanuatu; fieldwork data)

1a. *m̃a-u*            *mrawa*  
DRINK.CL-1SG coconut  
‘my coconut (to drink)’

1b. *ka-u*            *mrawa*  
FOOD.CL-1SG coconut  
‘my coconut (to eat)’

Use of (1a) is appropriate if the possessor intends to drink the coconut, while (1b) would indicate an intention to eat it. These variable-assignment classifier systems are well attested in Oceanic languages. Yet Franjeh (2016, 2018) observes that the system in North Ambrym is moving towards a fixed gender-like assignment system.

Gender is a familiar reference tracking device. Possessive classifiers can also function for reference tracking without a possessed noun. We investigate this using a storyboard experiment (c.f., Burton & Matthewson 2015), with speakers of six Oceanic languages, including Lewo and North Ambrym. By focusing on classifiers as reference tracking devices we can reveal their similarity to gender systems.

Pairwise comparisons of individual languages shows that North Ambrym’s system functions differently to the other languages in our sample, with the same classifier being used on average throughout a storyboard more than 90% of the time (Figure 1). The other languages in our sample retain typical relational systems, whereas North Ambrym’s system closely resembles a fixed gender-like system.

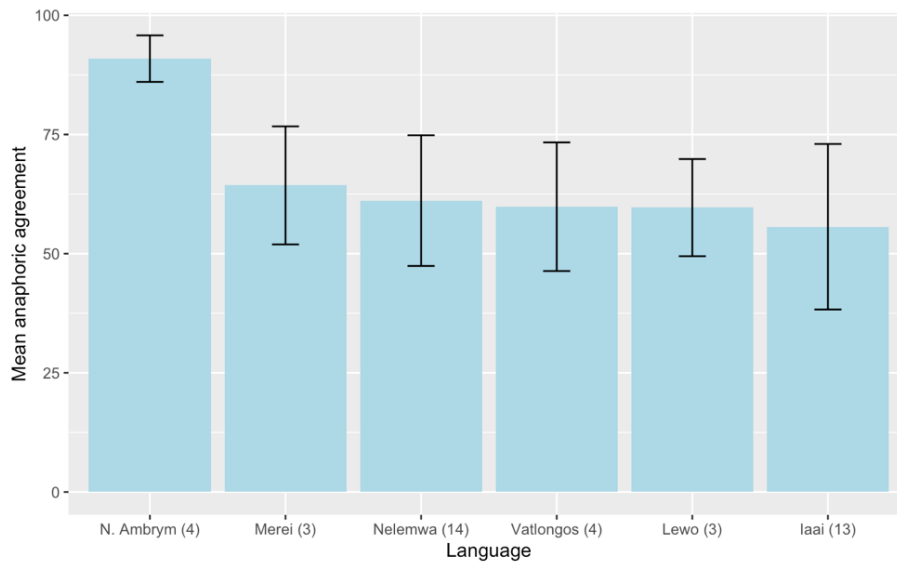


Figure 1: mean anaphoric agreement for each sample language (total number of classifiers given across all participants for each language in brackets). Standard error bars show 95% confidence intervals.

Gender and classifiers were once thought of as disparate systems of categorisation but are now understood as belonging to a single typological space (Fedden & Corbett 2017). Our findings



not only support a unified typology of nominal classification systems but provides critical evidence for the rise of agreement and gender systems from classifiers.

**Keywords:** possessive classifiers, reference tracking, grammatical gender, Oceanic, storyboards,

### References

- Burton, S. and Matthewson, L. 2015. Targeted Construction Storyboards in Semantic Fieldwork. In M. Ryan Bochnak and Lisa Matthewson (eds.), *Methodologies in Semantic Fieldwork* (pp. 135-156). Oxford: OUP.
- Fedden, S., Corbett, G G. 2017. Gender and classifiers in concurrent systems: Refining the typology of nominal classification. *Glossa: a journal of general linguistics*, 2(1:34), 1–47.
- Franjeh, M. 2018. North Ambrym possessive classifiers from the perspective of canonical gender. In S. Fedden, J. Audring and G. Corbett (eds.) *Non-canonical gender systems* (pp. 36-67). Oxford: OUP.
- Franjeh, M. 2016. Indirect Possessive Hosts in North Ambrym: Evidence for Gender. *Oceanic Linguistics* 55, 87-115

## Tukanoan highly grammaticalized classifiers are an expanded gender system

Jurriaan Wiegertjes  
*Surrey Morphology Group*

Many Northwest Amazonian languages have concurrent gender and classifiers, where gender is a more canonical and classifiers a less canonical system of nominal categorization (see Corbett and Fedden 2018; Fedden and Corbett forthcoming); in many Northwest Amazonian languages, classifiers are highly grammaticalized and can participate in nominal agreement (1; 2), though with fewer agreement targets than gender markers. This is characteristic of more canonical nominal categorization systems, and Northwest Amazonian classifier systems have been analysed as “emerging noun class systems” (e.g. Grinevald and Seifart 2004: 281; Seifart 2023). This would constitute an often proposed grammaticalization path for gender (e.g. Aikhenvald 2000; Fedden and Corbett 2017; but see Passer 2016). I propose the opposite development for the Tukanoan family: the morpho-syntactic category of classifiers developed out of the pre-existing inanimate gender, supplemented with semantically more transparent markers emerging from the lexicon.

- (1) VAÍMAJĀ  
*jiká-ro*                      *jutí-ro*                      *butí-ro*  
one.INAN-CLF:FOLDED    textile-CLF:FOLDED    white-CLF:FOLDED  
'one white shirt' (own fieldwork)
- (2) KUBEO  
*i=kũ*                      *hiađo=kũ*                      *kopo-i=ũ*  
DEM.PROX-CL:HOLLOW    tree.sp=CL:HOLLOW    break-ST=CL:HOLLOW  
'this broken canoe' (Chacon 2012: 240)

My evidence for the proposed development of classifiers out of the inanimate gender is: i) gender markers can be reconstructed to Proto-Tukanoan (Chacon 2021; 2022); ii) many classifiers in the family are language-internal developments or can only be reconstructed for a sub-branch; iii) the few classifiers that can be reconstructed to Proto-Tukanoan seem to be developments of Proto-Tukanoan gender markers; and iv) with the sole exception of Kubeo (Gomez-Imbert 1996), classifiers only occur with inanimate nouns.

The Proto-Tukanoan gender system at one point contained a three-way distinction: *\*-i* 'SG.ANIM'; *\*-ka* 'SG.INAN'; and *\*-je* 'PL', and later also *\*-o* 'SG.FEMININE' (Chacon 2021; 2022). I claim that, while reflexes of *\*-i*, *\*-o*, and *\*-je* still largely function as gender markers, the inanimate marker *\*-ka* was the source of the synchronic classifier *-ka/-a* 'CLF:ROUND' found across the family, where *\*-ka* was semantically narrowed in opposition to newly introduced classifiers sourced from the lexicon, in favour of semantic transparency, but only in certain grammatical positions – a process discussed by Seifart (2018). Thus, in several languages, inanimate *\*-ka* was retained in grammatical positions where gender markers but not classifiers are found, e.g. Retuarã neuter agreement (Strom 1992: 46), and Vaímajã has an inanimate numeral *jiká* 'one', which contrasts with feminine *jikó*, and non-feminine *jikí* (own fieldwork), and which allows further specification by means of a classifier (1). Likewise, agreement on numerals in Wa'ikhana and Kotiria is limited to forms corresponding to the Proto-Tukanoan inanimate gender marker (Stenzel 2013: 177–178; Balykova 2019; Wiegertjes 2023).

Typically, synchronic Tukanoan classifiers fill the same suffix slot as the earlier inanimate gender marker, contrasting with animate gender markers (3). Subsequently, the former

inanimate gender marker has obtained more specified semantics (viz. ROUNDED) in opposition to newly formed classifiers.

- (3) MÁÍHĪKÌ  
a. wèè tè-we  
house one-CLF:BUILDING  
'one house' (Farmer 2012: 106)
- b. tè-o nómiò-go  
one-SG.F woman-SG.F  
'one woman' (ibid.: 157)

This reconstruction explains classifiers' high degree of grammaticalization as retention rather than emergence, while simultaneously explaining the distinct morpho-syntactic distributions of the gender and classifier systems. The proposed reconstruction may also serve as a point of departure for further research on the development of similar gender and classifier systems in the Northwest Amazon.

**Key words:** gender; classifiers; Tukanoan languages; grammaticalization

### References

- Aikhenvald, Alexandra Y. (2000). *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.
- Balykova, Kristina. 2019. "What do oranges and hammers have in common? The classifier 'round' in Wa'ikhana and other East Tukano languages." *LIAMES* 19: 1–24. <http://dx.doi.org/10.20396/liames.v19i0.8655262>.
- Chacon, Thiago Costa. 2012. "The Phonology and Morphology of Kubeo: The Documentation, Theory and Description of an Amazonian Language." University of Hawai'i, Mānoa. PhD dissertation.
- . 2021. "Origem e convergência nos sistemas de classificação nominal em línguas tukano [Origin and convergence of gender and nominal classifier systems in Tukanoan languages]." Presented at Amazonicas VIII, Góias/Brasília. [https://www.youtube.com/watch?v=jPSQ9d-L\\_cg](https://www.youtube.com/watch?v=jPSQ9d-L_cg).
- . 2022. "Rumos, picadas e caminhos na mata densa: Diacronia dos sistemas de gênero na família tukano [Routes, trails and paths in the dense forest: Diachrony of the gender systems of the Tukano family]." *Cadernos de Etnolingüística* 10(3): 1–29.
- Corbett, Greville G., and Sebastian Fedden, "New approaches to the typology of gender." In Sebastian Fedden, Jenny Audring, and Greville G. Corbett (eds), *Non-Canonical Gender Systems*, 9–35. Oxford: Oxford University Press.
- Farmer, Stephanie. 2015. "Establishing reference in Máihiki." University of California, Berkeley. PhD dissertation.
- Fedden, Sebastian, and Greville G. Corbett. 2017. "Gender and classifiers in concurrent systems: Refining the typology of nominal classification." *Glossa: a journal of general linguistics* 2(1): 34. 1–47. <https://doi.org/10.5334/gjgl.177>.

- . Forthcoming. “Typology of gender and classifiers.” In Niels Schiller and Tanja Kupisch (eds.), *The Oxford Handbook of Gender and Classifiers*. Oxford: Oxford University Press.
- Gomez-Imbert, Elsa. 1996. “When animals become ‘rounded’ and ‘feminine’: Conceptual categories and linguistic classification in a multilingual setting.” In John J Gumperz and Stephen C Levinsohn (eds.), *Rethinking Linguistic Relativity*, 438–69. Cambridge: Cambridge University Press.
- Grinevald, Colette, and Frank Seifart. 2004. “Noun classes in African and Amazonian languages: Towards a comparison.” *Linguistic Typology* 8(2): 243–85. <https://doi.org/10.1515/lity.2004.007>.
- Passer, Matthias Benjamin. “The typology and diachrony of nominal classification.” Universiteit van Amsterdam. PhD dissertation. Utrecht: LOT, 2016.
- Seifart, Frank. 2018. “The semantic reduction of the noun universe and the diachrony of nominal classification.” In William B. McGregor and Søren Wichmann (eds.), *The diachrony of classification systems*, 9–32. Amsterdam: John Benjamins Publishing Company.
- . 2023. “Re-classification of nouns in the history of a nominal classification system: A case study on Amazonian Bora and Muinane.” Presented at Amazonicas IX, Universidad de los Andes, Bogotá, 5-9 June.
- Stenzel, Kristine. 2013. *A reference grammar of Kotiria (Wanano)*. Lincoln/London: University of Nebraska Press. <https://doi.org/10.2307/j.ctt1ddr99n>.
- Strom, Clay. 1992. *Retuamã syntax*. Studies in the Languages of Colombia 3. Dallas, Texas: Summer Institute of Linguistics.
- Wiegertjes, Jurriaan. 2023. “The development of Tukanoan classifiers.” Presented at Amazonicas IX, Universidad de los Andes, Bogotá, 5-9 June.

## Non-contextual uses of gender-number agreement markers and the emergence of orphan agreement markers in Atlantic languages

Denis Creissels

University of Lyon

[denis.creissels@univ-lyon2.fr](mailto:denis.creissels@univ-lyon2.fr)

<http://deniscreissels.fr>

It has been observed that, in the history of languages having grammatical gender systems, a gender may disappear completely from the noun system while surviving in the pronominal system, cf. e.g. Butt & Benjamin (1994) on neuter pronouns in Spanish. This presentation deals with languages whose history is not documented (the Atlantic languages of West Africa), but in which some pronouns and adverbs can be analyzed as keeping track of former genders that have been lost in the nominal system.

A common type of morphology-syntax mismatch in Atlantic gender systems is the existence of ORPHAN GENDER-NUMBER AGREEMENT MARKERS, affixes that occupy the same morphological slot as bona fide agreement markers but are not involved in agreement mechanisms.

For example, in Jóola Fóoñi, most of the inflected forms of *-ceen* ‘some’ can modify a noun, as in *a-seek a-ceen* ‘some woman’, *ko-seek ko-ceen* ‘some women’, *e-suk e-ceen* ‘some village’, *si-suk si-ceen* ‘some villages’, etc, or be used pronominally with reference to an antecedent, as in *a-ceen* ‘some X’ (X a noun of gender A/BK to be retrieved from the context), *si-ceen* ‘some Xs (X a noun of gender E/S to be retrieved from the context),<sup>1</sup> etc. However, two of the inflected forms of *-ceen* (*di-ceen* ‘something’ and *ni-ceen* ‘sometimes’) are semantically autonomous, cannot be explained with reference to the nouns for ‘thing’ and ‘time’ (since the regular noun-modifier constructions ‘some thing’ and ‘some time’ would be *w-aafɔ-ceen* and *Ø-waati e-ceen*), and more generally cannot be used as noun modifiers at all.

This phenomenon can be analyzed as having its origin in the development of NON-CONTEXTUAL USES of agreement markers. Forms inflected for agreement normally combine with an overtly expressed controller, or imply reference to a potential controller to be retrieved from the context. However, in Atlantic gender systems, they commonly also have NON-CONTEXTUAL uses. For example, in Wolof, forms consisting of the agreement marker *l-* ‘singular of gender L/Y’ and the stem *-ii* ‘this’ or *-an* ‘which?’ can modify nouns (*cin l-ii* ‘this pot’, *cin l-an* ‘which pot?’) or occur in contexts in which they are interpreted as ‘this one’ and ‘which one?’ with reference to a particular noun of gender L/Y to be retrieved from the context, but they can also express vague reference to inanimate entities without any contextual conditioning, as in (1), where the context plays no role in the interpretation of *l-ii* as ‘this (thing)’, and of *l-an* as ‘what?’.

---

<sup>1</sup> Genders are conventionally designated by means of pairs of capital letters that evoke the phonological form of their characteristic agreement markers in the singular and in the plural.

*L-ii l-an la?*  
*L-this L-which it.is*  
'What is this?'

In some cases, a possible explanation is that, in the absence of a particular controller suggested by the context, a hypernymous noun belonging to the gender in question acts as a default controller. However, synchronically, such an explanation is not always possible, which means that, diachronically, the non-contextual use of an agreement marker may be maintained after the hypernymous noun with reference to which the non-contextual use developed has disappeared from the lexicon or changed its meaning.

Once an agreement marker has acquired a non-contextual use, it may happen that not only the noun originally responsible for its non-contextual use, but more generally all its potential controllers disappear from the lexicon, or are assigned to other genders, and this is how agreement markers can become orphan agreement markers that are not involved in agreement mechanisms anymore, and independently contribute to the meaning of the words in which they can be found.

## References

- Butt, J. & C. Benjamin. 1994. Neuter article and pronouns. In: *A New Reference Grammar of Modern Spanish*. Springer, Boston, MA.
- Creissels, Denis & Konstantin Pozdniakov (eds.). 2015. *Les classes nominales dans les langues atlantiques*. Cologne: Rüdiger Köppe. 371-405.
- Corbett, Greville. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville. 2006. *Agreement*. Cambridge: Cambridge University Press.

# Changements dans les systèmes de classes des langues atlantiques Nord à travers la dissociation du genre et du nombre

Mots clés : atlantique-nord, badiaranke, genre, pluriel, érosion

Les langues atlantiques sont parlées en Afrique de l'Ouest. Elles forment une famille appartenant au phylum Niger-Congo. La famille atlantique, telle que délimitée par Pozdniakov et Segerer (sous presse), se décompose en deux branches : Nord et Bak. Ses langues se caractérisent en partie par le fait de présenter des systèmes de classification nominale incontestablement rattachables à des systèmes de classes nominales Niger-Congo (Creissels et Pozdniakov 2015), regroupé sous la notion de genre depuis Corbett (1991). Dans ces systèmes, l'opposition singulier/pluriel est marquée sur le nom et ses dépendants par des marques d'accords en genre-nombre morphologiquement inanalysables (Creissels 2001).

Cependant, les langues de la branche Nord attestent de systèmes moins conservateurs que les langues Bak. De plus, elles montrent davantage de divergences entre elles (Creissels 2001, 2015; Rochant, Allasonnière-Tang, et Cathcart 2022). Ces divergences s'expliquent par des voies de changement différentes, pouvant conduire à des réanalyses profondes. Certains de ces changements ont déjà fait l'objet d'une littérature détaillée, comme l'émergence d'un système de classe par suffixation en Niger-Congo (Greenberg 1977), comme en peut dans la famille atlantique.

Notre présentation se focalisera sur des changements plus spécifiques à la branche atlantique Nord. Elle portera sur la langue badiaranke du groupe Tenda-Jaad. Les langues de ce groupe (surtout les langues Tenda) présentent des systèmes de genre riches qui s'écartent peu du prototype des classes nominales Niger-Congo (Bassène 2015; Ferry 1991; Perrin 2015; Santos 1996). Néanmoins, nous montrerons que le système du badiaranke s'est modifié en s'écartant des systèmes des autres langues de la famille jusqu'à rendre questionnable son identification en tant que système de classe nominale, sans pour autant perdre son système de genre.

L'évolution de ce système est cohérente vis-à-vis d'autres langues de la famille. Nous émettons l'hypothèse que le système de classe d'une partie des langues atlantiques nord ont été modifié par un changement dans l'expression de la pluralité entraînant une dissociation de l'amalgame genre/nombre. Cette dissociation est aussi visible via les accords, y compris lorsque l'inventaire des modificateurs accordés est drastiquement réduit. Nous suggérons que ce changement peut être le résultat d'au moins deux voies différentes :

- via l'émergence d'une marque de pluriel depuis une marque d'associatif conduisant à des systèmes de genre hybrides (genre et nombre amalgamé et genre distinct du nombre) ;
- via la diminution du nombre de marques de classes de pluriel, pouvant aboutir à l'érosion de l'appariement en nombre conduisant à la dissociation genre-nombre.

Nous comparerons les deux voies, la première documentée au moins par les spécialistes des langues atlantiques pour les langues nyun-buy (Cobbinah & Lüpke 2014 ; Cobbinah 2017) et la deuxième à travers le cas du badiaranke. Nous montrerons que d'autres langues de la branche Nord ont également commencé à emprunter cette dernière voie. Cependant, différentes propriétés permettent à la fois de les relier entre elles (persistance de classe pour les humains) et de les différencier des autres systèmes attestés actuellement dans les langues de la branche Nord.

## **Références :**

Bassène, A. C. (2015). Les classes nominales du biafada. In D. Creissels & K. I. Pozdniakov (Eds.), *Les classes nominales dans les langues atlantiques, Grammatische Analysen afrikanischer Sprachen ; Gramatical analyses of African languages* (pp. 460–485). Köln: Rüdiger Köppe Verlag.

- Cobbinah, A. (2017). Suffixed plurals in Baïnonk languages: Agreement patterns and diachronic development. *Journal of African Languages and Linguistics*, 38(2), 145–185.
- Cobbinah, A., & Lüpke, F. (2014). When number meets classification: The linguistic expression of number in Baïnounk languages. In A. Sortch & G. Dimmendaal (Eds.), *Number – Constructions and Semantics. Case studies from Africa, Amazonia, India and Oceania* (pp. 199–220). Amsterdam / Philadelphia: John Benjamins.
- Corbett, G. G. (1991). *Gender*. Cambridge: Cambridge University Press.
- Creissels, D. (2001). Les systèmes de classes nominales des langues Niger-Congo : prototype et variations. *Linx*, (45), 157–166.
- Creissels, D. (2015). Typologie des systèmes de classes nominales dans deux groupes de langues atlantiques. In D. Creissels & K. I. Pozdniakov (Eds.), *Les classes nominales dans les langues atlantiques, Grammaticische Analysen afrikanischer Sprachen; Gramatical analyses of African languages* (pp. 3–53). Köln: Rüdiger Köppe Verlag.
- Creissels, D., & Pozdniakov, K. (Eds.). (2015). *Les classes nominales dans les langues atlantiques*. Köln: Rüdiger Köppe Verlag.
- Ferry, M.-P. (1991). *Thesaurus tenda: dictionnaire ethnolinguistique de langues sénégal-guinéennes, -níyàn (bassari), -nik (bedik), -mèỹ (konyagi)* (Vol. 326). Paris: Peeters Publishers.
- Greenberg, J. H. (1977). Niger-Congo Noun Class Markers: Prefixes, Suffixes, Both or Neither. In M. Mould & T. Hinnebush (Eds.), *Papers from the eighth Annual Conference on African Linguistics* (Vol. 7, pp. 97–104). African Studies Center & Department of Linguistics, University of California (UCLA).
- Perrin, L.-M. (2015). Les classes nominales en basari. In D. Creissels & K. I. Pozdniakov (Eds.), *Les classes nominales dans les langues atlantiques* (pp. 518–563). Köln: Rüdiger Köppe Verlag.
- Pozdniakov, K., & Segerer, G. (sous presse). A Genealogical classification of Atlantic languages. In F. Lüpke (Ed.), *The Oxford guide to the Atlantic languages of West Africa*. Oxford: Oxford University Press.
- Rochant, N., Allasonnière-Tang, M., & Cathcart, C. (2022). The evolutionary trends of noun class systems in Atlantic languages. In *Proceedings of the Joint Conference on Language Evolution (JCoLE)* (pp. 624–631). Max Planck Institute for Psycholinguistics.
- Santos, R. (1996). *Le mey: Langue ouest-atlantique de guinee* [Thèse pour le doctorat d'état ès Lettres, Université de la Sorbonne nouvelle - Paris III].



## From asymmetric grammatical treatment of nouns to gender: the case of Ubangi

Several studies have dealt with or even been dedicated to the question of how gender systems emerge but these attempts suffer in particular from the inaccessibility of the earliest stages in this complex diachronic process as well as the scarcity of language groups in which different historical stages can be observed synchronically side by side. Central Africa provides a unique laboratory in this respect, whereby genealogically related Ubangi languages in particular display a full set of different types of animacy-based noun classification. These range from the categorization of nouns outside the canonical domain of gender over intermediate stages to fully grammaticalized pronominal gender systems involving a sex distinction.

Four relevant noun-classification types, labelled A-D, can be characterized, whereby types B-D qualify as pertaining to the domain of gender:

- A preferential grammatical treatment of animate nouns without clear paradigmatic distinction (possibly involving pronominal anaphor),
- B obligatory pronominal anaphor for animate nouns vs. zero for inanimate nouns,
- C overtly distinct pronominal anaphor for animate and inanimate nouns,
- D overtly distinct pronominal anaphor for animate and inanimate nouns with additional gender distinctions (notably sex) and/or non-pronominal agreement.

In a type-A language, animate nouns have access to morpho-syntactic processes that are largely blocked for inanimate nouns, for example, overt number marking or resumption by proper anaphoric pronouns. A type-B language involves a grammaticalized asymmetrical treatment of nouns through regular pronominal anaphor for animates as opposed to referents lower on the nominal hierarchy displaying zero anaphor (or being resumed by proforms outside the pronoun paradigm). Such a "covert pronominal gender system" is shown in Figure 1 from Ngbandi (Ngbandic).



**Figure 1: The animacy-based pronoun system of Northern Ngbandi (Toronzoni 1989)**

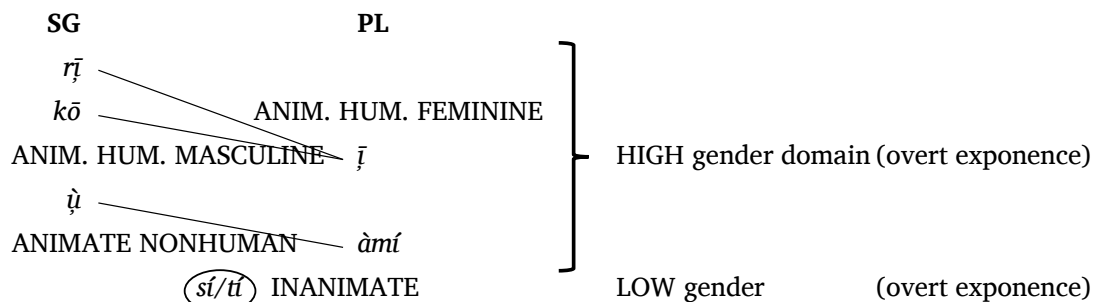
Type C is very similar to Type B the difference being that the low gender now also involves overt exponence, as shown in Figure 2 from Monzombo (Mundu-Baka) - this is a simple but overt gender system.

<b>SG</b>		<b>PL</b>
-----------	--	-----------



**Figure 2: The animacy-based pronoun system of Monzombo (Boyi 1983)**

A type-D language is shown in Figure 3 from Zande (Zandic), where the higher animate domain of the basic binary opposition is further elaborated into three genders.



**Figure 3: The pronoun system of Zande (Boyd n.d.)**

A sex distinction as in Zande is seen at the center of the gender domain in general, so that this and similar languages have been included in global surveys (cf. Corbett 2023). Since Zande is one endpoint in a full diachronic typological scale that starts out from noun classification without gender, Central Africa crucially informs our quest of assessing the origin of gender systems in Africa and globally.

Boyd, Raymond. n.d. Les pronominaux et démonstratifs du zande. unpublished ms.

<https://www.rblanguesdafrique.info/pronoms.pdf>

Boyi, Jean. 1983. Le munzombo (langue oubanguienne): étude du nom. Ph.D. thesis: Université de la Sorbonne Nouvelle Paris III.

Corbett, Greville G. 2013. Sex-based and non-sex-based gender systems. In Dryer, Matthew S. and Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/31>, Accessed on 2021-09-02.)

Toronzoni, Ngama-Nzombio Tra Ndele. 1989. Description du ngbandi, langue oubanguienne du nord-ouest du Zaïre. Ph.D. thesis: Université Libre de Bruxelles.

## Gender and declension at an early stage of gender loss

Traditional Norwegian dialects have three grammatical genders: Masculine, feminine, and neuter. Recent research has uncovered what seems to be an ongoing loss of the feminine grammatical gender in several urban varieties of Norwegian (Lødrup, 2011, Rodina & Westergaard, 2016, Busterud et al. 2019, Busterud et al., 2021). A core question emerges from these findings: How can the ongoing loss inform our understanding of the relationship between gender and declension? Results from a study investigating the status of grammatical gender in a Norwegian dialect with rich nominal morphology, the Voss dialect, show that the feminine gender is at an early stage of being replaced by the masculine among children and adolescents. Importantly, in the singular gender changes before the declension of the noun, and prenominal possessives change before postnominal possessives.

According to Hockett, “Genders are classes of nouns reflected in the behavior of associated words” (Hockett, 1958, p. 231). The Voss dialect has a three-gender distinction in the use of indefinite articles, anaphoric pronouns, and possessives (Aasen 1848, Vidsteen 1884, Heggstad 1932, Akselberg 2003). The definite suffix on feminine nouns ends in /-e/ (strong) or /-o/ (weak), and for masculine nouns the ending is /-en/.

The current study is an experimental study of the use of grammatical gender and nominal declension. It includes 64 participants, divided into three age groups: 24 children (age 10-12), 20 adolescents (age 15-19) and 20 adults (age 25-50). Based on previous methodology (Blom et al. 2008, Rodina & Westergaard 2015, Busterud et al. 2019), a picture-based elicitation task elicits indefinite forms, adjectives, definite forms, prenominal and postnominal possessives, and pronouns.

The results demonstrate that adults use the feminine indefinite article agreeing with feminine nouns 100% of the time (1a) and that the feminine article is partly being replaced by the masculine article (1b) to a various degree (ranging from 6,2 % to 97 %) by 29/44 children and adolescents.

- |     |    |                 |             |               |
|-----|----|-----------------|-------------|---------------|
| (1) | a. | <i>Ei</i>       | <i>kvit</i> | <i>dør.</i>   |
|     |    | DET.SG.F.INDEF  | white       | door.SG.INDEF |
|     |    | ‘A white door.’ |             |               |
|     | b. | <i>Ein</i>      | <i>kvit</i> | <i>dør.</i>   |
|     |    | DET.SG.M.INDEF  | white       | door.SG.INDEF |
|     |    | ‘A white door.’ |             |               |

The traditional suffixes associated with feminine nouns are used for 90 % (897/1002) of the tokens, regardless of the use of the feminine or masculine indefinite article. No participants use the definite suffix associated with masculine nouns on traditional feminine nouns without also using the masculine article.

Norwegian possessive pronouns can be prenominal and postnominal, and “mine” and your” agree with the gender of the noun. While all adults used traditional forms (2a), children and

adolescents replace the prenominal feminine possessive with a masculine in 31 % (131/425) of the tokens, but the postnominal stays feminine (2b).

- |     |    |   |                              |   |                              |                             |
|-----|----|---|------------------------------|---|------------------------------|-----------------------------|
| (2) | a. | <i>Mi</i><br>My.POSS.SG.F<br>'My cake'  | <i>kaka</i><br>cake.SG.INDEF | – | <i>kako</i><br>cake.SG.DEF.  | <i>mi</i><br>mine.POSS.SG.F |
|     | b. | <i>Min</i><br>My.POSS.SG.M<br>'My cake' | <i>kaka</i><br>cake.SG.INDEF | – | <i>kako</i><br>cake.SG.DEF.F | <i>mi</i><br>mine.POSS.SG.F |

The results reveal that the gender changes before declension in the singular, and that a dialect that has rich nominal morphology shows results in line with previous research on less rich dialects.

#### References

- Aasen, I. (1848). *Det norske Folkesprogs Grammatik*. Kristiania: Werner & Comp.
- Akselberg, G. (2003). Utviklinga av vossamålet ved tusenårsskiftet. I G. Akselberg, A. M. Bødal, & H. Sandøy, *Nordisk dialektologi*, pp. 197-226. Oslo: Novus.
- Blom, E., Polišenska, D., & Weerman, F. (2008). Articles, adjectives, and age of onset: the acquisition of Dutch grammatical gender. *Second Language Research* 24,3, pp. 297–331.
- Busterud, G., Lohndal, T., Rodina, Y., & Westergaard, M. (2019). The loss of feminine gender in Norwegian: A dialect comparison. *Journal of Comparative Germanic Linguistics*, pp. 141-167.
- Busterud, G., Lohndal, T., Rodina, Y., & Westergaard, M. (2021). Hvor mange genus er det i Trondheims-dialekten? *Maal og minne*, pp. 1-36.
- Enger, H-O (2004). On the relation between gender and declension: a diachronic perspective from Norwegian. *Studies in Language* 28, pp. 51-82.
- Heggstad, L. (1932). *Vossemaålet*. Molde: Molde prenteverk.
- Lødrup, H. (2011). Hvor mange genus er det i Oslo-dialekten? *Maal og minne*, pp. 120-136.
- Rodina, Y., & Westergaard, M. (2015). Grammatical gender in Norwegian: Language acquisition and language change. *Journal of Germanic Linguistics*, 27. 2, pp. 145–187.
- Rodina, Y., & Westergaard, M. (2016). Hvor mange genus er det i Tromsø-dialekten? *Maal og minne*, pp. 159-189.
- Rodina, Y., & Westergaard, M. (2021). Grammatical gender and declension class in language change: A study of the loss of feminine gender in Norwegian. *Journal of Germanic Linguistics* 33(3), pp. 235-263.
- Vidsteen, C. (1884). *Oplysninger om Vossemaalet*. Bergen: Griegs bogtrykkeri.

### Humanity, Animacy and “the Rest”: diatopic variation in gender-marking and its consequences for Proto-Dogon

Anonymous

The Dogon language family, composed of twenty or so languages, spoken primarily across the Bandiagara Escarpment of central Mali, holds an unresolved position within Niger-Congo (Hochstetler, Durieux & Durieux-Boon 2004: 16–17). Variably classified as an independent and ancient branch (Blench 2005; Güldemann 2018) or as a subbranch of Mande or Gur, the interpretation of Noun Class markers within Dogon and its Proto Language weighs into its phylogenetic association with either of these higher phyla and the reconstruction of Noun-Classes or Genders within either.

Looking synthetically at the Dogon languages, half of those accounted for make a morphological distinction between human or other animate nouns and inanimates (“the rest”). Among those expressing a morpho-syntactic distinction, 60% distinctly distinguish humans from all non-humans. Eight of the documented languages make no formal distinction, though several of these, including Togo Kan, Tomo Kan, and Nanga are claimed to have lexicalised traces of archaic morphological distinctions (Heath 2015; 2016; Dyachkov Under Review, etc.). Hantgan (*In Preparation*) has also argued for the preservation of a more extensive Noun Class system in Kindige Bondu So.

Figure 1: Distribution of Noun-Class Marking (Human, Animate, Non-Animate: singular and plural) across the Dogon Languages

Typology	-n sg, -m-plural				-n sg, -ŋ pl.	m sg ∅ pl.		∅ sg, -g plural					∅ sg, -M plural		Other			
	Jamsay	Yomo So	Tommo So	Toro Tegy	Donno So	Bankan Tey	Ben Tey	Nanga	Penange	Bunoge	Tiranige	Ampari	Tebul Ure	Yanda Dom	Dogul Dom	Togo Kan	Tomo Kan	Bondu So
Hum vs. Non-Hum	x	x	x	x	x										x			
Anim. vs. Inanim.						x	x						x	x				x
None								L	x	x	x	x				L	L	

Currently, the phonological expression of morphological animacy categories presents a complex linguistic puzzle. In much of North Eastern and North Central Dogon a labial nasal /-m/, /-mu/, /-mbo/ is used to indicate Human or Animate Plurals and a core of North-Eastern and Eastern Cliffs languages likewise employ an alveolar nasal /-n/, /-nV/ to indicate Human and Animate Singulars. Ben Tey however, indicates animate singulars with /-m/, leaving the plural unmarked, and it is unclear whether this Ben Tey /-m/ arose through a phonological change of Proto-Dogon root-final \*/-n/ → /-m/, through the reinterpretation and re-use of morphological markings, or whether branches of the Dogon Family have independently innovated Noun Class markings unrelated to their supposed Niger-Congo ancestor. Here we present the results of an ongoing comparative study within Dogon, made possible through extensive fieldwork and compiling of the Dogon Comparative Wordlist (Heath et al. 2016; 2023).

Considering the synchronic distribution of Animacy and Human markings, several pronouns animate singular \*mu, animate plural \*bu, seem to have been inherited from the proto-language. The widespread presence of Human singular marking /-n/, /-nɛ/, /-no/ and their overlap with animate categories, likewise points towards an early origin of Human marking at least within certain subbranches. We present several plausible typologically-driven phylogenies based upon the distribution of these potential gender categories and their overt morphological markers.

### References

- Blench, Roger. 2005. A Survey of Dogon Languages in Mali: Overview. *Ogmios : Newsletter of the Foundation for Endangered Languages* 26. 14–17.
- Dyachkov, Vadim. Under Review. *A Grammar of Tomo Kan Dogon* (Institute of Linguistics RAS). UCL Press.
- Güldemann, Tom. 2018. Historical linguistics and genealogical language classification in Africa. In *The Languages and Linguistics of Africa*, 58–444. De Gruyter Mouton.  
<https://doi.org/10.1515/9783110421668-002>.
- Hantgan, Abbie. In Preparation. *Bondu So Noun Classes*. Paris: LLACAN.
- Heath, Jeffrey. 2015. *A Grammar of Togo Kan (Dogon language family, Mali)*. Ann Arbor, Michigan: University of Michigan.
- Heath, Jeffrey. 2016. *A Grammar of Nanga (Dogon language family, Mali)*. Ann Arbor, Michigan: University of Michigan.
- Heath, Jeffrey, Laura McPherson, Kirill Prokhorov & Steven Moran. 2016. Dogon Comparative Wordlist. Jena: Max Planck Institute for the Science of Human History.  
<http://cdstar.shh.mpg.de/bitstreams/EAEA0-C97A-A1D2-2E76-0/a.xls>.
- Heath, Jeffrey, Steven Moran, Robert Forkel, Kirill Prokhorov, Abbie Hantgan & Johann-Mattis List. 2023. CLDF dataset derived from Heath et al’s “Dogon Comparative Wordlist” from 2016. CLDF. Jena: Max Planck Institute for the Science of Human History.  
<https://doi.org/10.5281/zenodo.8238983>.
- Hochstetler, J. Lee, J. A. Durieux & E. I. K. Durieux-Boon. 2004. *Sociolinguistic Survey of the Dogon Language Area* (SIL Electronic Survey Reports 2004-004). Dallas, Texas: SIL International.  
<http://www.sil.org/silesr/abstract.asp?ref=2004-004>.



**Langues & Langage  
à la croisée des Disciplines**

**1ère Rencontre annuelle LLcD**

Colloque : 9-11 septembre, 2024

Ecole d'été : 12-14 septembre, 2024

Sorbonne Université - 28 Rue Serpente, 75006 Paris, France



## **Workshop 7**

**Morph2Read: From morphological knowledge  
to reading comprehension in L1 and L2  
learners**

**Des connaissances morphologiques à la  
compréhension de la lecture chez des  
apprenants en L1 et L2**

## **Explorer la connaissance morphologique dans les mots préfixés et suffixés : le cas des locuteurs francophones plurilingues**

Despoina STEFANO  
*Université Paul-Valéry Montpellier 3 – DIPRALANG (UR 739)*

La compétence morphologique joue un rôle crucial dans l'acquisition et le traitement du lexique, aussi bien dans la langue L1 que dans la L2. Cette aptitude découle de la compréhension morphologique du mot, impliquant à la fois sa reconnaissance et sa compréhension. Étant donné que le mot revêt une forme et porte du sens, la reconnaissance et la compréhension de celui-ci impliquent à la fois sa forme et son sens (ex. Corbin, 1987/1991), en particulier dans le cas des mots complexes.

Il est essentiel de comprendre que cette compétence, ainsi que les connaissances morphologiques associées, sont fondamentales pour la compréhension de la lecture, compétence qui peut être difficile à développer. Cette compétence est étroitement liée au traitement lexical (L1 et L2) des locuteurs. Par conséquent, les différents profils de locuteurs, en termes de compétence linguistique et de niveau scolaire, présentent des disparités quant à leur compétence morphologique et, par extension, à leurs performances linguistiques.

De nombreuses études, s'appuyant sur des protocoles expérimentaux tels que l'amorçage masqué, se sont concentrées sur divers aspects des mots morphologiquement complexes. Ces aspects incluent les mots comportant des préfixes, des suffixes. Ces différents aspects de ce type des mots (i.e. préfixés et suffixés) concernent les caractéristiques différentes qui influencent le traitement du langage écrit chez les locuteurs d'une L1 ou d'une L2, par exemple, la fréquence du préfixe/suffixe (Giraud & Grainger, 2001 ; Giraud & Voga, 2013), celle de la saillance du préfixe/suffixe (ex, grec L2 : Voga, Anastassiadis-Syméonidis & Nikolaou, 2018 ; italien : Giraud & Dal Maso, 2016), la transparence sémantique de préfixe (Diependaele, Sandra, & Grainger, 2009).

Notre recherche examine le traitement morphologique chez les stagiaires de l'École de la Deuxième Chance à travers deux expériences utilisant le protocole d'amorçage masqué (Forster & Forster, 2003), menées auprès de locuteurs francophones plurilingues dont le français n'est pas nécessairement la L1. Ceux-ci n'ont pas tous suivi un parcours scolaire en France, où ils résident lors des passations d'expérience, et ils se sont engagés dans une formation visant l'insertion professionnelle. Nous avons proposé deux expériences d'amorçage masqué, basées sur une tâche de décision lexicale impliquant le même type de



participants, d'âge moyen 19,5 ans, qui doivent déterminer si le mot présenté est un mot existant ou non. Les expériences proposées comprennent une expérience portant sur les mots préfixés *pré-/pro-* (ex. *préfigurer, proportion*) et *com-/con-/co-* (ex. *compatriote, contour, coopérer*) et une autre sur les mots suffixés *-iste* et *-ique* (ex. *cycliste, diplomatique*), les deux avec une SOA de 48ms. L'objectif de ces expériences est d'étudier la représentation des mots dans le lexique mental et les effets entre l'amorce et la cible, par exemple l'effet morphologique (ex. *conserver – conservateur, pluraliste - pluralisme*). Les deux expériences sont comparables dans la mesure où nous utilisons le même type de protocole, avec une période d'attente entre le stimulus et la réponse (SOA) identique, et en ciblant la même population de locuteurs pour laquelle il n'existe pas beaucoup de données expérimentales de ce type. Les données observées suite aux résultats ressortant de ces expériences sont aussi comparables, soulignant le besoin d'un renforcement de la conscience et de la compétence morphologique de ce public présentant des fragilités. Ces données peuvent amener à des réflexions sur des perspectives soutenant des stratégies pour l'enseignement et le développement linguistiques utilisant des méthodes tant déductives qu'inductives (Mill, 1866) comme par exemple des stratégies d'apprentissage du vocabulaire (Efthymiou, 2013) qui peuvent être appliquées avec ajustements selon le public cible. De plus, à l'aide des outils numériques d'évaluation et de suivi des connaissances des élèves en lecture nous pouvons réfléchir à mettre en place des pratiques afin de renforcer l'apprentissage et le développement des connaissances linguistiques comme c'est le cas pour l'hybridation en langues (Nissen, 2019).

Cependant, nous soulignons également l'importance de tester une gamme variée de suffixes, de préfixes et de structures morphologiques avec ce type de profil de locuteurs/lecteurs, avant de tirer des conclusions sur le type spécifique de compétence morphologique qui sous-tend les effets du traitement morphologique.

Mots clés : morphologie dérivationnelle, étude expérimentale, amorçage masqué, plurilinguisme, école de la deuxième chance

## **Bibliographie**

- Corbin, D. (1987/1991). *Morphologie dérivationnelle et structuration du lexique*, vol. 2. Tübingen / Villeneuve d'Ascq: Max Niemeyer Verlag / Presses Universitaires de Lille.
- Diependaele, K., Sandra, D., & Grainger, J. (2009). Semantic transparency and masked morphological priming: the case of prefixed words. *Memory & cognition*, 37(6), 895–908. <https://doi.org/10.3758/MC.37.6.895>
- Efthymiou, A. (2013). *L'enseignement du vocabulaire à l'école primaire – Théorie et applications [Η διδασκαλία του λεξιλογίου στο δημοτικό σχολείο : Θεωρία και εφαρμογές]*. Thessaloniki : Epikentro.
- Ellis, N. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143-188.
- Forster, K. & Forster, J. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavioral Research Methods: Instruments & Computers* 35, 116-124.
- Giraud, H. & Dal Maso, S. (2016). Suffix perceptual salience in morphological masked priming. *Lingue e Linguaggio*, 15(1), 83–106. <https://www.rivisteweb.it/doi/10.1418/83655>
- Giraud, H., & Grainger, J. (2001). Priming complex words: Evidence for supralexicalexical representation of morphology. *Psychonomic Bulletin and Review*, 8(1): 127-131. <https://doi.org/10.3758/BF03196148>
- Giraud, H. & Voga, M. (2013). Prefix units in the mental lexicon, dans N. Hathout, F. Montermini et J. Tseng (Eds.), *LINCOM Studies in Theoretical Linguistics 51 (LSTL 51) : Morphology in Toulouse. Selected Proceedings of Décembrettes 7 (Toulouse, 2-3 December 2010)*, (pp. 61-78). LINCOM academic publishers.
- Mill, J. S. (1866). *Système de logique déductive et inductive : exposé des principes de la preuve et des méthodes de recherche scientifique*, (Peisse L. Trad., 6<sup>ème</sup> éd.), Vol. 2. Paris : Librairie Philosophique de Ladrangé.
- Nissen, E. (2019). *Formation hybride en langues : articuler présentiel et distanciel*. Paris: Didier.
- Voga, M., Nikolaou, G. & Anastassiadis-Symeonidis A. (2018). Prefix conceptual salience in L2 acquisition and processing: a reading-time study within context. In J. Audring, N. Koutsoukos, F. Masini, & W. Sandler (Eds.), *Online Proceedings of the 11th Mediterranean Morphology Meeting (MMM)*, 11, 81-92. University of Cyprus, Nicosia. <https://doi.org/10.26220/mmm.2875>

## **Processing morphology while reading: An eye-tracking study with typical and atypical adult readers**

**Introduction.** A body of research has evidenced that morpheme-based reading is a useful strategy for less skilled readers and children with dyslexia (Deacon et al., 2017; Angelelli et al. 2017). Several studies have assessed the role of morphology in child populations mainly deploying reading-aloud tasks of isolated words (a.o., Burani et al. 2008), whereas text reading has been comparatively less investigated. With this study, we addressed the question of whether morphological complexity also aids *adult* readers with and without dyslexia, and if it is equally relevant in reading texts or isolated words.

**Method.** A reading-aloud task with eye-movement recordings was administered to 25 Italian adults diagnosed with dyslexia (mean age = 23;6) and 37 Italian typical readers (mean age 22;8). The testing procedure consisted of two phases: 1) Text reading, where participants were asked to read aloud two four-sentence texts, and 2) word lists reading, with lists composed of nine “target” words taken from texts (see Figure 1). Words in texts and lists belonged to one of the four experimental conditions, namely the Simple Word (SW, e.g., *universo*, “universe”), the Complex Word (CW, e.g., *venditore*, “seller”), the Simple Nonword (SNW, e.g., *carcibuso*) and the Complex Nonword (CNW, e.g., *carcitore*). For the statistical analysis, a series of linear mixed effects regression models were run on behavioral and eye movement metrics, focused on group-level differences. Behavioral analysis included the investigation of the dependent variable reading duration, as a function of the two experimental groups (dyslexia and control), the morphology condition (simple vs. complex), and the word condition (word vs. nonword). Eye movement analyses examined local (i.e., word/interest area-based) metrics of eye movement events during reading (i.e., total fixation duration, number of fixations, first fixation duration, saccade amplitude) as a function of the groups, morphology, and word conditions, and modality (text vs. list).

**Results.** Unsurprisingly, adults with dyslexia showed higher reading time, longer fixations, and a greater number of fixations on target words in texts and lists. Both typical and atypical readers exploited morphological cues, which turned out to be relevant when reading pseudowords in texts. However, adults with dyslexia took greater advantage of morphological complexity in pseudo-word text decoding.

**Discussion.** Our findings provide strong evidence for the pivotal role of derivational morphology in both typical and atypical reading processing. Derivational suffixes are essential functional morphemes that aid reading by carrying syntactic-semantic features that allow the efficient syntactic parsing of a text. Interestingly, these morphological cues are especially exploited by readers with dyslexia, who can capitalize on their ability to parse words into morphemic units and rely on the role that suffixes play in bringing morphosyntactic coherence into a text.

### **References.**

- Angelelli, P., Marinelli, C. V., De Salvatore, M., & Burani, C. 2017. Morpheme-based reading and spelling in Italian children with developmental dyslexia and dysorthography. *Dyslexia*, 23(4), 387-405.
- Burani, C., Marcolini, S., De Luca, M., & Zoccolotti, P. 2008. Morpheme-based reading aloud: Evidence from dyslexic and skilled Italian readers. *Cognition*, 108(1), 243-262.

Deacon, S. H., Tong, X., & Francis, K. 2017. The relationship of morphological analysis and morphological decoding to reading comprehension. *Journal of Research in Reading*, 40(1), 1-16.

## Organisation et fonctionnement du lexique bilingue et plurilingue : un modèle fondé sur le transfert des connaissances morphologiques

Le lexique mental bilingue et plurilingue (désormais LMB) présente des similitudes avec le lexique monolingue, à la fois sur le plan de l'organisation et sur celui des processus responsables de son fonctionnement. Comment pourrait-il en être autrement pour décrire ce qui sous-tend la compétence de type *one mind two languages* (Grosjean, 2001), où un même contenu sémantique correspond à deux formes différentes du point de vue formel ?

Notre contribution présente un modèle de LMB fondé sur des données d'amorçage masqué *cross-script* (grec-français), qui attribue une large place :

a) aux processus généraux que Bybee (2010 :7) décrit comme *domain-general*, c'est-à-dire qui décrivent à la fois le langage et d'autres aspects du comportement humain : la catégorisation, la compression d'information (*chunking*), le stockage étendu en mémoire (*rich memory storage*), l'analogie et l'association multimodale. Le principe de compositionnalité par ex., central à la formation morphologique, serait lié au *chunking* ;

b) à l'acceptation que les effets cognat et non-cognat (récurrents) observés correspondent à la somme d'activations/inhibitions, telles que définies dans le modèle d'activation interactive de McClelland & Rumelhart (1981), complété par un niveau sémantique ;

c) à une organisation de type morphologique, observable lorsque l'amorce (ex. ιδεατό /idéató/ 'idéal') active la cible morphologiquement reliée (*idée*), et induit des effets de la même amplitude que l'amorçage cognat. Cela prouve que ces deux représentations sont interconnectées, *via* les *clusters lexicaux* décrits par Bybee (1988). Une telle organisation n'obéit pas à la logique du découpage en morphèmes (Corbin, 1987/1991 ; Bybee, 2010 ; Booij, 2010) : « Le mot n'est pas physiquement démembré (*physically dismembered*), mais ses parts sont néanmoins identifiées » (Bybee 1988 : 127). Ces 'parts' correspondent à des formants de nature diverse, entre autres :

1/ des morphèmes lexicaux, ex. des bases comme *athlèt-* dans *athlète, athlétisme, pentathlon*, etc., tous cognats grec-français (et d'autres langues), dont la taille de la famille morphologique influence le traitement ;

2/ des morphèmes non-lexicaux, comme les suffixes productifs *-iste/-isme* qui induisent des effets significatifs d'amorçage morphologique inter-langue en l'absence d'une base-mot (ou '0-base'), suggérant que le contact de la représentation lexicale (L1 vers L2 et inversement) s'établit grâce à des connexions qui forment une sorte de 'famille morphologique à noyau vide', sans doute fondée sur la base de la série.

3/ des préfixes grâce auxquels émerge une co-activation entre unités lexicales non-cognat, qui diffèrent sur le plan de la forme, mais qui présentent une structure morphologique commune (et sont non-décomposables), i.e., des effets traduction et morphologique concomitants et de même amplitude (29ms\* et 34ms\*) induits par l'amorce *συνεπώς* /sine'pos/ 'en conséquent' sur la cible *conséquence*. Ce type d'effet, combiné à la compétition exercée par des familles de mots (formées sur une même base-mot), ne peut qu'avoir lieu à un niveau post-lexical. Ces résultats compromettent largement l'approche sous-lexicale (Diependaele et al. 2011) et

pointent le fait que les transferts morphologiques, ainsi qu'une organisation de type morphologique représentent le fondement de cette architecture lexico-sémantique commune (Schoonbaert et al., 2009 :569).

#### References

- Booij, G. (2010). *Construction Morphology*. Oxford: Oxford University Press.
- Bybee, J. (1988). Morphology as lexical organisation. In Michael Hammond et Michael Noonan (eds.), *Theoretical Morphology. Approaches to modern linguistics*, 119-142. San Diego: Academic Press.
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.
- Corbin, D. (1987/1991). *Morphologie dérivationnelle et structuration du lexique*, vol. 2. Tübingen/Villeneuve d'Ascq: Max Niemeyer Verlag / Presses Universitaires de Lille.
- Diependaele, K., Duñabeitia, J.A., Morris, J. & Keuleers, E. (2011). Fast morphological effects in first and second language word recognition, *Journal of Memory & Language*, 64, 344-358.
- Schoonbaert, S., Duyck, W., Brysbaert, M. et Hartsuiker, R. (2009). Semantic and translation priming from a first language to a second and back: Making sense of the findings, *Memory & Cognition*, 37(5), 569-586.
- Grosjean, F. (2001). The bilingual's language modes. In J. Nicol (Ed.). *One Mind, Two Languages: Bilingual Language Processing*. Oxford: Blackwell. 1-22.
- McClelland, J. L. et Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88, 375-407.

**Proposition de communication au workshop WS7 Des connaissances morphologiques à la compréhension de la lecture chez des apprenants en L1 et L2**

**Langues & Langage à la croisée des Disciplines. 1<sup>ère</sup> Rencontre annuelle LLcD**

**Titre de la communication : Conscience morphologique et compréhension en lecture chez les étudiants de français L1 : Lien direct, lien indirect et modération par le niveau scolaire.**

La capacité à manipuler et analyser la dimension morphologique des mots (c'est-à-dire la conscience morphologique, désormais CM) est un facteur prédictif important de l'apprentissage de la lecture, en particulier de la compréhension de la lecture (Freitas Jr et al., 2018 ; Kirby et al., 2012 ; Liu et al., 2023 ; Rassel et al., 2021 ; Ruan et al., 2018). Certains modèles d'apprentissage de la lecture proposent deux modes de contribution de la CM à la compréhension écrite (désormais CE) : la première, indirecte, passe par la reconnaissance des mots en favorisant un accès lexical de haute qualité qui, à son tour, soutiendrait la CE (Ehri, 2014 ; Perfetti, 2007) ; la seconde, directe, qui rendrait compte d'une contribution par l'intégration du sens et de la syntaxe soutenant les processus de CE de haut niveau (Perfetti & Stafura, 2014). Ces liens ont récemment été modélisés dans le modèle *MPF* de Levesque et al. (2020). Cependant, les preuves empiriques éprouvant l'existence de ces deux voies obtiennent des résultats confus, en particulier entre la 2<sup>e</sup> et 4<sup>e</sup> année d'apprentissage (Deacon et al., 2014 ; Levesque et al., 2017, 2019 ; Oliveira et al., 2020 ; Proctor et al., 2013). Par ailleurs, d'autres résultats suggèrent que la contribution unique de la CM à la CE est universelle et stable, alors que le lien indirect par la lecture de mots peut varier selon l'opacité orthographique des langues (Desrochers et al., 2017 ; Lee et al., 2023 ; Manolitsis et al., 2019) ou être soumis à un *shift* développemental (Deacon et al., 2014) : Cette médiation pourrait être moins importante dans les langues opaques ou disparaître avec l'automatisation de la reconnaissance des mots écrits. A notre connaissance, cet effet développemental n'a jamais été testé empiriquement. Nous souhaitons contribuer à clarifier ces résultats contradictoires en présentant une étude transversale portant sur des enfants français de 3<sup>e</sup> et 5<sup>e</sup>. Nous examinerons 1/ si les effets directs et indirects (à travers la fluidité de lecture de mots) de la CM sur la CE sont observés; 2/ si le niveau scolaire affecte la relation entre CM et CE.

Nous avons testé 137 élèves francophones natifs de 3<sup>e</sup> et 5<sup>e</sup> années au développement typique et évalué la CM à l'aide d'une tâche d'analogie de mots, la fluidité de la lecture à l'aide du test de lecture de mots EVALEC (Sprenger-Charolles et al., 2010) et la CE à l'aide d'un test de comparaison du sens des phrases (Ecalte, 2011). Les résultats montrent une relation directe entre CM et CE, mais pas de relation indirecte via la fluidité de la lecture de mots. L'année scolaire modère la relation entre CM et fluidité en lecture de mots. En 3<sup>e</sup> année, la CM contribue à la fluidité, mais pas en 5<sup>e</sup> année. Malgré l'effet de médiation non significatif, ces résultats révèlent un schéma développemental selon lequel la CM aide les enfants à lire les mots en début d'apprentissage tout en jouant un rôle à plus long terme sur le processus général de compréhension.

**Mots-clés :** Conscience morphologique, compréhension écrite ; fluidité en lecture de mots.

**Key words:** Morphological awareness; Reading Comprehension; Reading Fluency;

## Références bibliographiques:

- Deacon, S.H., Kieffer, M.J. & Laroche, A. (2014). The relation between morphological awareness and reading comprehension: Evidence from mediation and longitudinal models. *Scientific Studies of Reading*, 18, 432–451. <https://doi.org/1.1080/10888438.2014.926907>
- Desrochers A., Manolitsis G., Gaudreau P., Georgiou G. (2017). Early contribution of morphological awareness to literacy skills across languages varying in orthographic consistency. *Reading and Writing*, 31(8), 1695–1719. <https://doi.org/1.1007/s11145-017-9772-y>
- Ecalte, J. (2011). *TeCoPé. Test de Compréhension de Phrases écrites*. Eurotests Editions. [www.eurotests.fr](http://www.eurotests.fr)
- Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading*, 18, 5–21. <https://doi.org/1.1080/10888438.2013.819356>
- Freitas Jr., P. V., Mota, M. M. P. E. & Deacon, S. H., (2018). Morphological awareness logical, word, reading, and reading comprehension in Portuguese. *Applied Psycholinguistics*, 39(1), 1-19. <https://doi.org/1.1017/S0142716417000479>
- Kirby, J., Deacon, H., Bowers, P., Izenberg, L., Wade-Woolley, L. & Parrilla, R. (2012). Children morphological awareness and reading ability. *Reading and Writing: An Interdisciplinary Journal*, 25, 389-41. <https://doi.org/1.1007/s11145-010-9276-5>
- Lee, J., Wolters, A. & Young-Suk Grace Kim Y.S. G. (2023). The Relations of Morphological Awareness with Language and Literacy Skills Vary Depending on Orthographic Depth and Nature of Morphological Awareness, *Review of Educational Research*, 9(4), 528-558. <https://doi.org/10.3102/00346543221123816>
- Levesque, K., Breadmore, H., & Deacon S. H. (2020). How morphology impacts reading and spelling: Advancing the role of morphology in models of literacy development. *Journal of Research in Reading*, 44, 10-26. <https://doi.org/1.1111/1467-9817.12313>
- Levesque, K., Kieffer, M. J., & Deacon, S. H. (2019). Inferring meaning from meaningful parts: The contributions of morphological skills to the development of children’s reading comprehension. *Reading Research Quarterly*, 54(1), 63–8.
- Levesque, K.C., Kieffer, M.J. & Deacon, S.H. (2017). Morphological awareness and reading comprehension: Examining mediating factors. *Journal of Experimental Child Psychology*, 160, 1–2. <https://doi.org/1.1016/j.jecp.2017.02.015>
- Liu, Y., Margriet Anna Groen, M. A. & Kate Cain, K. (2024). The association between morphological awareness and reading comprehension in children: A systematic review and meta-analysis. *Educational Research Review*, 42, <https://doi.org/10.1016/j.edurev.2023.100571>
- Manolitsis, G., Georgiou, G. K., Inoue, T., & Parrilla, R. (2019). Are morphological awareness and literacy skills reciprocally related? Evidence from a cross-linguistic study. *Journal of Educational Psychology*, 111(8), 1362–1381. <https://doi.org/1.1037/edu0000354>
- Oliveira, M., Levesque, K., Deacon, S. H., & da Mota, M. E. P. E. (2020). Evaluating models of how morphological awareness connects to reading comprehension: A study in Portuguese. *Journal of Research in Reading*, 43(2), 161-179. <https://doi.org/1.1111/1467-9817.12296>
- Perfetti, C.A. & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18(1), 22–37. <https://doi.org/1.1080/10888438.2013.827687>
- Proctor, C.P., Silverman, R.D., Harring, J.R. & Montecillo, C. (2012). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish–English bilingual children in elementary school. *Reading and Writing: An Interdisciplinary Journal*, 25, 1635–1664. <https://doi.org/1.1007/s11145-011-9336-5>.
- Rassel, A., Facon, B., & Casalis, S. (2021) Morphological awareness and learning to read: impact of socio-economic status in French third graders. *Journal of Research in Reading*, 44: 228–246. <https://doi.org/10.1111/1467-9817.12321>
- Ruan, Y., Georgiou, G. K., Song, S., Li, Y., & Shu, H. (2018). Does writing system influence the associations between phonological awareness, morphological awareness, and reading? A meta-analysis. *Journal of Educational Psychology*, 110(2), 180–202. <https://doi.org/10.1037/edu0000216>
- Sprenger-Charolles, L., Colé, P., Piquard-Kipffer, A., & Leloup, G. (2010). *EVALEC, Batterie informatisée d’évaluation diagnostique des troubles spécifiques d’apprentissage de la lecture*. Ortho Edition.



## ENHANCING MORPHOLOGICAL AWARENESS IN TYPICAL AND ATYPICAL READERS: EFFECTS OF A PRIMARY SCHOOL INTERVENTION ON MORPHOLOGICAL AND READING SKILLS

Marta Tagliani, Maria Vender & Chiara Melloni

*University of Verona*

**Theoretical Background.** In psycholinguistic research, there is growing interest in the role of morphology and (meta)morphological skills in facilitating the reading processes (Arnbak & Elbro, 2000; Casalis et al., 2004; Melloni & Vender, 2022; Vender & Melloni, 2021). Such skills are particularly relevant for individuals diagnosed with developmental dyslexia: indeed, research has shown that morphological skills (i.e., the ability to recognize and break down written words into morphemes) play an essential role in reading in dyslexic individuals (both adults and children) and are used, even into adulthood, for better decoding and comprehension of the written text (Casalis et al., 2004; Vender & Melloni, 2021). Based on the scientific evidence regarding the positive effect of morphological awareness development on reading skills in dyslexic and typical-developing subjects (Deacon et al., 2014; Cain & Oakhill, 2011; Tannenbaum et al., 2006) we implemented an innovative teaching methodology to strengthen the morphological abilities of children with both typical and atypical development.

**Methodology and Materials.** 40 fifth-grade primary school children (Bilingual Italian/Venetan = 34; Monolingual Italian = 6; DSA or poor readers = 10, age range 10;0-11;1, F = 18, M = 21) participated in the study: half of them were randomly assigned to the experimental group, which took part in the morphological intervention, while the remaining students did not receive any intervention and constitute the control group. The morphological intervention lasted 10 hours (5 meetings). Each session involved teaching aiming to increase the children's morphological awareness and morphological decoding. Practical activities included i) decoding of the suffix; ii) analysis of derived and simple words in isolation and in short sentences (e.g., manual decomposition of derived words into blocks consisting of root + suffix); iii) identification of derived words in short texts. The training covered the analysis of 15 high-frequency Italian suffixes (Gaeta & Ricca, 2003) for nouns and adjectives, and was organized in thematic units (e.g., agentive, locative, and action nouns, relational and possibility adjectives).

All children were tested in a series of pre- and post-tests to evaluate the effectiveness of the intervention on morphological competence and its impact on reading abilities. Morphological skills were tested by administering three original tasks modelled after the properties of Italian verbal morphology (Piccinin & Dal Maso, 2022, 2023) building on tests widely used for morphological assessment in English: i) lexical knowledge task (adapted from Deacon et al., 2017); ii) test of morphological structure (adapted from Carlisle, 2000); iii) non-word suffix choice (adapted from Nagy et al., 2003). Reading skills were assessed through standardized reading tasks and a novel task that we designed to address the role of morphological decoding in simple and

derived words and nonwords. Children were also tested in their phonological competence (Vender et al., 2020) by means of a nonword repetition task.

**Results and Conclusion.** Comparison of pre- and post-assessment revealed that all children (with both typical and atypical development) who participated in the morphological intervention improved significantly in the morphological analysis test ( $p = .042$ ) and in the morphological structure test ( $p < .001$ ). Moreover, they showed enhanced skills in their phonological competence ( $p < .001$ ). Children with atypical development improved in word ( $p < .01$ ) and nonword ( $p < .001$ ) reading speed after the training, although an improving trend can also be observed for typical readers ( $p = .056$ ). No significant improvements were instead observed in children who did not receive any intervention.

These findings show that activities aimed to enhance morphological skills significantly affect students' linguistic (morphological and phonological) abilities, with positive effects also on their reading outcomes. Crucially, these positive effects extend to the reading performance of all children, with typical and atypical development, but are particularly robust for the latter, confirming the crucial role of morphological skills in reading achievements.

**Keywords:** morphological awareness, developmental dyslexia, intervention, reading, literacy

## References

- Arnbak, E., & Elbro, C. (2000). The effects of morphological awareness training on the reading and spelling skills of young dyslexics. *Scandinavian journal of educational research*, 44(3), 229-251.
- Cain, K., & Oakhill, J. (2011). Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of learning disabilities*, 44(5), 431-443.
- Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing*, 12, 169-190.
- Casalis, S., Colé, P., & Sopo, D. (2004). Morphological awareness in developmental dyslexia. *Annals of dyslexia*, 54, 114-138.
- Deacon, S. H., Kieffer, M. J., & Laroche, A. (2014). The relation between morphological awareness and reading comprehension: Evidence from mediation and longitudinal models. *Scientific Studies of Reading*, 18(6), 432-451.
- Deacon, S. H., Tong, X., & Francis, K. (2017). The relationship of morphological analysis and morphological decoding to reading comprehension. *Journal of Research in Reading*, 40(1), 1-16.
- Gaeta, L., & Ricca, D. (2003). Frequency and productivity in Italian derivation: A comparison between corpus-based and lexicographical data. *Rivista di Linguistica*, 15, 63-98.
- Melloni, C., & Vender, M. (2022). Morphological awareness in developmental dyslexia: Playing with nonwords in a morphologically rich language. *PLoS ONE*, 17(11): e0276643.

- Nagy, W., Berninger, V., Abbott, R., Vaughan, K., & Vermeulen, K. (2003). Relationship of morphology and other language skills to literacy skills in at-risk second-grade readers and at-risk fourth-grade writers. *Journal of educational psychology*, 95(4), 730.
- Piccinin, S., & Dal Maso, S. (2022). Sviluppare la consapevolezza morfologica: linee guida per un intervento didattico. *Studi di glottodidattica*, 7(3), 85-94.
- Piccinin, S., & Dal Maso, S. (2023). Investigating factors affecting reading comprehension in second-generation pupils: the role of morphological skills. *Lingue e linguaggio*, 22(1), 117-142.
- Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific studies of reading*, 10(4), 381-398.
- Vender, M., & Melloni, C. (2021). Phonological awareness across child populations: How bilingualism and dyslexia interact. *Languages*, 6(1), 1-20.
- Vender, M., Delfitto, D., & Melloni, C. (2020). How do bilingual dyslexic and typically developing children perform in nonword repetition? Evidence from a study on Italian L2 children. *Bilingualism: Language and Cognition*, 23(4), 884-896.

## L'intégration des connaissances morphologiques à la compréhension de texte : Un modèle théorique cognitif fonctionnel

Hélène Giraud - *Laboratoire CLLE (UMR5263) – Université Toulouse Jean Jaurès & CNRS – Toulouse, France.*

Serena Dal Maso – *LaTec – Université de Vérone – Italie*

Sabrina Piccinin – *LaTec – Université de Vérone – Italie*

D'un point de vue linguistique, la morphologie étudie la forme des mots, leur structure, la manière dont ils sont construits et les éléments qui le constituts. L'étude de la morphologie nous renseigne sur les rapports d'interdépendance entre le lexique (la forme et le sens des mots) et la syntaxe (les relations grammaticales entre les mots d'un syntagme). Dans les langues alphabétiques, les mots du lexique peuvent être décrits comme des unités abstraites appelées des lexèmes, qui se réalisent concrètement dans le langage sous la forme de mots.

Par exemple, le lexème [ROND<sub>Adj</sub>] produit en français les mots formes : *rond, ronde, ronds, rondes, rondeur, rondin, rondement, arrondir, arrondissement, etc...*Ces mots formes comportent un thème/un radical (i.e., *rond-*) qui véhicule le sens du lexème (i.e., 'dont la forme extérieure constitue un cercle') et sa classe grammaticale (i.e., un adjectif) dont un mot forme donné est issu. Sont associés au thème un ou plusieurs affixes ou exposants (e.g., *-eur, -in, -e, -es, -s,*) qui peuvent modifier la catégorie grammaticale du lexème de base (e.g., [ROND<sub>Adj</sub>] => *arrondir*<sub>Verbe</sub>), le sens qu'il acquiert dans le langage ('action de rendre rond') et son rôle au niveau de la syntaxe (en matière d'accord et de gouvernement) au regard des traits et valeurs morpho-syntaxiques que le mot forme porte (i.e., verbe à l'infinitif). Les informations linguistiques contenues dans les mots inventorient aussi bien leur forme phonologique/orthographique, leur sens, que leur rôle syntaxique, ces éléments contribuant donc à associer au signifiant son sens inhérent (dans le lexique) et contextuel (dans la phrase).

Du côté de l'acquisition de la lecture, les connaissances morphologiques sont envisagées comme un terme générique intégrant plusieurs dimensions (décodage, analyse sémantique et syntaxique des mots, analyse relationnelles et distributionnelles entre les mots morphologiquement reliées, e.g., Nagy, Carlisle, & Goodwin, 2013) et elles seraient acquises de façon implicite via l'exposition au langage ou apprises de façon explicite au travers d'un enseignement formel (e.g., Ellis, 2008). Si les premières connaissances dépendent de facteurs liés au profil sociolinguistique des élèves (i.e., langue maternelle, âge, niveau scolaire, statut socio-économique), les secondes font partie des habilités qui relèvent d'une pédagogie d'enseignement ciblée sur la morphologie, dont l'objectif vise à faire réfléchir les élèves sur les dimensions syntagmatique (i.e., constructions morphologiques) et paradigmatiques des mots (i.e., relations sémantiques et formelles entre les mots issus d'un même lexème ou portant un même affixe ou exposant), ces habilités renvoyant à la notion de « conscience morphologique » (développée par Carlisle, 2000 ; Nagy, Carlisle & Goodwin, 2014).

De nombreuses recherches ont depuis permis de mettre en évidence que les connaissances morphologiques contribuent significativement à la lecture, l'écriture et à la compréhension des mots, les lecteurs possédant de bonnes compétences morphologiques présentant de bonnes performances en lecture. Ces études suggèrent par ailleurs que ces connaissances induiraient également des bénéfices en cascade sur la compréhension de texte (e.g., Deacon, Tong & Francis, 2017 ; Levesque & Deacon, 2022).

L'impact de la morphologie reste cependant sous-représentée dans les modèles du développement de la lecture proposés jusqu'alors dans la littérature (voir Coltheart, 2006 pour une revue). Des travaux récents ont tenté de proposer d'intégrer cette dimension pour expliquer la manière dont les différentes connaissances morphologiques pourraient interagir avec les autres composantes de la lecture constituées par les connaissances linguistiques et du système d'écriture, les processus d'identification des mots, ceux de la compréhension du texte lu, cet ensemble de connaissances étant mises en lien

avec les connaissances générales (e.g., Levesque, Breadmore & Deacon, 2022 ; Rastle, 2018). La réflexion menée demeure cependant largement circonscrite aux mécanismes qui s'opèrent sur les mots (i.e., le lexique), celle portant sur leur transfert bénéfique au niveau des mécanismes de la compréhension de texte demeurent relativement vague (e.g., Levesque et coll., 2022 : 18).

Dans cette proposition de communication, nous nous appuyons sur les données comportementales de la littérature (e.g., Piccinin & Dal Maso, 2023) et sur les résultats de nos travaux expérimentaux, actuellement en cours, dont l'objectif consiste à étudier les effets bénéfiques directs des connaissances morphologiques pour la compréhension de texte, aux niveaux littéral et figuré, chez des élèves de CE1, CE2 et CM1, de différents profils sociolinguistiques, scolarisés dans des écoles de type REP et non REP de l'académie de Toulouse.

Nous réfléchissons précisément à la manière dont les connaissances morphologiques codées au niveau du lexique mental contribuent à améliorer la compréhension de la lecture, laquelle est envisagée du point du décodage du texte lu et de celui sa compréhension efficace (i.e., la littéracie). L'architecture et les principes fonctionnels du modèle théorique proposé s'inspirent des modèles connexionnistes symboliques (e.g., McClelland & Rumelhart, 1981) rendant compte des processus cognitifs du traitement morphologique (e.g., Grainger & Grainger, 2001 ; Grainger & Dal Maso, 2016) et de la compréhension de texte (e.g., Kintsch, 1988 ; 1998 ; Bianco et al., 2004 ; Cordier, 2012).

#### Références bibliographiques

- Audring, J. (2019). Mothers or sisters? The encoding of morphological knowledge, *Word Structure*, 12(3), 274-296.
- Bianco, M., Lima, L., & Sylvestre, E. (2004). Comment enseigner les stratégies de compréhension en lecture. Dans E. Gentaz et P. Dessus (Eds.), *Comprendre les apprentissages* (pp 48-68), Paris : Dunod.
- Booij, G.E. (2010). *Construction Morphology*. Oxford: Oxford University Press.
- Carlisle, J.F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing*, 12, 169–190.
- Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education*, 4(1), 5–17.
- Cordier, F. (2012). Chapitre 6. Compréhension du langage : un paradigme pour la cognition. In M. Denis (éd.), *La psychologie cognitive* (1-). Éditions de la Maison des sciences de l'homme.
- Deacon, S.H., Tong, X. & Francis, K. (2017). The relationship of morphological analysis and morphological decoding to reading comprehension. *Journal of Research in Reading*, 40(1), 1-16.
- Ellis, N.C. (2008). Implicit and explicit knowledge about language. In J. Cenoz and N. H. Hornberger (eds), *Encyclopedia of Language and Education, 2nd Edition, Volume 6: Knowledge about Language* (pp.1–13). Springer Science.
- Grainger, H., & Grainger, J. (2001). Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin & Review*, 8(1), 127-131.
- Grainger, H., & Dal Maso, S. (2016). The saliency of complex words and their structure: Which comes first? *Frontiers in Psychology*, 7:1778.
- Harris, K. R., Graham, S., Urdan, T., McCormick, C. B., Sinatra, G. M. & Sweller, J. (2011), *APA educational psychology handbook, Vol. 1. Theories, constructs, and critical issues*. American Psychological Association.
- Nagy, W. E., Carlisle, J. F., & Goodwin, A. P. (2013). Morphological knowledge and literacy Acquisition. *Journal of Learning Disabilities*, 47(1), 3–12.
- Kintsch, W. & van Dijk, T. (1978). Toward a Model of Text Comprehension and Production, *Psychological Review*, 85, 363-394.

- Levesque, K. C., Breadmore, H.L. & Deacon, H.S. (2021). How morphology impacts reading and spelling: advancing the role of morphology in models of literacy development. *Journal of Research in Reading*, 44, 10-26.
- Levesque, K. C. & Deacon, H. (2022). Clarifying links to literacy: How does morphological awareness support children's word reading development? *Applied Psycholinguistics*, 1-23.
- van Dijk, T. & Kinstch, W. (1983). *Strategies of discourse comprehension*. New York : Academic Press.
- Piccinin, S. & Dal Maso, S. (2023). Investigating factors affecting reading comprehension in second-generation pupils: the role of morphological skills. *Lingue e Linguaggio*, 1, 112-142.



**Langues & Langage  
à la croisée des Disciplines**

**1ère Rencontre annuelle LLcD**

Colloque : 9-11 septembre, 2024

Ecole d'été : 12-14 septembre, 2024

Sorbonne Université - 28 Rue Serpente, 75006 Paris, France



## **Workshop 8**

**Nouvelles pistes pour la morphologie "Mot et Paradigme"**

**Future directions in Word and Paradigm morphology**

## **The usefulness of the developmental sciences for WP models of morphology**

The essential challenge addressed by Word and Pattern (WP) morphological approaches is to identify general principles concerning the nature of word internal structure and their connection to the systemic organization in patterns of word relatedness. Cross-linguistically the word (broadly construed) is often the smallest meaningful unit, often consisting of reused pieces (re-)configured to serve distinct purposes. This allows us to discover patterns of morphological organization in the systems of relations between words. This approach is paradigmatic, because it identifies patterns that whole words participate in, and configurative, because, while the meaning of a word form is not necessarily a straightforward composition of meaningful parts, the meaning of the whole is associated with (re-)arrangements of its constitutive elements.

Certain present day WP insights about morphological organization reflect earlier intuitions about morphology as a “system of systems” in the works of Kruszewski, Paul, Wurzel. They are usage-based in that they conceptualize morphology as arising dynamically from the interaction of individuals over time. Each wordform, connected to others as part of a system of relations, bore some information with respect to the patterns in which it participates. This led to a nuanced notion of the diagnostic value of ‘principal parts’: if morphological organization is a web of relations among wordforms, we would expect all forms to be predictive to some degree or other. The empirical basis for these insights was small: these earlier linguists were unaware of the extraordinary variety of types of surface expression in the organization of words, and they developed their insights when the resources for quantitative and computational exploration did not exist.

A substantial body of empirical cross-linguistic research in modern WP supports early systems intuitions as well as confirming the value of the Low Conditional Entropy Conjecture (LCEC) (Ackerman & Malouf 2013) in both synchrony and diachrony. The interaction of communicating local agents converges on two interdependent patterns of morphological organization as an emergent effect typical of complex adaptive systems. Systematic internal differences permit (classes of) words to be distinguished from each other which facilitates paradigmatic organization in a network of implicational relations revealed by the LCEC. While this cross-linguistic convergence of emergent organization is itself astonishing and requires explanation, perhaps more astonishing is that it does so despite an evidently illimitable number of surface expression strategies to achieve it.

In this talk, we argue that early systems intuitions reinforced and clarified by new research in WP can benefit from greater attention to the analysis of complex systems in the developmental sciences typified by evolutionary and developmental models of biology and psychology, process oriented models of psychology and economic models of bounded rationality. There have been limited moves in this direction employing, e.g., simulations (Ackerman & Malouf 2015, Round et al. 2022), evolutionary models (Newberry et al. 2017, Meinhardt et al 2022), and large language models (Muradoglu et al 2020, Misra & Kim 2023, Misra & Mahowald 2024). Extending these methods using insights from the developmental sciences offers the potential for a truly explanatory theory of morphology.



Ackerman, F. and R. Malouf. 2013. Morphological organization: The low conditional entropy conjecture. *Language*. 89:429-464.

Ackerman, F. and R. Malouf. 2015. The No Blur Principle effects as an emergent property of language systems. In *Proceedings of the 41st Annual Meeting of the Berkeley Linguistics Society*. Pages 1-14.

Meinhardt, E., F. Ackerman, and R. Malouf. 2022. Morphology gets more and more complex, unless it doesn't. In A. Sims, A. Ussishkin, J. Parker, and S. Wray (ed.), *Morphological Diversity and Linguistic Cognition*. Cambridge University Press

Misra, K., and N. Kim. 2023. Abstraction via exemplars? A representational case study on lexical category inference in BERT. In *BUCLD 48: Proceedings of the 48th annual Boston University Conference on Language Development*. Boston, USA.

Misra, K., and K. Mahowald. 2024. Language models learn rare phenomena from less rare phenomena: The case of the Missing AANNs. [arXiv:2403.19827](https://arxiv.org/abs/2403.19827)

Muradoglu, S., N. Evans, and E. Vylomova. 2020. Modelling verbal morphology in Nen. In *Proceedings of the 18th Annual Workshop of the Australasian Language Technology Association*, 3–53, Virtual Workshop. Australasian Language Technology Association.

Newberry, M., C. Ahern, R. Clark, J. Plotkin. 2017. Detecting evolutionary forces in language change. *Nature* 551: 223–226.

Round, E., S. Mann, S. Beniamine, E. Lindsay-Smith, L. Esher, M. Spike. 2022. Cognition and the stability of evolving complex morphology: an agent-based model, In *The Evolution of Language. Proceedings of the Joint Conference on Language Evolution (JCoLE)*, Kanazawa, Japan, 2022

## **Diachronizing WP morphology: quantifying the complexity effects of sound change vs analogy**

Borja Herce

*University of Zurich*

Clayton Marr

*Ohio State University*

While modern quantitative entropy-based methods have revolutionised the study of paradigm synchrony (see Ackerman & Malouf 2013, Beniamine 2018), they have had a smaller impact on the study of paradigm diachrony. Although simulations of morphological change (Round et al. forthcoming) and artificial language learning experiments (Smith et al. 2023) can give us insights into the nature of the different forces involved, they lack the ecological validity of real language data, so it is often difficult to interpret what the results of this literature mean for paradigmatic change in actual human languages. While in most languages deep history cannot be researched due to a lack of historical documentation, some languages do allow us to explore the diachrony of paradigms with quantitative methods like the ones which are routinely applied nowadays in the exploration of present-day inflectional systems.

Here we present a novel method which allows us to do this and, crucially, disentangle the contribution of different forces (namely regular sound change vs analogy) upon paradigmatic morphological complexity. This method first applies computerized forward reconstruction (Sims-Williams 2018, Marr & Mortensen 2023) to a large inflected lexicon of an ancient language to derive expected forms of a modern language. The forms expected under regular sound change can then be contrasted to actual forms. Here we explore change in the evolution of verbal morphology from Latin to French.

First, we annotate the most complete extant inflected lexicons of Latin (LatInfLexi, Pellegrini & Passarotti 2018) and French verbs (Vlexique, Bonami et al. 2013) for both lexeme-level and cell-level etymological relations. The intersection between the two lexicons resulted in 310 cognate lexemes and 38 cognate cells, and a total of 11,593 forms. We apply to the Latin forms the cascade DiaCLEF (Marr & Mortensen 2023), which sequences all regular sound changes that have taken place historically from Latin to French, and can hence generate expected French forms given Latin forms. The difference between the Latin system and the expected French system, thus, will correspond to the long-term effect of sound change upon paradigms, while comparison of expected French to actual French will correspond largely to the effect of morphological/analogical change. We perform calculations of morphological predictive structure and complexity of various kinds, mostly through the software Qumín (Beniamine 2018).

Our results are summarized in Table 1 below. Sound change and analogy seem to have the effect that is traditionally attributed to them (consider Sturtevant's 1947 Paradox) with respect to aspects of complexity like the number of inflection classes and the number of interpredictability areas within the paradigm. Their observed effects appear to be different,

however, with other aspects of morphological complexity. Most surprising is the fact that analogical change seemed not to reduce conditional entropy between cells, which has been the main measure to capture Paradigm-Cell-Filling-Problem complexity so far. In our dataset, analogy is observed to make morphology consistently more similar, thus creating or spreading not only total, but also partial allomorphic overlaps between lexemes, which increases the confusability of inflectional classes as measured by conditional entropies.

Complexity metric	Latin	Expected French	Actual French	Effect of sound change	Effect of analogy
Inflection classes	107	239	65	+123%	-73%
Interpredictability areas	14	28	14	+100%	-50%
Forms per verb	37.86	18.75	14.5	-50%	-23%
Conditional entropy	0.218	0.235	0.276	+8%	+15%

Table 1: **Complexifying** vs. **simplifying** effect of historical sound change and analogical change on various morphological aspects of inflectional systems.

## References

Ackerman, Farrell & Robert Malouf. 2013. Morphological organization: The low conditional entropy conjecture. *Language*, 89, 3: 429–464.

Beniamine, Sacha. 2018. Classifications flexionnelles. Étude quantitative des structures de paradigmes. Ph.D. thesis, Université Sorbonne Paris Cité – Université Paris Diderot (Paris 7). <https://qumin.readthedocs.io/index.html#>

Bonami, Olivier, Gauthier Caron, and Clément Plancq. 2013. Flexique: an inflectional lexicon for spoken French. *Technical documentation* [<http://www.llf.cnrs.fr/flexique/documentation.pdf>].

Marr, Clayton & Mortensen, David. 2023. Large-scale computerized forward reconstruction yields new perspectives in French diachronic phonology. *Diachronica*, 40(2), 238-285.

Pellegrini, Matteo & Marco Passarotti. 2018. LatInfLexi: an inflected lexicon of Latin verbs. *Computational Linguistics CLiC-it 2018*: 324-329.

Round, Erich, Louise Esher & Sacha Beniamine. Forthcoming. The natural stability of autonomous morphology: How an attraction–repulsion dynamic emerges from paradigm cell filling.

Smith, Kenny, Clem Ashton & Helen Sims-Williams. 2013. The relationship between frequency and irregularity in the evolution of linguistic structure: An experimental study. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 45, no. 45.

Sturtevant, Edgar H. 1947. *An Introduction to Linguistic Science*. New Haven: Yale University Press.

## The evolution of principal parts and paradigmatic leveling in Germanic

**Keywords:** Paradigm leveling, Germanic, phylogenetic comparative methods, principal parts

**Background.** Germanic strong verbs form their past tense via stem vowel ablaut patterns, following patterns of vowel alternation based on their principal parts. For example, Modern High German *trinken* ‘drink’ has the following principal parts: *trinken* (Infinitive), *trank* (3SG Past), *tranken* (1PL Past) and *ge-trunken* (Past Participle), displaying the pattern ABBC. Some studies link biases toward specific patterns to tense/aspect/mood semantics (Dammel et al. 2010), arguing for instance that the preference for the same vowel in past and participle forms in Dutch is due to the semantic overlap between the past tense form and the perfect periphrastic construction (De Smet and Van de Velde 2019).

This paper provides a systematic investigation of the relationship between TAM semantics and diachronic preferences toward ablaut patterns in the Germanic subgroup. Using phylogenetic comparative methods, we test the hypothesis that if the periphrastic perfect construction can be used to express the narrative past in a language, then verbs in that language will exhibit a diachronic tendency toward patterns that employ identical verb stem vowels for past and past participle forms.

**Data.** We take paradigm data for Germanic languages from UniMorph and Wiktionary, coding paradigms for individual verbs according to the vowel alternation patterns they display. From the paradigm data, we extracted four forms for each verb: the infinitive, 3SG past, 1PL past, and past participle (supine for North Germanic). Each verb was then coded according to a simplified tripartite scheme capturing their alternation patterns. We collected data for 107 strong verbs descending from Proto-Germanic for 14 modern and historical Germanic languages. We additionally code a binary character characterizing relevant TAM properties for all 14 languages in our sample, asking whether the periphrastic perfect construction can be used to express the narrative past.

**Model.** The phylogenetic analysis we carry out proceeds as follows: first, we infer transition rates between the states of the binary TAM character under a continuous-time Markov process using inferred phylogenies of the Indo-European language family, from which we extract Germanic languages (Heggarty et al. 2023, Chang et al. 2015); subsequently, we map the history of this character onto the phylogeny in a way that approximates the locations on the tree for which the periphrastic perfect construction expresses the narrative past, yielding two “regimes.”

Subsequently, we infer CTM transition rates between the 5 possible tripartite pattern states (ABC, ABB, ABA, AAB, AAA) for each verb in our data set, using phylogenies mapped with different TAM configurations. We allow the parameters of the CTM process characterizing transitions between pattern states to vary across regimes. We compare posterior transition rates across regimes to determine whether there is a higher long-term preference (as operationalized by the stationary distributions of each CTM process) for patterns in which past and past participle forms contain the same vowel (e.g., ABB, AAA), predicting that these preferences will be greater in regimes where the periphrastic perfect construction can be used to express the narrative past.

(488 words)

## References

- Dammel, A., Nowak, J., & Schmuck, M. (2010). Strong-Verb Paradigm Leveling in Four Germanic Languages: A Category Frequency Approach. *Journal of Germanic Linguistics*, 22(4), 337–359. doi: 10.1017/S1470542710000097
- De Smet, I., & Van de Velde, F. (2019). Reassessing the evolution of West Germanic preterite inflection. *Diachronica*, 36(2), 139–180. doi: 10.1075/dia.18020.des
- Heggarty, P., Anderson, C., Scarborough, M., King, B., Bouckaert, R., Jocz, L., Kümmel, M. J., Jügel, T., Irlinger, B., Pooth, R., Liljegren, H., Strand, R. F., Haig, G., Macák, M., Kim, R. I., Anonby, E., Pronk, T., Belyaev, O., Dewey-Findell, T. K., Boutilier, M., ... Gray, R. D. (2023). Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages. *Science* (New York, N.Y.), 381(6656), eabg0818. doi: 10.1126/science.abg0818
- Chang, W., Cathcart, C., Hall, D., & Garrett, A. (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language*, 91(1), 194-244. doi: 10.1353/lan.2015.0005

### Paradigmatic complexity as diachronic layering

The verbal paradigm of Seri (isolate, Mexico; [sei] iso-639-3) stands out as one of the most complex yet observed, in terms of the variety of forms and their unpredictability. E.g. its conditional entropy far exceeds that of any language measured by Stump & Finkel 2013 (Baerman 2016). The polyfunctionality of the morphological markers confounds a morpheme-based description, on account of their syncretic and disjunctive distribution in various cells of the paradigm. Consider the suffix *-ox*, a typical example (Moser & Marlett 2010):

(1) SINGULAR SUBJECT		PLURAL SUBJECT		
BASIC	PLURACTIONAL	BASIC	PLURACTIONAL	
iti:ket	iti:kt-o	iti:kt- <u>ox</u>	iti:kt-ołka	'be pregnant'
topχ <sup>w</sup> t	topχ <sup>w</sup> ot-i	topχ <sup>w</sup> ł-o	topχ <sup>w</sup> ł- <u>ox</u>	'be loose'
ti:nła	ti:nała	ti:nł- <u>ox</u>	ti:nał- <u>ox</u>	'make a sound'
k <sup>w</sup> ita:fp	k <sup>w</sup> ita:fip- <u>ox</u>	k <sup>w</sup> itato:fip- <u>ox</u>	k <sup>w</sup> itato:fip-ołka	'unfasten'

At the same time, such configurations call into question the efficacy of a WP analysis: as we approach the point where most forms need to be listed, have we not abdicated our responsibility?

We suggest here that a diachronic account will offer the most insight, and further argue that the evolutionary stages of the paradigm are still reflected in different configurations attested across the lexicon. This perspective should facilitate a more articulated and informative description even within an account which relies on listing of forms.

The different stages are represented schematically in Table 1a,b, extracted from the synchronic lexicon. Consider first Table 1a. Verbs and nouns share plural markers. A single plural form (row i) has a distributive and, with verbs, pluractional interpretation (Pasquereau & Cabredo Hofherr 2022). The first plural form can receive additional plural marking (row ii), indicating a plural subject (or plural possessor with nouns), thus e.g. *intita* > *intitat* > *intitatox* 'go away'. In row iii, distributivity/pluractionality marking is extended to plural subjects, realized by adding another layer of pluralization to the second plural form (*intitatox* > *intitatołka* 'go away'). Various derivational processes such as causativization can then 'import' plural marking into the singular subject subparadigm (row iv), e.g. plural subject *takłox* 'they are cut' > singular subject distributive *ita:kłox* 's/he cuts (them)', thus recalibrating the successive pluralization of forms. Typically (for semantic reasons?) with plural subjects there is no distinction between pluractional and distributive, but under the right conditions one can be induced, again realized through yet another layer of pluralization (row v). Table 1b shows the effects of the addition of a set of productive pluractional markers, one which does not directly participate in the system of cumulative pluralization. The verbal paradigm thus reflects two distinct principles of organization: hierarchical cumulativity (Table 1a) and more conventional morphosyntactic realization (Table 1b). The juxtaposition of the two lends the system its striking complexity.

Table 1a. no dedicated PLUR marking

SG SBJ			PL SBJ		
(BASIC)	PLUR	DISTR	(BASIC)	PLUR	DISTR
<i>i</i>	a	b			
<i>ii</i>	a	b	c		
<i>iii</i>	a	b	c	d	
<i>iv</i>	c	d	e	f	
<i>v</i>	a	b	c	d	e

Table 1b. with dedicated PLUR marking

SG SBJ			PL SBJ		
(BASIC)	PLUR	DISTR	(BASIC)	PLUR	DISTR
a	X	b		Y	
a		b	c		
a		b	c		
c		d	e		

## References

- Baerman, M. 2016. Seri verb classes: Morphosyntactic motivation and morphological autonomy. *Language* 92(4). 792–823.
- Marlett, S., & M.B. Moser. 2010. *Comcaac quih yaza quih hant ihiip hac = Diccionario seri-español-inglés*. Universidad de Sonora.
- Pasquereau, J., & P. Cabredo Hofherr. 2022. Two types of pluractionality in Seri. In *Proceedings of SULA 11*.
- Stump, G.T., & R. Finkel. 2013. *Morphological typology: From word to paradigm*. CUP.



## Pattern naturalness and stability in Romance stem alternations

Some Autonomous Morphological literature, and researchers working on (Romance) morphemes have stated that documented language change does not seem to show a bias against unnatural structures (Maiden 2016:49). Some even go as far as questioning the relevance of features and values in paradigm evolution (Blevins forthcoming). Most of the most successful patterns of stem alternation in Romance do indeed involve paradigm cells which are not coextensive with any particular morphosyntactic value (see Maiden 2018). However, there are various factors to consider: i) because there are many more ways of being disorderly than there are of being orderly, unnatural patterns are much more numerous and hence more probable by chance than natural patterns; and ii) naturalness, i.e. proximity to a natural class, is inherently a continuous and not a dichotomic dimension (see Saldana et al. 2022). Quantitative work that attempts to test for diachronic biases should consider both of these properties of (un)naturalness.

Here we test if there is a correlation, in Romance, between the naturalness of a pattern of stem alternation and its success. We coded by hand the paradigmatic and lexical distribution of all stem alternation patterns in all verbs in ODRVM (Maiden et al. 2010, Beniamine et al. 2020). The result is 2151 stem alternation patterns in total, 212 paradigmatically unique ones, distributed across 57 Romance languages (we selected the best documented varieties only, those with at least 15 complete documented paradigms), see Herce 2022. We also calculated a measure of the naturalness of each of these patterns (we use the number of values that need to be mentioned in its shortest description, and also the average percentage of shared morphosyntactic values between all cell pairs as possible operationalizations) and obtain a measure of the usage frequency of a pattern (we calculate the proportion of use that a pattern's cells represent over the complete paradigm, in the Latin corpus of Delatte et al. 1981 as a proxy for Romance frequencies).

Looking at the presence and absence of all patterns among our 57 Romance languages, and relying on the Romance family tree and methodology outlined in Cathcart et al. (2022), we build a phylogenetic model to explore the impact of pattern naturalness and pattern frequency on a pattern's long-term survival and the speed at which it emerges or is lost.

## References

- Beniamine, Sacha, Martin Maiden, and Erich Round. 2020. Opening the romance verbal inflection dataset 2.0: A CLDF lexicon. In *12th Conference on Language Resources and Evaluation*: 3027-3035. European Language Resources Association (ELRA).
- Blevins, James P. forthcoming. Two frameworks of morphological analysis. *Linguistic Analysis* [https://www.academia.edu/42103967/Two\\_frameworks\\_of\\_morphological\\_analysis](https://www.academia.edu/42103967/Two_frameworks_of_morphological_analysis)

Cathcart, Chundra, Borja Herce, and Balthasar Bickel. 2022. Decoupling speed of change and long-term preference in language evolution: insights from Romance verb stem alternations.

Delatte, Louis, Étienne Evrard, Suzanne Govaerts & Joseph Denooz. 1981. *Dictionnaire fréquentiel et index inverse de la langue latine*. L.A.S.L.A, Liege.

Herce, Borja. 2022. Quantifying the importance of morphomic structure, semantic values, and frequency of use in Romance stem alternations. *Linguistics Vanguard* 8, 1: 53-68.

Maiden, Martin. 2016. Some lessons from history: morphemes in diachrony'. In Luís and Bermúdez-Otero (eds.), *The Morphome Debate*: 33–63.

Maiden, Martin. 2018. *The Romance verb: Morphomic structure and diachrony*. Oxford: Oxford University Press.

Maiden, Martin, John Charles Smith, Silvio Cruschina, Marc-Olivier Hinzelin and Maria Goldbach. 2010. Oxford online database of Romance verb morphology. Available at: <http://romverbmorph.clp.ox.ac.uk/>.

Saldana, Carmen, Borja Herce, and Balthasar Bickel. 2022. More or less unnatural: Semantic similarity shapes the learnability and cross-linguistic distribution of unnatural syncretism in morphological paradigms. *Open Mind* 6: 183-210.

## An integrated approach to inflectional and derivational predictability. The case of Latin deverbal agent and action nouns

Matteo Pellegrini, Federica Iurescia, Eleonora Litta, Marco Passarotti  
CIRCSE Research Centre, Università Cattolica del Sacro Cuore, Milan

The issue of predictability of unknown forms from known ones has received a lot of attention in Word-and-Paradigm morphology. Most studies have dealt with inflectional predictions, in what is known as the Paradigm Cell Filling Problem (Ackerman et al. 2009). Different methodologies have been devised for a quantitative investigation of this issue, ranging from set-theoretic ones – like Principal Part Analysis (Stump & Finkel 2013) – to information-theoretic ones – using conditional entropy (Bonami & Boyé 2014).

Another commonplace topic of Word-and-Paradigm morphology is the extension of the notion of paradigm to derivation (see e.g. van Marle 1984, Štekauer 2014, Bauer 2019, Hathout & Namer 2019). In this spirit, the entropy-based approach to predictability has been applied to derivational predictions: Bonami & Strnadová (2019) show that patterns of interpredictability between derivationally related forms are comparable to those between inflectionally related forms. However, they use only the citation form of lexemes in their sample. While this is unlikely to be problematic for their study, in other cases this might not be so. Indeed, previous literature has shown that derivatives can be based on a stem different from the one of the citation form of their base (Bonami et al. 2009). Consequently, predictability can be expected to vary according to the form(s) considered. One example of this situation (cf. Aronoff 1994) is offered by Latin deverbal action nouns in *-io* and agent nouns in *-or*, that are (usually) based on the third stem, displayed e.g. in the supine, rather than on the present stem, displayed in the citation form, PRS.ACT.IND.1SG: e.g. PRS.ACT.IND.1SG *profundo* ‘pour’, SUP.ACC *profusum* → NOM.SG *profusio* (action), *profusor* (agent). While suffixation of *-or* to the third stem is the only derivational process available for the formation of agent nouns, for action nouns there are other options – for instance, suffixation of *-ium* – that require other stems – namely, the present stem (e.g., *impropero* ‘hasten’ → *improperium*). Therefore, this appears to be a promising case-study to investigate variation in predictability of derivatives from different inflected forms.

In this work, we build a paradigmatic system including both inflectionally and derivationally related forms, for a total of about 2,500 morphological families (Table 1). We take inflected forms of verbs displaying different stems from PrinParLat (Pellegrini 2023), and the corresponding agent and action nouns from WFL (Litta & Passarotti 2019). We then use Qumin (Beniamine 2018) to compute conditional entropy for all possible combinations of forms in our dataset.

Preliminary results (Table 2) confirm our expectation that action and agent nouns are more predictable from forms based on the third stem than on those based on the present stem, and that the difference is larger for agent nouns than for action nouns. They also reveal that forms based on the present stem are better predictors than those based on the perfect stem for action nouns, while for agent nouns results are closer. Such findings suggest that investigating derivational processes focusing only on citation forms might hinder relevant generalizations that can be drawn when other forms are considered.

**Table 1 – Some partial morphological families from our dataset**

V_prs.act. inf	V_fut.act. ind.3.sg	V_prf.act. ind.1.sg	V_prf.pass. ptcp.nom.n.sg	ActionN_ nom.sg	ActionN_ gen.sg	AgentN_ nom.sg	AgentN_ gen.sg
<i>abdicare</i> ‘indicate’	<i>abdicabit</i>	<i>abdicauī</i>	<i>abdicatum</i>	<i>abdicatio</i>	<i>abdicationis</i>	<i>abdicator</i>	<i>abdicatoris</i>
<i>profundere</i> ‘pour’	<i>profundet</i>	<i>profudi</i>	<i>profusum</i>	<i>profusio</i>	<i>profusionis</i>	<i>profusor</i>	<i>profusoris</i>
<i>improperare</i> ‘hasten into’	<i>improperabit</i>	<i>improperauī</i>	<i>improperatum</i>	<i>improperium</i>	<i>improperi</i>	<i>improperator</i>	<i>improperatoris</i>
<i>cupere</i> ‘desire’	<i>cupiet</i>	<i>cupiui/cupii</i>	<i>cupitum</i>	<i>cupido</i>	<i>cupidinis</i>	<i>cupitor</i>	<i>cupitoris</i>

**Table 2 – Results: matrix of conditional entropy of predicting form in column given form in row**

	V_prs. act.inf	V_fut.act. ind.3.sg	V_prf.act. ind.1.sg	V_prf.pass. ptcp.nom.n. sg	ActionN_ nom.sg	ActionN_ gen.sg	AgentN_ nom.sg	AgentN_ gen.sg
V_prs.act.inf		0.38	0.36	0.21	0.28	0.29	0.26	0.26
V_fut.act.ind.3.sg	0.1		0.3	0.19	0.3	0.3	0.19	0.19
V_prf.act.ind.1.sg	0.29	0.3		0.32	0.36	0.37	0.27	0.26
V_prf.pass.ptcp.nom.n.sg	0.26	0.31	0.41		0.23	0.23	0.05	0.05
ActionN_nom.sg	0.22	0.26	0.39	0.04		0	0.07	0.07
ActionN_gen.sg	0.22	0.26	0.39	0.04	0		0.07	0.07
AgentN_nom.sg	0.24	0.27	0.39	0.05	0.19	0.19		0
AgentN_gen.sg	0.24	0.27	0.37	0.04	0.19	0.19	0	

## References

- Ackerman, F., Blevins, J.P. & Malouf, R. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In Blevins, J. & Blevins, J.P. *Analogy in grammar: Form and acquisition*. Oxford: OUP.
- Aronoff, M. 1994. *Morphology by itself. Stems and inflection classes*. Cambridge: MIT Press.
- Bauer, L. 2019. Notions of paradigm and their value in word-formation. *Word Structure* 12(2), 153-175.
- Beniamine, S. 2018. *Classifications flexionnelles. Étude quantitative des structures de paradigmes*. PhD Thesis. Université Sorbonne Paris Cité-Université Paris Diderot (Paris 7).
- Bonami, O. & Boyé, G. 2014. De formes en thèmes. In Villoing, F., Leroy, S. & David, S., *Foisonnements morphologiques. Études en hommage à Françoise Kerleroux*. Paris: Presses Universitaires de Paris-Ouest, 17-45.
- Bonami, O., Boyé, G. & Kerleroux, F. 2009. L'allomorphie radicale et la relation flexion-construction. In Fradin, B., Kerleroux, F. & Plénat, M., *Aperçus de morphologie du français*. Saint-Denis: Presses Universitaires de Vincennes, 103-125.
- Bonami, O. & Strnadová, J. 2019. Paradigm structure and predictability in derivational morphology. *Morphology* 29(2), 167-197.
- Hathout, N. & Namer, F. 2019. Paradigms in word formation: what are we up to? *Morphology* 29(2), 153-165.
- Litta E. & Passarotti M. 2019. (When) inflection needs derivation: a word formation lexicon for Latin. In: Holmes, N., Ottink, M., Schrickx J. & Selig, M. (eds.) *Lemmata Linguistica Latina. Volume I: Words and Sounds*. Berlin / Boston: De Gruyter, 224-239.
- Pellegrini, M. 2023. Flexemes in theory and in practice: Modelling overabundance in Latin verb paradigms, *Morphology*, 33(3), 361-395.
- Štekauer, P. 2014. Derivational paradigms. In Lieber, R. & Štekauer, P. (eds.), *The Oxford Handbook of Derivational Morphology*. Oxford: OUP, 354-369.
- Stump, G. & Finkel R.A. 2013. *Morphological typology: From word to paradigm*. Cambridge: CUP.
- van Marle, J. 1984. *On the Paradigmatic Dimension of Morphological Creativity*. Dordrecht: Foris.

## It takes a family

### How morphological families shape the interpretation of derived words

Theoretical approaches to word formation come into two broad varieties: in rooted-tree approaches, which arguably constitute the mainstream, every derived word is characterized by its relation to a single *base*; any property shared by two members of a derivational family results from their having the same ultimate base, whether this is considered to be a morpheme, a stem, or a word. In paradigmatic approaches (van Marle, 1984; Bochner, 1993; Hathout & Namer, 2022), words participate in a rich network of morphological relations, not limited to those between a derivative and its formal base.

A main argument for pursuing a paradigm-based approach is the existence of situations where a member of a derivational family other than its formal base is the clear motivator for the meaning of a derived word. For example, Bonami & Guzman Naranjo (2023) provide distributional evidence that on average, nouns in *-ist* (e.g. *socialist*) relate semantically more readily to corresponding nouns in *-ism* (*socialism*) than to their formal bases (*social*). However, most other pairs of derivatives in their study (e.g., *-eur/-ion*) failed to reveal a comparable effect, *prima facie* questioning the prevalence of transversal paradigmatic relations.

Our presentation starts from the observation that the paradigmatic approach is distinctive in allowing the features of a word to be predicted *jointly* by *multiple members* of its derivational family, insofar as no single member need bear a privileged semantic relation to any other member. Joint predictability has already been demonstrated for word forms (Bonami & Strnadová, 2019). It is also evident in the case of polysemous multimotivation, when a derivative has multiple meanings relating to different members of its family: Strnadová (2014) discusses examples such as *senatorial*, which can mean both ‘relative to the Senate’ and ‘relative to senators’. We hypothesize that, more generally, a better approximation of word meaning can be predicted through leveraging the full power of multimotivation.

As a concrete case study, we will focus on French verbs in *-iser* (*staliniser*) and *-ifier* (*russifier*) having both a noun and an adjective in their derivational family, relying on a dataset of 617 triples from Bonami & Thuilier (2019), using methods from distributional semantics (Boleda, 2020). Manual examination of the data suggests that there are relevant examples: for instance, *platoniser* and *cartésianiser* have parallel meanings (‘behave as Plato/Descartes would’) despite being respectively formally based on a noun or related adjective.

Using a general methodology first put forth by Marelli & Baroni (2015), we will train parallel statistical models to predict the meaning of a verb from that of the corresponding noun, the adjective, whichever of the two is formally closest to the verb, and both. We hypothesize that joint prediction will outperform all alternatives. We will close with suggestions as to how the simultaneous availability of multiple family members as joint predictors of a word’s meaning can help address other long-standing problems in the semantics of word formation.

## References

- Bochner, Harry. 1993. *Simplicity in generative morphology*. Berlin: Mouton de Gruyter.
- Boleda, Gemma. 2020. Distributional semantics and linguistic theory. *Annual Review of Linguistics* 6(1). 213–234. doi:10.1146/annurev-linguistics-011619-030303.
- Bonami, Olivier & Matías Guzman Naranjo. 2023. Distributional evidence for derivational paradigms. In Sven Kotowski & Ingo Plag (eds.), *The semantics of derivational morphology: theory, methods, evidence*, 219–258. Berlin: De Gruyter.
- Bonami, Olivier & Jana Strnadová. 2019. Paradigm structure and predictability in derivational morphology. *Morphology* 29(2). 167–197. doi:10.1007/s11525-018-9322-6.
- Bonami, Olivier & Juliette Thuilier. 2019. A statistical approach to affix rivalry: French *-iser* and *-ifier*. *Word Structure* 12(1). 4–41.
- Hathout, Nabil & Fiammetta Namer. 2022. Paradis: a family and paradigm model. *Morphology* 32(2). 153–195. doi:10.1007/s11525-021-09390-w.
- Marelli, Marco & Marco Baroni. 2015. Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review* 122. 485–515.
- van Marle, Jaap. 1984. *On the paradigmatic dimension of morphological creativity*. Dordrecht: Foris.
- Strnadová, Jana. 2014. Multiple derivation in French denominal adjectives. In Sandra Augendre, Graziella Couasnon-Torlois, Déborah Lebon, Clément Michard, Gilles Boyé & Fabio Montermini (eds.), *Proceedings of the 8th décembrettes* (Carnets de grammaire 22), 327–346. Toulouse: CLLE-ERSS.

## **Towards an abstractive theory of morphophonological alternations: Empty morphs, neutralisation and paradigm predictability in Wubuy**

Peter Nyhuis  
University of Melbourne

Recent Word and Paradigm analyses of the implicational structure of inflectional paradigms have given attention to the ways that static phonological properties of stems and affixes may be diagnostic of a form's inflection class membership (Wilmoth & Mansfield 2021), thereby reducing the uncertainty associated with predicting an unknown inflected form from a known one (Ackerman et al. 2009). However, the place of dynamic phonology in abstractive models (Blevins 2006) – e.g., systematic alternations involving, *inter alia*, assimilation, deletion, epenthesis, or their interactions – is less well understood (though cf. Lévêque & Pellard 2023). As a result, a significant dimension of patterned variation among word forms remains largely unaccounted for in WP morphology. Meanwhile, formal phonology, which does address such patterns, has remained overwhelmingly committed to a 'constructive' conception of words' structure, whereby the phonological component of the grammar acts upon strings of invariant underlying morphemes and adjusts them according to well-formedness criteria.

To that end, the present paper demonstrates the affordances of Relational Morphology (RM: Jackendoff & Audring 2020) for bridging the divide, showing that the basic machinery of the formalism offers a fully declarative account of phonological alternations, without morphemes and without underlying representations. In RM, general patterns, including phonological ones, are expressed via schemas, deduced through analogical comparison of what is the same and what is different between full surface word forms, stored in the lexicon, and linked to one another on the basis of recurrent parts of structure. Those linkages, in turn, organise words (and schemas) into paradigms.

I apply the formalism to an otherwise problematic case in the phonology of Wubuy (Gunwinyguan, northern Australia: Heath 1984) involving the phonologically-conditioned empty morph *ɲu-*, which is inserted preceding a stop-initial stem when it follows a nasal-final inflectional prefix, or a derivational prefix of any kind. It is thus sensitive to a two-sided context, and its interactions at each side – with continuant fortition, nasal place assimilation, and nasal deletion – are opaque. While some dimensions of this problem, taken in isolation, can be handled with constructive approaches, a satisfying account of the whole ensemble is out of reach. By virtue of its abstractive assumptions, however, RM can address all of these interacting processes simultaneously, using the same analogical procedures required for more familiar purposes (such as lexical inflection classes). The crux of the analysis – as depicted in the schemas in (2)–(3), which generalise over the (partial) paradigm in (1) – is that prefix and stem allomorphs subcategorise for (i.e. predict) each other on the basis of their phonological realisation *across different contexts*. These distinct patterns of phonological realisation organise words into what I call 'phonological classes', which are orthogonal to the lexical inflectional classes that govern suffixal tense-aspect inflection in Wubuy. One of the most significant results here is that the model, despite having only one level of phonological representation, provides a compelling formal mechanism for analysing opaque rule interactions – an ability normally assumed to be the sole preserve of models which have more than two levels (Baković 2007). In addition to extending the empirical reach of RM in particular and WP approaches more broadly, the analysis presented here hints at some of the gains that formal phonology may accrue by reconsidering the morphemic assumptions it is to a large extent built on.

(1)	Alternating class:				Stop class:			
	/ju.ɫ-i:/ 'carry-NPST2'		/wult̪-i:/ 'cut-NPST2'		/cu.ɫ-i:/ 'push-NPST2'		/karac-i:/ 'jump-NPST2'	
1sg.R	ŋa-	ju.ɫ-i:	ŋa-	wult̪-i:	ŋa-	cu.ɫ-i:	ŋa-	karac-i:
12du.R	na-	ju.ɫ-i:	na-	wult̪-i:	na-	cu.ɫ-i:	na-	karac-i:
3pl.R	wuru-	ju.ɫ-i:	wuru-	wult̪-i:	wuru-	cu.ɫ-i:	wuru-	karac-i:
1sg.IRR	ŋaŋ-	cu.ɫ-i:	ŋaŋ-	kult̪-i:	ŋaŋu-	cu.ɫ-i:	ŋaŋu-	karac-i:
12du.IRR	naŋ-	cu.ɫ-i:	naŋ-	kult̪-i:	naŋu-	cu.ɫ-i:	naŋu-	karac-i:
2sg.R	nuŋ-	cu.ɫ-i:	nuŋ-	kult̪-i:	nuŋu-	cu.ɫ-i:	nuŋu-	karac-i:

(2) Alternating class: inflectional prefix-stem schemas

- a. Semantics  $[F([PRED(\dots\gamma)]_\alpha)_\delta]_x$   
Morphosyntax  $\{X_\alpha, MOOD_\delta, Pro_\gamma\}_x$   
Phonology  $[...V_{\delta,\gamma} [ \begin{matrix} *+SON* \\ *+CONT* \\ PLACE_\varepsilon \end{matrix} ] \dots]_\alpha]_x$
- b. Semantics  $[F([PRED(\dots\zeta)]_\alpha)_\eta]_x$   
Morphosyntax  $\{X_\alpha, MOOD_\eta, Pro_\zeta\}_x$   
Phonology  $[...* \begin{matrix} +NASAL \\ PLACE_\varepsilon \end{matrix} ]_{\eta,\zeta} [ \begin{matrix} *-SON* \\ *-CONT* \\ PLACE_\varepsilon \end{matrix} ] \dots]_\alpha]_x$

(3) Stop class: inflectional prefix-stem schemas

- a. Semantics  $[F([PRED(\dots\gamma)]_\beta)_\delta]_x$   
Morphosyntax  $\{X_\beta, MOOD_\delta, Pro_\gamma\}_x$   
Phonology  $[...V_{\delta,\gamma} [ [-SON] \dots]_\beta]_x$
- b. Semantics  $[F([PRED(\dots\zeta)]_\beta)_\eta]_x$   
Morphosyntax  $\{X_\beta, MOOD_\eta, Pro_\zeta\}_x$   
Phonology  $[...*\eta u_{\eta,\zeta} [ [-SON] \dots]_\beta]_x$

## References

- Ackerman, Farrell, James P. Blevins & Robert Malouf (2009). 'Parts and Wholes: Implicative Patterns in Inflectional Paradigms'. *Analogy in Grammar: Form and Acquisition*. Ed. by James P. Blevins & Juliette Blevins. Oxford: Oxford University Press.
- Baković, Eric (2007). 'A Revised Typology of Opaque Generalisations'. *Phonology* 24.2, pp. 217–259.
- Blevins, James P. (2006). 'Word-Based Morphology'. *Journal of Linguistics* 42.3, pp. 531–573.
- Heath, Jeffrey (1984). *Functional Grammar of Nungubuyu*. Canberra: Australian Institute of Aboriginal Studies.
- Jackendoff, Ray & Jenny Audring (2020). *The Texture of the Lexicon: Relational Morphology and the Parallel Architecture*. Oxford, New York: Oxford University Press.
- Lévêque, Dimitri & Thomas Pellard (2023). 'The Implicative Structure of Asama Verb Paradigms: A Quantitative Study of Segmental and Suprasegmental Alternations'. *Morphology* 33, pp. 261–286.
- Wilmoth, Sasha & John Mansfield (2021). 'Inflectional Predictability and Prosodic Morphology in Pitjantjatjara and Yankunytjatjara'. *Morphology*, pp. 355–381.



## **Assessing the import of morphology in gender assignment: A computational study of French**

Contemporary Word and Paradigm approaches to morphology (Blevins 2016) have paid comparatively little attention to the way subsequences of a word signal morphosyntactic properties (see Beniamine & Carroll 2023 for a notable exception). In this talk, we will address this issue through the lens of gender assignment.

The French grammatical gender system is a prime example of a case where phonological, morphological and semantic properties of nouns jointly partially predict gender assignment (Tucker et al. 1977, Corbett 1991). It remains unclear however what the ultimate causes of gender predictability are, given the strong correlations between predictors. In an unpublished study, Bonami et al. (2019) use computational classifiers to start addressing the issue. While that study was promising, it was inherently limited by requiring hard coding of the phonological properties of interest (the last three segments in a word plus word length), leading to poor interpretability of the results.

In this study we revisit the issue by training an LSTM (Hochreiter & Schmidhuber 1997), a type of recurrent neural network designed to capture dependencies over long sequences, to predict the gender of a noun from its form. The trained network examines words character by character starting from the end, and assigns at each point a probability for each gender given what has been seen so far. This allows us to probe which subsequences of segments are informative of gender, without imposing a morphological segmentation or a set window size.

We relied on two interconnected resources: *Flexique* (Bonami et al. 2014), a normalized phonological inflectional lexicon based on the wordset of *Lexique* (New et al. 2007); and *Echantinom* (Bonami & Tribout 2021), a morphologically annotated database of French nouns randomly sampled from *Lexique*. We excluded nouns with homographs in the other gender (both common gender nouns such as *artiste* ‘male or female artist’ and true homonyms such as *livre<sub>F</sub>* ‘pound’ vs. *livre<sub>M</sub>* ‘book’), and used the remaining 4534 nouns from *Echantinom* for test purposes and the 24880 other nouns from *Flexique* for training. We used PyTorch to train separate LSTMs on orthographic forms and phonemic forms.

The orthographic model reached an accuracy of 0.91, while the accuracy of the phonemic model was 0.80. This is basically in line with previous results by Bonami et al. (2019), and confirms that gender is highly predictable in French based on form only.

We then examined the extent to which models trained on form information capture morphological properties of words as documented in *Echantinom*. The models are more accurate on suffixed (0.98 with graphemes vs. 0.93 with phonemes) than unsuffixed (0.86 vs. 0.72) nouns, and also more confident in their prediction (average probability of predicted gender 0.98 vs. 0.94 for suffixed nouns, 0.92 vs. 0.83 for unsuffixed nouns). In the talk, we will examine character-by-character predictions of the models and show how the predictive power of derivational suffixes is exploited by the model in the absence of morphological segmentation.

## References

- Blevins, J. 2016. *Word and Paradigm Morphology*. Oxford: Oxford University Press.
- Bonami, O., G. Caron, and C. Plancq. 2014. Construction d'un lexique flexionnel phonétisé libre du français. *Actes du quatrième Congrès Mondial de Linguistique Française*, 2583–2596.
- Bonami, O., M. Guzman Naranjo, and D. Tribout. 2019. The role of morphology in gender assignment in French. Presented at the *Second International Symposium on Morphology*, Paris, September 2019.
- Bonami, O., and D. Tribout. 2021. Echantinom: a hand-annotated morphological lexicon of French nouns. *Proceedings of the Third International Workshop on Resources and Tools for Derivational Morphology*, 42–51.
- Corbett, G.. 1991. *Gender*. Cambridge: Cambridge University Press.
- Hochreiter, S., and J. Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 1735–1780.
- New, B., M. Brysbaert, J. Veronis, and C. Pallier. 2007. The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics* 28:661–677.
- Tucker, G. R., W. E. Lambert, and A. Rigault. 1977. *The French speaker's skill with grammatical gender: An example of rule-governed behavior*. Berlin : Mouton De Gruyter.

# Addressing the Problem of Syncretism in Russian Nominal Paradigms for the Task of Unsupervised Learning of Morphology

## 1 Introduction

Identifying a paradigm poses a challenge for unsupervised learning of morphology (Wiemerslage et al., 2022). When vector representations need to be organized as a paradigm of unknown size, multiple problems are to be handled, such as paradigm gaps, overabundance and syncretism. In this paper we advance a proposal to handle syncretism, using Russian data to illustrate the problem.

## 2 Syncretism and Paradigm

Syncretism is extremely prominent in the Russian nominal system: all lemmata have syncretic forms in their paradigm, and syncretic patterns vary depending on the gender, morphophonology, and animacy of the noun. Combined with a high number of cases (six in traditional grammar, 15 according to Zaliznyak 2002), this allows for a large number of possible sets of features for a given form. In this study we used seven cases: in addition to the six traditional ones (nominative, genitive, accusative, dative, ablative, prepositional) we included the partitive case. Multiplied by two number feature values, this means that 16383 ( $2^{14} - 1$ ) case-number feature combinations would be possible if syncretism could arise randomly.

## 3 Data visualisation

Using a finite-state morphological tool for Russian (*pymorphy2*, Korobov 2015), we have generated and tagged forms for 11515 nouns. To capture syncretism, we have annotated each unique form with sets of case-number features for which it can stand, which led to 40 different labels. For each unique form, we have obtained a 300-dimensional fastText vector (Bojanowski et al., 2017) and for each lexeme we calculated an average vector over all its forms. Then we have computed difference vectors for each form (lexeme vector - form vector) and visualised the resulting vectors using PCA (reduced to 50 dimensions) and t-SNE (van der Maaten & Hinton 2008, reduced to two dimensions) algorithms (cf. Shafaei-Bajestan et al. 2022 for English number and Nikolaev et al. 2022 for Finnish nominal inflection). As can be seen in Figure 1, the resulting space is divided into morphologically motivated regions, although the clusters cannot be visually separated. This is resolved by restricting the noun forms to high-frequency ones at the cost of losing eleven labels for lower-frequency syncretic patterns (Figure 2). Multiple clusters with the same labels are explained by different gender values (Figure 3).

## 4 Feature clustering

To evaluate the presented approach, we have run learning and clustering algorithms on the full range of data with the aim to classify unique forms by the set of forms they represent. The accuracy of a supervised learning algorithm (linear classifier, 80%/20% test/train split) is 86.75%, the best result of a k-Means clustering (evaluated using a Minimum Cost Maximum Flow algorithm, since clusters need to be mapped to the corresponding labels) is 81.2%. This is remarkable since Edmiston 2020 report a 0.15 F1 score for a k-Means clustering versus 0.87 F1 score for a linear classifier for Russian case (based on RuBERT data, Kuratov & Arkhipov 2019).

## 5 Discussion

In this paper, we propose a novel approach to syncretism for the task of unsupervised extraction of morphological features, illustrated using Russian data. In particular, we classify difference vectors and capture syncretism by introducing set labels. By the comparison of the unsupervised and supervised evaluation, we demonstrate that this approach lays on the right path towards unsupervised extraction of morphological features and paradigm structure despite the challenges of syncretism.

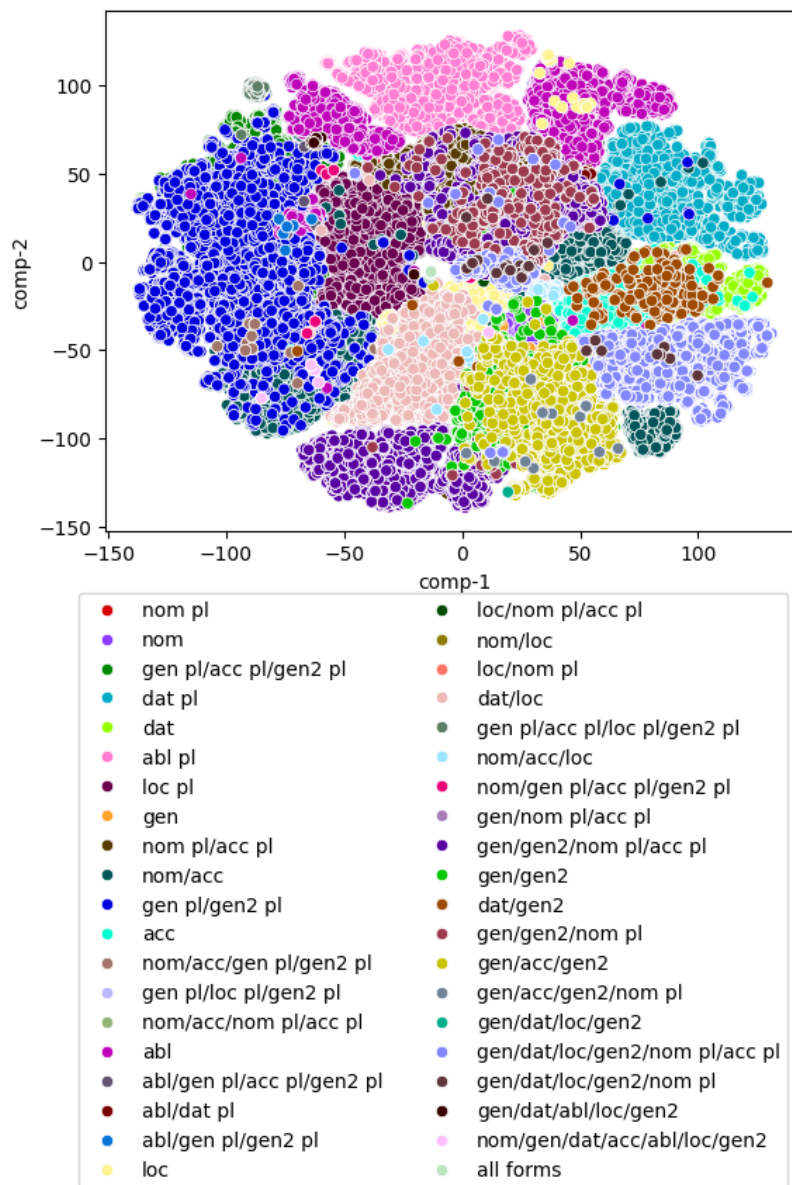


Figure 1: Visualization of difference vectors for the nominal paradigms in Russian including syncretic forms (PCA to 50 dimensions and t-SNE to two dimensions).

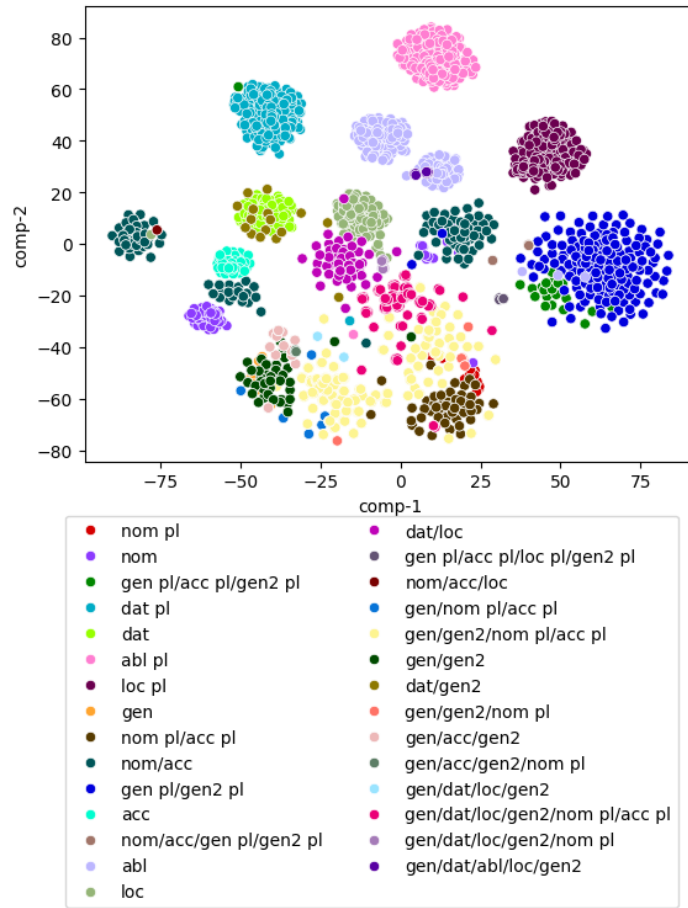


Figure 2: Visualization of difference vectors for the nominal paradigms of frequent (200 ipm or higher) Russian nouns including syncretic forms.

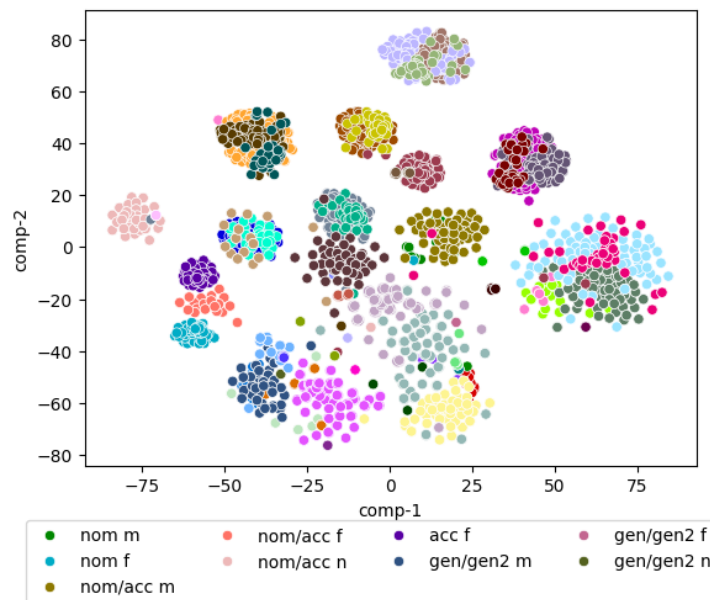


Figure 3: Visualization of difference vectors for the nominal paradigms of frequent (200 ipm or higher) Russian nouns including syncretic forms, separated by gender.

## References

- Bojanowski, P., E. Grave, A. Joulin & T. Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5. 135–146.
- Edmiston, D. 2020. A systematic analysis of morphological content in BERT models for multiple languages. *arXiv preprint arXiv:2004.03032* .
- Korobov, M. 2015. Morphological analyzer and generator for Russian and Ukrainian languages. In *Analysis of Images, Social Networks and Texts: 4th International Conference, AIST 2015, Yekaterinburg, Russia, April 9–11, 2015, Revised Selected Papers 4*, 320–332. Springer.
- Kuratov, Y. & M. Arkhipov. 2019. Adaptation of Deep Bidirectional Multilingual Transformers for Russian Language.
- van der Maaten, L. & G. Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9(11).
- Nikolaev, A., Y.-Y. Chuang & R. H. Baayen. 2022. A generating model for Finnish nominal inflection using distributional semantics. *Mental Lexicon* 17(3). 368 – 394. <https://search.ebscohost.com/login.aspx?direct=true&db=cms&AN=173324800&site=ehost-live>.
- Shafaei-Bajestan, E., M. Moradipour-Tari, P. Uhrig & R. H. Baayen. 2022. Semantic properties of English nominal pluralization: Insights from word embeddings. *arXiv* .
- Wiemerslage, A., M. Silfverberg, C. Yang, A. D. McCarthy, G. Nicolai, E. Colunga & K. Kann. 2022. Morphological Processing of Low-Resource Languages: Where We Are and What's Next. *arXiv* doi: 10.48550/arxiv.2203.08909.
- Zaliznyak, A. 2002. Russkoe imennoe slovoizmenenie s prilozheniem izbrannykh rabot po sovremennomu russkomu iazyku i obshchemu iazykoznaniiu [Russian nominal inflection with additional selected works on contemporary Russian language and general linguistics]. *Moscow: Iazyki slavianskoi kul'tury [Languages of slavic culture]* .

## Multiple periphrastic relations in the morphology of French

Gabrielle Aguila-Multner and Berthold Crysmann

French verbal inflection makes extensive use of periphrasis to express both temporal categories, like past, and non-temporal categories, like passive. In addition to the difficulties posed by such phenomena, such as non-compositionality and the integration of grammatical function change to inflectional morphology, the French system presents two further challenges: firstly, we often find periphrases that involve more than one ancillary element. This can arise both from periphrasis stacking, i.e. the regular combination of two (or more) periphrastic relations, and from multiple periphrasis, i.e. a kind of extended exponence where expression of a category requires the simultaneous presence of two ancillary elements. Secondly, pronominal affixes interact extensively with the system: as shown by Miller (1992), so-called pronominal ‘clitics’ in French are lexical affixes, and constitute an integral part of the verbal paradigm (cf. Stump, 1980; Bonami & Boyé, 2007). Their distribution in verbal periphrases varies from local realisation on the lexical verb (e.g. “near” future *aller*+infinitive) to non-local, i.e. ‘clitic climbing’ to the auxiliary (in analytic tenses with *avoir/être*). As suggested by Aguila-Multner & Crysmann (2020); Aguila-Multner (2023) ‘climbing’ can be fruitfully understood as an instance of periphrasis that goes piggy-back on an independent periphrastic dependency (TAM or voice). In the extreme case, combination of periphrases can yield forms with four or more syntactic words that jointly realise a single cell in the paradigm.

- (1) elle y a eu été distribuée  
SUBJ.3SG.F LOC have.IND.3SG have.PTCP be.PTCP distribute.PTCP.F.SG  
‘it used to be distributed there’

In this talk, we discuss the complex interaction of grammatical function change, multiple periphrasis relations and pronominal affixation in French. In line with previous work on inflectional periphrasis (Vincent & Börjars, 1996; Ackerman & Webelhuth, 1998; Ackerman & Stump, 2004; Popova & Spencer, 2013), we assume that morphology must have control over the entire paradigm, whether expression is synthetic or analytic. We argue that the complexity of the French dataset calls for an adequately explicit formal model of morphology, able to intertwine grammatical function change with realisation, and to deal with combinations of periphrastic forms. Building on Bonami (2015), we integrate a theory of inflectional periphrasis into Information-based Morphology (Crysmann & Bonami, 2015), treating periphrasis as delegation of exponence to an ancillary element by means of inside-out feature percolation. To inflect a lexeme given a paradigm cell, morphology may utilise not only synthetic realisation, but also periphrastic exponence, i.e. selection of an ancillary element. Given (i) the number of possible concurrent periphrastic relations a French verb can enter, (ii) the dependence of e.g. clitic climbing on an existing tense or voice periphrasis, as well as (iii) proper nesting for stacked periphrases, we need precise control over the interaction. We show how these interactions can be captured by giving inflection rules not only control over the periphrastic dependencies they launch themselves, but also over the word’s entire set of periphrastic relations, thereby accounting for the (synthetic vs. periphrastic) distribution of pronominal affixes, as well as constraining the combination of multiple periphrastic relations.

### References

- Ackerman, Farrell & Gregory T. Stump. 2004. Paradigms and periphrastic expression: A study in realization-based lexicalism. *Projecting morphology* 111–157.
- Ackerman, Farrell & Gert Webelhuth. 1998. *A theory of predicates*. CSLI Publications.
- Aguila-Multner, Gabrielle. 2023. *The morphosyntax of French complex predicates: Clitic climbing and periphrasis*: Université Paris Cité dissertation.

- Aguila-Multner, Gabrielle & Berthold Crysmann. 2020. French clitic climbing as periphrasis. *Linguisticae Investigationes* 43. 23–61.
- Bonami, Olivier. 2015. Periphrasis as collocation. *Morphology* 25(1). 63–110. doi:10.1007/s11525-015-9254-3.
- Bonami, Olivier & Gilles Boyé. 2007. French pronominal clitics and the design of Paradigm Function Morphology. In *Proceedings of the fifth Mediterranean Morphology Meeting*, 291–322.
- Crysmann, Berthold & Olivier Bonami. 2015. Variable morphotactics in Information-based Morphology. *Journal of Linguistics* 52(2). 311–374. doi:10.1017/S0022226715000018.
- Miller, Philip H. 1992. *Clitics and constituents in phrase structure grammar*. Garland, New York.
- Popova, Gergana & Andrew Spencer. 2013. Relatedness in periphrasis: A paradigm-based perspective. In Marina Chumakina & Greville G. Corbett (eds.), *Periphrasis: The role of syntax and morphology in paradigms*, 191–225. Oxford: Oxford University Press.
- Stump, Gregory. 1980. An inflectional approach to French clitics. *Ohio State University Working Papers in Linguistics* 24. 1–54.
- Vincent, Nigel & Kersti Börjars. 1996. Suppletion and syntactic theory. In *Proceedings of the 1st LFG colloquium*, 448–62.



## Community structure in inflectional networks: Exploring inflectional class structure

Maria Copot      Andrea Sims  
*The Ohio State University*

Word-and-paradigm morphology frames the lexicon as a network of words, entering in relationships of predictability with each other based on the way morphological exponents are distributed in the lexicon (Wurzel, 1984; Blevins, 2016). But while we can quantify uncertainty over these distributions (e.g. Ackerman & Malouf, 2013; Baayen et al., 2019), we lack tools to directly query their structure. Extending Sims (2020) and Beniamine (2021), we propose a methodology for querying the structure of relationships between word forms more directly.

The method is rooted in network theory, mathematical tools that centre the relational nature of object systems. We use the French verbal lexicon as a test case. We model the lexicon as a bipartite network (Gross & Yellen 2006), a structure with two partitions of nodes and edges connecting nodes across partitions (Figure 1). This allows us to capture similarities and differences between lexemes' inflectional behaviour. We globally segment all lexemes in a phonetically-transcribed inflected lexicon (Bonami et al., 2014), splitting each desinence into sequences of triphones. The triphones, together with their morphosyntactic tags, constitute nodes in one partition (N = 4432); nodes in the other partition are lexemes (N = 5249). For example, the method links /fɪni-sje/ 'finish.IND.IPFV.2PL' and /mãʒ-je/ 'eat.IND.IPFV.2PL' by virtue of sharing the triphone je#\_IND.IPFV.2PL, but simultaneously acknowledges that /finisje/ has inflectional markers that /mãʒje/ does not. The distributional properties of the resulting network can be analysed.

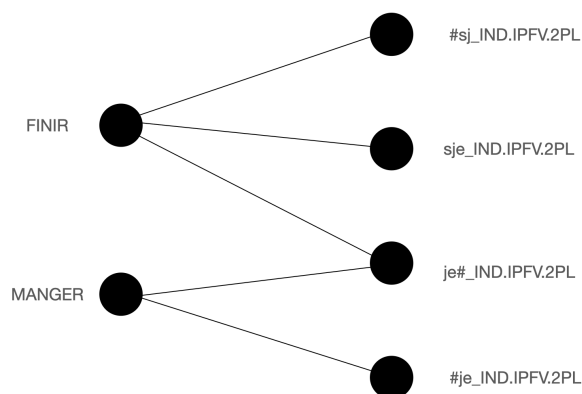


Figure 1: Illustration of the structure of the bipartite network

To demonstrate how queries over this structure provide insight, we show how inflection class grouping can be reframed in network terms as a community detection problem (Fortunato & Hric, 2016). When analyzing a system into inflectional classes by hand, linguists make decisions regarding which inflectional patterns to treat as equivalent and the extent to which differences should be abstracted over. This process inevitably results in a lossy representation

of the language’s inflectional structure. Approaches that aim for a lossless representation while simultaneously providing abstraction (Brown & Hippiisley, 2012) do so based on the assumption that inflectional classes have a hierarchical structure.

We show that classification of lexemes into inflectional classes in network terms -- specifically, based on the distribution of lexemes’ edges -- allows us to test whether the hierarchy assumption emerges from the data, while providing a lossless representation. Applying the Louvain algorithm (Blondel et al., 2008) to the network with different resolutions yields inflectional class sets of different granularities based on the same network structure (Table 1). We test the assumption of inflection class hierarchy by calculating whether two members of the same inflectional class at some resolution are assigned to the same inflectional class in the next-most coarse-grained resolution (the hierarchy coefficient in Table 1), iterating over pairs of words. We find that while broadly true, hierarchical organisation is not an absolute property that emerges.

The research programme of representing morphological systems as networks provides a more direct way to capture the relational nature of the lexicon implied by WP morphology, and has promising applications for morphological description, typology, and quantitative endeavours.

Resolution Parameter	N. of inflection classes	Hierarchy Coefficient
0.1	1	
0.2	4	1.000
0.3	5	0.967
0.4	7	0.958
0.5	7	0.923
0.6	9	0.988
0.7	9	0.957
0.8	8	0.982
0.9	9	0.993
1.0	9	0.976
1.1	10	0.942
1.2	18	0.987
1.3	28	0.896
1.4	24	0.991
1.5	36	0.993
1.6	60	0.922
1.7	363	0.575

**Table 1:** The number of inflection classes yielded by applying the Louvain algorithm to the network with different resolutions, and the hierarchy coefficient of each inflection class set compared to the next-most coarse-grained resolution.

## References

Ackerman, F., & Malouf, R. (2013). Morphological organisation: the low conditional entropy conjecture. *Language*, 89(3), 429–464. <https://doi.org/10.1353/lan.2013.0054>

- Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., and Blevins, J. (2019). The discriminative lexicon: a unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*, 2019, 4895891. <http://dx.doi.org/10.1155/2019/4895891>
- Beniamine, S. 2021. One lexeme, many classes: Inflection class systems as lattices. In Crysmann, B. & Sailer, M. (Eds.), *One-to-many relations in morphology, syntax and semantics*. Berlin: Language Science Press, 23-51. <http://dx.doi.org/10.5281/zenodo.4729789>
- Beniamine, S., Bonami, O., & Sagot, B. (2018). Inferring Inflection Classes with Description Length. *Journal of Language Modeling*, 5, 465-525. <http://dx.doi.org/10.15398/jlm.v5i3.184>
- Blevins, J. P. (2016). *Word and Paradigm morphology*. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199668984.013.22>
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008, P10008. <http://dx.doi.org/10.1088/1742-5468/2008/10/P10008>
- Bonami, O., Caron, G., & Plancq, C. (2014). Construction d'un lexique flexionnel phonétisé libre du français. In Neveu, F., Blumenthal, P., Hriba, L., Gerstenberg, A., Meinschaefer, J., & Prévost, S. (Eds.), *Actes du quatrième congrès mondial de linguistique française*, 2583–2596. <https://doi.org/10.1051/shsconf/20140801223>
- Brown, D. P., & Hippiusley, A. (2012). *Network Morphology: A defaults-based theory of word structure*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511794346>
- Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, 659, 1-44. <https://doi.org/10.1016/j.physrep.2016.09.002>
- Gross, J. L., & Yellen, J. (2006). *Graph theory and its applications*. Boca Raton, FL: Chapman & Hall. <https://doi.org/10.1201/9780429425134>
- Sims, A. D. (2020). Inflectional networks: Graph-theoretic tools for inflectional typology. In *Proceedings of the Society for Computation in Linguistics*, 3, 88-98. 10.7275/c1f4-pg94
- Wurzel, W. U. (1984). *Flexionsmorphologie und Natürlichkeit. Ein Beitrag zur morphologischen Theoriebildung*. Berlin: Akademie Verlag. <https://doi.org/10.1515/9783050080130>

**Keywords:** word-and-paradigm morphology, network theory, inflectional class structure, inflectional class hierarchy, French



## **Langues & Langage à la croisée des Disciplines**

### **1ère Rencontre annuelle LLcD**

Colloque : 9-11 septembre, 2024

Ecole d'été : 12-14 septembre, 2024

Sorbonne Université - 28 Rue Serpente, 75006 Paris, France



## **Workshop 9**

### **Cooccurrences et marquage discursif**

## L'expression de la surprise. Phrases préfabriquées et marqueurs discursifs

Nous étudions les combinaisons de marqueurs discursifs (MD) et de phrases préfabriquées présentes dans les interactions (PPI), lorsqu'elles marquent la surprise.

Les PPI sont des phrases toutes faites fréquemment utilisées dans les conversations comme *y'a pas de quoi, et puis quoi encore !, c'est clair*. Ces phrases prêtes à l'emploi présentent plusieurs propriétés : elles sont souvent non compositionnelles, sont particulièrement fréquentes dans les dialogues et sont associées à des actes de langage. Elles sont rarement traduites de façon littérale et posent souvent des difficultés dans l'apprentissage d'une langue étrangère. Par ailleurs, les PPI présentent une assez grande ambiguïté et peuvent être associées à des actes de langage divers. Ainsi, *ça va* a un emploi de salutation rituel, mais plutôt un emploi expressif quand il exprime l'exaspération.

Les MD sont envisagés ici dans une perspective large qui inclut les connecteurs et les particules énonciatives. Ils désignent des expressions (mots ou syntagmes) qui servent à organiser le discours : soit pour en marquer les articulations logico-rhétoriques (*mais, donc, la preuve, quand même, autrement dit, etc.*), soit pour manifester les attitudes épistémiques et affectives du locuteur (*tiens, ah, etc.*), ou la gestion de l'interaction (*écoute, d'accord, etc.*) (Dostie 2004). Leurs propriétés sont régulièrement débattues (xx 2023, à par.). Les MD peuvent apparaître seuls ou combinés à d'autres MD (*mais enfin, bon ben, non mais oh, etc.*) (cf. Dargnat 2022, Cuenca et Crible 2019).

Nous souhaitons observer le fonctionnement d'un sous-ensemble de PPI renvoyant à l'expression de la surprise, champ émotionnel particulièrement fertile, où l'on peut dénombrer plus d'une centaine d'expressions comme *tu plaisantes ? c'est dingue ! les bras m'en tombent !* (Tutin et Grossmann 2023). Parmi ces PPI, on repère des régularités aux plans pragmatique, sémantique et syntaxique. Par exemple, la surprise est souvent associée au schéma pragmatique de la plaisanterie (*tu plaisantes ? c'est une blague ? sans rire ?, etc.*) ou au schéma de l'observation d'un phénomène hors-norme (*c'est dingue ! comme c'est curieux !, etc.*).

L'association de ces PPI de surprise à des MD assez spécifiques est fréquente. Notre hypothèse est que ces MD, simples ou eux-mêmes déjà complexes, permettent de renforcer ou de préciser certaines fonctions pragmatiques des PPI. Ainsi des MD comme *non mais* en initiale de tour de parole ou en antéposition renforcent l'effet de surprise (*non mais tu rigoles*), alors que d'autres MD postposés comme *ou quoi* (*vous plaisantez ou quoi ?*) vont associer à l'expression de la surprise un désaccord marqué. L'observation des cooccurrences entre PPI et MD permet ainsi de mieux comprendre le fonctionnement interactionnel des PPI, mais aussi des MD eux-mêmes. Par ailleurs, les MD peuvent constituer des indices utiles pour opérer une désambiguïté en contexte des PPI.

Concernant les données, nous recourons à plusieurs corpus communs à deux projets (en particulier ESLO, CFPP, TCOF, Wikipédia)<sup>1</sup> ainsi qu'à des dialogues extraits de romans contemporains. La méthodologie fait appel à deux environnements : *Le Lexicoscope* (Kraif 2019) pour le relevé des PPI, et *Unitex-GramLab* pour l'identification des MD. Le calcul statistique des cooccurrences privilégiées de MD et PPI exploite certaines mesures d'association, entre autres les mesures directionnelles telles qu'elles sont définies par exemple par Brezina (2018).

### Références

- Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide*, Cambridge, Cambridge University Press.
- Cuenca, M.-J. & Crible, L. (2019). « Co-occurrence of discourse markers in English: From juxtaposition to composition ». *Journal of Pragmatics* n° 140, p. 171-184.
- Dargnat, M. (2022). « *Mais enfin* : construction et association », *Langages* n° 225, p. 49-63.

---

<sup>1</sup> Certains sont repris dans la base ORFEO, mais pas dans leur totalité.

Dostie, G. (2004). *Pragmaticalisation et marqueurs discursifs. Analyse sémantique et traitement lexicographique*, Liège, De Boeck/Duculot.

Kraif, O. (2019). « Explorer la combinatoire lexico-syntaxique des mots et expressions avec le LEXICOSCOPE », *Langue française* n° 203, p. 67-82.

Tutin, A. & Grossmann, F. (2023). « Les phrases préfabriquées exprimant la surprise : vers l'élaboration de schémas sémantico-syntaxiques et pragma-sémantiques rendant compte des régularités », *Studii de lingvistică* n° 13-2, p. 145-171.

*Unitex-GramLab*, suite logicielle pour l'analyse de corpus : <https://unitexgramlab.org/fr>

\*\*\*

xx (2023). à compléter après évaluation.

xx (à par.). à compléter après évaluation.

## Modèles de cooccurrences collostructionnels pour l'étude des expressions préfabriquées de l'oral

Olivier Kraif

Univ. Grenoble Alpes, LIDILEM, F-38000 Grenoble, France

olivier.kraif@univ-grenoble-alpes.fr

Cette communication propose une réflexion autour des modèles de cooccurrences pertinents pour l'étude de constructions productives dans le champ des expressions préfabriquées de l'oral (Tutin & Grossmann, 2023 ; Pausé, et al. 2022). Nous commencerons par définir différents modèles de cooccurrence en reprenant notamment l'opposition entre cooccurrence de surface et cooccurrence syntaxique (Evert, 2008). Le modèle syntaxique, popularisé avec les *wordsketches* du Sketch Engine (Kilgariff, et al. 2004), a prouvé son intérêt dans l'étude des collocations et de la phraséologie (Seretan 2011, Bartsh et Evert 2013), et est aujourd'hui accessible dans une variété d'outils au delà du Sketch Engine, comme TXM (avec l'extension appropriée, Rainsford & Heiden 2014) ou le Lexicoscope 2.0 (Kraif 2019). Partant du constat que le modèle syntaxique et le modèle surfacique des *ngrams* sont complémentaires (Kraif et Tutin, 2015), nous chercherons à déterminer, à travers une étude de cas, comment combiner au mieux des patterns surfaciques caractéristiques (pour des constructions telles que *c'est pas + adjectif*, comme dans *c'est pas vrai*, *c'est pas possible*, *c'est pas croyable*) avec des relations de dépendances pour observer les marqueurs discursifs caractéristiques de l'environnement de ces constructions de l'oral (*non mais*, *ça alors*, etc.). Nous introduirons notamment de nouveaux observables spécifiquement dédiés à l'étude des collostructions (Stefanowitsch & Gries 2003, Gries 2019), tels que les « tableaux de réalisation » implantés dans le Lexicoscope. On pourra ainsi distinguer les collostructions endogènes (collocation d'une unité lexicale à l'intérieur d'une construction) et exogènes (collocation avec une unité extérieure à la construction). Nous examinerons enfin différents indices statistiques dédiés à la mesure de la force d'association, en évaluant l'apport des indices directionnels comme suggérés par Gries (2019).

**Mots clés :** phraséologie, expressions préfabriquées de l'oral, collostruction

### Références

- Evert, S. (2008). Corpora and collocations, in A. Lüdeling and M. Kytö (eds.), *Corpus Linguistics. An International Handbook*, article 58. Mouton de Gruyter, Berlin, 1212–1248.
- Gries, S. Th. (2019) 15 years of collostructions: Some long overdue additions/corrections (to/of actually all sorts of corpus-linguistics measures). *International Journal of Corpus Linguistics*, 24(3):385-412
- Kilgariff, A., Rychly, P., Smrz, P., Tugwell, D. (2004), "The Sketch Engine", *Proceedings of the Eleventh EURALEX International Congress*. Lorient, France, p. 105–116.
- Kraif, O. (2019). Explorer la combinatoire lexico-syntaxique des mots et expressions avec le Lexicoscope. In Max Silberstein (dir.), *Langue française*, N° 203, Armand Colin, p. 67-82.
- Pausé, M.-S., Tutin, A., Kraif, O., Coavoux, M. (2022). Extraction de Phrases Préfabriquées des Interactions à partir d'un corpus arboré du français parlé : une étude exploratoire. *Congrès Mondial de Linguistique Française - CMLF 2022*, Jul 2022, Orléans, France.
- Rainsford T., Heiden S. (2014). « Key Node in Context (KNIC) Concordances: Improving Usability of an Old French Treebank », in *Actes de la 4e édition du Congrès Mondial de Linguistique Française (CMLF)*. Berlin, 19-23 juillet 2014, Vol. 8, 2707-2718.
- Seretan, V. (2011) . *Syntax-Based Collocation Extraction*. Text, Speech and Language Technology, Springer.
- Stefanowitsch, A., & Gries, S. Th. (2003). Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243. <https://doi.org/10.1075/ijcl.8.2.03ste>
- Tutin, A., Kraif O. (2016). Routines sémantico-rhétoriques dans l'écrit scientifique de sciences humaines : l'apport des arbres lexico-syntaxiques récurrents, in A. Tutin & F. Satri (coord.) *Phraséologie et genres de discours : patrons, motifs, routines*, Lidil n° 53, Grenoble : Ellug.

Tutin, A., Grossmann, F. (2023). Les phrases préfabriquées exprimant la surprise : vers l'élaboration de schémas sémantico-syntaxiques et pragma-sémantiques rendant compte des régularités. *Studii de lingvistică*, 13 (2), pp.145 - 171.



## L'extraction des marqueurs de discours dans les corpus

De plus en plus de travaux s'intéressent aux cooccurrences des marqueurs de discours (MD) comme *mais bon, alors donc, oui mais quand même tu vois*, etc. (voir par exemple Haselow 2018, Crible et Degand 2021, Dargnat 2022, Dargnat et Jayez 2021). Il serait souhaitable d'étudier ces cooccurrences sur de grands corpus pour avoir une vision plus précise de leur composition et de leur importance relative. Cependant, ce projet se heurte à au moins deux problèmes classiques. Premièrement, l'ambiguïté lexicale ou catégorielle de nombre de MD. Celle-ci a principalement trois manifestations : en dehors de leur emploi comme MD, certains mots peuvent fonctionner comme adjectifs et/ou noms (par exemple *bon* et *bien*), d'autres comme adverbes (*bien*), d'autres encore comme constructions verbales (par exemple la série des *tu vois, tu sais*, etc.). Deuxièmement, le fait que l'annotation manuelle sur de grands corpus est coûteuse et pas toujours bien évaluable.

L'extraction automatique à l'aide d'étiqueteurs probabilistes ou par apprentissage profond donne des résultats inégaux. Les tests effectués sur une dizaine de modèles montrent que les étiqueteurs dissimilent correctement les emplois pour des mots comme *bon* mais ne sont pas appropriés dans beaucoup d'autres cas. Cela tient principalement à deux facteurs. (1) les corpus d'apprentissage favorisent les étiquettes les plus courantes, et les annotations plus exigeantes (French SUD guidelines) sont encore minoritaires. (2) Le statut de MD est souvent parenthétique, au sens de Heine et al. (2021), c'est-à-dire, pour ce qui nous concerne, en dehors des relations de dépendance. Par exemple *tu vois* comme MD est à distinguer de *tu vois que/comment/pourquoi ...*, *ce que tu vois* ou *tu vois bien*. Un étiquetage de *tu vois* comme interjection dans un corpus d'entraînement ignore ces différences et conduit à surestimer le nombre d'occurrence de *tu vois* comme MD. Inversement, une analyse de *tu vois* comme pronom+verbe va rater les occurrences comme MD.

Je présenterai les résultats d'une autre approche reposant sur l'analyse de motifs par automates finis à l'aide du logiciel Unitex-Gramlab (<https://unitexgramlab.org/>) et reposant sur trois techniques. (1) Un traitement par élimination, où les motifs qui excluent l'interprétation MD sont d'abord reconnus et marqués pour isoler différenciellement les occurrences authentiques de MD, le tout au moyen de séquences de transducteurs (= automates annotant le texte). (2) Une organisation des séquences pour traiter les MD complexes qui contiennent des mots susceptibles d'être des MD (hiérarchie de préemption). (3) L'adjonction de scripts pour affiner les résultats, notamment pour les MD comme *après*, où l'emploi comme préposition (donc non MD) est difficile à isoler. Il y a à l'heure actuelle une séquence de 597 automates qui explorent 7 corpus (4 relevant de l'oral et 3 de l'écrit) pour annoter les occurrences de 726 MD. Un des avantages de l'approche est de permettre d'ajouter des MD et des corpus en ne recalculant que les MD pertinents. Outre les retombées pour étudier statistiquement les combinaisons de MD, on peut utiliser les données produites comme corpus d'entraînement pour des étiqueteurs, ce qui fait l'objet d'un projet complémentaire en cours.

### Références

- Crible, L et Degand, L. 2021. Co-occurrence and ordering of discourse markers in sequences: A multifactorial study in spoken French. *Journal of Pragmatics* 177, 18-28.  
Dargnat, M. 2022. *Mais enfin* : construction et association. *Langages* 225, 49-63.  
Dargnat, M. et Jayez, J. 2021. *Mais* et associés. *Lexique* 29, 211-228.  
French SUD guidelines : <https://surfacesyntacticud.github.io/guidelines/fr/>  
Haselow, A. 2018. Discourse marker sequences: Insights into the serial order of

communicative tasks in real-time turn production. *Journal of Pragmatics* 146, 1-18.  
Heine, B. et al. 2021. *The Rise of Discourse Markers*. Cambridge U.P.

## Ordre, force et variation en genre des co-occurrences de marqueurs discursifs

Ludivine Crible

*Université de Gand*

Les co-occurrences de marqueurs discursifs (MDs) du type *et donc* ou *mais enfin* ont fait l'objet d'un nombre croissant d'études (par exemple Waltereit 2007, Cuenca & Marín 2009, Dostie 2013, Koops & Lohmann 2015, Haselow 2019) visant à déterminer les principes expliquant l'ordre et/ou la force d'attraction des MDs au sein de ces séquences. Cet intérêt peut s'expliquer par la grande diversité formelle et fonctionnelle des MDs, résultant ainsi en un phénomène multifactoriel. La syntaxe, la fonction pragmatique ou encore la prosodie ont été avancées par ces auteurs comme autant de variables pertinentes dans l'étude des co-occurrences de MDs. La variation en genre de parole a, quant à elle, reçu moins d'attention dans cette perspective. Dans la lignée de ces travaux, et pour pallier cette lacune, notre question de recherche est la suivante : à quel(s) besoin(s) communicatif(s) répondent les co-occurrences de MDs ? En d'autres termes, pourquoi utiliser deux MDs plutôt qu'un ? Nous y répondrons grâce à plusieurs analyses qualitatives et quantitatives basées sur corpus oraux.

Cette contribution fait d'abord la synthèse de nos travaux sur ce sujet, en classant d'une part les facteurs responsables de l'ordre des MDs en co-occurrence, et d'autre part les différents types de séquences. Pour cela, nous ferons intervenir des variables à la fois formelles et fonctionnelles comme la catégorie syntaxique des marqueurs ou la (les) fonction(s) exprimée(s) par la séquence. Des analyses systématiques sur corpus oraux en anglais et en français révèlent que la catégorie d'origine du marqueur (conjonction, adverbe, etc.) joue un rôle primordial pour expliquer tant l'ordre des MDs que leur degré d'association.

Dans un second temps, nous explorons les co-occurrences de MDs dans un corpus récent de dialogues hautement collaboratifs en français, dans lesquels les interlocuteurs participent à un jeu de description d'images. La nature-même de la catégorie des MDs semble différente dans ces données par rapport à d'autres interactions orales car elle s'ouvre à des expressions appelées par ailleurs *backchannels*, marqueurs de feedback ou d'alignement, telles que *d'accord*, *okay* ou *c'est clair*. De cette ouverture de la catégorie découlent des séquences de co-occurrences plus fréquentes et plus longues que dans les corpus précédemment analysés. Nous rapporterons les résultats quantitatifs de cette analyse comparative sur plusieurs genres de français oral et proposerons une interprétation communicative et interpersonnelle de ces différences. Il semble en effet que, dans un genre très interactif et collaboratif, la cooccurrence de MDs répond à un besoin constant d'exprimer son alignement avec l'interlocuteur et de faciliter le succès de la tâche en cours. En conversation générale, toutefois, c'est la multifonctionnalité qui semble expliquer le recours à des séquences de MDs (besoin d'exprimer une fonction nouvelle ou de renforcer une fonction sous-spécifiée), et cet aspect se trouve reflété dans la syntaxe (ordre et catégorie grammaticale des MDs co-occurents).

**Mots-clés** : marqueurs discursifs ; co-occurrence ; variation ; corpus

## Références

Cuenca, M. J. & Marín, M. J. (2009). Co-occurrence of discourse markers in Catalan and Spanish oral narrative. *Journal of Pragmatics*, 41, 899–914.

Dostie, G. (2013). Les associations de marqueurs discursifs. De la cooccurrence libre à la collocation. *Linguistik*, 62(5), 15–45.

Haselow, A. (2019). Discourse marker sequences: Insights into the serial order of communicative tasks in real-time turn production. *Journal of Pragmatics*, 146, 1–18.

Koops, C. & Lohmann, A. (2015). A quantitative approach to the grammaticalization of discourse markers. *International Journal of Corpus Linguistics*, 20(2), 232–259.

Waltereit, R. (2007). A propos de la genèse diachronique des combinaisons de marqueurs. L'exemple de bon ben et enfin bref. *Langue française*, 154, 94–109.

## S'associer pour mieux contrer

Combinaison de marqueurs discursifs dans le processus argumentatif de la concession

### Questions de recherche

L'analyse emprunte à deux recherches plus larges, l'une assez générale traitant de la combinaison des marqueurs discursifs (Md) en français (xxx), et l'autre plus ciblée traitant de la concession dans une perspective expérimentale (xxx). Nous proposons ici une étude exploratoire de la combinaison des Md habituellement liés à la relation de discours de concession. Cette combinaison, ou association, sera interrogée dans deux directions :

- Dans quelle(s) mesure(s) les marqueurs de concession peuvent-ils se combiner entre eux ?
- Avec quel(s) autre(s) type(s) de Md les marqueurs de concession se combinent-ils ?

### Approche et méthode

Nous situons notre approche dans le cadre de la linguistique de corpus et de la modélisation du sens.

– La première étape du travail est définitoire. Elle consiste : (a) à identifier les schémas dits concessifs dans une perspective discursive (cf. Asher & Lascarides 2003, Taboada & Mann 2006, Renkema 2009, Taboada, Stede & Das 2017) et dans une perspective argumentative (Walton et al. 2008, Freeman 2011, Sanders et al. 2021), et (b) à lister les marqueurs de concession en français, sur la base des grammaires, des études et des listes disponibles (cf. Morel 1996, Roze et al. 2012, Abeillé et al. 2021).

– La deuxième étape concerne la préparation des données. Nous utilisons plusieurs corpus oraux et écrits mis à disposition par la communauté (références dans la bibliographie). Ils ont été récupérés et uniformisés afin d'être exploitables avec le logiciel *Unitex-GramLab* (xxx 2023). Dans cet environnement, nous avons développé un dictionnaire enrichi comportant plus de 700 Md, dont les marqueurs de concession, et nous avons généré et appliqué des cascades d'automates finis pour identifier au plus près les candidats Md et laisser de côté les emplois non discursifs d'items polyfonctionnels (par ex. *bon* adjectif vs *bon* Md ; *tu parles* verbe vs *tu parles* Md). Une procédure de transduction nous a permis de les annoter directement dans les données brutes (Boczon 2023).

– La troisième étape est en cours et constitue l'analyse statistique des associations de Md et la sélection des associations les plus fréquentes. Une fois les Md accessibles dans les données, nous lançons des mesures d'association (cf. Gries 2013, Brezina 2018, xxx 2022) en déclarant les marqueurs de concession comme pivots et l'ensemble des Md comme associés potentiels. Ceci doit nous permettre de répondre quantitativement aux deux questions posées ci-dessus et de pointer les cas les plus fréquents.

– La quatrième étape n'est pas encore réalisée. Il s'agira d'une proposition de représentation des deux types de combinaison envisagés : (a) la combinaison de deux marqueurs de concession (par ex. *mais + quand même/pourtant*), et (b) la combinaison d'un marqueur de concession et d'un autre Md (par ex. *et + pourtant, si + toutefois*). Nous étudions actuellement les possibilités offertes par le cadre interactionnel de Ginzburg (2012) pour décrire les marqueurs et leurs combinaisons au moyen de structures de traits typées.

### Bibliographie

- Abeillé A. & Godard D. (dir.) 2021. *Grande Grammaire du Français*. Arles : Actes Sud/Imprimerie Nationale.
- Asher N. & Lascarides A. 2003. *Logics of conversation*. Cambridge : Cambridge University Press.
- Boczon K. 2023. *Text Mining for Discourse Markers*. Mémoire de stage, Master 2 TAL, Université de Lorraine-IDMC, Nancy, 25 p.
- Brezina V. 2018. *Statistics in Corpus Linguistics : A Practical Guide*. Cambridge: Cambridge University Press.
- Cuenca M.-J. & Crible L. 2019. «Co-occurrence of discourse markers in English: From juxtaposition to composition ». *Journal of Pragmatics* 140, 171-184.

- Freeman J. B. 2011. *Argument Structure: Representation and Theory*. Cham : Springer.
- Ginzburg J. 2012. *The Interactive Stance. Meaning for Conversation*. Oxford : Oxford University Press.
- Gries S. T. 2013. « 50-something years of work on collocations : what is or should be next... ». *International Journal of Corpus Linguistics* 18 (1), 37-165.
- Morel M.-A. 1996. *La Concession en français*. Paris : Ophrys.
- Peldszus A. & Stede M. 2016. « Rhetorical structure and argumentation structure in monologue text ». *Proceedings of the 3rd Workshop on Argumentation Mining*, ACL, Berlin, August 2016, 103-112.
- Roze C., Danlos L. & Müller P. 2012. « LEXCONN : A French Lexicon of Discourse Connectives. *Discours* 10, 15 p.
- Stede M., Taboada M. & Das D. 2017. *Annotation Guidelines for Rhetorical Structure*. Version 1.0.
- Renkema J. 2009. *The Texture of Discourse*. Amsterdam-Philadelphia : John Benjamins.
- Sanders T., Demberg V., Hoek J., Scholman M. C.J., Asr. F. T., Zufferey S. & Evers-Vermeul J. 2021. « Unifying dimensions in coherence relations: How various annotation frameworks are related ». *Corpus Linguistics and Linguistics Theory* 17/1, 1-71.
- Taboada M. & Mann W. 2006. « Rhetorical Structure Theory: Looking Back and Moving Ahead ». *Discourse Studies* 8/3, 423-459.
- Unitex-GramLab* : suite logicielle pour l'analyse de corpus : <https://unitexgramlab.org/fr>
- Walton D. N., Reed C. & Macagno F. 2008. *Argumentation Schemes*. Cambridge University Press.

### Corpus

*Écrit* : extraits de *Le Monde*, *Le Monde Diplomatique*, Wikipedia.

*Oral* : extraits de CFPP (Corpus de Français Parlé Parisien), CLAPI (Corpus de Langue Parlée en Interaction), CRFP (Corpus de Référence du Français Parlé), DÉCLICS (Corpus du projet Dispositif d'Études CLIniques sur les Corpus de Santé), ESLO (Enquêtes Sociolinguistiques à Orléans), FRA80 (Corpus de Français des Années 80), MPF (Multicultural Paris French), TCOF (Traitement de Corpus Oraux en Français).

---

# Phrasèmes interactifs intégrant le morphème interro-exclamatif ”-tu” en français québécois parlé (1970-2020). ”On s’entend-tu”, ”ça se peut-tu”, ”tu penses-tu”, ”tu veux-tu ben”

Gaétane Dostie\*<sup>1</sup>

<sup>1</sup>Université de Sherbrooke (CANADA) – Canada

## Résumé

Les variétés européennes du français disposent de trois procédés interrogatifs pour les questions directes totales (auxquelles on peut répondre par *oui* ou *non* ; entre autres, Lefeuve & Rossi-Gensane 2015, Delaveau 2021). Il s’agit de l’intonation montante (*tu viens ?*), de l’inversion du verbe et du pronom clitique (*viens-tu ?*) et du morphème *est-ce que* figurant en début d’énoncé (*est-ce que tu viens ?*). À cela s’ajoute, en français québécois, le morphème *-tu* (*tu viens-tu ?*), qui provient vraisemblablement d’une réanalyse de la particule postverbale *-ti* (elle-même issue de *-t-il* ; Foulet 1921, Picard 1992, Elsig 2009), assimilée au pronom d’adresse de deuxième personne du singulier (*tu* ; Léard 1995). Les spécialistes situent le passage progressif de *-ti* à *-tu* vers les années 1940-1950 (entre autres, Léard 1995). De nos jours, *-tu* est largement utilisé dans les questions directes totales (notamment, Bally 2022, Bergeron-Maguire, Dostie Lefeuve 2024). Ce morphème s’emploie également dans les énoncés exclamatifs qui expriment, par définition, le haut degré, à l’image d’autres morphèmes interrogatifs tels *qui*, *comme*, *combien* et *qu’est-ce que*. Ces derniers ne sont pas en soi intensifs mais l’énoncé exclamatif les dote d’une telle valeur (*Grande grammaire historique du français* 2020, p. 743-744). Il en est de même de *-tu*, qui prend une teinte intensive lorsqu’il est inséré dans un énoncé exclamatif (*elle est-tu fine rien un peu !* (‘elle est vraiment fine’)).

L’attention se focalisera dans cette communication sur une série de phrasèmes intégrant, dans leur signifiant, le morphème interro-exclamatif *-tu*, comme *ça a-tu du bon sang*, *je le sais-tu* et *tu veux-tu ben me dire*. Ces phrasèmes remplissent diverses fonctions dans les interactions verbales à l’oral, notamment expressive, conative et métadiscursive, à l’instar de ce qu’on a appelé diversement ailleurs, entre autres, *énoncés liés* (Fónagy 1982, Bidaud 2002), *phrases (figées) situationnelles* (Náray-Szabó 2006, Klein et Lamiroy 2016), *phrases préfabriquées des interactions* (Tutin 2019, 2022), sans oublier pour certains de ces phrasèmes, *marqueurs discursifs* (Hansen & Visconti éds sous presse).

La liste des phrasèmes que nous étudierons sera établie par un dépouillement systématique du *Corpus de français parlé au Québec* (2006-2019). Les corpus oraux disponibles sur le site du *Fonds de données linguistiques du Québec* seront également consultés afin de compléter l’inventaire. Les enregistrements à la base de ces derniers corpus datent, pour les plus anciens, du début des années 1970 et, pour les plus récents, de la décennie 2010-2020. Étant donné que la généralisation du morphème interro-exclamatif *-tu* est relativement récente

---

\*Intervenant

dans les énoncés ” libres ”, comme cela a été précisé plus haut, nous nous attendons à ce que les phrasèmes interactifs intégrant ce morphème soient plus fréquents et plus diversifiés dans les corpus les plus récents.

L'exposé sera axé sur les questions suivantes.

- Nous examinerons les items linguistiques qui participent à la formation des phrasèmes inventoriés (notamment les pronoms et les verbes) afin de dégager leurs propriétés sémantico-syntaxiques communes. Par exemple, nous observons une possibilité de ” remodelage ” des marqueurs discursifs qui exploitent la construction ” verbe de perception/cognition + pronom de 2e personne ” (comme *sais-tu*, *vois-tu* et *penses-tu* qui deviennent *tu sais-tu*, *tu vois-tu* et *tu penses-tu*). Cette ” adaptabilité ” des phrasèmes initiaux repose sur la plus ou moins grande fixité de leurs constituants et sur leur plus ou moins grande compositionnalité sémantique. Ainsi, *tu parles* n'a pas pour pendant \**tu parles-tu*, ce qui est cohérent avec le fait qu'il s'agit vraisemblablement d'une locution forte dans la typologie melchukienne des phrasèmes (entre autres, Mel'ëuk, 2023), alors que *vois-tu*, à titre comparatif, serait une semi-locution. De ce fait, ce dernier phrasème serait également moins contraint d'un point de vue morphosyntaxique. Ainsi, il présente des variantes, à savoir *tu vois*, *voyez-vous* et, donc aussi dans le cas qui nous occupe ici, *tu vois-tu*. La variante \**vous voyez-tu* est pour sa part exclue, en raison d'une contrainte historique : la réanalyse de *-ti* en *-tu* par assimilation avec le pronom d'adresse *tu* entraîne des blocages dans les énoncés ” libres ” (p. ex. ?*vous venez-tu* ; entre autres, Picard 1992, Elsig 2009), qui persistent lorsqu'il y a phraséologisation.

- En lien avec ce qui précède, nous chercherons à déterminer le statut phraséologique des phrasèmes étudiés en nous référant à la typologie melchukienne (cf. en particulier, les distinctions entre la locution forte, la semi-locution et la locution faible). Cette approche vaudra pour les phrasèmes qui mettent à contribution des constructions comportant des variables lexicalement instanciées (comme cela se produit avec les marqueurs discursifs, qui sont en nombre restreint). De manière complémentaire, nous étudierons quelques constructions schématiques (plus abstraites ou générales), qui comportent des variables dont l'instanciation repose sur des listes ouvertes (Legallois 2018, Golberg 2019, Carlier & Prévost 2021). Ainsi, la construction < *ce* + *être-tu* ADJ axiologique > permet de générer une liste ouverte de phrasèmes évaluatifs pour qualifier des situations tels, *c'est-tu naïseux* et *c'est-tu fou*.

- Chemin faisant, nous prêterons attention aux collocatifs qui apparaissent avec une certaine récurrence à gauche et à droite des diverses séquences phraséologiques examinées (comme *heille*, *ben* et *là*). Il s'agira ici de voir comment ces collocatifs se distribuent eu égard aux fonctions interactionnelles assumées par les phrasèmes comportant le morphème interrogatif *-tu*.

## Références

- Bally, A.-S., 2022, ” Les interrogatives totales en français québécois dans l'écrit SMS : à la croisée de l'oral et de l'écrit ”, SHS Web of Conferences, *Congrès Mondial de Linguistique Française - CMLF 2022*, 138. (<https://doi.org/10.1051/shsconf/202213812006>)
- Bergeron-Maguire, M., G. Dostie & F. Lefevre, 2024, ” Les procédés (morpho-)syntaxiques de l'interrogation directe totale en français québécois des années 2000 : *l'as-tu lu ?*, *tu l'as-tu lu ?*, *est-ce que tu l'as lu ?* ”, *Langue française*, 221, p. 21-38.
- Bidaud, F., 2002, *Structures figées de la conversation : analyse contrastive français-italien*, Bern, P. Lang.
- Carlier, A. & S. Prévost, 2021, ” Constructions, constructionnalisation et changement linguistique. Présentation ”. *Langue française*, 209/1, p. 9-22.
- Delaveau, A., 2021, ” Les phrases interrogatives indépendantes ”, dans A. Abeillé & D. Godard (éds), *La Grande Grammaire du français*, Arles, Actes Sud / Imprimerie nationale Éditions, tome 2, p. 1402-1410.
- Elsig, M., 2009, *Grammatical variation across Space and Time: the French interrogative system*, Amsterdam, John Benjamins.
- Fónagy, I., 1982, *Situation et signification*, Philadelphia, J. Benjamins.
- Foulet, L., 1921, ” Comment ont évolué les formes de l'interrogation ”, *Romania*, 47/186-187, p. 243-348.



- Goldberg, A., 2019, *Explain Me This: Creativity, Competition, and Partial Productivity of Constructions*, Princeton, Princeton University Press.
- Grande grammaire historique du français*, 2020, 2 volumes, C. Marchello-Nizia, B. Combettes, S. Prévost & T. Scheer, Berlin (éds), Berlin/Boston, De Gruyter Mouton.
- Hansen, M. B. M. & J. Visconti (éds), sous presse, *Manual of Discourse Markers in Romance*, Berlin, De Gruyter.
- Klein, J.-R. & B. Lamiroy, 2016, " Le figement : unité et diversité (...) ", *L'Information grammaticale*, 148, p. 15-20.
- Léard, J.-M., 1995, *Grammaire québécoise d'aujourd'hui. Comprendre les québécismes*, Montréal, Guérin universitaire.
- Lefevre, F. & N. Rossi-Gensane, 2015, " Interrogation ", dans P. Larrivée & F. Lefevre (éds), *Projet Français Contemporain Vernaculaire* (Fracov). (<http://www.univ-paris3.fr/index-des-fiches-227311.kjsp?RH=1373703153287>)
- Legallois, D., 2018, " La notion de construction ", dans D. Apothéloz et al., (éds). *Encyclopédie grammaticale du français*. [http://www.encyclogram.fr/notx/012/012\\_Notice.php](http://www.encyclogram.fr/notx/012/012_Notice.php)
- Mel'čuk, I. A. 2023, *General Phraseology: Theory and Praxis*, Amsterdam/Philadelphia, J. Benjamins.
- Náray-Szabó, M., 2006, " Pragmatique et sémantique des phrases figées situationnelles ", *Verbum Analecta Neolatina*, 8/2, p. 473-493.
- Picard, M., 1992, " Aspects synchroniques et diachroniques du *tu* interrogatif en québécois ", *Revue québécoise de linguistique*, 21/2, p. 65-75.
- Tutin, A., 2019, " Phrases préfabriquées des interactions : quelques observations sur le corpus CLAPI ", *Cahiers de lexicologie*, 114/1, p. 63-91.
- Tutin, A., 2022, " *Comment dirais-je ?*, *Que veux-tu ?*, *Comment ça va ?* ", *Lingvisticae Investigationes*, 45/2, p. 172-196.

#### Corpus

- Corpus de français parlé au Québec* (2006-2019), Université de Sherbrooke. (En ligne : <https://applis.flsh.usherbrooke.ca/cfpq/>) (Dernière consultation le 1er avril 2024)
- Fonds de données linguistiques du Québec*, Université de Sherbrooke. (En ligne : <https://fdlq.recherche.usherbrooke>. (Dernière consultation le 1er avril 2024)

**Mots-Clés:** phraséologie, énoncés liés des interactions, marqueurs discursifs, degré de compositionnalité sémantique, constructions, linguistique de corpus

## Références :

- André, V. 2005. *Oui non*: une pratique discursive sous influence. *Marges linguistiques* 9. 195-213.
- Avanzi, M., Bordal, G., Lacheret-Dujour, A., Obin, N., & Sauvage-Vincent, J. (2019). The annotation of syllabic prominences and disfluencies. In *Rhapsodie* (pp. 157-173). John Benjamins.
- Cohen, M. 1952. Emplois nouveaux de *oui* et *non* en français. *Bulletin de la Société de Linguistique de Paris*, 48, 40-51.
- Crible, L., & Degand, L. 2021. Co-occurrence and ordering of discourse markers in sequences: A multifactorial study in spoken French. *Journal of Pragmatics*, 177, 18-28.
- Dostie, G. 2012. *Ben* en tant que collocatif discursif. *Travaux de linguistique*. (2). 105-122.
- Garcia, C. 1982. Interaction et analyse du discours: étude comparative de débats entre adolescents. *Études de linguistique appliquée*. 46. 98-118.
- Lacheret-Dujour, A. 2019. Prosodic annotation of the Rhapsodie corpus: Expectations and issues. In *Rhapsodie* (pp. 147-155). John Benjamins.
- Lacheret-Dujour, A., & Victorri, B. 2019a. Segmentation into intonational periods. In *Rhapsodie* (pp. 175-211). John Benjamins.
- Lacheret-Dujour, A., Bordal, G., & Truong, A. 2019b. Derivation of the prosodic structure. In *Rhapsodie* (pp. 213-231). John Benjamins.
- Lefeuve, F. à par. Analyse topologique des marqueurs interactionnels *tu sais* et *tu vois*. *Langages* (soumission).
- Obin, N., Beliao, J., Veaux, C., & Lacheret, A. 2014. SLAM: Automatic stylization and labelling of speech melody. In *Speech prosody* (p. 246).
- Waltereit, R. (2007). À propos de la genèse diachronique des combinaisons de marqueurs. L'exemple de *bon ben* et *enfin bref*. *Langue française*, (2), 94-109.
- Corpus : Gadet F. 2023. MPF [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage) - [www.ortolang.fr](http://www.ortolang.fr), v4.1, <https://hdl.handle.net/11403/mpf/v4.1>.

## The discourse markers *vallë* and *a thua* in Albanian

Bujar RUSHITI

AAB University – Prishtinë (Kosovo)

LLF (Laboratoire de linguistique formelle) – Université de Paris

Over the last decade or so, there has been intensive research on discourse markers used (nearly) exclusively in interrogatives (cf. Bayer (2012), Csipak&Zobel (2014), von Fintel&Iatridou (2019), Eckardt (2020), Farkas (2022), Miyagawa (2022)). This paper provides the first in-depth discussion and analysis of the discourse markers *vallë* and *a thua* in Albanian. They are used in interrogative sentences in order to express the speaker’s attitude of wondering towards the truth of the proposition.

- (1) (Vallë), kush është (vallë) fajtori (vallë)?  
(vallë), who is (vallë) guilty (vallë)  
“Who is the guilty person I wonder?”
- (2) A thua kush është fajtori?  
Q say.2sg who is guilty  
“I am wondering who is the guilty person?”
- (3) (Vallë), a thua (vallë) kush është (??vallë) fajtori \*(vallë)?  
(vallë), Q say.2sg (vallë) who is (??vallë) guilty \*(vallë)  
“I am wondering who is the guilty person?”

(1) shows the various positions *vallë* can occur with no detectable meaning difference; (2) shows that *a thua* occurs in initial position and expresses the same attitude of wondering as *vallë*; (3) shows that the two of them can co-occur and the canonical position of *vallë* is either immediately before or immediately after *a thua*. Trying to detach *vallë* from *a thua* renders it marked as indicated by (??*vallë*) or ungrammatical as indicated by (\**vallë*). *Vallë* differs from *a thua* in that it can be used in different positions in the sentence whereas *a thua* can only be used in initial position, even when it co-occurs with *vallë*.

We adopt a synchronic perspective and apply a qualitative corpus methodology. All the data will be drawn from the Albanian National Corpus available at: [albanian.web-corpora.net](http://albanian.web-corpora.net).

Beyond a detailed investigation of their semantic/pragmatic function potential of *vallë* and *a thua* as well as their position and structural status in discourse, our corpus research intends to answer the following question:

- (a) when *vallë* and *a thua* co-occur as in (3), do they have an additive or a compositional meaning?

The results of the analysis will be interpreted within the pragmatic framework of Traugott (2022, among many others). We argue that *vallë* and *a thua* can be accounted for in terms of the subjectification hypothesis of Traugott (2022, among others), according to which *vallë* and *a thua* express the speaker’s epistemic stance. As Traugott argues, discourse markers become increasingly associated with speaker’s attitude towards (the truth of) the proposition. Thus, the meaning of discourse markers is to express the speaker’s subjective belief, state or attitude towards what is being said (cf. speaker’s perspective point (Talmy (2000))); and speaker’s subjective view; (Espinal et al. (2022)). This process of semanticization, which consists of the speaker’s point of view, is known as subjectification and it corresponds exactly to the meaning of the discourse markers *vallë* and *a thua* in Albanian.

## REFERENCES

- Bayer, J. (2012). From modal particle to interrogative marker: A study of German *denn*. *Bibliothek der Universität Konstanz*, 13-28.
- Csipak, E., & Zobel, S. (2014). A condition on the distribution of discourse particles across types of questions. In *Proceedings of NELS (Vol. 44, No. 1)*.83-94.
- Eckardt, R. (2020). Conjectural questions: The case of German verb-final *wohl* questions. *Semantics and Pragmatics*, 13, 9-1.
- Espinal, M. T., Real-Puigdollers, C., & Villalba, X. (2022). From a movement verb to an epistemic discourse marker: The diachronic change of Spanish *vaya*. *Linguistic Variation*.
- Farkas, D. F. (2022). Non-intrusive questions as a special type of non-canonical questions. *Journal of Semantics*, 39(2), 295-337.
- von Fintel, K., & Iatridou, S. (2019). Unasked questions. A paper presented at *WAFI*, 15.
- Talmy, L. (2000). *Toward a cognitive semantics: Concept structuring systems* (Vol. 1). MIT press.
- Traugott, E. C. (2022). *Ten lectures on a diachronic constructionalist approach to discourse structuring markers* (Vol. 27). Brill.



**Langues & Langage  
à la croisée des Disciplines**

**1ère Rencontre annuelle LLcD**

Colloque : 9-11 septembre, 2024

Ecole d'été : 12-14 septembre, 2024

Sorbonne Université - 28 Rue Serpente, 75006 Paris, France



**Workshop 10**  
**Phonologie empirique**  
**Empirical Phonology**

## Les oppositions de durée vocalique en syllabe finale ouverte en français de Suisse romande

Marie-Hélène Côté & Isabelle Racine  
*Université de Lausanne & Université de Genève*

En français de référence, les oppositions historiques de durée vocalique ont soit disparu, soit été réanalysées en oppositions de timbre. Les oppositions de durée se maintiennent cependant dans certaines variétés régionales, notamment en Suisse romande. Ces oppositions peuvent être caractérisées selon le contexte (syllabe fermée ou ouverte) et la fonction (lexicale ou morphologique, selon que l'opposition sert à distinguer des lemmes différents ou des formes différentes relevant du même lemme). Le tableau suivant illustre chacune des combinaisons de contexte et de fonction en Suisse romande.

	Syllabe fermée	Syllabe ouverte
Morphologique	<i>général</i> /a/ vs <i>générale</i> /a:/	<i>ami</i> /i/ vs <i>amie</i> /i:/
Lexicale	<i>faites</i> /ɛ/ vs <i>fête</i> /ɛ:/	<i>bout</i> /u/ vs <i>boue</i> /u:/

Ces oppositions de durée peuvent par ailleurs être renforcées en surface par la diphtongaison (fermante ou centralisante) (Andreassen et al., 2010) et/ou la fricativisation. *Vie* peut ainsi être réalisé [vi:]~[vi:j]~[vi:ə]~[vi:ç].

Métral (1977) et Schoch (1980) montrent que les oppositions de durée s'observent dans l'ensemble de la Suisse romande, mais avec des variations régionales dans l'identité des voyelles concernées et le mode de réalisation des oppositions. Ces études ne s'appuient cependant que sur des données déclarées et n'incluent qu'une partie des oppositions potentielles. Plus récemment, Grosjean & al (2007), Racine & Andreassen (2012) et Guignard (2019) ont analysé des données de production, mais en se concentrant sur une région particulière (Neuchâtel ou la Vallée de Joux), Racine & Andreassen ajoutant certains éléments de comparaison avec Nyon. Côté, Prikhodkine & Racine (à paraître) montrent que le /e:/ de *journée* est exclusif à la Suisse et absent chez les Haut-Savoyards de France voisine, alors que la diphtongaison est mobilisée par les Genevois-es, mais non les Neuchâtelois-es.

Cette présentation vise à fournir une première analyse globale de la production des oppositions de durée vocalique en syllabe finale ouverte, en tenant compte de la variation régionale pour l'ensemble des voyelles et des différents modes de réalisation. Elle s'appuie sur les données du projet Phonologie du Français Contemporain (PFC; Durand, Laks & Lyche, 2002; Detey et al., 2016) recueillies en Suisse. Ces données proviennent de 87 locuteurs-trices, réparti-e-s entre hommes et femmes et trois générations, et originaires de 5 localités situées dans 4 cantons :

- Genève (canton de Genève) : 23
- Neuchâtel (canton de Neuchâtel) : 13
- Vallée de Joux (canton de Vaud) : 14
- Chavornay (canton de Vaud) : 21
- Martigny (canton du Valais) : 16

L'analyse se concentre sur deux listes de mots spécialement établies pour les enquêtes en Suisse romande, qui comprennent notamment des paires minimales impliquant une opposition de durée en syllabe finale ouverte pour l'ensemble des voyelles concernées : /i/, /e/, /ɛ/, /a/, /y/, /ø/ et /u/. Tous les mots et voyelles cibles (N=2088) ont été segmentés dans Praat, ce qui permet de procéder à des analyses acoustiques et statistiques des différences de durée (en valeur absolue et en proportion de la durée du mot), des variations formantiques indicatrices d'une diphtongaison et des portions fricativisées, en fonction de la voyelle et de la région.

**Mots clés** : Voyelles, Français, Suisse, Durée

## Références

- Andreassen, H. N., Maître, R., & Racine, I. (2010). Le français en Suisse: éléments de synthèse. In S. Detey, J. Durand, B. Laks, & C. Lyche (Eds.), *Les variétés du français parlé dans l'espace francophone. Ressources pour l'enseignement* (pp. 212–231). Paris: Ophrys.
- Côté, M.-H., Prikhodkine, A., & Racine, I. (À paraître). Negotiating language resources in border areas: production and discourse of cross-border workers in Greater Geneva. In Y. Bürki & A. N. Garcia Agüero (Eds.), *Challenging Borders, Challenging Categorisations: from/between borders*. Berlin: De Gruyter.
- Detey, S. Durand, J., Laks, B., & Lyche, C. (Eds.). (2016). *Varieties of Spoken French*. Oxford: Oxford University Press
- Durand, J., Laks, B., & Lyche, C. (2002). La phonologie du français contemporain: usages, variétés et structure. In C. Pusch & W. Raible (Eds.), *Romanistische Korpuslinguistik - Korpora und gesprochene Sprache/Romance Corpus Linguistics – Corpora and spoken language* (pp. 93–106). Tübingen: Gunter Narr Verlag.
- Guignard, P. (2019). « À la Vall[e:]! » *Oppositions de longueur vocalique en syllabe finale ouverte à la Vallée de Joux (Vaud, Suisse)*. Mémoire de maîtrise, Lausanne: Université de Lausanne.
- Métral, J.-P. (1977). Le vocalisme du français en Suisse romande. Considérations phonologiques, *Cahiers Ferdinand de Saussure*, 31, 145–176.
- Racine, I. & Andreassen, H. N. (2012). A phonological study of a Swiss French variety: data from the canton of Neuchâtel. In R. Gess, C. Lyche, & T. Meisenburg (éds), *Phonological Variation in French: Illustrations from three Continents* (pp. 173–207). Amsterdam: John Benjamins.
- Schoch, M. (1980). *Résultats d'une enquête phonologique en Suisse romande*. Bulletin de la Section de linguistique de la Faculté des lettres de Lausanne, 2.

## Variation et codages inter-juges : le rôle de la variété native du codeur sur la qualité du codage à l'épreuve des données

Marc Chaliier, [marc.chalier@sorbonne-universite.fr](mailto:marc.chalier@sorbonne-universite.fr)

Sorbonne Université

Linda Bäumlér, [linda.baeumlér@univie.ac.at](mailto:linda.baeumlér@univie.ac.at)

Université de Vienne

Elisabeth Heiszenberger, [elisabeth.heiszenberger@univie.ac.at](mailto:elisabeth.heiszenberger@univie.ac.at)

Université de Vienne

La recherche en phonologie et en sociolinguistique requiert souvent le codage de vastes corpus par plusieurs codeurs (cf. Milroy & Gordon 2003). Ce faisant, il est généralement admis que les codeurs devraient présenter des profils spécifiques. Ainsi, l'on aura recours à des locuteurs natifs de la langue à coder (cf. Racine & Detey 2015), à des locuteurs entraînés au codage utilisé, surtout pour un codage prosodique (cf. Durand, Laks & Lyche 2009), ou encore – dans les études variationnistes – à des locuteurs natifs d'une variété en particulier. Des études récentes remettent cependant ces choix en question : Hay & Drager (2010) et Sloos (2013) montrent que la validité des codages par des linguistes formés peut être sujette à des biais liés à l'accent perçu dans les stimuli. Mahmoud (2020) observe que des codeurs natifs et non natifs peuvent produire des codages similaires, alors que Hall-Lew & Fix (2012) notent une influence des caractéristiques externes des locuteurs codés (notamment leur origine ethnique) sur les résultats des codages.

Cette communication vise à évaluer la nécessité d'avoir recours à des natifs de la variété étudiée lors du codage phonologique. Nous analysons les réalisations des voyelles /*ǣ*/ et /*ē*/ dans un corpus du français québécois (cf. Chaliier 2021), comparant les codages de locuteurs natifs de cette variété à ceux de locuteurs natifs d'une autre variété, le français parisien. Notre corpus comprend 20 présentateurs de journaux radiophoniques et télévisés du français québécois dans des situations de lecture (un texte et deux listes de mots lus à haute voix), dans lequel nous avons fait coder 1620 réalisations du /*ǣ*/ et 1742 réalisations du /*ē*/. Ces deux voyelles sont généralement antériorisées en français québécois, mais pas systématiquement et à des degrés différents dans l'usage des présentateurs de journaux télévisés, raison pour laquelle le codage de la première voyelle contient trois variantes possibles ([*æ̃*]/[*ã*] ~ [*ẽ*] ~ [*ǣ*]/[*ǽ*]) et celui de la deuxième deux variantes ([*ē*] ~ [*ĕ*]). Le codage a été réalisé auditivement par deux locuteurs québécois et deux locuteurs parisiens, ainsi que par une phonéticienne selon des indicateurs acoustiques préétablis. Les résultats ont été comparés statistiquement, notamment à l'aide du coefficient Kappa de Cohen (cf. Fleiss, Levin & Paik 2003).

Les résultats montrent que les codages des codeurs natifs de la variété québécoise ne sont pas significativement plus précis que ceux des codeurs non-natifs de cette même variété, les variantes centrales du /*ǣ*/ menant notamment à de fortes hésitations dans les deux groupes de codeurs et à un accord inter-juges relativement faible. Si ces résultats ne sont pas généralisables à d'autres variables étant données les spécificités articulatoires et acoustiques des voyelles nasales, ils montrent cependant qu'un entraînement systématique au codage et un protocole strict quant aux indicateurs phonétiques à prendre en compte pourraient se révéler plus déterminants que le profil des codeurs, et ce, non pas seulement pour le codage de variables prosodiques (cf. Durand, Laks & Lyche 2009, Avanzi *et al.* 2010), mais également pour des variables segmentales.



## Références

- Avanzi M., Simon A. C., Goldman J.-P., Auchlin A. (2010). « C-PROM. Un corpus de français parlé annoté pour l'étude des proéminences », in : *XXVIIIèmes journées d'étude sur la parole (JEP 2010), Mons, 25-28 mai 2010*, 73-76.
- Chalier, M. (2021). *Les normes de prononciation du français. Une étude perceptive panfrancophone*. Berlin & Boston : De Gruyter.
- Durand, J., Laks, B. & Lyche, C. (2009). « Le projet PFC. Une source de données primaires structurées », in : Durand, J., Laks, B. & Lyche, C. (éds.) : *Phonologie, variation et accents du français*. Paris : Lavoisier/Hermès Science, 19-61.
- Fleiss, J. L., Levin, B. & Paik, M. C. (2003). *Statistical Methods for Rates and Proportions*. New-York : John Wiley & Sons.
- Hay, J., & Drager, K. (2010). « Stuffed toys and speech perception ». *Linguistics* 48.4, 865-892.
- Hall-Lew, L. & Fix, S. (2012). « Perceptual coding reliability of (L)-vocalization in casual speech data », *Lingua* 122.7, 794-809.
- Mahmoud, M. (2020). « Native vs. Nonnative Raters in Second Language Pronunciation Assessment of Guttural Sounds », *Open Journal of Modern Linguistics* 10, 468-479.
- Milroy, L. & Gordon, M. (2003). *Sociolinguistics: Method and Interpretation*. Oxford : Wiley-Blackwell.
- Racine, I. & Detey, S. (2015). « Corpus oraux, liaison et locuteurs non natifs : de la recherche en phonologie à l'enseignement du français langue étrangère », *Bulletin suisse de linguistique appliquée* 102, 1-25.
- Sloos, M. (2013). « Halo! Don't always trust the native speaker », in : Conlan, C. (éd.) : *Evolving paradigms: Language and applied linguistics in a changing world*. Perth : Applied Linguistics Association of Australia, 325-350.

## Interactions prosodie-phonétique en français L2

Dans cette présentation, nous discutons des effets de la structure prosodique dans l'acquisition des phonèmes /i/-/y/-/u/ et /e/-/ø/-/o/ en français langue étrangère. Dans la plupart des modèles en acquisition de la phonologie L2, tels que le (*Revised*) *Speech Learning Model* (Flege & Ocke-Schwen 2021), le *Perceptual Assimilation Model* (Best 1995) ou la *L2 Intonation Learning Theory* (Mennen 2015), l'acquisition de la structure sonore (consonnes et voyelles) et celle de la structure prosodique sont vues comme deux dimensions indépendantes. Cependant, les recherches en L1 (Cho 2011, Georgetown & Fougeron 2014, *inter alia*) examinant les interactions entre la structure prosodique et les événements phonétiques ont montré que les segments sont hyper-articulés selon la position hiérarchique où ces derniers sont produits (syllabes inaccentuées vs syllabes accentuées vs positions où les frontières droites des syntagmes intonatifs sont réalisées).

Selon la théorie du renforcement prosodique (*prosodic strengthening*), les segments produits dans des positions prosodiques fortes sont réalisés avec une intensification de leurs traits distinctifs (ils sont plus canoniques). Par exemple, la voyelle /y/ est articulée avec un renforcement du trait articulaire de labialisation (elle est plus arrondie) lorsqu'elle est produite en fin d'un énoncé interrogatif (*Quelqu'un l'a vu ?*) ou lorsqu'elle est accentuée (*Quelqu'un a vu mon stylo ?*) par rapport à une position prosodiquement faible comme c'est le cas d'une syllabe inaccentuée (*Quelqu'un a des allumettes ?*).

Ces interactions ont été largement sous-représentées dans les recherches en acquisition de L2 : la structure accentuelle a-t-elle des effets sur l'acquisition du trait /±arrondi/ pour les voyelles en FLE ? Ces voyelles sont-elles mieux réalisées lorsqu'elles sont produites à la frontière droite des syntagmes intonatifs chez les apprenants de FLE ?

L'objectif de cette présentation est triple. Premièrement, nous présentons les résultats de trois études analysant ~15k voyelles orales /±arrondies/ produites dans différentes positions prosodiques par 30 apprenants en FLE niveau intermédiaire (B1) ayant différentes L1 (espagnol, italien et anglais). Nous avons calculé différentes métriques acoustiques : distances Euclidiennes entre le centre acoustique et les segments analysés sur les plans F1/F2/F3, ainsi que le calcul du chevauchement acoustique via des scores *Pillai*. Ces métriques nous ont permis d'évaluer la catégorisation phonétique-acoustique des voyelles non natives. Notre objectif est d'examiner si les interactions entre la composante phonologique et la structure prosodique ont des effets positifs dans l'acquisition des voyelles qui sont souvent vues comme difficiles pour les apprenants. Nos résultats montrent que les unités positionnées dans une position haute de la hiérarchie prosodique ont des effets positifs pour la catégorisation de ces voyelles. Tout particulièrement, nous observons que les frontières mélodiques droites bornant les syntagmes intonatifs du type LH(H)% déclenchent une meilleure catégorisation des voyelles /±arrondies/ dans l'espace acoustique chez les apprenants que celles en position interne dans l'énoncé (LH\*). Nous discutons dans quelle mesure ces résultats peuvent avoir des implications théoriques dans les modèles formels en phonologie de L2. Enfin, nous proposons que les interactions prosodie-phonétique peuvent avoir également des implications pédagogiques pour améliorer la prononciation en L2.

### Références

- Best, C.T. (1995). A direct realist view of cross-language speech perception. In W. Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 171-204), Timonium, MD, York Press.

- Cho, T. (2011). The phonetics–prosody interface in laboratory phonology. In N. C. Kuma, B. Botma, & K. a Nasukawa (eds.), *The continuum companion to phonology* (pp. 243-368). London/New York: Continuum.
- Flege, J. E. & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In R. Wayland (ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3-83). Cambridge University Press.
- Georgeton, L & Fougeron, C. (2014). Domain-initial strengthening on French vowels and phonological contrasts: Evidence from lip articulation and spectral variation, *Journal of Phonetics* 44, 83–95.
- Mennen, I. (2015). Beyond segments: towards a L2 intonation learning theory (LILt). In E. Delais- Roussarie, M. Avanzi & S. (eds.), *Prosody and Language in Contact. Prosody, Phonology and Phonetics* (pp. 171-188), Springer, Berlin-Heidelberg.

## **Analyse exploratoire des corrélats rythmiques du français lu chez les apprenants chinois de FLE**

Cette étude explore le rythme du français chez les apprenants chinois de Français Langue Etrangère (FLE). Si le rythme du français a fait l'objet de nombreuses études, celui du chinois mandarin reste sujet à débat. La divergence des structures syllabiques entre le mandarin et le français est attestée : le français utilise principalement CV, CCV, CCCV, CVC et V, tandis que le mandarin privilégie V, VV, VVV, CV, CVV et CVVV (Pillot-Loiseau & Xie, 2018, p. 6-7). Grabe & Low (2002) considèrent le chinois mandarin et le français comme des langues syllabiques. Mais plusieurs recherches montrent des divergences sur lesquelles nous reviendrons dans cette étude. Par exemple, Mok & Dellow (2008) notent que bien que le cantonais et le mandarin de Pékin aient tous deux un rythme syllabique, le rythme de la parole naturelle en mandarin semble plus variable que celui du cantonais. D'autres études semblent remettre en cause les typologies rythmiques observées par certains. Pour notre part, nous postulons que les apprenants sinophones produisent, en FLE, un rythme plus proche de celui du mandarin que celui du français. Notre étude se base sur un échantillon de locuteurs composé de 22 apprenants chinois et 10 natifs français, sur une tâche de lecture. Les fichiers audio ont été étiquetés automatiquement en intervalles syllabiques, consonantiques et vocaliques avec WEBMAUS (Kisler et al., 2017). Ils ont été ensuite annotés manuellement avec Praat (Boersma & Weenink, 2013). Les mesures rythmiques ont été calculées avec CORRELATORE (Mairano, 2011). Nous avons retenu les mesures suivantes :  $\Delta V$ , %V (Ramus, 1999), VarcoC, VarcoV (Dellwo & Wagner, 2003), rPVI\_C, rPVI\_V, nPVI\_V, nPVI\_ (Grabe & Low, 2002). Cette étude dans laquelle nous présenterons les résultats préliminaires, devrait aider à mieux comprendre le rôle de la langue maternelle dans la gestion des corrélats rythmiques chez les apprenants sinophones de FLE.

### **Références :**

- Benton, M. A., Dockendorf, L., Jin, W., Liu, Y., & Edmondson, J. A. (2007). THE CONTINUUM OF SPEECH RHYTHM 1: COMPUTATIONAL TESTING OF SPEECH RHYTHM OF LARGE CORPORA FROM NATURAL CHINESE AND ENGLISH SPEECH.
- Boersma, P., Weenink, D. (2013). Praat: Doing phonetics by computer, logiciel téléchargé le 7 janvier 2013 du site Internet [<http://www.fon.hum.uva.nl/praat/>].
- Dellwo, V., & Wagner, P. (2003, janvier 1). Relationships between rhythm and speech rate. 15th International Congress of the Phonetic Sciences, Barcelona, pp. 471-474.

Grabe, E., & Low, E. (2002). Durational variability in speech and the Rhythm Class Hypothesis. In *Papers in Laboratory Phonology: Vol. Vol. 7* (p. 515-546). <https://doi.org/10.1515/9783110197105.2.515>

Mairano, P. (2011). RHYTHM TYPOLOGY: ACOUSTIC AND PERCEPTIVE STUDIES [Phdthesis, Università degli studi di Torino]. <https://theses.hal.science/tel-00654261>

Mok, P. (2009). On the syllable-timing of Cantonese and Beijing Mandarin. *Chinese Journal of Phonetics*, 148-154.

Mok, P. P. K., & Dellwo, V. (2008). Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. *Speech Prosody 2008*, 423-426. <https://doi.org/10.21437/SpeechProsody.2008-93>

Pillot-Loiseau, C., & Xie, H. (2018). Transfert rythmique du chinois mandarin au français dans l'apprentissage du Français Langue Étrangère : Acoustique et perception. *SHS Web of Conferences*, 46, 09001. <https://doi.org/10.1051/shsconf/20184609001>

Ramus, F. (1999). La discrimination des langues par la prosodie : Modélisation linguistique et études comportementales.

Thomas Kisler, Uwe D. Reichel, and Florian Schiel (2017): Multilingual processing of speech via web services, *Computer Speech & Language*, Volume 45, September 2017, pages 326–347.

## Prosodic aspects of discourse particles in L2 learners of French

Svenja Krieger (University of Konstanz)

Linguistic phenomena at the interfaces are often considered to be vulnerable to language influence (CLI) in research on second language (L2) acquisition, especially if pragmatics is involved (Sorace 2004). Many studies have shown such a vulnerability for the syntax-pragmatics interfaces (see e.g., Trotzke et al. 2020 on the positioning of discourse markers in L2 learners, i.e., words that are used on a discursive level instead of a referential or textual level, Bayer, 2009). Compared to this interface, the pragmatics-prosody interface, in particular the prosodic realization of discourse markers in L2 learners has received considerably less attention (e.g., Staples & Fernández 2019).

The present study aims at filling this gap by focusing on the pragmatic meaning of the French discourse marker *alors* ('so'), *bon* ('well'), and *donc* ('thus') as well as their prosodic properties. In French, the discourse marker *alors* is most often used to express introduction, additional information, and conclusion, *bon* usually expresses additional information, and *donc* additional information and conclusion (e.g., Lee et al. 2019). From a prosodic perspective, French discourse markers are mostly produced with medium/high F0 level and a plateau, especially if they function as an introduction marker (e.g., Lee et al. 2020). By contrast, English (as well as German) particles usually show a rising-high-falling F0 (Lee et al. 2020). This difference between repertoire in French and German make discourse markers potentially susceptible to CLI (see also Mennen's 2015 prediction of the L2 Intonational Learning Theory, LILt).

In this study, I examine: i) which pragmatic meanings *alors*, *bon*, and *donc* encode in the speech of L2 speakers and ii) which prosodic properties they use in their production of discourse markers. For this purpose, a study on the corpus HABLA ("Hamburg Adult Bilingual Language" Kupisch et al. 2012) is carried out. This corpus includes semi-structured interviews with 18 advanced German L2 speakers of French (mean age = 33 years, 13 f, 5m). The data was analyzed with respect to the pragmatic functions of the discourse particles, pauses that precede or follow the particles, and F0, using Praat (Boersma & Weenink 2015). For the statistical analyses, generalized linear mixed-effects regression models in R were run. Preliminary results of five L2 learners reveal that if the particles were used (*alors*, *bon*, and *donc*) with a pause before and after the particles, L2 learners mostly produced a falling movement of f0 ( $\beta=0.70$ ,  $SE=0.11$ ,  $z=6.42$ ,  $p<.001$ ), suggesting CLI from German (in line with the prediction by the systemic dimension of the LiLT theory, see also Pešková, 2023 for a

similar line of argument on German and Czech L2 learners of Italian and Spanish). Even though, the findings indicate that the prosody of L2 learners' discourse particles is vulnerable, they are able to use the discourse-complex structures of particles.

In the presentation, I will provide results for all L2 learners including a detailed analysis of the pragmatic functions and prosodic characteristics of the three discourse markers.

## References

- Bayer, J. (2009). Discourse particles in questions. In: Proceedings of GLOW Asia.
- Boersma, P., & Weenink, D. (2015). *Praat: doing phonetics by computer (version 6.1.05)*.
- Kupisch, T., Barton, D., Bianchi, G., & Stangen, I. (2012). The HABLA-Corpus (German-French and German-Italian). *Multilingual corpora and multilingual corpus analysis*, 163-179.
- Lee, L., Bartkova, K., Jouvét, D., Dargnat, M., & Keromnes, Y. (2019). Can prosody meet pragmatics? Case of discourse particles in French. In: *ICPhS 2019-International Congress of Phonetic Sciences*.
- Lee, L., Jouvét, D., Bartkova, K., Keromnes, Y., & Dargnat, M. (2020). Correlation between prosody and pragmatics: case study of discourse markers in French and English. In: *INTERSPEECH 2020*.
- Pešková, A. (2023). *L2 Spanish and Italian intonation: Accounting for the different patterns displayed by L1 Czech And German learners. (Open Romance Linguistics 5)*. Language Science Press.
- Sorace, A. (2004). Native language attrition and developmental instability at the syntax-discourse interface: Data, interpretations and methods. *Bilingualism: Language and cognition*, 7(2), 143-145.
- Staples, S., & Fernández, J. (2019). Corpus linguistics approaches to L2 pragmatics research. In *The Routledge handbook of second language acquisition and pragmatics* (pp. 241-254). Routledge.
- Trotzke, A., Bidese, E. & Moroni, M. C. (2020). German discourse particles in the second language classroom: Teasing apart learning problems at the syntax-pragmatics interface. *Pedagogical Linguistics* 1 (2), 184\_210.



**Langues & Langage  
à la croisée des Disciplines**

**1ère Rencontre annuelle LLcD**

Colloque : 9-11 septembre, 2024

Ecole d'été : 12-14 septembre, 2024

Sorbonne Université - 28 Rue Serpente, 75006 Paris, France



## **Workshop 11**

**Exprimer la quantité extrême : une kyrielle de  
variantes**

**Expressing extreme quantity: a myriad of  
variants**



# L'origine de l'élatif lituanien *daũg* 'beaucoup'

Daniel Petit (ENS & EPHE, Paris)

## Résumé

L'expression de l'élatif en lituanien passe par une forme *daũg* 'beaucoup', qui peut fonctionner comme adverbe, comme déterminant ou comme forme nominale (accompagnée du génitif). Au sein de la famille des langues baltes, à laquelle appartient le lituanien, cette forme n'est pas entièrement isolée, puisqu'elle trouve un correspondant en letton (*daũdz* 'beaucoup') et indirectement en vieux-prussien (anthroponyme *Daugis*), et elle paraît même disposer d'une certaine antiquité balto-slave, puisqu'il existe une forme apparentée en polonais (*dužo* 'beaucoup', *duży* 'grand, fort'), peut-être empruntée à un modèle slave oriental. Les autres langues baltes disposent de formes élatives différentes plus couramment utilisées, comme *loti*, *lioti* en letton ('très, beaucoup', emprunté au slave) et *tūlan*, *toūlan* en vieux-prussien.

L'origine des expressions élatives peut se fonder sur (1°) d'anciens adjectifs exprimant un haut degré ('grand, fort'), comme irlandais *mór* (< 'grand'), gotique *mikils* (< 'grand'), ou 'plein', comme grec πολύ *polú*, vieux-haut-allemand *filu* (d'une racine indo-européenne \**pelh<sub>1</sub>*- 'remplir'); (2°) des substantifs signifiant 'masse, multitude', comme irlandais *imde* (< 'masse'), breton *kalz* (< 'tas'), sanskrit *bhūri*- (< 'masse'); (3°) des racines verbales signifiant 'croître', 'gonfler', comme vieux-prussien *tūlan*, *toūlan* (d'une racine indo-européenne \**tū*- 'gonfler'); (4°) diverses expressions métaphoriques, comme « un beau coup » (français *beaucoup*).

L'origine de l'élatif lituanien *daũg* 'beaucoup' demeure obscure. La forme lituanienne et son correspondant letton *daũdz* sont tirés d'un substantif neutre \**daugi*, dont la préhistoire est inconnue. Il a été rattaché soit à une racine indo-européenne \**d<sup>h</sup>eug<sup>h</sup>*- 'fabriquer, produire' (grec τεύχω *teúk<sup>h</sup>ō* 'fabriquer', gotique *daug* 'être utile'), soit à une racine homonyme \**d<sup>h</sup>eug<sup>h</sup>*- 'être joyeux' (lituanien *džiaũgtis* 'se réjouir'), certains savants pensant même que les deux racines n'en font qu'une. L'idée qu'un adverbe 'beaucoup' puisse provenir d'une racine exprimant la 'joie' pourrait être appuyée par le parallèle du sanskrit *maháyati* qui signifie à la fois 'renforcer' et 'réjouir'.

Les différentes hypothèses étymologiques concernant l'élatif lituanien *daũg* 'beaucoup' seront présentées au cours de l'exposé et évaluées au regard de leur plausibilité typologique. Un point qui mérite d'être pris en considération est la frontière entre élatif et excessif, qui a pu connaître des évolutions remarquables dans l'histoire des langues baltes. On observe que l'excessif ('trop') est exprimé en lituanien par un préfixe *pér-*, par ex. *pérdidelis* 'trop grand' (: *didelis* 'grand'). Si l'on tire ce préfixe d'un ancien sens élatif, comme le suppose le parallèle du latin *perclārus* 'très clair' (: *clārus* 'clair'), on peut admettre que la forme \**daugi* constitue un remplacement lexical pour le sens élatif à la place du préfixe *pér-* lui-même passé au sens excessif. Par ailleurs, il est connu que l'élatif peut être exprimé en lituanien par une forme *labaĩ* 'très' (adverbe de l'adjectif *lābas* 'bon'), avec la différence que *daũg* est quantitatif ('beaucoup'), tandis que *labaĩ* est qualitatif ('très'). Alors que le préfixe *pér-* ne distinguait pas entre sens quantitatif et sens qualitatif, la langue a établi cette

distinction par des éléments de nature hétérogène, le substantif *daũg* et l'adverbe *labaĩ*, dans des conditions qu'il conviendra de clarifier. Les langues baltiques pourront ainsi apporter une contribution intéressante à la typologie des élatifs dans les langues indo-européennes et de leur évolution diachronique.

Conférence LLcD 2024

Workshop 11 « Exprimer la quantité extrême : une kyrielle de variantes »  
Proposition de communication

« Expression de la grande quantité en moyen-haut-allemand et en vieux-russe :  
approche sémantique et pragmatique »

L'angle d'attaque de cette analyse est onomasiologique, contrastif et historique : il s'agit d'examiner les signes linguistiques mis en œuvre pour exprimer le grand nombre dans deux états anciens de langues indo-européennes issues de sous-familles différentes, le vieux-russe et le moyen-haut-allemand. Cette analyse contrastive en synchronie historique portera sur deux traditions textuelles bien représentées dans ces langues, à la charnière entre le XIIe et le XIIIe s., à savoir la chronique et l'épopée ; afin de définir un *tertium comparationis* au niveau onomasiologique, on s'intéressera plus précisément aux récits de batailles dans ces textes (qu'il s'agisse du grand nombre des guerriers, des chevaux, des armures, des victimes, des otages, des prisonniers, des veuves, soit d'êtres animés ou inanimés, humains ou non humains).

A partir d'un état des lieux morphosyntaxique des constructions attestées dans chaque langue pour exprimer la quantité, on se demandera si des spécificités morphosyntaxiques sont observables concernant l'expression de la très grande quantité, et si ces spécificités offrent des points de comparaison entre les deux langues. Au-delà des systèmes linguistiques, il s'agira ensuite, dans une approche sémantique et pragmatique, de déterminer les préférences des locuteurs, selon les traditions textuelles et les contextes d'emploi, pour désigner le grand nombre : les désignations collectives au singulier sont-elles plus fréquentes que l'emploi du pluriel indéfini, comme le montre Desportes (2000) pour l'allemand ancien (sans privilégier la quantité extrême) ? Dans les chaînes anaphoriques, une désignation collective peut-elle être reprise par une désignation individualisante au pluriel ? Observe-t-on des divergences d'emploi selon le caractère animé ou humain des êtres quantifiés ? Outre l'opposition entre le choix du singulier et celui du pluriel, dont on examinera les effets de sens en contexte, on s'intéressera aux métonymies et aux métaphores attestées pour exprimer le grand nombre, et, au niveau de l'énoncé, aux marqueurs de modalisation qui déterminent fréquemment ces désignations, notamment lorsqu'elles sont numérales, ainsi qu'aux commentaires suscités par l'expression de la quantité extrême (comme les hyperboles, ou bien les *topoi* de l'incomptable et/ou de l'inénarrable). L'approche contrastive devra évaluer si les ressemblances observées entre les pratiques désignatives dans ces deux langues s'expliquent par les traditions textuelles communes, au-delà des divergences entre les systèmes linguistiques correspondants.

Cette étude de corpus portera d'une part sur deux chroniques, la *Chronique des temps passés* (dite du moine Nestor, XIIe s.), et la *Chronique rimée de Livonie* (XIIIe s., témoin des premiers contacts entre locuteurs germanophones, princes de la Rus' et Lituaniens) ; d'autre part sur des textes épiques, l'épopée *Le Dit de l'ost d'Igor* (anonyme, XIIe s.), la *Chanson des Nibelungen* (anonyme, début XIIIe s.) et le *Roman d'Eneas* (Heinrich von Veldeke, fin XIIe s.). Au niveau méthodologique, les relevés seront réalisés à l'aide de corpus électroniques (par exemple le corpus TITUS pour le moyen-haut-allemand et le Corpus national de la langue russe pour le vieux-russe), mais également manuellement, pour des sondages qualitatifs.

Cette étude pilote est destinée à être étendue aux autres vernaculaires européens de cette époque, et notamment aux langues slaves occidentales et méridionales, ainsi qu'aux langues scandinaves, qui ont selon Haspelmath (2001 : 1505) de nombreux traits structurels communs. Au-delà de l'origine indo-européenne commune, il s'agira d'évaluer si les traditions textuelles

communes à ces langues (et notamment celle des sagas) ont favorisé des contacts linguistiques anciens, qui pourraient être traités en linguistique typologique aréale dans le cadre des aires de convergence linguistique.

#### Bibliographie indicative

- AKIMOVA, Tatiana G. & all., 1996, *Teorija funkcional'noj grammatiki. Kačestvennost'. Količestvennost'*, Saint-Pétersbourg : Nauka.
- BOSVELD-DE SMET, Léonie, 2001, « Le pluriel et le massif : une paire unique », in D. Amiot, W. de Mulder, N. Flaux (dir.), *Le Syntagme nominal : syntaxe et sémantique*, Arras : Artois, 147-161.
- DESPORTES, Yvon, 2000, « Der Artikel im Mittelhochdeutschen. Lässt sich Paul Valentins Modell des Artikelsystems im heutigen Deutsch auf das Mittelhochdeutsche übertragen? », in Y. Desportes (Hrsg.), *Zur Geschichte der Nominalgruppe im älteren Deutsch*, Festschrift für Paul Valentin, Germanistische Bibliothek 5, Heidelberg: Winter, 213-253.
- GAJLOMAZOVA, Elena S., 2011, *Grammatičeskaja kvanifikacija ob'ektov i faktov: lingvopragmatičeskij aspekt*, Rostov sur le Don: Akademlit.
- GARDELLE, Laure, 2023, « Lions, flowers and the Romans: exception management with generic and other count plurals », in *Reference : from Conventions to Pragmatics*, Studies in Language Companion Series, Amsterdam : John Benjamins, 71-87.
- HASPELMATH, Martin, 2001, « The European linguistic area: Standard Average European », in M. Haspelmath, E. König, W. Oesterreicher, W. Raible (eds.), *Language Typology and Language Universals*, vol. 2, Berlin : De Gruyter, 1492-1510.
- KLEIBER, Georges, 2014, *Les noms à la croisée du massif et du comptable*, Langue Française n°183, Paris : Armand Colin.
- LAMMERT, Marie, 2014, « Référence collective massive versus référence plurielle indéfinie », in G. Kleiber, 2014, 87-99.
- NIKOLAEVA, Tat'jana M., 1997, «*Slovo o polku Igoreve*». *Poètika i lingvistika teksta*, Moscou : izdatel'svto Indrik.

#### Corpus électroniques

- Nacional'nyj korpus russkogo jazyka, drevnerusskij podkorpus, 2003-2024, [ruscorpora.ru](http://ruscorpora.ru)
- TITUS, Thesaurus Indogermanischer Text- und Sprachmaterialien, Thesaurus Indogermanischer Text- und Sprachmaterialien, Johann Wolfgang Goethe Universität, Frankfurt am Main, [titus.uni-frankfurt.de](http://titus.uni-frankfurt.de)

## Une syntaxe saisie en pleine transition : la syntaxe des numéraux de 1 à 5 en russe littéraire moderne entre diachronie et synchronie.

L'analyse d'un syntagme au nominatif apparemment aussi anodin que

эти два больших стола  
*éti dva bol'six stola* « Ces deux grandes tables »

se révèle particulièrement incohérente — et déroutante pour qui apprend le russe. La grammaire du russe moderne fait en effet du cardinal au nominatif masculin *dva* (« deux ») la tête du syntagme numéral ; elle présente le dénombré *stola* (« tables », masculin en russe) comme un génitif singulier régi par *dva*, mais avec lequel *dva* s'accorde en genre (ici accord au masculin-neutre). Il semblerait que le démonstratif et l'adjectif dussent logiquement s'accorder avec le substantif dénombré présenté comme un génitif singulier. Or il n'en est rien : l'adjectif *bol'six* (« grandes »), intercalé entre le cardinal *dva* et le substantif masculin dénombré *stola*, est certes au génitif, mais au génitif *pluriel*, alors que le démonstratif *éti* (« ces »), que l'usage place *avant* le numéral cardinal *dva*, est au *nominatif* pluriel. En synchronie, tout semble donc n'être qu'incohérence : mélange des cas (nominatif et génitif dans un syntagme numéral qui lui-même est au nominatif), mélange des nombres avec un génitif adnuméral réputé *singulier* alors que dans notre exemple l'adjectif et le démonstratif, quel que soit leur cas, sont eux au pluriel.

La littérature sur la description synchronique de la syntaxe des numéraux en russe moderne est pléthorique (Mel'čuk 1985, Babby 1987, Rappaport 2002, entre autres). S'astreignant à une description strictement synchronique, elle s'oblige le plus souvent à bien des circonvolutions pour décrire des phénomènes pourtant limpides dès que l'on sort de la synchronie. Notre propos sera de mettre en lumière l'articulation trop souvent oubliée entre diachronie et synchronie en montrant que ces incohérences synchroniques ne sont que l'illustration d'un cas typique où, à un moment donné de l'histoire de la langue, tout un pan du système (ici la syntaxe des numéraux) est surpris en pleine transition.

### En slave commun et en vieux russe

1 *Odin* (*edin* en vieux russe littéraire « un ») est adjectif. Il s'accorde en nombre (singulier), en cas et plus tard en sous-genre avec le substantif dénombré et les éventuels adjectifs épithètes qui qualifient le dénombré. Pour *odin*, la situation perdure, inchangée, jusqu'en russe moderne.

2 *Dva* (« deux », masculin) et *dvě* (« deux », féminin et neutre) est lui aussi adjectif : il s'agit d'une forme duelle qui s'accorde en genre, cas et nombre (duel) avec le substantif dénombré, tête de syntagme numéral, avec lequel s'accordent aussi, également au duel, les éventuels adjectifs et déterminants.

*Trie* (« trois », masculin), *tri* (« trois », féminin et neutre) ainsi que *četyre* (« quatre », masculin), *četyri* (« quatre », féminin et neutre) se comportent également comme des adjectifs et ils s'accordent en genre, en cas et en nombre — cette fois bien sûr au pluriel — avec le substantif dénombré, là encore tête de syntagme, avec lequel s'accordent également les adjectifs, démonstratifs et possessifs éventuels.

Les choses changent à partir de cinq : 5 (*pjat*), 6 (*šest*), etc., substantifs de la troisième déclinaison en *-i* à paradigme accentuel particulier. Ces cardinaux sont la tête du syntagme numéral, le substantif dénombré n'est que leur complément de nom et se met au génitif pluriel en conservant normalement ce cas quel que soit le cas du cardinal tête de syntagme. Les adjectifs s'accordent au génitif pluriel avec le substantif dénombré. Le syntagme numéral, à partir de 5, a donc en vieux russe une structure équivalente à celle du français « un million de nouveaux francs ».

Le système cohérent du vieux russe s'est désagrégé sous l'effet d'un double mouvement : pour les masculins, l'emploi de la forme du nominatif-vocatif-accusatif duel, homonyme — pour la majorité des substantifs masculins — de celle du génitif singulier, s'est étendue aux nombres 3 et 4. Inversement, pour les féminins, c'est la forme du pluriel utilisée pour 3 et 4 qui s'est étendue à 2 et a remplacé celle du duel. Or, pour la plupart des substantifs féminins, les formes du nominatif pluriel sont homonymes des formes du génitif singulier (à l'accent près pour de nombreux substantifs). Du fait de ces différents syncrétismes, dans un syntagme numéral lui-même au nominatif, la forme adnumérale du substantif a été réinterprétée en synchronie comme une forme de génitif singulier. Les neutres, dont le nominatif pluriel est le plus souvent homonyme lui aussi du génitif singulier (à l'accent près dans une grande quantité de substantifs) ont suivi le chemin des féminins, étendant l'usage de leur nominatif pluriel en <.a> au nombre 2. Mais comme le morphe <.a> du nominatif pluriel et du génitif singulier neutre est à l'écrit identique à l'ancien morphe du nominatif duel masculin réinterprété lui-même comme un génitif singulier dans le syntagme numéral, les neutres, ayant après 2 la même désinence <.a> que les masculins, ont semblé s'aligner sur ces derniers, à tel point que la frontière entre les genres s'en est trouvée modifiée. En vieux russe, 2 était *dva* pour les seuls masculins et *dvě* pour les masculins et les neutres, alors qu'en russe moderne *dva* est utilisé pour les masculins et les neutres et *dve* pour les seuls féminins. Les autres distinctions de genre qui existaient pour 3 et 4, qui opposaient également en vieux russe une forme uniquement masculine à une forme féminine et neutre, ont été abolies (pour 3 c'est l'ancienne forme du féminin-neutre qui s'emploie pour tous les genres, alors que pour 4 c'est au contraire celle du masculin qui s'est imposée aux trois genres).

**En russe moderne** la forme du substantif dénombré que l'on a après 2, 3 et 4 dans un syntagme numéral lui-même au nominatif est présentée comme un génitif singulier par la grammaire traditionnelle et scolaire. L'appellation « génitif adnuméral », parfois employée, a au moins le mérite de souligner le caractère particulier de ce « génitif » dont les propriétés prosodiques le distinguent effectivement dans certains cas d'un vrai génitif. Mais il est plus juste de voir dans ces formes adnumérales un nombre particulier, en l'occurrence un nominatif paucal ou adnuméral, c'est-à-dire *un duel étendu* à 3 et 4.

En russe moderne, lorsque l'ensemble du syntagme numéral est à un cas oblique, la tête est désormais incontestablement le substantif dénombré, qui se met au pluriel et avec lequel s'accordent tous les autres éléments : du numéral aux adjectifs et déterminants éventuels.

À partir de 5, la structure du vieux russe (du type « un million de francs ») ne se maintient donc qu'au nominatif. Aux cas obliques, la situation est la même qu'avec 2, 3 et 4 : le substantif dénombré désormais tête de syntagme se met au cas voulu par la fonction du syntagme numéral dans la proposition ; le numéral devient adjectif et se met

au même cas que le dénombré, avec lequel s'accordent, au pluriel, adjectifs et déterminants<sup>1</sup>.

Le cas le plus étrange en apparence est celui de l'emploi du génitif *pluriel* de l'adjectif lorsque ce dernier qualifie un nom dénombré par 2, 3 ou 4 réputé être au « génitif singulier » dans un syntagme numéral lui-même au nominatif (cf. exemple ci-dessus *dva bol'six stola*). Nous lui avons consacré une étude particulière (Azam 2011) dans laquelle nous avons montré que sous sa forme actuelle, un énoncé comme « deux grandes tables » était une réorganisation de l'énoncé « des grandes tables [Gén. pluriel] il y [en] a deux tables » avec suppression des éléments redondants et modification de l'ordre des mots pour des raisons pragmatiques

Le système vieux-russe hérité du slave commun était parfaitement logique. Mais le système actuel qui émerge sur ses ruines n'a pas encore totalement acquis sa cohérence propre. Le syntagme numéral est donc l'un des domaines du russe contemporain où s'observe le mieux l'affleurement — au sens quasi-géologique — de la diachronie en synchronie. Seul un détour par la description diachronique permet en effet rendre compte de ce qui, sans elle, ne demeurerait qu'une somme d'aberrations arbitraires.

- O. AZAM, « Structure métainformative de l'énoncé et choix de la forme de l'adjectif qualificatif après dva (dve), tri et četyre en russe littéraire classique et contemporain », *Revue des Études Slaves* Année 2011 82-4 pp. 691-707.
- O. AZAM, « L'accord du syntagme numéral sujet et du nexus (ou prédicat) en russe littéraire moderne : les critères du choix et leur hiérarchisation », colloque international The 6th Meeting of the Slavic Linguistics Society, SLS 2011, Université de Provence, 1-3 septembre 2011.
- L. BABBY, « Case, prequantifiers and discontinuous agreement », *Natural Language And Linguistic Theory*, Vol. 5, No. 1 (Feb., 1987), pp. 91-138.
- Istoricheskaja grammatika drevnerusskogo jazyka / Pod red. V. B. KRYS'KO. T. IV: O. F. ŽOLOBOV. Čislitel'nye. M.: Azbukovnik, 2006. 360 c. [Grammaire historique du vieux russe. Sous la réd. de V.B. KRYSKO et O.F. ŽOLOBOV. Tome IV. Les numéraux, Moscou, 2006.]*
- I. A. MEL'ČUK, « Poverxnostnyj sintaksis russkix čislovyx vyraženij », *Gesellschaft zur Förderung Slawistischer Studien*, Vienne, 1985.
- G. RAPPAPORT, Numeral phrases in Russian : A minimalist approach, *Journal of Slavic Linguistics*, 2002, Vol. 10, No. 1/2, pp. 327-340.
- Русская грамматика / Н. Ю. Шведова (гл. ред.). — Т. 1: Фонетика. Фонология. Ударение. Интонация. Словообразование. Морфология. — М.: Наука, 1980. [Grammaire russe de l'Académie des Sciences, éd. 1980]*
- Русская грамматика / Н. Ю. Шведова (гл. ред.). — Т. 2: Синтаксис. М.: Наука, 1980. [Grammaire russe de l'Académie des Sciences, éd. 1980]*

---

<sup>1</sup> Sauf pour certains numéraux empruntés comme *million*, *milliard* et, dans certains cas, le russe *tysjača* « millier », qui conservent leur statut de purs substantifs et la construction du type un « million de nouveaux francs » à tous les cas. Mais il s'agit là de grands nombres qui dépassent le cadre de notre propos.

## Expression de la quantité extrême dans les langues finno-ougriennes (hongrois, mordve erzya)

Krisztina Hevér-Joly

Sorbonne Université, Faculté des Lettres

Dans la langue hongroise on identifie traditionnellement le nom collectif par un groupe d'êtres (vivants ou inanimés). Le nom collectif est également utilisé dans la description grammaticale pour désigner plusieurs groupes de noms. D'une part, il est utilisé pour les noms dont le sens inclut la notion de majorité (*nép* « un peuple », *csapat* « un groupe, une troupe », *tömeg* « une foule », *erdő* « la forêt »). Ces noms peuvent dénoter plus d'une chose, mais ils nécessitent un prédicat au singulier (*A nép elégedetlen.* « Le peuple est mécontent. » *A tömeg tüntet.* « La foule manifeste. »). Le terme nom collectif s'applique également aux mots formés, dont le nom, dans sa forme, indique déjà le groupe d'individus identiques. Tels sont les noms formés par le suffixe *-s*. (ho. *cserje* < *cserjés* « une broussaille », *kukorica* < *kukoricás* « champ de maïs », *fenyő* < *fenyves* « une pinède », *tölgy* < *tölgyes* « chênaie »). En langue mordve erzya les mêmes noms collectifs sont les mots composés (*kurakšpulo* « une broussaille » : *kurakš* « buisson » et *pulo* « un endroit boisé, bosquet, forêt », *pičepulo* « une pinède » dont *piče* « pin ». Les langues finno-ougriennes présentent donc les similitudes et les différences dans l'expression de quantité.

Dans une approche contrastive comme champ d'analyse on propose de faire état de la variété morphologique que revêt l'expression de la quantité extrême dans les langues finno-ougriennes (hongrois et mordve erzya). Selon notre hypothèse, l'expression de quantité extrême varie en fonction des époques mais montre les similitudes sémantico-conceptuelles entre les langues de mêmes familles de langues. Notre corpus hongrois et mordve erzya est composé des contes folkloriques et de textes journalistiques analysé par le paradigme du T.A.L (Traitement Automatique des Langues).

Nous allons étudier par exemple la notion de quantité exprimée par les adjectifs, suivi de l'examen de l'échelle de quantification (ho. *nevetséges fizetés* « un salaire ridicule », *ésszerű (normális) fizetés* « un salaire raisonnable », *szép fizetés* un beau salaire, *csillagászati (mértékű) jövedelem* « revenu astronomique ». Dans les contes hongrois on trouve : *szép pénz* « beaucoup d'argents ». Nous allons montrer également les différences dans les expressions de quantité : ho. *nagy vendégsereg* « une grande foule d'invités », mdE *lamo inžt'* « beaucoup d'invités ».

Cette communication a pour objectif ultime d'associer la typologie morphologique et la sémantique conceptuelle. L'analyse permet de montrer le contraste qui existe entre les concepts sémantiques et les formes grammaticales dans les corpus et dans les langues différentes d'une même famille de langue.

### Bibliographie :

- Keresztes, László. 1990. *Chrestomathia Morduinica*. Budapest : Tankönyvkiadó.  
Keszler, Borbála (éd.). 2000. *Magyar grammatika* [Grammaire du hongrois]. Budapest : Nemzeti Tankönyvkiadó.  
Lammert, M. (2014). Référence collective massive vs référence plurielle indéfinie. *Langue française*, 183, 87-99. <https://doi.org/10.3917/lf.183.0087>

Mots de clés : approche contrastive, hongrois, erzya, sémantique



## How to express extreme quantity through Italian adverbial constructions: a proposal study

Martina Verdelli  
University of Bergamo

### Abstract

The aim of this work is to explore the linguistic expression of extreme large quantity by focusing on some Italian constructions characterised by semi-fixed syntactic patterns and a constructional semantics (cf. Fillmore et al., 1998; Hilpert, 2014; Mellado Blanco, 2020; for Italian, Piunno, 2018; Casadei, 2024; Piunno & Pisano, 2024). More specifically, the expressions under investigation are represented by adverbial constructions (AdvCs) with the pattern [a + X<sub>NOUN <plural></sub>] as shown in the following examples:

AdvCs meaning ‘a lot’	Examples
a badilate (lit. ‘at shovelfuls’)	mangio lenticchie <i>a badilate</i> ‘I eat a large quantity of lentils’
a fiumi (lit. ‘at rivers’)	Il sangue scorreva <i>a fiumi</i> ‘The blood was flowing in large quantity’
a mucchi (lit. ‘at piles’)	Per le strade ci sono cadaveri, <i>a mucchi</i> ‘There are large quantities of corpses lying on the streets’
a ondate (lit. ‘at waves’)	I guai arrivano sempre <i>a ondate</i> ‘Troubles always come in large quantities’
a pacchi (lit. ‘at packages’)	proprio io che dovrei saperne <i>a pacchi</i> ‘me, of all people, who should know a lot about it’
a palate (lit. ‘at spadefuls’)	facciamo soldi <i>a palate</i> ‘we make a large quantity of money’
a valanghe (lit. ‘at avalanches’)	le firme che stanno arrivando <i>a valanghe</i> ‘signatures that are arriving in large quantities’
a vagonate (lit. ‘at wagonloads’)	ne possono vendere <i>a vagonate</i> ‘they can sell a large quantity of them’

Table 1. Examples of AdvCs from itTenTen20

The AdvC given above shows a semi-rigid pattern composed by a fixed lexical element (the preposition *a*) followed by an empty slot that must necessarily be filled by a plural noun, and whose referent is, in most cases, a container or its content (cf. Piunno, 2018: 197). From a semantic point of view, they represent partially lexicalized structures conveying idiomatic meanings related to abundance, whose pattern can be represented as: [V<sub>PROCESS</sub> + [a + N<sub>CONTAINER plural</sub>]]. However, beyond sharing the expression of high quantity, these constructions can also be considered evaluative strategies used to convey the semantic value of intensification (cf. Bolinger, 1972; Anscombe & Tamba, 2013; Napoli & Ravetto, 2017; Rainer, 2015; Piunno, 2023; Micheli, forthcoming).

The goal of my work is twofold. Firstly, to propose a preliminary study of Italian AdvCs concerning quantity intensification. The study will be driven by data extracted from a corpus of contemporary Italian by using the *SketchEngine interface* (Kilgarriff et al., 2014). Each construction will be analyzed both on the syntactic and the semantic level. Secondly, I will focus on the multiword [a + fiumi]. Although this pattern acts as a quantitative modifier, it also expresses an iterative value linked to the repetition of the action expressed.

In particular, the work will address the following research questions:

- a. What kind of nouns can fill the slot of the AdvCs? Are there any semantic differences involved by noun changes?
- b. Which is the relation between the expressions of quantity and quality in the usage of the AdvCs extracted? Is it always possible to draw a line between quantity and quality?

## References

- Anscombe, J.-Cl. & Tamba, I. (2013). Autour du concept d'intensification. *Langue française* 177(1), 3-8. <https://doi.org/10.3917/lf.177.0003>
- Bolinger, Dwight. (1972). *Degree Words*. De Gruyter.
- Casadei, Federica. (2024). La costruzione fraseologica italiana [a tutto/a N]. In F. Casadei, & S. Kösters Gensini (Eds.), *I fraseologismi schematici. Questioni descrittive e teoriche* (pp. 83-113). Aracne.
- Fillmore, C. J., Kay, P. & O'Connor, M. C. (1998). Regularity and idiomatity in grammatical constructions: The case of "let alone". *Language*, 64(3), 501-538.
- Hilpert, Martin. (2014). *Construction Grammar and its application to English*. Edinburgh University Press
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. & Suchomel, V. (2014). The Sketch Engine: Ten Years On. *Lexicography*, 1(1), 7-36.
- Mellado Blanco, Carmen. (2020). Esquemas fraseológicos y construcciones fraseológicas en el continuum léxico-gramática. In E. Tabares, C. Sinner, E. T. Montoro (Eds.), *Clases y categorías en la fraseología de la lengua española* (pp. 13-36). Peter Lang.
- Micheli, M. Silvia. (forthcoming). Evaluative morphology. In Ledgeway, A. & Maiden, M. (Eds.), *The Oxford handbook of Italian language*. Oxford University Press.
- Napoli, M. & Ravetto, S. (Eds.) (2017). *Exploring intensification. Synchronic, diachronic and cross-linguistic perspectives*. John Benjamins.
- Piunno, Valentina. (2018). *Sintagmi preposizionali con funzione aggettivale e avverbiale*. LINCOM.
- Piunno, Valentina. (2023). Intensifying constructions in Italian. Types, function and representation. In J-P Colson (Ed.), *Phraseology, constructions and translation. Corpus-based, computational and cultural aspects* (pp. 399-408). Presses universitaires de Louvain.
- Piunno, V. & Pisano, S. (2024). Costruzioni a schema fisso in alcune varietà diatopiche d'Italia. Prospettive teoriche e lessicografiche. *Linguistik Online* 125(1), 213-233. <https://doi.org/10.13092/lo.125.10793>

Rainer, Franz. (2015). Intensification. In P. O. Müller, I. Ohnheiser, S. Olsen, F. Rainer (Eds.), *Word-formation: An international handbook of the languages of Europe* (pp. 1339-1351). De Gruyter Mouton. <https://doi.org/10.1515/9783110246278-033>

**Keywords**

extreme quantity, partially lexically specified constructions, intensification, Italian, multiword expressions

« Il y en avait tant, tant, tant ! » -

### Comment les enfants acquièrent-ils l'expression de la quantité extrême ?

Dans la recherche sur le développement du langage chez l'enfant, l'acquisition d'expressions quantitatives a déjà fait l'objet de plusieurs études (cf. entre autres Barner et al. 2009, Hurewitz et al. 2006, Bloom/Wynn 1997). L'acquisition de mécanismes d'intensification est également étudiée depuis peu (cf. entre autres Hendriks et al. 2024, Hendriks/Van Goethem 2021, Bonacchi 2017), en mettant toutefois principalement l'accent sur les différences entre l'acquisition de la langue première (L1) et celle de la langue seconde (L2).

En revanche, les études qui intègrent les deux domaines de recherche et qui s'intéressent donc spécifiquement à l'acquisition d'expressions quantitatives intensifiées sont beaucoup plus rares. La communication proposée ici a pour objectif de contribuer à combler cette lacune de recherche et d'étudier le développement de l'expression de la quantité extrême chez des enfants germanophones, francophones et italophones (L1). Pour ce faire, elle utilise en priorité des données tirées de la base de données CHILDES, mais aussi d'autres corpus tels que PLPNat, le *Saarbücker Korpus*, etc.

La communication part des conclusions des travaux de recherche effectués jusqu'à présent, selon lesquelles il existe des préférences propres à chaque langue en ce qui concerne l'intensification linguistique qui sont repérables dès le début de l'acquisition. Ainsi, dans les langues germaniques, ce sont surtout les structures d'intensification morphologiques, par exemple par des préfixes comme *sau*-ADJ ou *mords*-ADJ en allemand, qui sont privilégiées, alors que dans les langues romanes, ce sont les structures d'intensification syntaxiques, réalisées par exemple à l'aide des adverbes comme *molto* + ADJ en italien ou *très* + ADJ en allemand ou bien à l'aide des reduplications (cf. Hendriks et al. 2017) qui sont utilisées plus fréquemment. D'après Hendriks et al. 2024, on trouve en outre des intensifications plus fréquentes mais plus faibles dans les données linguistiques des enfants germanophones, tandis que les intensifications sont plus rares mais plus fortes dans les langues romanes.

La contribution proposée ici vise donc à vérifier si ces résultats restent valables dans le domaine de l'intensification des expressions quantitatives. Une attention particulière sera accordée à la comparaison des différentes phases de développement dans l'acquisition du langage chez l'enfant et au rôle du soutien gestuel. Il est supposé que (1) les différences énumérées existent également dans ce domaine entre l'allemand et les deux langues romanes étudiées, mais que (2) le français et l'italien se distinguent considérablement l'un de l'autre. Sur la base des premiers examens de corpus, on a pu constater une tendance légèrement plus forte à

l'intensification morphologique en italien. On s'attend également à ce que (3) le soutien gestuel joue un rôle important, en particulier chez les enfants à un stade précoce de l'acquisition du langage, et qu'il soit plus largement utilisé que d'autres mécanismes d'intensification pour exprimer des quantités extrêmes.

#### Bibliographie

- Barner, David, Katherine Chow & Shu-Yu Yang. 2009. Finding one's meaning: A test of the relation between quantifiers and integers in language development. *Cognitive Psychology* 58, 195-219.
- Bloom, Paul & Karen Wynn. 1997. Linguistic cues in the acquisition of number words. *Journal of Child Language* 24, 511-533.
- Bonacchi, Silvia. 2017. What does reduplication intensify? The semantics and pragmatics of reduplicated forms in Italian and their equivalents in German. Maria Napoli & Miriam Ravetto (éd.), *Exploring Intensification: Synchronic, diachronic and cross-linguistic perspectives*, Amsterdam / New York, 289-304.
- Hendriks, Isa, Kristel Van Goethem & Natacha Buntinx. 2024. Intensification strength in speech: Language-specific preferences and differences between first and additional language. *Dutch Journal of Applied Linguistics* 13 (2024).
- Hendriks, Isa & Kristel Van Goethem. 2021. Intensifying constructions in second language acquisition: a diasystematic-constructionist approach. Hans C. Boas & Steffen Höder (éd.), *Constructions in Contact 2: Language change, multilingual practices, and additional language acquisition*, Amsterdam / New York, 375-428.
- Hendriks, Isa, Kristel Van Goethem, Fanny Meunier & Philippe Hiligsmann. 2017 = Language-specific tendencies towards morphological or syntactic constructions. A corpus study on adjective intensification in L1 Dutch, L1 French and L2 Dutch. *Nederlandse taalkunde* 22/3, 389-420.
- Hurewitz, Felicia, Anna Papafragou, Lila Gleitman & Rochel Gelman. 2006. Asymmetries in the Acquisition of Numbers and Quantifiers. *Language Learning and Development* 2/2, 77-96.

## Spatial metaphors of exhaustive quantification in Chinese

The conceptualization of extreme quantity comes in a variety of linguistic forms at a variety of linguistic levels (Keenan 2006; Acquaviva 2008; Corbett 2000; Cabredo Hofherr & Doetjes 2021; Gardelle 2019; Rothstein 2019). In this presentation, I report a corpus analysis of a productive group of semi-formulaic constructions in Chinese which use concepts of spatial opposites metaphorically to encode extreme quantity in the sense of exhaustive quantification. Based on quantitative usage data, I analyze the form-function pairing of the spatial metaphors of exhaustive quantification and answer three research questions:

- 1) What are the structural properties of the spatial metaphors of exhaustive quantification?
- 2) What are the most productive lexical categories that correlate with spatial opposites in the constructions in question? In other words, does the conceptual mapping from spatial opposites onto exhaustive quantification pertain to actions or entities?
- 3) What are the pragmatic functions of the spatial metaphors of exhaustive quantification?

The preliminary results show that Chinese makes extensive use of spatial encoding of exhaustive quantification, as evidenced by both high type and token frequencies of the semi-formulaic constructions. There is consistent mapping from spatial opposites to exhaustive quantification across multiple semi-formulaic constructions. This metaphorical conceptualization of exhaustive quantification pertains to both actions and entities. The metaphorical encoding of exhaustive quantification points to a strong experiential basis of the construal of extreme quantity in Chinese. The pragmatics of the spatial metaphors of exhaustive quantity is associated with the pragmatic process of intensification (cf. Adler & Asnès 2013). The findings of the study support the argument of embodied pragmatics.

### Selected bibliography:

- Acquaviva, Paolo. 2008. *Lexical Plurals: A Morphosemantic Approach*. Oxford: OUP
- Adler, Silvia / Asnès, Maria, « Qui sème la quantification récolte l'intensification », *Langue française*, 2013/1 (n°177), 9-22.
- Cabredo Hofherr, Patricia & Doetjes, Jenny (eds.), 2021, *The Oxford Handbook of grammatical Number*, Oxford: OUP.
- Corbett, Greville, 2000, *Number*, Cambridge: Cambridge University Press.
- Gardelle, Laure, 2019, *Semantic Plurality: English Collective Nouns and Other Ways of Denoting Pluralities of Entities*, Amsterdam: John Benjamins.
- Keenan, Edward L. 2006. Quantifiers: semantics. In Keith Brown (ed.), *Encyclopaedia of language and linguistics* Vol. 10, 302–308. Amsterdam: Elsevier.
- Kuhn, Jeremy. 2019. Pluractionality and distributive numerals. *Language and Linguistics Compass* 13(2). e12309
- Matthewson, Lisa (ed.). 2008. *Quantification: A cross-linguistic perspective*. Bingley: Emerald.
- Perrin, Laurent, 2015, *L'intensification dans l'hyperbole et la litote*, *Travaux neuchâtelois de linguistique*, 61-62, 33-51.
- Quer, Josep. 2019. Reduplication revisited: verbal plurality and exhaustivity in the visual-gestural modality. *Sensos-e VI* (1). sensos-e.v6i1.3459
- Rothstein, Susan, 2017, *Semantics for Counting and Measuring*, Cambridge: Cambridge University Press.

## Downward entailing properties of the Partitive of Negation from a typological perspective

**Partitive markers (PMs)** are primarily used to express the notion of true partitivity, that is, the proportional relation between a subset and a definite superset, for example, *a cup of the milk in the fridge* (Seržant 2021a: 885).

However, PMs have also been observed to expand beyond their prototypical function. Between 14% and 18% of world languages (Miestamo 2014; Seržant 2021a), employ partitive markers also together with the expression of negation, a phenomenon which is referred to (in the context of Eurasian languages) as the **partitive of negation (PartNeg)**, particularly w.r.t. Uralic (Lees 2015), Balto-Slavic (Arkadiev 2016), and Early Germanic (Bucci 2023) languages. The PartNeg does not express a partitive relationship between a subset and a superset. Instead, it serves as a quantificational expression, completely negating the indefinite quantified entity from a scalar perspective. Consider the example below: in (1a), the presence of the partitive marker implies that no type of chocolate has been bought whatsoever, whereas the use of the definite article suffix in (1b) implies some level of discourse retrievability, such as ‘the chocolate we know of’.

Basque, isolate (de Rijk 1998: 440)

- 1) a. *gaur ez dur txokolate-rik erosi*  
today NEG AUX chocolate-PART bought  
‘Today, I have not bought **any chocolate**’
- b. *gaur ez dur txokolate-a erosi*  
today NEG AUX chocolate-DEF.ART bought  
‘Today, I have not bought the chocolate’

To explain this feature, I propose the hypothesis that PMs exhibit a negative polarity feature known as **downward entailingness (DE)** when found in PartNeg constructions with an indefinite direct object. DE is a logical function whereby the “expressions denoting sets can be substituted for expressions denoting subsets *salva veritate*” (Giannakidou 2011: 1669): for instance, the quantifier phrase in the sentence *John didn’t buy any book* can be replaced by *any English book* because “English book” is a subset of “any book”. Similarly, in example (1a) above, “any chocolate” could be replaced by any other specific subset, while preserving the truth value of the sentence.

While the hypothesis that PMs might exhibit DE properties is not new in the literature (cf. Etxeberria 2021 for Basque), a typological survey of the combination of the two phenomena is still missing. This oversight will be addressed in this paper. First, it will be shown how PMs can exhibit an array of DE properties (e.g. quantification, domain widening, and strengthening, cf. Kadmon & Landman 1993; Chierchia 2013) in PartNeg constructions will be presented from a dataset of 55 languages. What emerges from the analysis is that such features are absent once the PartNeg becomes an obligatory feature in a given language. It will further be argued that this analysis has consequences for the interpretation of the diachrony of PMs, since the recruitment of PMs in the expression of negation via Jespersen’s cycle (already observed by Devos & van der Auwera 2013; Vossen & van der Auwera 2014; Seržant 2021b) can only be understood if DE properties are taken into account as well: the emergence of DE properties in PartNeg constructions seems in fact to coincide with the **rise of indefinite and non-referential properties associated with the generalized pseudo-partitive constructions**, that is the phase when PMs begin to express plain quantification rather than partitivity (cf. Seržant 2021b).

## References

- Arkadiev P. (2016). “Case Alternations in Five Finnic Languages: Estonian, Finnish, Karelian, Livonian and Veps”, Axel Holvoet & Nicole Nau (eds.), *Argument Realization in Baltic*, 37–81.
- Bucci G. (2023). “The contribution of Germanic to the expansion of partitive-related phenomena in the prehistoric Circum Baltic area”, *Transactions of the Philological Society* 121 (3), 427–443.
- Budd P. (2014). “Partitives in Oceanic languages”, Luraghi S. & Huumo T. (eds.), *Partitive Cases and Related Categories*, 523–562.
- Chierchia G. (2013). *Logic in grammar: Polarity, free choice, and intervention*, Oxford: OUP.
- Devos M. & van der Auwera J. (2013). “Jespersen Cycles in Bantu: Double and Triple Negation”, in *Journal of African Languages and Linguistics* 34 (2), 205–264.
- Giannakidou A. (2011). “Negative and Positive Polarity Items”, in Maienborn C., von Stechow P., Portner P. (eds.), *Semantics: An International Handbook of Natural Language Meaning*, Berlin: Mouton de Gruyter, 1660–1712.
- Kadmon N. & Landman F. (1993). “Any”, *Linguistics and Philosophy* 16 (4), 353–442.
- Lees A. (2015). *Case Alternations in Five Finnic Languages: Estonian, Finnish, Karelian, Livonian and Veps*, Leiden/Boston: Brill.
- Luraghi S. & Kittilä S. (2014). “Typology and Diachrony of Partitive Case Markers”, Luraghi S. & Huumo T. (eds.), *Partitive Cases and Related Categories*, 17–62.
- Miestamo M. (2014). “Partitives and Negation: A Cross-Linguistic Survey”, Luraghi S. & Huumo T. (eds.), *Partitive Cases and Related Categories*, 63–86.
- de Rijk R. (1998 [1996]). “On the origin of the partitive determiner”, reprinted in Rudolf de Rijk (ed.), *De lingua Vasconum: selected writings*, 435–448.
- Seržant I. (2015). “The Independent Partitive as an Eastern Circum-Baltic Isogloss”, *Journal of Language Contact* (8): 341–418.
- Seržant I. (2021a). “Typology of Partitives”, *Linguistics* 59 (4), 881–947.
- Seržant I. (2021b). “Diachronic Typology of Partitives”, Sleeman P. & Giusti G. (eds.), *Partitive Determiners, Partitive Pronouns and Partitive Case*, 111–168.
- Vossen F. & van der Auwera J. (2014). “The Jespersen cycles seen from Austronesian”, Maj-Britt Mosegaard Hansen and Jacqueline Visconti (eds.), *The diachrony of negation*, 47–82.
- Witzlack-Makarevich A. & Seržant I. (eds.), (2018). “Differential Argument Marking: Patterns of Variation”, Berlin: LSP.



Conférence invitée dans le cadre du workshop « exprimer la quantité extrême : une kyrielle de variantes » (workshop 11)

### « Les degrés de grammaticalisation dans les binominaux anglais et français désignant des grandes quantités »

Laure Gardelle, Université Grenoble Alpes  
Wiltrud Mihatsch, Université de Tübingen

D'un point de vue translinguistique, les binominaux du type *N1 de N2* sont une source courante de quantificateurs se référant à de grandes quantités, ainsi l'anglais *a lot of* ou le français *un tas de* (et le désormais largement opaque *beaucoup de*). Les éléments nominaux ne sont pas la seule source de quantificateurs (on trouve aussi des adjectifs comme *plein de*, ou encore des adverbes, ainsi dans *énormément de*), mais c'est sur les syntagmes binominaux que se concentrera la présente étude.

L'objectif de notre exposé est d'étudier les indices de grammaticalisation dans ces binominaux en français et en anglais, en comparant des quantificateurs pleinement établis avec des binominaux partiellement lexicaux et des binominaux lexicaux utilisés sporadiquement pour quantifier. L'Échelle de Quantification de Gardelle (2019) guidera notre analyse. Dans un premier temps, nous rassemblerons les séquences (*déterminant*) *N1 de* recensées dans les dictionnaires monolingues comme véhiculant une quantité, comme *a lot of* (*Oxford English Dictionary* et *Grand Robert*). Nous examinerons de plus près à cette occasion les sources lexicales : à côté des noms collectifs (*flopée*), on trouve des noms de configurations (*tas*), de portions (*lot*), de contenants (*bags*), de mesure (*tonnes*) ou des noms abstraits (*wealth, quantité*). Il s'agira de voir dans quelle mesure on trouve des similitudes entre les deux langues, et de tenter d'expliquer de manière unifiée ce qui peut conduire à une réanalyse comme quantificateur.

Nous procéderons ensuite à une analyse d'une sélection d'occurrences dans les corpus TenTen les plus récents en nous concentrant sur différentes propriétés qui permettent de distinguer les quantificateurs faiblement et fortement grammaticalisés, reflétant la perte du statut de noyau ainsi que les spécifications lexicales de N1. Parmi elles, citons :

- Différences dans l'utilisation des déterminants
- Préférences de N1 en matière de nombre et effets du pluriel (cf. *a lot of* beaucoup de)
- Restrictions sémantiques sur N2 indiquant une généralisation de N1
- Utilisation ou omission de la préposition
- Accord du verbe

L'échelle ainsi obtenue sera ensuite appliquée à l'analyse de métaphores présumées exprimant une quantification (telles que *légions de, avalanche de*), afin de déterminer si les métaphores peuvent être formellement distinguées des quantificateurs grammaticaux ou si les schémas de construction et l'analogie brouillent les pistes.

#### Bibliographie

Brems, Liselotte. 2003. Measure Noun constructions: An instance of semantically-driven grammaticalization. *International Journal of Corpus Linguistics* 8.2: 283-312.

- De Vries, Hanna. 2017. Two kinds of distributivity. *Natural Language Semantics* 25: 173-197.
- Delbecque, Nicole & Katrien Verveckken. 2014. Conceptually driven analogy in the grammaticalization of Spanish binominal quantifiers. *Linguistics* 52.3: 637-684.
- Doetjes, Jenny. 2006. Adverbs and quantification: Degrees versus frequency. *Lingua* 117: 685-720.
- Francis, Elaine J. 1998. When form and meaning come apart: quantificational nouns, predicate nominals, and locative subjects in English. *Chicago Linguistic Society* 34.1: 159-170.
- Gardelle, Laure. 2019. *Semantic Plurality. English collective nouns and other ways of denoting pluralities of entities*. Amsterdam: John Benjamins.
- Klockmann, Heidi. 2017a. *The Design of Semi-lexicality. Evidence from case and agreement in the nominal domain*. Utrecht: LOT.
- Klockmann, Heidi. 2017b. The semi-lexicality of quantificational nouns in English pseudopartitives. Com-Syn Talk, Leiden University, 20 April 2017.
- Mihatsch, Wiltrud. 2016. Collectives, object mass nouns and individual count nouns: Nouns between lexical and inflectional plural marking. *Lingvisticae investigationes* 39. 289–308.
- Mihatsch, Wiltrud. 2023. General introduction: Taxonomic nouns and their derived functions in Germanic, Romance and Slavic languages. In Wiltrud Mihatsch et al. (eds.), *Type Noun Constructions in Slavic, Germanic, and Romance Languages - Semantics and pragmatics on the move*. Berlin, Boston: De Gruyter Mouton, 1-52.
- Schwarzschild, Roger. 2011. Stubborn Distributivity, Multiparticipant Nouns and the Count/Mass Distinction. In Suzi Lima, Kevin Mullin, and Brian Smith (eds.), *Proceedings of NELS 39*, Amherst, MA: GLSA, 661-678.



**Langues & Langage  
à la croisée des Disciplines**

**1ère Rencontre annuelle LLcD**

Colloque : 9-11 septembre, 2024

Ecole d'été : 12-14 septembre, 2024

Sorbonne Université - 28 Rue Serpente, 75006 Paris, France



## **Workshop 12**

**Traitement automatique du langage et analyse  
de la variation**

**Natural language processing and linguistic  
variation analysis**

**(Vers) l'adaptation des modèles à des données hors du domaine d'entraînement pour l'analyse syntaxique du français en diachronie: Exemple du français des îles Anglo-Normandes du 16ème siècle**

Natasha Romanova  
*CRISCO, Université de Caen*  
Rayan Ziane  
*CRISCO, Université de Caen*

La mise à disposition récente de plusieurs outils neuronaux d'analyse syntaxique (*parsers*) a considérablement facilité la tâche de l'annotation syntaxique (en catégories et en fonctions) des corpus du français en diachronie. Il reste néanmoins de nombreux défis pour la constitution de corpus en diachronie longue annotés au formalisme Universal Dependencies (UD) (de Marneffe et al., 2021).

Sur le site du projet UD les corpus validés qui peuvent être utilisés pour l'entraînement sont disponibles uniquement pour la période médiévale (Profiterole, ancien corpus SRCMF) et le français moderne (Sequoia et GSD). La distance entre le corpus d'entraînement et le corpus cible peut concerner le genre et la variation diachronique et diatopique. Par ailleurs, les différents corpus présentent des incohérences de principes d'annotation (notamment, le statut des modaux) et de segmentation en phrase.

Le premier volume du registre Crime (Archives de la Cour Royale à Guernesey) contient les témoignages dans les affaires criminelles traitées sur l'île entre 1563 et 1569, dont des affaires d'infanticide, de sorcellerie et de piraterie (Ogier, 2010). L'annotation syntaxique de la transcription de quelques 42000 mots du manuscrit à l'aide d'outils automatiques présente des difficultés liées au caractère dialectal et archaïsant de la langue et au genre. Le greffier qui a retranscrit les dépositions à la fin du 16ème siècle a utilisé des graphies non-standard (la graphie "cyeulx" pour la préposition "chez" est systématiquement identifiée par les outils comme le pluriel du nom commun "ciel"), un vocabulaire marqué d'influences normandes voire anglaises et des traits grammaticaux régionaux (le pronom "no" normand). En ce qui concerne le genre, Crime I a des caractéristiques qui le distinguent de la langue littéraire, la présence de formules administratives et de phrases longues (33,8 *tokens* en moyenne, avec plusieurs dizaines de phrases mesurant plus de 80 *tokens*). Dans le corpus Profiterole (textes littéraires, y compris des textes en vers), en revanche, les phrases sont de 11,8 *tokens* en moyenne, ce qui explique une tendance à une performance réduite, surtout dans le domaine de détection de têtes syntaxiques, quand les modèles entraînés sur ce corpus sont confrontés à des phrases plus longues.

Un modèle entraîné sur le corpus Profiterole nous a permis de commencer l'analyse de Crime I ; plusieurs cycles de correction et de réentraînement ont néanmoins été nécessaires pour adapter le modèle aux données (Peng et al., 2022; Miletic et al., 2018). Dans cette intervention nous présenterons la méthode empirique utilisée pour l'annotation du texte qui a consisté à un entraînement progressif par longueur de phrase, en mélangeant des échantillons de phrases courtes

et des phrases longues corrigées manuellement pour de différentes étapes de peaufinage du modèle de base.

Nous concluons par une évaluation de la performance de différents systèmes de parsing” (Renwick & Kraif, 2024), notamment des modèles entraînés sur le corpus Profiterole par différents analyseurs syntaxiques (HOPS, BertForDeprel, UDPipe) à la base de différents modèles pré-entraînés (Conneau et al., 2020; Devlin et al., 2019; Grobol et al., 2022) et fine-tunés avec différents jeux de phrases corrigées, échantillonnées par taille et complexité.

Mots-clés : *parsing* automatique ; adaptation de modèles ; variation linguistique

## Références

Candito M. & Seddah D. (2012) Le corpus Sequoia : annotation syntaxique et exploitation pour l’adaptation d’analyseur par pont lexical. *Actes de TALN 2012. Grenoble*. 15 pp. <https://inria.hal.science/file/index/docid/698938/filename/canditoseddah-taln2012-final.pdf>.

Conneau A., Khandelwal K., Goyal N., Chaudhary V., Wenzek G., Guzmán F., Grave E., Ott M., Zettlemoyer L., & Stoyanov V. (2020). Unsupervised Cross-lingual Representation Learning at Scale (arXiv:1911.02116). *arXiv*. <https://doi.org/10.48550/arXiv.1911.02116>.

de Marneffe M.-C., Manning C. D., Nivre J., & Zeman D. (2021). Universal Dependencies. *Computational Linguistics*, 47(2), pp. 255-308. [https://doi.org/10.1162/coli\\_a\\_00402](https://doi.org/10.1162/coli_a_00402).

Devlin J., Chang M.-W., Lee K., & Toutanova K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (arXiv:1810.04805). *arXiv*. <https://doi.org/10.48550/arXiv.1810.04805>.

Grobol L. & Crabbé B. (2021). Analyse en dépendances du français avec des plongements contextualisés (French dependency parsing with contextualized embeddings). *Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles*. Volume 1: conférence principale, pp. 106-114. <https://aclanthology.org/2021.jeptalnrecital-taln.9/>

Grobol L., Regnault M., Ortiz Suarez P., Sagot B., Romary L., & Crabbé B. (2022). BERTrade: Using Contextual Embeddings to Parse Old French. In Calzolari N., Béchet F., Blache P., Choukri K., Cieri C., Declerck T., Goggi S., Isahara H., Maegaard B., Mariani J., Mazo H., Odijk J. & Piperidis S. (Eds.), *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 1104-1113. European Language Resources Association. <https://aclanthology.org/2022.lrec-1.119>

Crime I, Guernsey Greffe (manuscrit).

Guillaume B., de Marneffe G. & Perrier G. (2019). Conversion et améliorations des corpus de français annotés en Universal Dependencies. *TAL* 60 (2): 71-95. <https://aclanthology.org/2019.tal-2.4/>

Guiller K. 2020. *Analyse syntaxique automatique du pidgin-créole du Nigeria à l'aide d'un transformer (BERT): Méthodes et Résultats*. Mémoire de Master, Sorbonne Nouvelle. <https://github.com/kirianguiller/BertForDeprel>

Miletic A., Fabre C. & Stosic D. (2018). De la constitution d'un corpus arboré à l'analyse syntaxique du serbe. *Revue TAL : traitement automatique des langues*, 59 (3): 15-39. <https://aclanthology.org/2018.tal-3.2/>

Ogier M. D. (2010). The Glimpses of the Obscure : The Witch Trials of the Channel Islands. IN MacShane A. & Walker G. (eds.) *The extraordinary and the everyday in early modern England, essays in celebration of the work of Bernard Capp*, New York, Palgrave Macmillian, pp. 177-191.

Peng Z., Gerdes K. & Guiller K. (2022). Pull your treebank up by its own bootstraps. *Journées Jointes des Groupements de Recherche Linguistique Informatique, Formelle et de Terrain (LIFT) et Traitement Automatique des Langues (TAL)*, Marseille, pp.139-153. <https://hal.science/hal-03846834v1/file/504.pdf>

Prévost S. & Stein A. (2013) Syntactic annotation of medieval texts: the Syntactic Reference Corpus of Medieval French (SRCMF). IN. P. Bennett, M. Durrell, S. Scheible & R. Whitt. *New Methods in Historical Corpus Linguistics*. Narr Verlag, pp. 275-282.

Prévost S., Grobol L., Dehouck M., Lavrentiev A. & Heiden S. (2024) Profiterole: Un corpus morpho-syntaxique et syntaxique de français médiéval. *Corpus*, 25. 16 pp. URL: <https://journals.openedition.org/corpus/8538>.

Renwick A. & Kraif O. (2024) Annotation de textes d'états de langue anciens: pour le redéploiement de l'existant. *Corpus*, 25. 23 pp. URL: <http://journals.openedition.org/corpus/8286>.

### Latin treebanker's pain: suboptimal recall of key syntactic constructions

It is well-known that especially in languages with no strict word order and rich inflection, the recall of various syntactic constructions that corpus linguists are typically interested in does not increase at the same pace as latest parsers report higher and higher performance scores. As a matter of fact, even the state-of-the-art parsers that perform highly in terms of conventional metrics often underachieve with linguistic constructions that scholars consider staple subjects of linguistic inquiry, even when these constructions are relatively simple for human interpreter and relatively frequent in training data.

While it is obvious that the probability of predicting a combination of two or more linguistic tags is lower than that of a single tag, the poor recall of many human-relevant constructions may also partly arise from the fact that productive and transparent syntactic rules produce lots of new and diverse instances, which may differ considerably from the ones present in training data, especially when relatively little data is available, as is the case with many historical language varieties.

I plan to scrutinize the phenomenon from a corpus syntactician's perspective with existing Latin treebanks and language models trained thereon. I report scores for successfully parsed (morpho)syntactic constructions (e.g., accusative with infinitive, ablative absolute, various agreement phenomena) that are crucial for any corpus-linguistic analysis of Latin. I examine how much the increase in normal performance metrics, such as PoS, UAS, LAS, MLAS, etc., tends to improve the recall of the mentioned constructions in a given test data parsed using different parsers and UD language models, as well as using some basic post-processing procedures, such as median prediction vote and weighting. The discussion also necessarily touches on the out-of-domain problem and (other) challenges of not-that-big data.

While NLP tools often promise to make it possible to answer "questions so far unasked", the flip side is that they do not often make it possible to answer questions that the research community does ask. In addition to NLP-to-NLP research, I would advocate more attention to how NLP development could benefit linguistics in a more traditional sense of the word. In practice, this would mean incorporating the parsers into a pipeline of customized pre- and post-processing steps, possibly using simple rule-based parsers alongside deep learning techniques and large language models to guarantee the correct analysis of at least the most fundamental basic constructions, such as the subject and object of predicate verbs and the subordination of predicates in complex sentences.

### References

- W. Berkmans. 2019. SynthIA – Latin syntax through Artificial Intelligence. (a rule-based parser combined with a Bi-LSTM neural network) <http://www.latinsynthia.nl/aboutSynthia.html>
- P. Burns. 2023. LatinCy: Synthetic Trained Pipelines for Latin NLP. <https://arxiv.org/pdf/2305.04365>
- M. Fantoli & M. de Lhoneux. 2022. 'Linguistic Annotation of Neo-Latin Mathematical Texts: A Pilot-Study to Improve the Automatic Parsing of the Archimedes Latinus', in Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages, ELRA, 129–134.
- F. Gamba & D. Zeman. 2023. 'Universalising Latin Universal Dependencies: a harmonisation of Latin treebanks in UD', in Proceedings of the Sixth Workshop on Universal Dependencies (UDW, GURT/SyntaxFest 2023), 7–16.
- H. Pinkster. 2021. The Oxford Latin Syntax: Volume II: The Complex Sentence and Discourse, Oxford Academic.

M. Straka. 2018. UDPipe 2.0 Prototype at CoNLL 2018 UD Shared Task, in Proceedings of CoNLL 2018: The SIGNLL Conference on Computational Natural Language Learning, Association for Computational Linguistics, Stroudsburg, PA, USA, 197–207.



# Building Quantitative Contrastive Grammars from Syntactic Treebanks

Santiago Herrera<sup>1</sup>, Ioana-Madalina Silai<sup>1</sup>, Caio Corro<sup>2</sup>, Bruno Guillaume<sup>3</sup>, Sylvain Kahane<sup>1,4</sup>

<sup>1</sup>*Modyco, Université Paris Nanterre, CNRS*

<sup>2</sup>*Sorbonne Université, CNRS, ISIR*

<sup>3</sup>*LORIA, Inria, Université de Lorraine*

<sup>4</sup>*Institut Universitaire de France*

**Keywords:** contrastive grammar, quantitative grammar, grammar extraction, sparse logistic regression

Our goal is to develop corpus-driven contrastive grammars, that is, to identify the most salient patterns of a given language with respect to another language or to a set of languages, and more generally, for a given corpus with respect to another corpus (for a survey in contrastive studies, see Hasselgård, 2020). To achieve this, we use treebanks, and we compare them to induce their specific characteristics. We obtain a set of linguistic rules or tendencies ordered by importance.

Growing interest in quantitative linguistics has led to the adoption of continuous features inferred from corpora (cf. Levshina, 2019; Baylor et al., 2024) and the extraction of fine-grained patterns directly from linguistic data to answer complex linguistic questions (Chaudhary et al., 2020, 2022; Herrera et al., 2024). For example, corpora are used not only to assess whether the dominant order for a language is subject-verb (SV), but also to quantitatively capture the cases where the order is inverted. Extracting quantitative patterns and rules is particularly important in the case of comparison between language varieties, where most differences are due to variation in frequency (Blanche-Benveniste, 1997).

We built two different experimental scenarios to capture inter-corpus variation:

- Comparison between a written French corpus (GSD: Guillaume et al., 2019) and spoken French corpus (ParisStories: Kahane et al., 2021).
- Comparison between three Romance language: French (GSD), Spanish (AnCora: Taulé et al., 2008) and Romanian (RTT: Barbu Mititelu, 2018).

For each scenario, we use gold treebanks (manually annotated or corrected) to ensure the linguistic quality of our results. We employ sample techniques to ensure the comparability of the treebanks. Whenever possible, we harmonize the annotations between the treebanks.

We rely on the Universal Dependencies (UD) treebank collection (de Marneffe et al., 2021), conceived to maximize language comparability. Some of the treebanks used are from the Surface Syntactic UD (SUD) (Gerdes et al., 2018, 2022) collection. We use the converted UD versions of them which are notably uniform due to the conversion.

We follow the work of Herrera et al. (2024), inspired in turn by previous works (Chaudhary et al., 2020, 2022) which proposes a general formalization of a corpus-driven syntactic rule and

uses sparse logistic regressions to extract and rank quantitative syntactic and morpho-syntactic patterns from treebanks. In these works, rule extraction is defined as a classification task, where the linguistic phenomenon of interest (e.g. agreement, word order, case marking) is the target to predict.

We adapt the latter approach to extract the most salient syntactic patterns characterizing a treebank among a set of treebanks. Given two patterns S (scope) and Q (question), we look for patterns P (predictors) that trigger Q better than S does. For instance, if we want to know what characterizes the French verbs with respect to Romanian verbs, S will be a pattern selecting all verbs of the sample and Q the language/treebank we want to predict. We look for contexts P that favor one language/treebank over the other (such as French verbs having more subjects because Romanian is a pro-drop language).

As a result, we obtain (1) patterns that are unique to one treebank and (2) patterns that are present in both treebanks, but which are more salient in one of them. The extracted patterns are of linguistic interest and capture linguistic variation between the two treebanks. They also reveal problems due to annotation variation, providing a way to capture the difference in annotation.

## References

- Barbu Mititelu, V. (2018). *Modern Syntactic Analysis of Romanian*, pages 67–78. Editura Universității Alexandru Ioan Cuza, Iași.
- Baylor, E., Ploeger, E., and Bjerva, J. (2024). Multilingual gradient word-order typology from Universal Dependencies. In Graham, Y. and Purver, M., editors, *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 42–49, St. Julian’s, Malta. Association for Computational Linguistics.
- Blanche-Benveniste, C. (1997). La notion de variation syntaxique dans la langue parlée. *Langue française*, 115(1):19–29. Included in a thematic issue: *La variation en syntaxe*.
- Chaudhary, A., Anastasopoulos, A., Pratapa, A., Mortensen, D. R., Sheikh, Z., Tsvetkov, Y., and Neubig, G. (2020). Automatic extraction of rules governing morphological agreement. In Webber, B., Cohn, T., He, Y., and Liu, Y., editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5212–5236, Online. Association for Computational Linguistics.
- Chaudhary, A., Sheikh, Z., Mortensen, D. R., Anastasopoulos, A., and Neubig, G. (2022). Autolex: An automatic framework for linguistic exploration.
- de Marneffe, M.-C., Manning, C. D., Nivre, J., and Zeman, D. (2021). Universal Dependencies. *Computational Linguistics*, 47(2):255–308.
- Gerdes, K., Guillaume, B., Kahane, S., and Perrier, G. (2018). SUD or Surface-Syntactic Universal Dependencies: An annotation scheme near-isomorphic to UD. In Lynn, T. and Schuster, S., editors, *Universal Dependencies Workshop 2018*, Brussels, Belgium.

Gerdes, K., Guillaume, B., Kahane, S., and Perrier, G. (2022). Starting a new treebank? Go SUD! Theoretical and practical benefits of the Surface-Syntactic distributional approach. In *SyntaxFest Depling 2021 - 6th International Conference on Dependency Linguistics*, pages 35–46, Sofia, Bulgaria. Association for Computational Linguistics.

Guillaume, B., de Marneffe, M.-C., and Perrier, G. (2019). Conversion et améliorations de corpus du français annotés en Universal Dependencies. *Revue TAL : traitement automatique des langues*, 60(2):71–95.

Hasselgård, H. (2020). Corpus-based contrastive studies: Beginnings, developments, and directions. *Languages in Contrast*, 20(2):184–208.

Herrera, S., Corro, C., and Kahane, S. (2024). Sparse logistic regression with high-order features for automatic grammar rule extraction from treebanks. In Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S., and Xue, N., editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 15114–15125, Torino, Italia. ELRA and ICCL.

Kahane, S., Caron, B., Strickland, E., and Gerdes, K. (2021). Annotation guidelines of UD and SUD treebanks for spoken corpora: A proposal. In Dakota, D., Evang, K., and Kübler, S., editors, *Proceedings of the 20th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2021)*, pages 35–47, Sofia, Bulgaria. Association for Computational Linguistics.

Levshina, N. (2019). Token-based typology and word order entropy: A study based on universal dependencies. *Linguistic Typology*, 23(3):533–572.

Taulé, M., Martí, M. A., and Recasens, M. (2008). AnCorà: Multilevel annotated corpora for Catalan and Spanish. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., and Tapias, D., editors, *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

## Measuring the accessibility of personal pronouns through surprisal in a UD-parsed parallel corpus

*Luigi Talamo, Julius Steuer, Dietrich Klakow and Annemarie Verkerk  
(Saarland University)*

Personal pronouns share with other nominal constructions, such as common nouns and demonstratives, the basic function of selecting and identifying a referent. Their semantics, however, are quite peculiar in that they (i) depend on context and (ii) are defined by the interlocutor role (speaker, addressee, other) (Croft 2022). As proxies of something “that is already in the hearer’s knowledge” (Croft 2022: 77), are personal pronouns constantly accessible? And from a cross-linguistic perspective, do different languages use personal pronouns in a similar way (to simply refer to given entities) or not? To capture the accessibility of personal pronouns in textual discourse, we propose to employ the information-theoretic measure of surprisal, which characterizes the information value of a word given its preceding context (Hale 2016). Since personal pronouns are likely to be used frequently in order to keep referents in ‘in the reader’s knowledge’, we expect their surprisal to be low. However, there are some intervening factors that complexify this apparently simple scenario, both from a cross-linguistic perspective as well as from the perspective of a single language. Regarding the former, the obligatory expression of pronominal subjects (Dryer 2013) and their morpho-phonological form (Siewerska 2004) will impact surprisal. As for the latter, we expect differences in surprisal between first, second, and third person pronouns, as well as between singular and plural forms.

Our data comes from the mini Parallel Corpus of Indo-European Prose Plus (mini-CIEP+: anonymized), from which we sample 16 languages from 7 linguistic families. Mini-CIEP+ is automatically annotated using Stanford Stanza (Qi et al. 2020) with Universal Dependencies (UD: Marneffe et al. 2021) pre-trained models. It contains excerpts from 10 books in the genres of fiction and diary, for an average of 110,000 tokens and 5,600 sentences.

We calculate language model surprisal with mGPT (Shliazhko et al., 2022), a multilingual transformer model pre-trained on instances of pronouns from all languages in miniCIEP+. We fit Bayesian generalized non-linear multivariate multilevel models with surprisal values as the dependent variable and frequency of the pronouns and morpho-syntactic parameters (syntactic role, number and person) as independent variables. These are extracted from grammars and the UD annotation. We also include parameters that are not present in all languages in the sample, such as the morphological form (clitic vs independent) and the phonological form (stressed vs unstressed), as well as gender and clusivity. Figure 1 shows the posterior densities of a multilanguage model including all languages of the sample and a fixed context of seven preceding tokens. This model includes logged relative frequency, position of the pronoun with respect to the predicate, syntactic role (subject vs. non-subject), person, and number (singular vs. non-singular), with random intercepts for language. As expected, relative frequency has a positive correlation with surprisal, meaning that high frequent pronouns are less surprising;

furthermore, postverbal pronouns have less surprisal when compared to preverbal pronouns and both first and second person pronouns are less surprising than third person pronouns. Lastly, non-singular pronouns show more surprisal than singular pronouns.

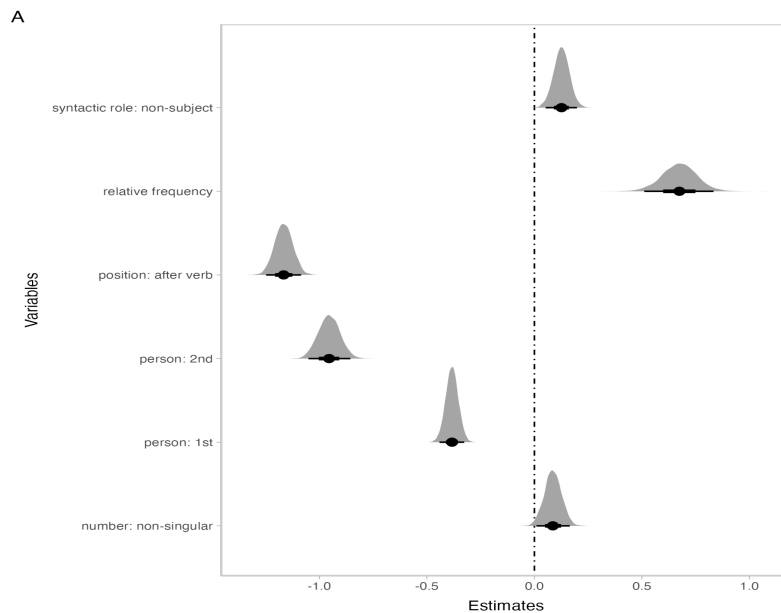


Figure 1: Posterior densities of the fixed effects of the multi-language model, with surprisal with context size seven. Main intercepts are excluded as they are removed much further from 0 (Surprisal with context size seven: Estimate 5.20, CI: 4.14, 6.25).

## References

- Croft, William (2022). *Morphosyntax. Constructions of the world’s language*. Cambridge: Cambridge University Press.
- Dryer, Matthew S. (2013). “Expression of Pronominal Subjects (v2020.3)”. In: *The World Atlas of Language Structures Online*. Ed. by Matthew S. Dryer and Martin Haspelmath. Zenodo. DOI: 10.5281/zenodo.7385533. URL: <https://doi.org/10.5281/zenodo.7385533>.
- Hale, John (2016). “Information-theoretical Complexity Metrics”. In: *Language and Linguistics Compass* 10.9, pp. 397–412. DOI: <https://doi.org/10.1111/lnc3.12196>.
- Marneffe, Marie-Catherine de et al. (2021). “Universal Dependencies”. In: *Computational Linguistics* 47.2, pp. 255–308. ISSN: 0891-2017. DOI: 10.1162/coli\_a\_00402. eprint: [https://direct.mit.edu/coli/article-pdf/47/2/255/1938138/coli\\_a\\_00402.pdf](https://direct.mit.edu/coli/article-pdf/47/2/255/1938138/coli_a_00402.pdf). URL: [https://doi.org/10.1162/coli%5C\\_a%5C\\_00402](https://doi.org/10.1162/coli%5C_a%5C_00402).

Qi, Peng et al. (2020). “Stanza: A Python Natural Language Processing Toolkit for Many Human Languages”. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. URL: <https://nlp.stanford.edu/pubs/qi2020stanza.pdf>.  
Shliazhko, Oleh et al. (2022). mGPT: Few-Shot Learners Go Multilingual. DOI: 10.48550/ARXIV.2204.07580. URL: <https://arxiv.org/abs/2204.07580>.

**Keywords:** personal pronoun, surprisal, universal dependencies, parallel corpus, accessibility

## Étude de la variation textuelle par le prisme des entités nommées

Anonyme

20 avril 2024

L'évaluation des performances des systèmes de Traitement Automatique des Langues (TAL) reste un défi, et ce malgré le développement d'outils annoncés comme toujours plus performants. En effet, la variation observée au sein d'une langue ainsi que le manque de robustesse des outils face à celle-ci conduit les systèmes à ne pas se comporter comme attendu.

C'est le cas pour la tâche de Reconnaissance d'Entités Nommées (REN), c'est-à-dire l'extraction d'expressions linguistiques qui réfèrent sans équivoque à un objet de l'univers du discours (noms de personnes, de lieux, d'organisations...).

La variation linguistique peut influencer sur les performances de cette tâche suivant plusieurs dimensions : le genre textuel<sup>1</sup> du texte considéré (poésie, prose, parole transcrite, encyclopédie...), le domaine de spécialité, intégrant la notion de langue de spécialité (les entités pertinentes dans le domaine médical ne le sont pas nécessairement dans le domaine juridique), mais encore l'évolution diachronique de la langue.

L'évaluation – de quatre outils (SpaCy, CasEN, Flair et CamemBERT) sur cette tâche – réalisée par Millour *et al.* (2024) sur le corpus varié FENEC<sup>2</sup> a permis d'observer des différentiels de performances importants non seulement entre les différents genres textuels, mais aussi entre les différents outils. En effet, ces derniers n'ont pas les mêmes comportements face à des textes de genres et d'époques différents.

En REN, tous les mots d'un texte ne sont pas des entités à identifier. Un outil est donc évalué sur les frontières des segments annotés et sur l'étiquette qui lui est associée, générant deux types d'erreurs différents (segmentation et typage). Que nous apprend l'étude de ces erreurs, d'une part sur la nature des entités concernées, et d'autre part sur le fonctionnement des outils qui les commettent ?

---

<sup>1</sup>Nous ne souscrivons pas pleinement à l'utilisation du terme "genre", généralement peu défini, mais l'utilisons ici au sens qui lui a été donné par Millour *et al.* (2024).

<sup>2</sup>*FrEnch Named-entity Evaluation Corpus*, disponible ici : <https://github.com/alicemillour/FENEC>.

En nous inspirant du travail de Fu *et al.* (2020), nous proposons d'identifier les propriétés de chaque entité nommée du corpus FENEC (propriétés intrinsèques – au niveau de l'entité, telle que sa **longueur** –, et extrinsèques – au niveau de la phrase ou du corpus, telles que sa **fréquence** ou son **ambiguïté** –) et d'étudier la corrélation entre propriétés des entités nommées et performances des systèmes. La projection des entités enrichies de leurs propriétés à l'aide d'une méthode de réduction de dimension<sup>3</sup> nous permettra de déduire une typologie des entités nommées présentes dans le corpus varié FENEC. Nous étudierons dans un premier temps dans quelle mesure les propriétés des entités et des "genres textuels" sont corrélées.

Dans un second temps, nous proposerons une analyse fine et détaillée des erreurs faites par les quatre systèmes présentés plus haut, afin d'identifier leurs forces et leurs faiblesses face aux dimensions de variation des entités identifiées.

Nous faisons l'hypothèse qu'une étude conjointe des entités nommées du corpus FENEC et des erreurs des systèmes évalués nous permettra de prédire leurs performances selon les types d'entités nommées présentes dans un texte.

## References

- Jinlan Fu, Pengfei Liu, and Graham Neubig. Interpretable multi-dataset evaluation for named entity recognition. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6058–6069, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.489. URL <https://aclanthology.org/2020.emnlp-main.489>.
- Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2020.
- Alice Millour, Yoann Dupont, Karën Fort, and Liam Duignan. Unveiling Strengths and Weaknesses of NLP Systems Based on a Rich Evaluation Corpus: the Case of NER in French. In *LREC-COLING 2024*, Turin, Italy, May 2024. URL <https://hal.science/hal-04534593>.

---

<sup>3</sup>l'algorithme UMAP<sup>4</sup> McInnes *et al.* (2020)



## Model Interpretability as Error Analysis for Variation Research

Language variety classification has seen significant advances in its results with accuracies of up to 100%. However, results largely depend on the similarity of the varieties and the scope of classification (see, for example, Aepli et al., 2022, 2023). Additionally, given black-box approaches used for classification, we don't understand how models reach their classifications. This means that on the one hand we don't know if the classifications are indeed based on dialect features and on the other hand we can't trouble-shoot or influence classifications linguistically.

To that end, Xie et al. (2024) propose an approach to extract dialect features used for classification by training BERT-based dialect classifiers and using a post-hoc leave-one-word-out approach to detect lexical items that contribute most to the prediction probability of the sentence. These words can be assumed to be most characteristic of a particular dialect, which in turn allows us to understand if the relevant features are indeed dialectal in nature and advance models by proper error analysis.

For this, we first test this approach on a corpus collected from the social media platform Jodel with data from Austria, Germany and Switzerland (Hovy & Purschke, 2018, Purschke & Hovy, 2019). We extend the original method, which is based on binary classification, to multiclass classification and extract dialect-specific features for various numbers of classes. This allows us to evaluate the features most relevant for the dialect classification. Additionally to Xie et al.'s original analysis, we also look at incorrect classifications and the features relevant to these decisions to understand the model classifications. Especially for this type of social media data we may find that the gold standard assumed by a posts geolocation may be somewhat contorted as people do not write the dialect of the region they are posting from.

Moreover, we work with Bosnian-Croatian-Montenegrin-Serbian social media data (Rupnik et al., 2023, Miletic & Miletic, 2024) as well as with reference corpora of historical German (<https://www.deutschdiachrondigital.de/>) to further evaluate how model interpretability is beneficial for error analysis. Overall, this advances our understanding of the reasoning behind model classifications and offers a way to analyse the classification errors more effectively.

## References

- Aepli, N., Anastasopoulos, A., Chifu, A.-G., Domingues, W., Faisal, F., Găman, M., Ionescu, R. T., & Scherrer, Y. (2022). Findings of the VarDial Evaluation Campaign 2022. *Proceedings of the Ninth Workshop on NLP for Similar Languages, Varieties and Dialects*, 1–13.
- Aepli, N., Cöltekin, C., Van Der Goot, R., Jauhiainen, T., Kazzaz, M., Ljubescic, N., North, K., Plank, B., Scherrer, Y., & Zampieri, M. (2023). Findings of the VarDial Evaluation Campaign 2023. *Tenth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2023)*, 251–261.
- Hovy, D., & Purschke, C. (2018). Capturing Regional Variation with Distributed Place Representations and Geographic Retrofitting. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 4383–4394.
- Miletić, A. & Miletić, F. (2024). A Gold Standard with Silver Linings: Scaling up Annotation for Distinguishing Bosnian, Croatian, Montenegrin and Serbian. *Proceedings of the 4th Workshop on Human Evaluation of NLP Systems (HumEval 2024)*, ELRA.
- Purschke, C., & Hovy, D. (2019). Lörres, Möppes, and the Swiss. (Re)Discovering regional patterns in anonymous social media data. *Journal of Linguistic Geography*, 7(2), 113–134.
- Rupnik, P., Kuzman, T. & Ljubešić, N. (2023). BENCHiĆ-lang: A Benchmark for Discriminating between Bosnian, Croatian, Montenegrin and Serbian. *Tenth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2023)*, 113–120.
- Xie, R., Ahia, O., Tsvetkov, Y., & Anastasopoulos, A. (2024). *Extracting Lexical Features from Dialects via Interpretable Dialect Classifiers* (arXiv:2402.17914). arXiv.



**Langues & Langage  
à la croisée des Disciplines**

**1ère Rencontre annuelle LLcD**

Colloque : 9-11 septembre, 2024

Ecole d'été : 12-14 septembre, 2024

Sorbonne Université - 28 Rue Serpente, 75006 Paris, France



**Workshop & Tutoriel**  
**Workshop & Tutorial**

Workshop & Tutorial

## Research infrastructures and FAIR principles in linguistics

Interdisciplinary perspectives, theoretical implications and practical applications



Organised by

Eva SOROLI, Associate Professor  
[efstathia.soroli@univ-lille.fr](mailto:efstathia.soroli@univ-lille.fr)  
University of Lille & UMR 8163 STL lab, France  
Webpage: <https://pro.univ-lille.fr/efstathia-soroli>

Christophe PARISSE, Researcher  
[cparisse@parisnanterre.fr](mailto:cparisse@parisnanterre.fr)  
INSERM, UMR 7114 Modyco & U. Paris Nanterre  
Webpage: <https://cv.hal.science/christophe-parisse>

The event is supported by the European Research Infrastructure [CLARIN](#).

---

### Description of the workshop

#### Target Topics

Research infrastructures, FAIR principles, Corpus linguistics, Crosslinguistic studies, Comparative Linguistics, Second Language Acquisition and Teaching, Historical Linguistics, Comparative Anthropological research, Discourse analysis, Resources, Tools.

#### State of the art

Digital technologies and information systems are now ubiquitous in Human and Social Sciences (HSS) impacting research, education and nearly every form of expression and creation [1]. Digital humanities –a discipline at the intersection of Computer science and traditional HSS (history, philosophy, linguistics, literature, anthropology, etc.) often considered as a “trans-discipline” that uses tools, digital methods, devices and heuristics borrowed from information science [2]– faces more and more challenges related to the increasing volume of data, corpora, applications, software, and to their great heterogeneity [3]. In this context, where great amounts of data become available in various formats and open science practices are generalized, the role of infrastructures becomes central in data sharing and standardization of practices for every interdisciplinary endeavor [4].

In the field of Language Sciences, this digital and interdisciplinary aspect is highly prominent and inherently integrated into methods and practices used in Corpus linguistics, Computational linguistics, Natural language processing and Experimental linguistics. Such a digital approach, although initially focused on large corpora of digitized texts focused on fine-grained quantitative analyses on word occurrences, calculation of frequencies of specific linguistic phenomena, and the typology of literary styles, now extends to other areas such as discourse analysis, textometry, modeling, and corpus-based teaching [5-7], enriched by techniques inspired by data mining, experimental simulation, machine learning and deep neural networks [8].

#### Objectives

The aim of this Workshop is to discuss the challenges associated with the surge in the volume of linguistic data, the diversification of approaches in this domain, and the pressing need for sharing knowledge, practices and resources. To address these challenges effectively, it is crucial to avoid duplication of efforts by adhering to common research infrastructures and to FAIR principles, thereby ensuring that linguistic resources are findable, accessible, interoperable and reusable [9]. While these principles are universally applicable across scientific disciplines, their significance has become increasingly relevant for Linguistics – a discipline characterized by a high variability in data collection practices and important heterogeneity in the formats used [10].

### (Sub)disciplines involved and scope of the workshop

More specifically, during this Workshop, we will present the most important research infrastructures (national and European) that support the work of linguists through platforms offering access to data, advanced tools for their annotation, exploration, comparison and analysis, such as Ortolang, Corli, Humanum, RnMSH, Dariah and CLARIN [e.g., 11-14]. The first part of the Workshop will focus on those research infrastructures and support centers which actively contribute to the development of precious resources and the sharing of knowledge, not only for linguists but also for anyone working in Human and Social sciences, interested in such resources (researchers, engineers, speech and language therapists, educators).

The second part of the Workshop will be devoted to case studies, tutorials, and examples of database and tool usage (e.g., use of parallel and comparable corpora) in specific research areas (e.g., study of L2 acquisition, typology research, historical linguistics, teaching, etc.). The goal is to provide hands-on demonstrations showcasing the diverse applications and wide-ranging theoretical implications of digital resources and methodologies. Discussions will include proposed solutions to hypothesis testing, and to challenges in maintaining consistent coding principles, in utilizing common technologies and in standardizing sharing practices for improved replicability, interoperability, and collaborative efforts [15].

The final portion of the day (upon registration) will focus on a tutorial introducing basic tools for spoken discourse analysis using the CLARIN Talkbank knowledge Centre and the CLAN software [16].

### References

- [1] Clement, T.E. & Carter, D. (2017), Connecting theory and practice in digital humanities information work. *Journal of the Association for Information Science and Technology*, 68: 1385-1396.
- [2] Mounier, P. (2010). Manifeste des Digital Humanities. *Journal des anthropologues*, 122-123 | 447-452.
- [3] Burnard, L. (2012). Du literary and linguistic computing aux digital humanities : retour sur 40 ans de relations entre sciences humaines et informatique In : *Read/Write Book 2 : Une introduction aux humanités numériques [en ligne]*. Marseille : OpenEdition Press (généré le 03 mars 2022). Disponible sur Internet : <<http://books.openedition.org/oep/242>>.
- [4] Joffres, A., Priddy, M., Morselli, F., Lebarbé, Th., Granier, X., Bertrand, P., Rodier, P., Melka, F., Camlot, J., Sinclair, S., Fatiha, I., Abela, C., Chayani, M., Parris, Ch., Poudat, C., Ginouvès, V., Sinatra, M., et al. (2019). "Building community" at the national and/or international level in the context of the Digital Humanities. *Digital Humanities Conference, 2019*, Utrecht, The Netherlands.
- [5] Meunier, J. G. (2020). La rencontre du sémiotique et du «numérique» : Le rôle d'une modélisation conceptuelle. *Semiotica*, 2020 (234), 177-198.
- [6] Longhi J. (2020). Proposals for a Discourse Analysis Practice Integrated into Digital Humanities: Theoretical Issues, Practical Applications, and Methodological Consequences. *Languages* 5(1): 5.
- [7] Boulton, A., & Landure, C. (2016). Using corpora in language teaching, learning and use. *Research and Teaching Languages for Specific Purposes*, 35(2). <https://doi.org/10.4000/apliut.5433>
- [8] Suissa, O., Elmalech, A., & Zhitomirsky-Geffet, M. (2021). Text analysis using deep neural networks in digital humanities and information science. *Journal of the Association for Information Science and Technology*, 73( 2): 268– 287.
- [9] Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3: 160018
- [10] Cimiano, Ph., Chiarcos, Ch., McCrae, J. & Gracia, J. (2020). *Linguistic Linked Data: Representation, Generation and Applications*. Springer International Publishing.
- [11] Erhard, H. & Krauer, S. (2014). The CLARIN Research Infrastructure: Resources and Tools for E-Humanities Scholars. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*, May 2014, 1525–31.
- [12] Olivier Baude, Adeline Joffres, Nicolas Larrousse, Stéphane Pouyllau. *Huma-Num : Une infrastructure française pour les Sciences Humaines et Sociales. Stratégie, organisation et fonctionnement*. DH 2017, Aug 2017, Montréal, Canada.
- [13] Parris, Ch. & Poudat, C. (2023). CORLI CLARIN K Centre: development and perspectives. *CLARIN Annual Conference 2023*, 16-18 octobre 2023, Leuven, Belgique.
- [14] Soroli, E. (2021). CORLI, the French Knowledge Centre for Corpora, Languages and Interaction. In Lenardic, J., F. Frontini & D. Fiser (eds.) *Tour de CLARIN*, volume IV: 40-45.
- [15] Branco, A., Calzolari, N., & Choukri, K. (2018). LREC 2018 workshop proceedings: 4REAL 2018 Workshop on Replicability and Reproducibility of Research Results in Science and Technology of Language, 12 May 2018, Miyazaki, Japan. *European Language Resources Association*.
- [16] Mac Whinney, B. & Fromm, D. (2022) Language Sample Analysis with Talkbank: An update and review. *Frontiers in Communication*, 7: 865498.

---

## Overview of the planned slots

### Introduction

Slot 1: Eva Soroli & Francesca Frontini (CLARIN-ERIC)

Slot 2: Olivier Baude & Nicolas Lrousse (HumaNum)

### Coffee break

Slot 3: Chiara Chelini & Cécile Fabre (RnMSH)

Slot 4: Edward Gray (DARIAH-ERIC)

Slot 5: Christophe Parisse (CORLI, Ortolang)

### Lunch break

Slot 6: Annemarie Verkerk et al. (Univ. of Saarland, Germany)

Slot 7: Matthias Urban (Univ. of Tübingen & DDL, Lyon)

Slot 8: Thomas Gaillat et al. (Univ. of Rennes)

Slot 9: Antonio Balvet (Univ. of Lille & STL)

### Coffee break

Slot 10: Christophe Parisse, Alina Tsikulina, Adélie Buclier & Eva Soroli

Tutorial Session on discourse analysis with the CLAN (Computerized Language Analysis) software.  
(registration required)

### Cocktails

---

## Abstracts and useful information

### **Efstathia Soroli (University of Lille, STL, CLARIN) & Francesca Frontini (CLARIN-ERIC)**

*CLARIN The European research infrastructure for linguistic data resources and technology*

CLARIN, the European research infrastructure for linguistic data resources and technology, established in 2012, provides researchers with a platform enabling easy and sustainable access to language data and processing tools. CLARIN offers advanced services for corpus compilation, comparison, annotation, and analysis, facilitates cross-linguistic studies, supports format interoperability, and aims to enhance the potential for comparative research following FAIR principles through a distributed network of over 70 data repositories and knowledge centers in 25 participating countries. With the development of this international network and the offer of a secure, unified interface, CLARIN ensures continuous resource sharing among all stakeholders involved in building, exploiting, and using language corpora. Its' mission is to ensure that knowledge and expertise are not fragmented but organized and accessible to anyone interested in such resources (researchers, engineers, educators). This presentation will cover: (a) how a dataset can be accessed/constructed/shared through the CLARIN platform (e.g., via the Language families and the Virtual language observatory services) (b) the ways it can be cited (e.g., through the Virtual collections tool); (c) used/reused (e.g., through the Language resource inventory); and (d) processed (see Switchboard tools). An illustrative example will be provided based on the ParlaMint project, supported by CLARIN, which provides parliamentary corpora in 17 languages. These corpora are designed with the same encoding standards (TEI ParlaMint) and can be utilized with the same tools and interface for comparative analysis. We will demonstrate that the value of this approach lies in sharing various language datasets, linguistic tools and collections. The goal is to enable researchers and anyone interested in these resources to access a comprehensive range of services, to navigate seamlessly between databases and tools, while maintaining consistent encoding standards, employing common language technologies, and sharing a unified commitment compatible with open science practices and FAIR principles.

### **Olivier Baude & Nicolas Lrousse (CNRS, Huma-Num, UAR 3598).**

*The role of the Huma-Num infrastructure in supporting language sciences*

Huma-Num IR\* is a French national infrastructure dedicated to supporting research project in SSH (Social Science and Humanities) in accordance with the principles of Open Science. More particularly for language sciences, this support takes several forms: in addition to the standard services offered by Huma-Num tailored to each step of the research data lifecycle (<https://documentation.huma-num.fr/media/services-HN-en.png>), some specific support is

provided for this community at very different levels. Here are a few examples. At the core infrastructure level, Huma-Num is providing specific tools such as automatic speech recognition (ASR) system. At the national level, Huma-Num has labelled and financed the CORLI Consortium, which enables targeted actions like identifying and developing tools, providing guidance for good practices, organizing trainings etc. Huma-Num also supports the development of national repositories dedicated to language resources, ORTOLANG and COCOON, in particular by offering its long-term preservation service (<https://documentation.huma-num.fr/en/parteneriat-hn-cines-en>) for the data hosted. At the European level, Huma-Num rely on its international network and its experience of the European ecosystem in order to play the role of facilitator, notably with the European CLARIN infrastructure

**Chiara Chelini (RnMSH, UAR 3603) et Cécile Fabre (MSHS-T, UAR 3414)**

*The French national networks of humanities and social sciences centres, their platforms and the role of linguistics.*

The national network of humanities and social sciences centres (*Réseau national des maisons des sciences de l'homme* - RnMSH) is a French infrastructure, a scientific cooperation consortium and a unit of the National Centre of Scientific Research (CNRS). The concept of « maison » (house) in the French name is difficult to translate into English and shows the idea of providing mutualised services, platforms and practices to the laboratories affiliated to the « houses ». The network federates 22 *maisons* all over the French territory which cover about the 70% of social sciences and humanities laboratories, in all disciplines. The network, managed by a collegial direction, is a place where collective intelligence and best practices can spread. Several interdisciplinary working groups and innovative actions are boosted by the network for the *maisons*, among which five networks of technological platforms, which, depending on their specific applications, are called: audio-visio, cogito, data, scripto, spatio. In 2023 the RnMSH launched a label campaign and 53 platforms have been labelled in 20 MSH, according to specific criteria. Implications of linguistics is mainly found inside cogito and scripto platforms, for instance in experimental research testing the cognitive bases of language acquisition and pathologies, and in digitalisation and textual analysis of different kinds of texts (modern or ancient) conducted with digital humanities methods. The 22 *maisons* are also an effective relay of the Very Large Research Infrastructures devoted to Social Sciences and Humanities and supported by the CNRS. Within this panel we would like to present a broad overview of the RnMSH missions and actions, and how the infrastructure can be useful to researchers in linguistics.

**Edward J. Gray (Dariah-EU infrastructure)**

*DARIAH: Helping humanities researchers - including linguists - confront the digital turn*

DARIAH-EU, founded in 2014, is European Research Infrastructure Consortium (ERIC) dedicated to the arts and humanities. As a research infrastructure, DARIAH's goal is to enable and empower researchers to conduct their research with the help of state-of-the-art tools and methods, in doing so in an internationally-connected way. This talk will examine how DARIAH creates a forum for knowledge exchange and collaboration for researchers, while also providing platforms such as the SSH Open Marketplace and DARIAH Campus which empower researchers to find the SSH tools and training materials that they need. It will also discuss working groups that give a forum for researchers across disciplines and countries to come together to confront common problems, such as Lexical Resources or Multilingual DH. It will also discuss our work with our sister infrastructure, CLARIN, in common initiatives such as the SSH Open Cluster. For instance, the SSH Open Marketplace, which was developed during the SSHOC project, was built in a Work Package led by DARIAH but with help from CLARIN colleagues, ensuring that at every step of the way it remains viable for linguistic uses. Together with CESSDA, CLARIN and DARIAH actively maintain the SSH Open Marketplace, ensuring that the data quality remains high and that the marketplace continues to adapt to meet the needs of researchers. DARIAH, as an international network, is also involved in other partnerships, such as a Cooperating Partnership with Princeton University which gave rise to a NEH-funded "New Languages for NLP: Bringing Linguistic Diversity in the Digital Humanities" workshop that is ongoing, with resulting training materials being uploaded into DARIAH-Campus so that everyone can benefit.

**Christophe Parisse (University of Paris Nanterre, Modyco, Corli & Ortolang)**

*CORLI and ORTOLANG: Providing tools and guidance for sharing and reusing open science language data*

Language sciences have a long tradition in data sharing at least for two main reasons: data collection is a costly procedure, and each language production is unique and non-repeatable. Therefore, recording and preserving language datasets is essential, especially for linguists. While language corpora have existed for more than 40 years,

in France, language data were not consistently centralized, leading to great diversity of formats. This has changed in the past decade thanks to the CORLI initiative and infrastructures such as ORTOLANG. The goal of the CORLI network is to help researchers to use common formats and standards, and to save data in publicly available infrastructures such as ORTOLANG. The goal of ORTOLANG is to make data deposit easy, to support data management and processing, to preserve data in the short and long terms, and thus facilitate dissemination. This paper presents the work done on spoken language corpora (similar work is ongoing for written language in CORLI and ORTOLANG). Creating and disseminating spoken language data involves three main phases and the main aim is to help the researcher to handle these phases as efficiently as possible. The first phase involves corpus constitution, which includes collecting, formatting, and describing the data. This process typically involves utilizing specialized tools, the output of which is then converted in a standardized TEI for Spoken Language format. The second phase is depositing and providing adequate metadata – e.g., through the TEIMETA tool developed by CORLI. The third phase involves using or reusing the data for research purposes. This is usually done by using specific tools such as TXM (a textometric tool), any specific tools using R or Python, or more simple tools that make it easy to browse the data, such as the export feature of our TEI tool, TEICORPO, which produces adequate formats enhancing interoperability for language data.

**Annemarie Verkerk, Luigi Talamo & Andrew Dyer (Univ. of Saarland, Germany)**

*Compiling and annotating mini-CIEP+: a sharable parallel corpus of prose*

In this talk we present mini-CIEP+, a sharable multilingual corpus of prose. mini-CIEP+ consists of the first part of ten different popular works of prose across many different languages (version 1.0 contains 35 languages, with more planned for the future). It is therefore almost entirely parallel, while a few subcorpora are comparable rather than parallel. Subcorpora typically contain 5750 sentences and almost 125K tokens. Subcorpora have dependency grammar annotation based on the Universal Dependencies standard (de Marneffe et al., 2021). It is shareable due to recent developments in German law, which allow researchers to share up to 15% of copy-righted material with a select group of people. We discuss these developments and detail the conditions under which mini-CIEP+ can be shared. We additionally describe future plans for further annotation and analysis of mini-CIEP+. These future plans centre around two major efforts: (1) annotating the corpus for information status using a newly built annotation scheme and (2) generating sentence alignment for the corpus, with the ultimate aim of annotation projection across subcorpora. The talk will highlight preliminary results for both projects.

de Marneffe, M.-C., Manning, C. D., Nivre, J., & Zeman, D. (2021). Universal Dependencies. *Computational Linguistics*, 47(2), 255–308. [https://doi.org/10.1162/coli\\_a\\_00402](https://doi.org/10.1162/coli_a_00402)

**Matthias Urban (Univ. of Tübingen, Germany & DDL, CNRS)**

*Using cross-linguistic data formats (CLDF) databases for historical linguistics and comparative anthropological research: the example of CINWA, a database of names of cultivated plants in South American Indigenous communities*

In this presentation, I will illustrate how cross-linguistic data formats (CLDF, Forkel et al. 2018) can be used to construct intuitively accessible online databases of different types of cross-linguistic data, and store them according to FAIR principles. After briefly introducing some of CLDF's design principles, I will focus on presenting the CINWA database for names of cultivated plants in South American Indigenous communities (Aguilar Panchi et al. 2023, Urban et al. 2023) as an example of a CLDF database. CINWA contains names for cultivated plants in Indigenous languages of South America, which we use in ongoing work on linguistic aspects of the transition from hunting and gathering to sophisticated forms of agriculture that occurred at different times in different parts of pre-Columbian South America. I will discuss some of the decisions we made in designing the database and extracting data, highlighting in particular the value of (i) consistency in the generation of characters with diacritics in the orthographies of Indigenous languages and (ii) full documentation of the original glosses in the lexicographic sources that we consulted, and then demonstrate the functionalities of CINWA available to users through the online interface available at <http://www.cinwa.org/>

Panchi, A., Michelle, E., Lee, S. & Brodetsky, E. (2022). CINWA – Database of Cultivated plants and their names in the indigenous languages of South America. [www.cinwa.org](http://www.cinwa.org).



- Forkel, R., List, J.-M., Greenhill, S. J., Rzymiski, Ch., Bank, S., Cysouw, M., Hammarström, H., Haspelmath, M., Kaiping, G. A., & Gray, R. D. (2018). Cross-Linguistic Data Formats, advancing data sharing and re-use in comparative linguistics. *Scientific Data* 5: 180205. <https://doi.org/10.1038/sdata.2018.205>
- Urban, M., Michelle, E., Panchi, A., Lee, S. & Brodetsky, E. (2023). CINWA (database of terminology for cultivated plants in indigenous languages of northwestern South America): introducing a resource for research in ethnobiology, anthropology, historical linguistics, and interdisciplinary research on the neolithic transition in South America. *Language Resources and Evaluation* 57: 1787-1817. <https://doi.org/10.1007/s10579-022-09628-x>

**Cyriel Mallart (LIDILE, Univ. de Rennes), Andrew Simpkin (NUI Galway, National Univ. of Ireland), Rémi Venant (LIUM, Université du Mans), Nicolas Ballier (CLILLAC-ARP, Univ. de Paris), Bernardo Stearns (Insight Centre for Data Analytics, Galway), Jen-Yu Li & Thomas Gaillat (LIDILE, Univ. De Rennes).**  
*Linguistic interoperability within a unified architecture: the case of Analytics for Language Learning*

Modern approaches to quantitative linguistics rely on large datasets. These datasets are representations of linguistic observations made up of features of various dimensions. Analyses rely on these data representations. As a result, the question of their construction is essential. In the case of quantitative research methods, datasets are built automatically. This includes the sequential use of various types of NLP tools. These tools are the components of software architectures that ensure data interoperability (Rehm et al., 2020; Sérasset et al., 2009). In this paper, we present the data architecture used in the Analytics for Language Learning (A4LL) project (Gaillat, 2022a) whose aim is to produce linguistic analytics for foreign language teachers. The system transforms learner writings into linguistic feature sets which are used to produce linguistic indicators of L2 writing performance. The system relies on an architecture composed of NLP tools. Multilingual Linguistic annotation is provided by UDPipe (Straka et al., 2016) and Stanza (Qi et al., 2020) and stored in a database. Several tools exploit this annotation to compute textual measures of different linguistic dimensions. Syntactic complexity ratios are extracted with the use of TAASSC (Kyle, 2016). Text cohesion ratios are computed with TAACO (Crossley et al., 2016). Noun-Verb collocations are extracted (Li et al., 2022) and their association strength is computed by NLTK (Bird & Loper, 2004), yielding a number of scores. Specific linguistic structural complexity indicators are computed on by language models predicting small groups of words related to each other by the same paradigmatic relationships (Gaillat, 2022b). Some lexical sophistication indicators are computed via BERT-trained language models (Devlin et al., 2019). All these indicators are filtered out as part of a graphical visualization module of the architecture. Overall this process relies on interoperable data representations. The architecture is hosted on a virtual machine of the Huma-Num consortium. The Docker technology supports a modular implementation of the system in which all modules act as standalone programmes with the aforementioned roles. These programmes are related to each other thanks to APIs and a data transfer queuing system. Automation and interoperability are the cornerstones of quantitative data processing.

- Bird, S. & Loper, E. (2004). NLTK: The Natural Language Toolkit. The Companion Volume to the Proceedings of 42st Annual Meeting of the Association for Computational Linguistics, 214–217.
- Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*, 48(4), 1227–1237.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (arXiv:1810.04805). arXiv. <https://doi.org/10.48550/arXiv.1810.04805>
- Gaillat, T. (2022a). Language learning analytics: Designing and testing new functional complexity measures in L2 writings. Proceedings of the 11th Workshop on NLP for Computer Assisted Language Learning, 55–60.
- Gaillat, T. (2022b). Language learning analytics: Designing and testing new functional complexity measures in L2 writings. 55. <https://doi.org/10.3384/ecp190006>
- Kyle, K. (2016). Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication [Dissertation, Georgia State University]. [https://scholarworks.gsu.edu/alesl\\_diss/35](https://scholarworks.gsu.edu/alesl_diss/35)
- Li, J.-Y., Gaillat, T., & Richard, É. (2022). Exploring the use of dependency parsing in automatic erroneous collocation extraction in learner English. LCR2022, Padua, Italy. <https://www.aclweb.org/anthology/2020.mwe-1.13>
- Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A Python Natural Language Processing Toolkit for Many Human Languages (arXiv:2003.07082). arXiv. <https://doi.org/10.48550/arXiv.2003.07082>
- Rehm, G., Bontcheva, K., Choukri, K., Hajič, J., Piperidis, S., & Vasiljevs, A. (Eds.). (2020). Proceedings of the 1st International Workshop on Language Technology Platforms. European Language Resources Association. <https://www.aclweb.org/anthology/2020.iwlt-1.0>
- Sérasset, G., Witt, A., Heid, U., & Sasaki, F. (2009). Multilingual language resources and interoperability. *Language Resources and Evaluation*, 43(1), 1–14.
- Straka, M., Hajič, J., & Straková, J. (2016). UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization,

**Antonio Balvet (STL, University of Lille, France)**

*Teaching linguistics with Clarin resources: preliminary results for syntax*

The Clarin infrastructure has been actively promoting the development of language resources, reference corpora and Natural Language Processing, for over a decade. As a result, academic and corporate researchers have now at their disposal a wide range of reference corpora for over 100 languages, from the Universal Dependencies corpora repository (available at <https://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-5150>), together with state-of-the-art NLP tools such as the UDPipe part-of-speech tagger/lemmatizer and functional dependency parser (Straka M. et al, 2016), available as a web application (<https://lindat.mff.cuni.cz/services/udpipe/>) or as a web service (<https://lindat.mff.cuni.cz/services/udpipe/api/>). The domain of Computer-Assisted Language Learning (CALL) has actively been experimenting with digital resources to foster -and optimize- language learning, as early as seminal works such as (Bitzer et al., 1961). In more recent approaches, such as (Aldabe et al., 2006; Borin & Saxena, 2005; Lee & Seneff, 2007; Mitkov et al., 2006; Perez-Beltrachini et al., 2012; Smith et al., 2010) and (Heck & Meurers, 2022) several projects have explored different aspects of using reference corpora and NLP tools, to automatically derive vocabulary questions and grammar exercises. In this presentation, we will address the topic of how to integrate reference corpora, language resources and NLP tools to teach linguistics. In other words, we will address computer-assisted metalinguistic competence acquisition, rather than language acquisition per se. We will present preliminary results of an ongoing research project aiming at deriving formative and evaluative self-correcting quizzes and other activities for syntax, by relying on existing reference corpora and NLP tools available from the Clarin infrastructure. We will present a software platform that enables teachers to generate Moodle compatible syntax quizzes and other activities, by leveraging annotations present in manually-revised reference corpora, as well as from NLP tools such as UDPipe. Our platform aims both at reducing manual edition to a minimum and at overcoming the subjectivity (and errors) associated with manually created exercises. Here, we report experiments conducted on French, Spanish and English, although the approach can be extended to other languages, since CONLL-U corpora are readily available from [universaldependencies.org](http://universaldependencies.org) for over 100 different languages.

---

**TUTORIAL SESSION (registration required\*)**

**Christophe Parisse, Alina Tsikulina, Adélie Buclier & Eva Soroli**

*Basic tools tutorial for spoken discourse analysis with TalkBank and CLAN (Computerized Language Analysis)*

This tutorial introduces the CLARIN K-centre Talkbank and the CLAN program, designed within this project for linguistic analysis, and will cover three main parts: transcription, linguistic (semantic and morphosyntactic) annotation, and automatic analysis. Participants will learn CLAN's utility (e.g., for studying discourse, language learning, disorders). The program facilitates transcription, media linking, annotation and aids researchers and clinicians in hypothesis testing by computing indices like MLU, TTR, and IPSyn (among others). The examples provided will focus on child language data but are adaptable to other language (re)appropriation types like second language acquisition, acquired and developmental language disorders. The tutorial aims to empower young researchers with practical skills for their language studies.

\*Registration: This last tutorial session on basic tools for discourse analysis is limited to a maximum of 15 participants. To register, please send an email to [efstathia.soroli@univlille.fr](mailto:efstathia.soroli@univlille.fr) by providing the following: **Name**, **Institution**, your **field/discipline** and your **research experience level** (e.g., Early-Stage Researcher (ESR) in the first 4 years of their research careers from the date they embarked on a doctorate; Experienced researcher (ES) in possession of a doctoral degree and within their first 5 years of their career). Accepted participants will be notified by e-mail and will be asked to download the CLAN software to their machines before coming to this tutorial session with their laptop. Please follow the instructions here: <https://dali.talkbank.org/clan/>.

Registrations for this tutorial will be open **until: August 16, 2024**. Notifications will be sent by the end of August.



**Langues & Langage  
à la croisée des Disciplines**

**1ère Rencontre annuelle LLcD**

Colloque : 9-11 septembre, 2024

Ecole d'été : 12-14 septembre, 2024

Sorbonne Université - 28 Rue Serpente, 75006 Paris, France



**General session  
Session générale**



## References

- Givón, Talmy. 2016. 'The Diachrony of Pronominal Agreement: In Ute and maybe elsewhere'. In *Cyclical Change Continued*, edited by Elly Van Gelderen, pp. 251–286. Amsterdam: Benjamins.
- Hansen, Maj-Britt Mosegaard. 2018. 'The Expression of Clause Negation: From Latin to Early French'. In *Latin tardif, français ancien: Continuités et ruptures*, edited by Anne Carlier and Céline Guillot-Barbance, pp. 269–297. Berlin: De Gruyter.
- Jespersen, Otto. 1917. *Negation in English and other languages*. København: Høst & Søn.
- Ladusaw, William A. 1993. 'Negation, indefinites, and the Jespersen cycle'. *Proceedings of the nineteenth annual meeting of the Berkeley Linguistics Society*, edited by Joshua S. Guenter, Barbara A. Kaiser and Cheryll C. Zoll, pp. 437-446. Berkeley: Berkely Linguistics Society.
- van der Auwera, Johan, Olga Krasnoukhova and Frens Vossen. 2022. 'Intertwining the negative cycles'. In *The negative existential cycle*, edited by Ljuba Veselinova and Arja Hamari. pp. 611-650. Berlin: Language Sciences Press.
- Van Gelderen, Elly. 2023. *The linguistic cycle. Economy and renewal in historical linguistics*. New York: Routledge.
- Willis, David. 2011. 'Negative polarity and the quantifier cycle: Comparative diachronic perspectives from European languages'. *The evolution of negation. Beyond the Jespersen cycle*, edited by Pierre Larrivée and Richard P. Ingham, pp. 285-323. Berlin: De Gruyter Mouton.

# Modal combination patterns in Latin: A corpus analysis by genre

Paola Marongiu  
*University of Neuchâtel*

This presentation focuses on the combination of modal expressions in a corpus of Latin texts. This phenomenon is observed when a modal marker modifies another modal marker by embedding it into its own scope. Consider example 1.

(1) *Certe pugnare abdicatis licet.*

‘**Certainly** the disinherited **are allowed** to fight’.

(Seneca the Elder, *Controversiae* 1, 8, 4)

In this sentence, the modal adverb *certe* ‘certainly’ embeds both the modal verb *licet* ‘it is allowed’ and its argument clause *abdicatis pugnare* into its own semantic scope. The value of epistemic certainty expressed by *certe* is applied to the value of authority (permission) expressed by *licet*.

The reference framework for this study is described in Dell’Oro (2022), which builds on the cognitive-functional framework presented in Nuyts (2016). The definition of modality as the association of the domains of possibility, necessity, probability and volition to a state of affairs, and the main modal categories (epistemic, deontic acceptability, deontic authority and dynamic) are drawn from this reference framework.

The combination of modalities has been investigated in various languages e.g. English (Coates 1983); French (Huot 1974; Hütsch 2020); Dutch (Nuyts 2004); Japanese (Narrog 2009); Spanish (Kratochvílová 2018) among others. So far, no study has been devoted to this phenomenon in Latin or any other historical language. Investigating the combination of modalities at the sentence level would shed new light on how modal hierarchies and scopal differences observed in other languages work in Latin (Dik 1997; Cinque 2006; Nuyts 2006 among others). This study aims at determining i) the extent to which this phenomenon occurs in Latin compared to other modern languages; ii) its main features in Latin (i.e., which markers and types of modality can be involved in the combination and in what measure); iii) if it is specific to one or more literary genres.

In order address these research questions, I conducted a corpus-based study to investigate modal combination patterns in context. The corpus used for the analysis contains texts of early and classical Latin, belonging to five different genres: theatre, letter, treatise, historiography, and oratory, for a total of around 700k tokens. The corpus was parsed by using the Stanza parser (2020). The sentences containing multiple modal markers were automatically extracted from the corpus by searching for the presence of (at least) two markers in the same sentence, based on a list of target lemmas. The sentences extracted were

then manually annotated by type of modality and presence or absence of an embedding relation between the occurring markers.

Results show that the phenomenon is quite rare in the sample examined (around 3% of the sentences in the corpus), in compliance with the results obtained for other languages. Based on the embedding properties of each type of modality annotated, it is possible to conceive a preliminary hierarchy of modalities, where epistemic modality is able to embed all the other types of modality. Dynamic modality shows smaller scopal properties, whereas deontic modality is only present in embedded roles. Epistemic modality is the most frequent type observed in the combinations for all genres, and it appears primarily in an embedding position (as shown in example 1). The combination of modalities is observed to the largest extent in the genre of letters, followed by treatises, while it is almost absent in historiography. Future work will aim at verifying these results on a bigger corpus, including more genres and diachronic stages of the language.

Keywords: modality, modal hierarchies, Latin, corpus studies

## Acknowledgements

This paper is based on my PhD thesis *Co-occurrence of modal markers in Latin: a quantitative and qualitative analysis* (in preparation). My PhD is part of the project *WoPoss: A world of possibilities. Modal pathways over an extra-long period of time: the diachrony of modality in the Latin language*. My PhD as well as the WoPoss project are funded by the Swiss National Science Foundation (SNSF project n° 176778). The project is hosted at the Institut des sciences du langage—ISLa (University of Neuchâtel).

## References

- Cinque, Guglielmo (2006). *Restructuring and functional heads: the cartography of syntactic structures*. New York: Oxford University Press.
- Coates, Jennifer (1983). *The semantics of the modal auxiliaries*. London, Canberra: Croom Helm.
- Dell’Oro, Francesca (2022). *WoPoss guidelines for annotation*. Revised version. Zenodo. DOI: 10.5281/zenodo.6417878.
- Dik, Simon C (1997). *The theory of Functional Grammar, Part 1: the structure of the clause*. Ed. by Kees Hengeveld. Vol. 1. Berlin, New York: Mouton de Gruyter.

Huot, Hélène (1974). *Le verbe devoir: étude synchronique et diachronique*. Paris: Klincksieck.

Hütsch, Annalena (2020). *L'usage des verbes modaux en français et en allemand. Étude contrastive de la combinatoire adverbiale sous l'éclairage quantitatif*. PhD thesis. Neuchâtel: Université de Neuchâtel, Faculté des Lettres et Sciences Humaines, Institut des Sciences du Langage et de la Communication.

Kratochvílová, Dana (2018). *Modality in Spanish and Combinations of Modal Meanings*. Charles University: Karolinum Press.

Narrog, Heiko (2009). *Modality in Japanese: the layered structure of the clause and hierarchies of functional categories*. Amsterdam, Philadelphia: John Benjamins Publishing.

Nuyts, Jan (2004). *Over de (beperkte) combineerbaarheid van deontische, epistemische en evidentieële uitdrukkingen in het Nederlands*. Antwerp: Antwerp University. <https://repository.uantwerpen.be/docman/irua/cd8cd9/49314.pdf>

Nuyts, Jan (2006). *Modality: Overview and linguistic issues*. In William Frawley (ed.) *The expression of modality*. Berlin, New York: Mouton de Gruyter.

Nuyts, Jan (2016). *Analyses of the modal meanings*. In Jan Nuyts and Johan van der Auwera (eds.), *The Oxford handbook of modality and mood*. Oxford: Oxford University Press.

Qi, Peng, Zhang, Yuhao, Zhang, Yuhui, Bolton, Jason & Manning, Christopher D. (2020). *Stanza: A Python Natural Language Processing Toolkit for Many Human Languages*. Association for Computational Linguistics (ACL) System Demonstrations, 101-108.



## GENERAL HUMAN NOUNS IN MEDIEVAL FRENCH: BORDERLINE GRAMMATICALIZATION?

Zinaida Geylikman

*EDA, Université Paris Cité*

### Research question:

Over past decades, a number of studies have been dedicated to general human nouns (=GHN) – common nouns referring to humans without specifying any additional biological or social feature. Just as other general nouns, they can be used as hyperonyms/classifiers and tend to be grammaticalized overtime (Haspelmath, 1997; Mihatsch, 2017). The literature is particularly abundant on Modern and Contemporary French material (18th – 21th century) (Cappeau, 2018; Cappeau & Schnedecker, 2014, 2018), sometimes in comparison to other Romance, Germanic or Slavic languages (Roque Amaral & Mihatsch, 2016; Rössler, 2018; Mostrov & Aleksandrova, 2017).

The collected data on GHN such as *homme*, *personne*, *individu*, *gens*, *mec* shows that beyond the apparent synonymy this lexicon is marked by an important diversity in respect to immediacy/distance perspective and text-type distribution (Cappeau & Schnedecker, 2014; Cappeau, 2018). Moreover, Mihatsch (2017) argues that GHN that are most frequently used in common language tend to partially desemanticize and appear in ‘typically grammatical’ roles, such as counting units, templates for modification (ex. *des gens formidables* ‘amazing people’) and nominalization (ex. *personne en situation de handicap* ‘person with a disability’), or equivalents of indefinite and impersonal pronouns.

In the present talk we will investigate the use of three GHN – *homme*, *gent/gens*, *personne* – in Medieval French (9th – 15th century) in order to determine whether the patterns that characterize them in Contemporary French can be traced back to the Medieval period.

### Data and methods:

The study will be conducted on the BFM22 material which is a representative corpus of Medieval French texts from 9th to 15th century (7328715 tokens). The absolute frequency of the studied items per 5 subcorpora is represented in Table 1:

	9-11 <sup>th</sup> century	12 <sup>th</sup> century	13 <sup>th</sup> century	14 <sup>th</sup> century	15 <sup>th</sup> century
total tokens	12009	1595623	2069099	1734284	1917700
<i>homme</i>	44	2757	3468	3195	3692
<i>gent/gens</i>	10	2322	3695	3909	3764
<i>personne</i>	0	81	264	607	688

Table 1.

Following the framework suggested by the GHN research, we will proceed to a quantitative and qualitative corpus analysis of *homme*, *gent/gens* and *personne*. The dataset will be annotated semi-automatically, while random samples per item and per century will undergo manual annotation. We will consider several parameters of different nature:

- semantic features: generic/specific reference, meaning and connotation if applicable
- syntactic function: subject/object/presentative
- presence/absence of determiners and/or modifiers
- distribution by text-type
- distribution by represented speech segments (direct and indirect speech, see Marchello-Nizia, 2012).

The talk will demonstrate that:

- the three items can be considered as GHN since the Medieval period;
- but for some synonymous uses (ex. *hommes/gens* + *de* + NOUN), each item has its specificity both on morphosyntactic level and in respect to text-type distribution;
- the precise set of lexical combinations is proper to the studied period (ex. *religieuse personne* “clergyman”), but on a more general level the three items tend to appear in the same ‘grammatical’ roles as in Contemporary French: counting units, template for modification/nominalization, pronoun equivalents.

Thus, on the example of French diachrony, the talk will suggest for GHN a set of inherent features that resist linguistic evolution.

### Keywords

- general human nouns
- historical semantics
- grammaticalization
- Medieval French
- corpus linguistics

### References

- BFM - Base de Français Médiéval [En ligne]. Lyon : ENS de Lyon, Laboratoire IHRIM, 2022, <txm.bfm-corpus.org>.
- Cappeau, P. (2018). (Pro)Noms d’humains à l’oral. *Linx*, 76, 57–84. <https://doi.org/10.4000/linx.2391>
- Cappeau, P., Schnedecker, C. (2014). Gens, personne(s), individu(s) : trois saisies de l’humain. *SHS Web of Conferences*, 8, 3027–3040. <https://doi.org/10.1051/shsconf/20140801274>
- Haspelmath, M. (1997). *Indefinite Pronouns*. Oxford: Oxford University Press.
- Marchello-Nizia, C. (2012). L’oral représenté : un accès construit à une face cachée des langues ‘mortes’. In C. Guillot, B. Combettes, A. Lavrentiev, E. Opperman-Marsaux, & S. Prévost (Eds.), *Le changement en français. Études de linguistique diachronique*, (pp.247–264). Peter Lang.
- Mihatsch, W. (2015). La sémantique des noms généraux ‘être humain’ français et allemands. In W. Mihatsch, & C. Schnedecker (Eds.), *Les noms d’humains : une catégorie à part ?* (pp.55-83). Steiner.
- Mihatsch, W. (2017). Les noms d’humains généraux aux limites de la grammaticalisation. *Syntaxe et Sémantique*, 18(1), 67–99. <https://www.cairn.info/revue-syntaxe-et-semantique-2017-1-page-67.htm>

- Mostrov, V., & Aleksandrova, A. (2017). Les noms d'humains généraux «passés sous silence»: du français vers le bulgare, *JATEPress*, 363-375. <https://uphf.hal.science/hal-03269475>
- Roque Amaral, E. T., & Mihatsch, W. (2016). Le nom français personne en comparaison avec le portugais brésilien pessoa et l'allemand Person – des noms en voie de pronominalisation? *SHS Web of Conferences*, 27, 12015. <https://doi.org/10.1051/shsconf/20162712015>
- Rössler, S. (2018). Leute (allemand) et gens (français): Du nom au pronom? *Linx*, 76, 115–132. <https://doi.org/10.4000/linx.2411>
- Schnedecker, C. (2018). Le nom d'homme est-il un nom général? *Linx*, 76, 23–56. <https://doi.org/10.4000/linx.2506>

## Variabilité en grammaire et vernacularité : les adverbes du type *en neuf* et *à blanc* en français

Dans la synchronie contemporaine, la langue française n'affiche pas une variation linguistique importante en comparaison avec des périodes antérieures compte tenu du processus de nivellement qu'elle a parcouru. Ce nivellement semble aussi toucher le domaine adverbial, en particulier un schéma morphosyntaxique rare issu du latin : l'adverbe 'prépositionnel', formé de deux éléments catégoriels – une préposition et un adjectif – selon le schéma < PREP + ADJ >.

Des travaux récents montrent que ce schéma, d'apparence vernaculaire, est présent dans toutes les langues romanes, malgré une tendance largement partagée à imposer le schéma 'adjectif *-ment(e)*', associé à l'usage normatif (Hummel *et al.* 2019). Il a été formé dès le latin tardif : Solari Jarque (2022) en identifie plus de 120. C'est néanmoins surtout au cours de l'analytisation des langues romanes qu'il a pu prospérer. Selon le discours métalinguistique, davantage lexicographique que grammatico-graphique, l'ancien et le moyen français affichent une grande diversité d'adverbes prépositionnels ainsi qu'un nombre élevé de variantes morphologiques, structurelles, sémantiques et/ou graphiques (cf. TL, DMF, FEW). C'est le cas pour le type *à certes*, largement ancré et représenté entre le 11<sup>e</sup> et les 16/17<sup>e</sup> siècles, qui coexiste d'abord avec une variante à double préposition, *adecertes* 'certes, réellement, vraiment' (11-14<sup>e</sup> siècles) (FEW II, 610b I.3). Comme attendu, la variabilité du schéma est également marquée à l'époque contemporaine en périphérie, dans les variétés dialectales au sein de l'espace galloroman (p.ex. *à certes*, *de continu*, *de grout*) mais aussi outremer comme au Québec (ex. *en grand* 'beaucoup'). Pourtant, en français de référence, il se limite pour l'essentiel au maintien d'un petit nombre d'adverbes comme *en gros*, surtout en emploi lexicalisé. Cette évolution est-elle due à la vernacularité de l'adverbial sous étude ?

La présente communication s'attache à l'analyse de l'histoire d'une quarantaine de types d'adverbes construits avec des adjectifs hérités du latin comme afr.-frm. *à certes*, *en gros*, *à/en plein*, *à/de/en bref*, dial. *de continu* ou frm. *au sérieux*. L'analyse identifie plus de trois cents adverbiaux différents dans l'histoire du français dans la francophonie du Nord. Ceci constitue un ensemble considérable compte tenu du nombre d'adverbes en français : Marchello-Nizia en compte entre 100 et 150 dans l'histoire du français en excluant les 'locutions adverbiales' et les adverbes en *-ment* (dans GGHF, 912).

L'analyse retrace une variabilité décroissante depuis l'ancien français jusqu'aux variétés galloromanes modernes, décroissance qui accompagne un refoulement à des variétés conservatrices et/ou périphériques. Elle montre aussi que les adverbiaux déjà présents en latin sont en français de référence parmi les rares à tendre vers le principe d'économie langagière 'une forme / un sens', comme IN BREVI (afr.-mfr. *à/de/en bref*, frm. *en bref*, cf. TL, DMF, TLF). Au contraire, dans des variétés périphériques comme au Québec, les adverbiaux montrent une très grande variabilité morphologique et sémantique : *en bref* y est (toujours) polysémique, et d'autres continueurs du latin connaissent à la fois des variantes morphologiques et sémantiques. Cette variabilité peut être interprétée comme un manifeste d'une histoire essentiellement informelle.

## Bibliographie

- DMF : *Dictionnaire du Moyen Français 1330 – 1500*. Nancy : ATILF-CNRS/Université de Lorraine, version 2016 [<sup>1</sup>dmf 2003, <sup>2</sup>dmf 2007, <sup>3</sup>dmf 2012] [en ligne : <http://www.atilf.fr/dmf/>].
- FEW : Wartburg, Walther von (1922-2002). *Französisches Etymologisches Wörterbuch. Eine Darstellung des galloromanischen Sprachschatzes*. 25 vol. Bonn *et al.* : Klopp *et al.*
- GGHF : Marchello-Nizia, Christiane, Bernard Combettes, Sophie Prévost & Tobias Scheer (eds.) (2020). *Grande grammaire historique du français*, Berlin / Boston: De Gruyter.
- Hummel, Martin, Chircu, Adrian, García Sánchez, Jairo, García Hernández, Benjamín, Koch, Stefan, Porcel Bueno, David et Wissner, Inka (2019). « Prepositional adverbials in the diachrony of Romance : a state of the art » [en ligne : DOI 10.1515/zrp-2019-0001]. *Zeitschrift für romanische Philologie* 135 (4), 1080-1137.
- Solari Jarque, Nicolás (2022). *Las formaciones adverbiales con preposición y adjetivo (tipos de pleno, in serium) en latín*. Thèse doctorale 2022, Université de Alcalá de Henares.
- TL : Tobler-Lommatsch, Adolf (1925-2008). *Altfranzösisches Wörterbuch*. Wiesbaden, puis Stuttgart : Steiner.
- TLF : Imbs, Paul et Quemada, Bernard (1971-1994). *Trésor de la langue française. Dictionnaire de la langue du XIX<sup>e</sup> et XX<sup>e</sup> siècle (1789-1960)*. CNRS Éditions / Gallimard : Paris [en ligne : <http://atilf.atilf.fr/tlf.htm>].

## **Corrélation entre l'émergence des prépositions fonctionnelles et la perte de la flexion casuelle en italien et en français : perspectives diachronique et comparée**

Axelle Domingues  
*Sorbonne Université*

Retracer l'histoire d'une langue [...] c'est, pour une large part, rendre compte des changements qui ont affecté les moyens de codage des fonctions, les fonctions elles-mêmes, et la relation entre les deux. (Prévost & Dufresne, 2020)

### **1. Hypothèse**

Il a souvent été avancé qu'il existe une corrélation causale entre la perte de la flexion casuelle et l'émergence de prépositions fonctionnelles dans l'évolution du latin aux langues romanes, malgré certaines critiques de nature qualitative pour le français (Schøsler, 2020). Cependant, cette hypothèse n'a pas encore été testée quantitativement avec les outils statistiques disponibles aujourd'hui. Notre projet de communication vise à combler cette lacune à travers une étude diachronique large de 1200 à 1840, comparant le rythme d'émergence des prépositions françaises *à* et *de* et des prépositions italiennes *a*, *di* et *da*, en y incluant leurs formes fusionnées avec l'article défini. Étant donné que l'italien a perdu la flexion casuelle dès la période pré-littéraire, alors que le français offre une opposition bicasuelle jusqu'au milieu du 14<sup>e</sup> siècle, l'hypothèse formulée est que l'italien connaît un développement plus précoce du paradigme des prépositions fonctionnelles que le français (Fagard, 2006 ; Iliescu, 2013 ; Pinkster, 1990 ; Schøsler, 2020).

### **2. Méthodologie et traitement de données**

Ce travail s'appuie sur les corpus numériques de Frantext et MIDIA (*Morfologia dell'italiano In DIChronia*) dont ont été extraites 8 154 890 occurrences selon la périodisation suivante : 1200-1375, 1376-1532, 1533-1691, 1692-1840.

Nous avons intégré la fréquence textuelle relative de ces mégadonnées dans des codes exécutables que nous avons rédigés en Python dans l'environnement de programmation et d'exécution *Google Colaboratory* basé sur *Jupyter Notebook*. Nous y avons ajouté diverses bibliothèques logicielles : *Pandas* pour les outils de structure et d'analyse de données, *Matplotlib* et *Seaborn* pour ceux de visualisation et *Numpy* pour le traitement d'*arrays*, c'est-à-dire l'unité de chaque trame. Nous obtenons ainsi une modélisation du changement linguistique avec le traitement de *bigdata* sans échantillonnage.

### **3. Résultats et discussion**

La figure 1 montre que la fréquence relative des prépositions fonctionnelles est constamment plus élevée dans MIDIA que dans Frantext. De plus, le paradigme italien évolue moins que le français, ce qui nous mène à penser qu'il était déjà développé en 1200, tandis que le français rattrape son retard au fil du temps. Cette observation met en évidence une corrélation entre la suppression des cas dans les langues romanes et l'évolution des paradigmes prépositionnels.

Les figures 2 et 3 prennent en compte la complexité du système morphologique par rapport aux conventions graphiques. La figure 2 distingue les prépositions contractées traditionnelles des prépositions, tandis que la figure 3 regroupe, sous l'étiquette de 'prépositions contractées révisées', les formes fusionnées et les suites des prépositions *à*, *de*, *a*, *di* et *da* avec un article défini. La tendance à la fusion étant plus prononcée en italien qu'en français, notre étude propose une reconfiguration des sous-paradigmes prépositionnels. L'analyse des limites de l'interface entre morphosyntaxe et phonétique nous amène à ignorer la séparation graphique. Cette dernière figure visualise le rapprochement progressif des différents sous-paradigmes italiens et français jusqu'à ce que l'ensemble des quatre cercles se chevauchent. Elle fait apparaître encore plus clairement un retard initial du français par rapport à l'italien et un mouvement de rattrapage en diachronie.



Figure 1 (ligne du haut) : Évolution comparée du paradigme prépositionnel en italien (MIDIA) et en français (Frantext) de 1200 à 1840 sur une FTR (Fréquence Textuelle Relative) d'un million d'occurrences.

Figure 2 (ligne du milieu) : Évolution comparée des prépositions et des prépositions contractées en italien (MIDIA) et en français (Frantext) de 1200 à 1840 sur une FTR (Fréquence Textuelle Relative) d'un million d'occurrences.

Figure 3 (ligne du bas) : Évolution comparée des prépositions et des prépositions contractées révisées en italien (MIDIA) et en français (Frantext) de 1200 à 1840 sur une FTR (Fréquence Textuelle Relative) d'un million d'occurrences.



**Mots clés** : préposition, morphosyntaxe, linguistique de corpus, linguistique comparée, linguistique diachronique.

### **Bibliographie**

- Fagard, B. (2006). *Les prépositions: du latin aux langues romanes modernes*. In *Modèles linguistiques*, Association Modèles linguistiques/Éditions des dauphins, 53 (27(1)), 91-110.
- Iliescu, M. (2013). *Varia Romanica : Universaux linguistiques, analyse du discours et caractère variationnel de la Romania* (Vol. 13). Frank & Timme GmbH.
- Pinkster, H. (1990). The development of cases and adpositions in Latin. In *Unity in diversity. Papers presented to Simon C. Dik on his 50th birthday* (pp. 195-209). Dordrecht: Foris.
- Prévost, S. & Dufresne, M. (2020). Chapitre 2 L'approche du changement linguistique dans la Grande Grammaire Historique du Français. In C. Marchello-Nizia, B. Combettes, S. Prévost & T. Scheer (Ed.), *Grande Grammaire Historique du Français (GGHF)* (pp. 15-36). Berlin, Boston: De Gruyter Mouton.
- Schøsler, L. (2020). Chapitre 30 Catégories variables : noms, adjectifs, pronoms et déterminants. In C. Marchello-Nizia, B. Combettes, S. Prévost & T. Scheer (Ed.), *Grande Grammaire Historique du Français (GGHF)* (pp.664-674). Berlin, Boston : Gruyter Mouton.

## Paradigmatic leveling in the nominal inflection of Modern Greek in Southern Italy

Michail I. Marinis

*The Ohio State University*

In nominal inflection of Modern Greek in Southern Italy<sup>1</sup>, all masculine nouns traditionally formatted their plural sub-paradigm according to -at least- two distinct *inflectional classes* (compare I.C.a.1 & I.C.a.2, in Table 1). According to Marinis (2020), this situation has changed today, as all the masculine nouns realize their plural sub-paradigm using a unified set of inflectional suffixes (compare I.C.b.1 & I.C.b.2, in Table 1) previously used exclusively in I.C.a.1. At the same time, the set of inflectional suffixes previously marked the plural sub-paradigm of I.C.2, has not been lost from the system but is now specialized for the formation of feminine nouns.

	previously		today		
	I.C.a.1	I.C.a.2	I.C.b.1	I.C.b.2	I.C.b.3
	m	m/f	m	m	f
NOM.SG.	-o	-a/-i	-o	-a/-i	
NOM.PL.	-i	-e <sub>(s)</sub>	-i		-e <sub>(s)</sub>
ACC.PL.	-u		-u		
GEN.PL.	-o				

*Table 1: Schematic representation of the evolution of masculine nouns through restructuring of their inflectional paradigm*

In this paper, (a) I investigate the mechanisms (analogy, inter- and intra-paradigmatic leveling, reanalysis) through which the plural sub-paradigm of masculine nouns has been changed in Italian nominal inflection. (b) I look at the nature of the causes leading to language change, showing that it has been achieved by both intra-linguistic and inter-linguistic factors.

In the former, I classify factors such as the tendency of the system to evolve to a state where each form corresponds to only one meaning/function (for example, Hock, 1991; Hock

---

<sup>1</sup> Greek Dialects system spoken in South Italy in two dialectal enclaves, in Puglia (the system is called *Griko*) and Calabria (the system is called *Greko*). In 1999, the Greek of Southern Italy was included in the **UNESCO Red Book of Endangered Languages**. In December of the same year, the official Italian state recognized the Puglian Greek and Calabrian Greek as **national minority languages**, by the Law on the protection of historical linguistic minorities (Norme in materia di tutela delle minoranze linguistiche storiche - 482/1999). For a more detailed description of the system, see Ledgeway (2013).

& Joseph, 1996). With the latter, I classify factors such as the *intensive language contact* (Thomason, 2001) of Italiot with romance systems, since Italiot was found (a) in long-term isolation from the other Modern Greek varieties and at the same time (b) in intense linguistic contact with Standard Italian, and also with the local Romance system of the Lecce region (Salentino) and of the Calabria (Calabrese, Boveze).

I will show that the leading cause of the language change is intra-linguistic, which is also evidenced by the presence of the phenomenon in other Greek dialects, as in those being in long-term isolation (for example, Tsakonian). At the same time, intense contact of Greek system(s) with Romance contributed to the prevalence of the phenomenon in Greek of South Italy but not in the rest of the Modern Greek Dialects, where the phenomenon is also scarcely present. Also, one can find extensive syncretic patterns, described by Marinis (2024) as morphomic.

The data of this research comes from fieldwork I collected personally in the Greek-speaking areas of Southern Italy in the summer of 2016 and 2018, and the analysis of the system's historical evolution is based on the available written sources.

### **Keywords**

inflectional classes, language contact, paradigmatic leveling, Modern Greek Dialects, Greko

### **Selected References**

- Hock, H. H. (1991) [1986]. *Principles of Historical Linguistics*. Berlin: Mouton de Gruyter.
- Hock, H. H. & B. D. Joseph. (1996). *Language History. Language Change, and Language Relationship: An Introduction to Historical and Comparative Linguistics*. Berlin: Mouton de Gruyter.
- Ledgeway, A. (2013): Greek Disguised as Romance? The Case of Southern Italy. In Janse, M., Joseph, B., & A. Ralli (Eds.), *Current Issues in Modern Greek Dialects and Linguistic Theory, 184–228*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Marinis, M. I. (2020). Klisi kai organosi klitikon paradeigmaton [Inflection and Inflectional Paradigms]. Unpublished doctoral dissertation. University of Patras. <https://doi.org/10.12681/eadd/46842>
- Marinis, M. I. (2024). Morphological interpretations of syncretism in the panorama of Greek. *Folia Linguistica*. <https://doi.org/10.1515/flin-2024-2021>
- Thomason, S. (2001). *Language Contact: An Introduction*. Edinburgh and Washington, Edinburgh University Press & Georgetown University Press.

## Micro-diachronie en langue des signes française (LSF) : quels changements lexicaux ?

Mots clés : micro-diachronie, sociolinguistique, langue des signes française, changements lexicaux

Le milieu des années 1970 marque le début de nouvelles actions menées par et pour les sourds, donnant ainsi une visibilité à la communauté sourde et à la langue des signes française (LSF) (Kerbourc'h, 2012). Le décroisement de la LSF a permis sa diffusion et sa pratique par davantage de locuteurs, sourds comme entendants. Elle n'en reste pas moins une langue minorisée puisque bien qu'institutionnalisée par la loi de 2005 considérant la LSF comme une langue à part entière, elle ne bénéficie pas d'une pleine reconnaissance (Bertin, 2010). Notre travail de recherche porte sur les changements lexicaux en LSF entre les années 1970 et aujourd'hui. Parmi les domaines qui ont connu une expansion lexicale, nous avons retenu celui du militantisme. En effet, celui-ci est un lieu de rencontres et d'échanges où les locuteurs ont parfois besoin de nouveaux termes pour s'exprimer. Nous nous questionnons sur les changements et leurs sources d'influences ainsi que sur le rôle de ces évolutions dans la reconnaissance et l'institutionnalisation de la LSF.

Les changements lexicaux que nous étudions sont associés aux locuteurs, à leurs pratiques et aux diverses influences de la société sur les locuteurs et leurs habitudes, nous positionnant ainsi dans une approche sociolinguistique (Boutet et Heller, 2007 ; Calvet, 2013). Dans la mesure où ce travail porte sur des changements qui s'étalent sur une période de cinquante ans, il se situe en micro-diachronie (Siouffi, Steuckardt, Wionet, 2012 ; Abouda, Skrovec, 2022/2). L'aspect sociolinguistique de cette recherche et le fait de travailler sur le thème du militantisme avec notamment des communautés minoritaires (féminisme, LGBT, par exemple), fait également intervenir la théorie de l'intersectionnalité (Crenshaw, 1990 ; Fassin, 2015; Jaunait, Chauvin, 2012/1).

Afin de pouvoir observer les changements lexicaux sur la période choisie, un double corpus est en cours de constitution. Il est constitué, d'une part, d'archives vidéo et, d'autre part, d'entretiens semi-directifs auprès de personnes sourdes signeuses. Les vidéos proviennent de sources multiples. Nous travaillons à partir des émissions de chaînes publiques (l'œil et la main) et des vidéos disponibles dans les bases de l'INA (fonds de l'International Visual Theatre et de Jacques Sangla), sur des vidéos publiées sur les réseaux sociaux par des particuliers ou des associations, et sur des documents transmis par des particuliers que nous avons contactés. Les entretiens semi-directifs sont réalisés en LSF avec des personnes sourdes signeuses de toutes générations confondues, nous permettant à la fois d'avoir des signeurs témoins du réveil sourd et de jeunes signeurs. Les données que nous obtiendrons au terme du recueil serviront à identifier et analyser les signes employés et les usages sur une période de plusieurs décennies. La prise en compte de tout un ensemble de facteurs (profil sociolinguistique, situation de communication, variations régionales, etc.) offre la possibilité d'expliquer pourquoi certains changements ont eu lieu (DELAPORTE, 2005) ou pourquoi un signe a été préféré à un autre pour un même référent.

### Références bibliographiques :

ABOUDA L., SKROVEC M., « Micro-diachronie de l'oral. Présentation », In : *Langages*, n°226, Éd. : Armand Colin, pp.9-24

BERTIN F., 2010/1, « La langue des signes, une langue vivante comme les autres ? Petite histoire d'une grande question linguistique », *La nouvelle revue de l'adaptation et de la scolarisation*, n°49, Ed. de l'INSHEA, pp.13-19

CRENSHAW K., 1989, "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics", *University of Chicago Legal Forum*, Vol.1989, Issue 1, Article 8, pp.139-167

DELAPORTE Y., 2007, « Dictionnaire etymologique et historique de la langue des signes française. Origine et évolution de 1200 signes », Éd. : Fox, 680p.

FASSIN É., 2015/2, « Les langages de l'intersectionnalité », In : *Raisons politiques*, Éd. Presses de sciences Po, pp.5-7

JAUNAIT A., CHAUVIN S., 2012/1, « Représenter l'intersection. Les théories de l'intersectionnalité à l'épreuve des sciences sociales », In : *Revue française de politique*, Éd. : Presses de Sciences Po. Vol.62, pp.5-20

KERBOURC'H S., 2012, "Le mouvement sourd (1970-2006). De la langue des signes française à la reconnaissance sociale des sourds », Paris : l'Harmattan

SIOUFFI G., STEUCKARDT A., WIONET C., 2012, "Le marqueur après à l'oral : une approche micro-diachronique, variationniste et interactionnelle », In : *Langage*, n°226, Éd. : Armand Colin, pp.215-226

## Differential Object Marking in New Persian: Diachronic and Dialectal Variations

Agnès Lenepveu-Hotz  
University of Strasbourg

It is well known that in contemporary Persian there is a Differential Object Marking (DOM) in which the postposition *rā* marks the definite Direct Object, while the indefinite one remains unmarked (Lazard 1982, Bossong 1985); the Indirect Object is introduced by prepositions. From a diachronic and dialectological point of view, however, the situation is more complex, given that *rā* is also used with dative functions. The purpose of this paper is to examine how the recipient of ditransitive verbs is marked (within the theoretical framework of valency as developed in Lazard 1994) and how this marking has evolved, with consequences for the development of DOM. Based on a corpus of 15 texts dating from the 10th century to the present day, I will study two verbs of speech, *goftan* “to say” and *āmuxtan* “to teach”, and two verbs of transfer, *dādan* “to give” and *ferestādan* “to send”. In the oldest texts, the recipient is encoded by the postposition *rā* (1), whereas in contemporary Persian it is introduced by the preposition *be* “to” (2).

(1) *mār=rā*    *guft*                    “...”  
snake=D    say.PST.3SG  
“He said to the snake: ...”                    (‘Atīq Sūrābādī. 11th c.)

(2) *be kazariān*    *goft-am*                    “...”  
to PN            say.PST-1SG  
“I said to Kazarian: ...”                    (Zoyā Pirzād. 2004)

Nevertheless, in some dialectal varieties of contemporary Persian, such as Tajik, the postposition *rā* is still used to encode the recipient of some ditransitive verbs (3). It will then be fruitful to study these variations in other dialectal varieties, such as the Persian spoken in Sistan, Giv, Kabul and Hazarajat, and also in Tat (another Iranian language close to Persian).

(3) Tajik (Perry 2005)  
*rahim=ro*    *guft-and*            *ki*            *berun*    *bar-oy-ad*  
PN=D            say.PST-3PL    that            out            PREV-come.PRS-3SG  
“They told Rahim to come out.”

I also argue that the overall change in the marking of the recipient is intrinsically linked to the general evolution in the use of the postposition *rā*. In the 10th century, *rā* has Accusative and Dative values in almost identical proportions (for morphemes marking both Accusative and Dative, see other languages in Aissen 2003). However, *rā* has gradually lost its Dative value in favour of its Accusative value, which accounts for almost 80% of usage in texts after the 14th century. There is evidence to suggest that this shift is due to the increase in complex predicates, which create ambiguity between the Accusative and Dative values. In

(4), *hēč āfarīda=i digar=ē=rā* “any other creature” can be 1° the recipient of *dādan* “to give” (and the unmarked patient is *rāh* “the way”) or 2° the patient of the complex predicate *rāh dādan* “to let pass”. We will see that these cases of ambiguity also explain the attempt in the 15th century to use the circumposition *mar...rā* to indicate the Dative value of a phrase (Lenepveu-Hotz 2018).

(4) *az īn sabab hēč āfarīda=i digar=ē=rā ba x<sup>w</sup>ēš*  
 from this reason any creature=EZ other=INDF=A/D to REFL  
*rāh na-tavān-ad dād*  
 way NEG-can.PRS-3SG give.INF

“For this reason, he cannot give the way to / let any other creature pass.” (Juvainī. 1260)

Keywords: Differential Object Marking, Persian, Recipients, Dative/Accusative functions

#### References:

- Aissen, J. (2003). Differential Object Marking: Iconicity vs. Economy. *Natural Language and Linguistic Theory*, 21, 435–83.
- Bossong, G. 1985. *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Gunter Narr.
- Lazard, G. (1982). Le morphème *rā* en persan et les relations actanciennes. *Bulletin de la Société de Linguistique de Paris*, 77(1), 177–208.
- 1994. *L’actance*. Paris: Presses Universitaires de France.
- Lenepveu-Hotz, A. (2018). Specialization of an ancient object marker in the New Persian of the 15<sup>th</sup> century. In A. Korangy & C. Miller (Eds.), *Trends in Iranian and Persian Linguistics* (pp. 81–100). Berlin: Mouton De Gruyter.
- Perry, J. R. 2005. *A Tajik Persian reference grammar*. Leiden – Boston: Brill.

- Lazard, G. (1982). Le morphème *rā* en persan et les relations actanciennes. *Bulletin de la Société de Linguistique de Paris*, 77(1), 177–208.
- Bossong, G. 1985. *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Gunter Narr.
- Lazard, G. 1994. *L’actance*. Paris: Presses Universitaires de France.
- Aissen, J. (2003). Differential Object Marking: Iconicity vs. Economy. *Natural Language and Linguistic Theory*, 21, 435–83.
- Perry, J. R. 2005. *A Tajik Persian reference grammar*. Leiden – Boston: Brill.
- Lenepveu-Hotz, A. (2018). Specialization of an ancient object marker in the New Persian of the 15<sup>th</sup> century. In A. Korangy & C. Miller (Eds.), *Trends in Iranian and Persian Linguistics* (pp. 81–100). Berlin: Mouton De Gruyter.

## Spatial Applicatives in Georgian in Diachronic Perspective

Georgian is one of the Kartvelian languages, a family indigenous to the South Caucasus. Its uninterrupted literary tradition goes back to the V<sup>th</sup> century A.D., thus providing a large body of linguistic evidence, whereas its sisters (Megrelian, Laz, and Svan) are barely written and have been on linguistic record for only some 150 years now.

Applicatives in Georgian (c.f. Boeder 1969) feature several values, one of which is the so-called locative, or superessive/superlative applicative. To put it briefly, the locative applicative obligatorily marks the venue, either position of the state or target of the action, which is usually represented as an indirect object in the clause, cross-referenced on the verb and marked by the dative case in the NP:

- (1) *buz-i zroxa-s̄ (zurg-ze) a-ždeba.*  
 fly-NOM cow-DAT back-SUPERL VAL:SUP-X\_sits\_down  
 ‘The fly is sitting down **on** the cow(’s back).’
- (2) *buz-i zroxa-s̄ (zurg-ze) a-zis.*  
 fly-NOM cow-DAT back-SUPERESS VAL:SUP-X\_sits  
 ‘The fly is sitting **on** the cow(’s back).’

Here, *a-* is a pre-radical vowel (often referred to as “version marker”, “versionizer”, “valency operator”...) that contributes to the formation of applicatives *a-ždeba* and *a-zis* from the source verbs *ždeba* ‘X is sitting down’ and *zis* ‘X is sitting’ respectively. Similarly to its paradigmatic counterparts, the pre-radical vowel synchronically only appears on the finite verb, whereby neither participles nor masdars (verbal nominalizations) can carry it.

Following Q’ipšize (1914), Gippert (2000) provides an account in which he connects another pre-radical vowel (benefactive *u-*) with the prefixal part of the adjectival comparative/superlative grade circumfix, viz. *u-\_-es-*, thus proving that the use of such vowels is not diachronically restricted to the domain of the finite verb. Consequently, a recent study suggests that there exists a diachronic connection between the locative applicative marker *a-* in Georgian and the prefixal part in a number of “vowel-initial nouns based on verbal roots” (Tuite 2024), e.g. *a-dg-il-* ‘place’ (< *dg-* ‘stand, put’); *a-lag-* ‘place, position’ (< *lag-* ‘arrange’); *a-s-o* ‘limb, letter’ (< *sv-* ‘set, affix’); and in the triplet *a-xl-eb-s* ‘X touches Y’ – *s-a-xl-* ‘house, home’ – *a-xl-o-* ‘close, near,’ all derived from the same root *xl-*.

Probably due to certain reasons including a fairly firm slot-filling arrangement of the Kartvelian verb, the renowned etymologists (Klimov 1998; Fähnrich and Saržvelaže 2000; Fähnrich 2007) were reluctant to draw a connection between the locative applicative marker *a-* (*a-* in Georgian and Svan; *o-* in Megrelian and Laz) and the spatial preverb *a-* ‘up(ward)’ (*a(γ)-* in Georgian; (*g*)*e-* [< *o-*] and *o-* in Megrelian and Laz). However, certain works of late (e.g. Rostovtsev-Popiel 2012) demonstrated that these etymologists had sometimes appeared to be way too cautious (e.g. with respect to connecting Common Kartvelian *\*kwe* ‘down(ward)’ with Common Kartvelian *\*kwe* “affirmative”). Accordingly, another recent study (Chukhua and Rostovtsev-Popiel) acknowledges a possibility of a direct diachronic relationship between the spatial operators in issue and brings forth evidence that supports this claim.

This talk aims therefore to discuss possible scenarios of the development of spatial semantics in locative applicatives against the background of the semantics of spatial lexical preverbs in Georgian.



## Abbreviations

DAT – dative; NOM – nominative; SUP – superessive/superlative (on the verb); SUPERESS – superessive (on the noun); SUPERL – superlative (on the noun); VAL – valency operator.

## References

- Boeder, Winfried. 1969. Über die Versionen des georgischen Verbs. *Folia Linguistica* 2: 82–152.
- Chukhua, Merab and Alexander Rostovtsev-Popiel. 2024. Toward a Historical Grammar of Elevation in Kartvelian. To Appear in: *Voprosy jazykoznanja* 3.
- Fähnrich, Heinz. 2007. *Kartwelisches etymologisches Wörterbuch*. Leiden / Boston: Brill.
- Fähnrich, Heinz and Zurab Saržvelaḡe. 2000. *Kartvelur enata et'imologiuri leksik'oni* [Etymological Dictionary of the Kartvelian Languages]. Tbilisi: Pedagogical University Press.
- Gippert, Jost. 2000. The Formation of Comparatives in the History of Georgian. Part I: The Prehistory of Synthetic Comparatives. *Enatmecnierebis sak'itxebi* 1: 37–45.
- Klimov, Georgij. 1998. *Etymological Dictionary of the Kartvelian Languages*. (Trends in Linguistics Documentation. Vol. 16.) Berlin / New York: Mouton de Gruyter.
- Q'ipšize, Ioseb. 1914. *Grammatika mingrel'skago (iverskago) jazyka s xrestomatijeju i slovarëm* [Grammar of the Mingrelian (Iverian) Language with a Chrestomathy and Lexicon]. St. Petersburg: Imperial Academy of Sciences.
- Rostovtsev-Popiel, Alexander. 2012. Grammaticalized Affirmativity in Kartvelian. Frankfurt PhD Diss., Ms.
- Tuite, Kevin. 2024. Applicative Constructions in Kartvelian. In: Zuñiga, Fernando and Denis Creissels (Eds.). *Applicative Constructions in the World's Languages*. (Comparative Handbooks of Linguistics. Vol. 7.) Berlin / Boston: Mouton de Gruyter: 913–942.

### Une alternance en question : *avoir été* et *être allé* en français oral

Juliette Thuilier

Univ. Toulouse Jean Jaurès & CLLE

Gilles Corminboeuf

Univ. de Fribourg

Christophe Benzitoun

Univ. de Lorraine & ATILF

L'objectif de cette communication est d'étudier les conditionnements, notamment linguistiques, de l'alternance entre *avoir\_été* et *être\_allé* aux temps composés :

(1) j'**ai été** lui apporter un ballotin de pralines (CEFC)

(2) qu'est-ce qu'il **est allé** lui raconter encore ? (CEFC)

En (1) et (2), l'alternance est possible.

Si Vandeloise (2007) voit une nuance dans le profilage du procès, la plupart des travaux descriptifs n'identifient qu'une opposition de registre. Il ressort de l'étude de Sammons et al. (2015) sur le français parlé ontarien que la plupart des facteurs linguistiques sont non significatifs (complément prépositionnel vs infinitif, personne, temps verbal), à l'exception de la polarité (la négation attire *avoir\_été*) et, dans une moindre mesure, du nombre (le pluriel favorise *être\_allé*). Dans leurs données orales, Glikman et Patard (2022) observent une prédominance d'*avoir\_été* (53%) sur *être\_allé* (47%), mais moins nette que chez Sammons et al. (2015) (63,2% *avoir\_été* ; 36,8% *être\_allé*).

Nos données de français parlé sont issues de la partie orale du *Corpus d'Etude pour le Français Contemporain (CEFC)* (Benzitoun et al. 2016). Les critères de sélection de l'alternance que nous avons retenus diffèrent de ceux des études précédentes, dans la mesure où nous avons pris en compte les cas où le complément n'est pas réalisé segmentalement ou est réalisé pronominalement, ainsi que le cas particulier de *ça va* pour lequel certains locuteurs autorisent l'alternance aux temps composés. Nous avons trié les occurrences des deux verbes pour écarter les exemples d'*avoir\_été* qui n'autorisent pas l'alternance avec *être\_allé*. Nous avons ensuite annoté les données retenues, en caractérisant le verbe (*être* ou *aller*), le complément (GP, infinitif, pronom, absent), la personne, le nombre et l'animéité du sujet, ainsi que la présence/absence de la négation.

A l'heure actuelle, nous avons 379 occurrences triées. La distribution des deux formes verbales n'est pas concordante avec celles mentionnées dans la littérature (274 *être\_allé* (72,3%) et 105 *avoir\_été* (27,7%)), témoignant d'une possible volatilité du phénomène. Dans la sous-partie annotée (106 occ., 68,3% *être\_allé* / 31,7% *avoir\_été*), il n'y a aucun effet clair des variables linguistiques considérées. La polarité négative semble néanmoins légèrement favoriser *avoir\_été*, comme mentionné par Sammons et al. Mais nous disposons de trop peu de données pour statuer sur la significativité de cette tendance (10 *être\_allé* / 6 *avoir\_été*, soit 62,5% vs. 37,5%). Concernant le complément, conformément à la littérature, il n'y a pas de différence apparente selon que le complément est un GP ou un infinitif. Cependant, la réalisation pronominale semble attirer *être\_allé*, alors que l'absence du complément favoriserait *avoir\_été*. De nouveau, il faudrait plus de données pour confirmer ces premières tendances.

Pour la présentation au colloque, nous présenterons les résultats concernant près de **1600 occurrences annotées**. Cela nous permettra de statuer sur les premières tendances observées, et de documenter une éventuelle variation régionale. En effet, CFPB (Belgique) et CLAPI (région lyonnaise) présentent des distributions très différentes : CLAPI : 87,9% *être\_allé* / 12,1% *avoir\_été* ; CFPB 28,9% *être\_allé* / 71,1% *avoir\_été*.

#### Bibliographie

- Benzitoun C., Debaisieux J.-M., Deulofeu J. (2016), Le projet ORFÉO : un corpus d'étude pour le français contemporain, *Corpus* 15 ; 91-114 : <http://journals.openedition.org/corpus/2936>
- Glikman J., Patard A. (2022), *Être pour aller* en français d'Europe, *Congrès mondial de linguistique française (CMLF'22)*, publication électronique : [https://www.shs-conferences.org/articles/shsconf/abs/2022/08/shsconf\\_cmlf2022\\_02001/shsconf\\_cmlf2022\\_02001.html](https://www.shs-conferences.org/articles/shsconf/abs/2022/08/shsconf_cmlf2022_02001/shsconf_cmlf2022_02001.html)
- Sammons O., Nadasdi T., Mougeon R. (2015). 'Moving' through the past: Thirty years of *avoir été* in Ontario French, *Journal of French Language Studies* 25-3, 397-422.
- Vandeloise C. (2007). Le verbe ALLER : l'affranchissement du contexte d'énonciation immédiat, *Journal of French Language Studies* 17, 343-359.

When corpus linguistics helps solving a theoretical puzzle: the case of French relative *qui*

How to analyze the French *qui* appearing as a subject in (1) ?

- (1) L'homme qui est venu  
The man who came

The current admitted analysis taken up in the reference French Grammar GGF (Abeillé & al. 2021) is based on Kayne's (1972) proposal that this *qui* is in fact a complementizer and not a relative pronoun (see also Blanche-Benveniste 1980). The arguments are both morphologic and syntactic.

As for morphology Kayne points out that this *qui* doesn't show up basic pronominal properties (sensitivity to the [+/- human] distinction, phonetic reduction, incompatibility with prepositions). These arguments are convincingly challenged by Sportiche (2011). He points out that this *qui* can be aligned as a nominative weak relative pronoun with well admitted personal (clitics) pronouns, according to table 4 (2011 : p.96)

**Table 4**  
Extended comparison of the pronominal system and the tensed relative pronoun system

Case		Nom	Acc	Dat	Gen	Loc	Elative	Temp	Manner	Cause Reason
<i>Pronouns</i>										
Strong	[+human]	lui	lui	(à) lui	(de) lui	X	X	X	X	X
Strong	[-human]	ça	ça	(à) ça	(de) ça	là	de là	X	X	X
Weak	[0human]	il	le	lui	en	y	en	X	X	X
<i>Tensed restrictive relative pronouns</i>										
Strong	[+human]	<i>qui</i>	<i>qui</i>	à <i>qui</i>	de <i>qui</i>	X	X	X	X	X
Strong	[-human]	<i>quoi</i>	<i>quoi</i>	à <i>quoi</i>	de <i>quoi</i>	où	d'où	<i>quand</i>	<i>comment</i>	<i>pourquoi</i>
Weak	[0human]	<i>qui</i>	<i>que</i>	X	dont	où	dont	où	dont	%e <sup>a</sup>
C		* <i>que</i>	* <i>que</i> (?)	* <i>que</i>	* <i>que</i>	* <i>que</i>	* <i>que</i>	* <i>que</i>	% <i>que</i>	% <i>que</i>

<sup>a</sup> e denotes a silent form.

On the basis of Occam's razor, this solution by including the *qui* subject in relatives in an independent motivated existing category, avoids positing the ad hoc category of a French inflected complementizer. From a syntactic point of view, Kayne analysis is based on subject movement, which causes a wh-gap, leaving alone the complementizer *que* in the C constituent. The resulting structure is ungrammatical in French, which does not allow subject less finite clauses. The marker "-i" helps filling the gap in a somewhat ad-hoc way. It appears clearly that a no movement subject relative clause analysis alternatively proposed by Sportiche eliminates unnecessary machinery.

Nevertheless, if we extend the analysis to wider syntactic contexts, Kayne's proposal seems to prevail. Looking at the following examples :

- 2.a L'homme à qui je crois qu'il il faut parler  
2b homme que je crois qui a parlé  
2c \*l'homme qui je crois que a parlé

*Qui* in 2b appears in a prohibited position for relatives pronouns. Kayne's solution explains this fact straightforwardly, as he analyses *qui* as an inflected complementizer. Sportiche, referring to Koopman and Sportiche (2008) puts forward an alternative solution preserving *qui* as a weak relative. He proposes that cases of apparent long extraction of subjects actually

be cases of short extraction of the subject of a pseudorelative small clause, much like what happens in “*il voyait Jean qui dormait* (he saw John sleeping)”. Thus, we have the parallelism in (3)

(3) a. le garçon qu’il voyait [t qui dormait]

The boy he was seeing who was sleeping

b. le garçon qu’il disait [t qui dormait]

The boy he was saying who was sleeping

The problem is, as noticed by Melis (1988), that the small clause structure is not available for all the verbs that are possible in 3b, for instance volition verbs :

(4) \*Je le veux qui assiste à la conférence

L’homme que je veux qui assiste à la conférence

The man that I want that he attend the conference

Sportiche brings about a theory internal solution based on selected interventions effects allowing special case marking of extracted constituents.

In fact, the theoretical discussion comes short because it is undermined as suggests Sportiche himself by the lack of a consistent empirical basis (p. 109) : “No systematic empirical study has been made of the variation found in this area, and especially of how judgments correlate across paradigms and constructions”.

Our purpose in this presentation is to implement a consistent collection of authentic data by means of the French platform Orfeo that allows to query an oral corpus of more than 4 million words.

This empirical study will allow us to compare the lexical and structural constraints bearing on structures in (2) (3) and (4) with those already established for other long distance dependencies phenomena ( Deulofeu & Valli 2007, Berard & allii 2014), in order to decide if structures (2 b) are a subcase of them or a specific construction.

## References

- Abeillé A & Godard D, 2021, *La Grande Grammaire du Français*,: Chapitre XIII, ‘Les subordonnées relatives’ pp. 1469 à 1537
- Bérard L., Deulofeu J., Kahane S., 2014, A corpus-based study of bridge verbs in long distance dependencies in French: a specific construction. *Chimera: Romance Corpora and Linguistic Studies* 1, pp.137 - 155.
- Blanche-Benveniste, C. 1980. Divers types de relatives en français parlé, *TA Informations* 21 : 2, 16-25.
- Deulofeu J, & Valli, A. 2007, « Sur l’aspect normatif des descriptions linguistiques en français. Quels faits de langue faut-il retenir pour une description grammaticale satisfaisante ? » In Gilles Siouffi & Agnès Steuckard (eds) *Les linguistes et la norme*, Peter Lang, Berne
- Godard, D. 1988. *Syntaxe des relatives en français*, CNRS editions, Paris.
- Kayne, R. 1974-75. French relative que, *Recherches Linguistiques*, II 40-61 et III 27-92.
- Koopman, Hilda, and Dominique Sportiche. 2008. The que/qui alternation: New analytical directions. Ms., UCLA, Los Angeles, CA.
- Melis, L, 1988, Les propositions enchevêtrées, des complétives ou des relatives-interrogatives *Neuphilologische Mitteilungen* Vol. 89, No. 2, pp. 189-196
- Sportiche D, 2011, French relative qui, *Linguistic Inquiry*, Volume 42, Number 1, Winter 2011 83–124

Corpus oraux et variation diatopique  
Documenter le français parlé à Bruxelles  
Anne Dister  
UCLouvain – Saint-Louis - Bruxelles

Le *Corpus de français parlé à Bruxelles* (CFPB) a vu le jour en 2013 (Dister et Labeau 2017) ; il se veut le pendant bruxellois du *Corpus de français parlé parisien* (CFPP200, Branca-Rosoff *et al.* 2011). En effet, moyennant quelques adaptations, les intervieweurs utilisent le même guide d'entretien semi-dirigé dont le thème principal est celui de la relation des habitants à leur quartier, et plus généralement leur ville.

Le CFPB vient assurément combler un vide. En effet, il n'existait pas jusqu'à sa mise en place de corpus académique, comme c'est le cas par exemple à Montréal (Sankoff *et al.* 1976), documentant le français parlé dans la capitale belge. Dans cette communication, nous verrons notamment le terrain particulier, du point de vue linguistique, que constitue Bruxelles, ville officiellement bilingue, mais aussi ville multiculturelle, capitale des institutions européennes.

Références citées

- DISTER Anne et LABEAU Emmanuelle (2017). « *Le Corpus de français parlé à Bruxelles : origine, hypothèses et développements* », *Cahiers de l'AFLS* 21 (Association for French Language Studies).
- BRANCA-ROSOFF Sonia, FLEURY Serge, LEFEUVRE Florence et PIRES Matthew (2011). *Constitution et exploitation d'un corpus de français parlé parisien*, *Corpus* 10, pp. 81-98.
- SANKOFF David, SANKOFF Gillian, LABERGE Suzanne et TOPHAM Marjorie (1976). « Méthodes d'échantillonnage et utilisation de l'ordinateur dans l'étude de la variation grammaticale », *Cahiers de linguistique*, n° 6, pp. 85-125.

OFROM+ : documenter la variation du français en Suisse romande et dans la francophonie  
Mathieu Avanzi & Laure Anne Johnsen, Université de Neuchâtel

L'archive de parole OFROM (corpus oral du français en Suisse romande) a été mise au point dans le but de décrire les pratiques linguistiques de francophones bien ancrés en Suisse romande. Avec la récolte initiée en 2007, puis l'ouverture du site au grand public quelques années plus tard, la base a non seulement permis de disposer du premier corpus de français oral de Suisse romande, mais aussi de rattraper, grâce aux 1,3 millions de mots alignés texte/son qu'elle contient actuellement, le retard pris par rapport à d'autres pays francophones. Dans cette présentation, nous ferons état des derniers développements qu'a connus la base, à savoir : (i) ouverture à la francophonie, avec l'inclusion d'enregistrements réalisés en France, en Belgique mais aussi dans d'autres régions du monde (Antilles, Afrique), (ii) recueil et traitement de messages vocaux échangés sur WhatsApp, (iii) prise en compte de données produites dans d'autres contextes de parole (sermons, conférences, parole radiophonique, etc.). Nous présenterons les outils développés et une ou deux études de cas permettant de montrer ce que la base nous apprend sur le français parlé dans une société décentralisée.

**Marqueurs de haut degré et variation régionale :**  
**État des lieux des contraintes sémantico-syntaxiques sur les MDH en usage en Suisse romande**

Le haut degré peut être exprimé par un *intensifieur*, s'il modifie un terme gradable (adjectif, adverbe, nom, etc.) ou par un *quantifieur*, si ce degré concerne une quantité. Toutefois, cette frontière est souvent perméable (Gaatone, 2008, 2013). J'utiliserai donc par la suite le terme proposé par Dostie (2018) *marqueur de haut degré* (MHD) qui subsume les deux notions.

Dans les langues, les MHD sont très largement soumis à la variation diastratique, diaphasique, générationnelle (Zribi-Hertz, 2015) et régionale. C'est ce dernier aspect qui sera au centre de la présentation.

Souvent, la variation régionale est traitée d'un point de vue purement lexical ou au contraire d'un point de vue purement syntaxique. Pour les MHD, la variation régionale se manifeste naturellement par différents items lexicaux propres à une région, mais la particularité de ces MDH est que leur distribution diffère de celle des MDH standard. Ainsi, Corminboeuf & Avanzi (2020) mettent en avant que *monstre* peut avoir entre autres la fonction de spécifieur (1), d'ajout au nom (2), d'ajout au verbe (3) et d'ajout à l'adjectif (4).

(1) *Verbier les canonniers font monstre neige*

(2) *y'a la monstre fumée qui sort*

(3) *Je me suis monstre pris la tête*

(4) *elles coutent monstre cher*

Cette étude se focalise sur les MDH utilisés en Suisse romande, région dans laquelle plusieurs MHD régionaux sont en compétition, et repose sur une étude de corpus et une série d'expériences.

**Étude de corpus**

J'ai extrait du corpus OFROM (Avanzi et al., 2016) toutes les occurrences des MHD *monstre* (34), *pire* (10), *une chiée* (23), *beau* (12), *fin* (7) et *franc* (4). Si ces données écologiques permettent d'affirmer qu'un emploi existe dès qu'il est présent dans le corpus, l'absence en corpus est un maigre indice de l'agrammaticalité, tant le nombre total d'occurrences faible. En effet, Corminboeuf & Avanzi (2020) donne des occurrences attestées sur twitter dans lesquelles *monstre* a la fonction de spécifieur (1), et une recherche en ligne permet d'affirmer que *une chiée* peut modifier un adjectif (5), alors que ces configurations sont absentes d'OFROM (cf. fig. 1).

(5) *ça aurait fait une chiée belle scène de film*

Par ailleurs, il semble qu'une forte contrainte sémantique, voire lexicale pèse sur la tête des MHD *beau*, *fin* et *franc*. Mais là encore, la petite poignée d'occurrences trouvée ne permet pas de définir avec finesse les têtes possibles pour ces MHD.

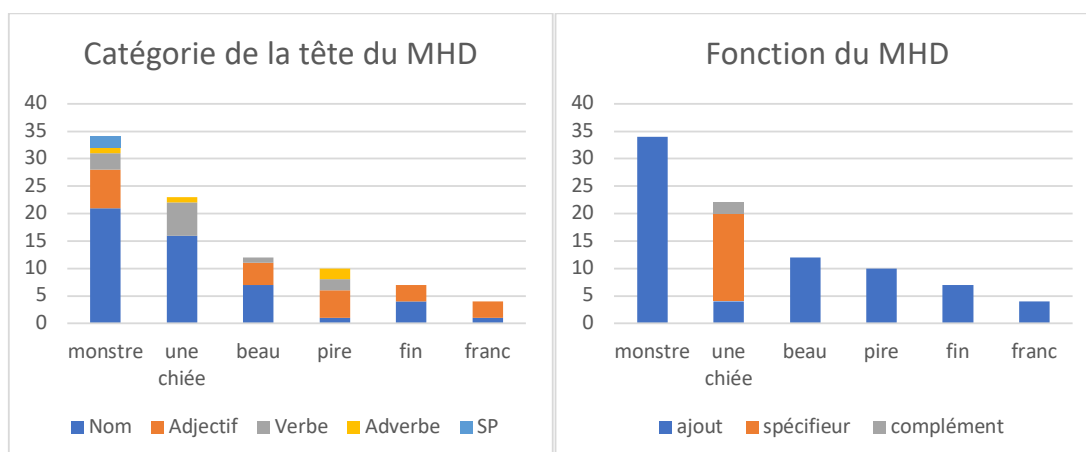


Figure 1 : Catégorie de la tête et fonction du MHD sur OFROM

### Études expérimentales

Pour pallier les limites de l'étude précédente, deux expériences de jugement d'acceptabilité sont sur le point d'être réalisées. Cette méthodologie est encore très rare pour documenter la variation régionale (Dagnac & Thuilier, 2020). La première expérience vise à déterminer les contraintes sémantiques pesant sur les têtes des MHD mentionnés au paragraphe précédent. La deuxième expérience vise à déterminer l'acceptabilité des différents MHD selon la fonction qu'ils occupent.

### Bibliographie

- Avanzi, M., Béguelin, M.-J., & Diémoz, F. (2016). De l'archive de parole au corpus de référence : La base de données orales du français de Suisse romande (OFRM). *Corpus*, 15. <https://journals.openedition.org/corpus/3060>
- Corminboeuf, G., & Avanzi, M. (2020). Monstre intensifieur en français. *Congrès Mondial de Linguistique Française CMLF 2020*, 78. <https://hal.science/hal-03321628/>
- Dagnac, A., & Thuilier, J. (2020). Français du Sud-Ouest et marquage différentiel de l'apostrophe : Une approche expérimentale. *L'Information Grammaticale*, 166, 29-35.
- Dostie, G. (2018). *Synonymie et marqueurs de haut degré. Sens conceptuel, sens associatif, polysémie*. Classiques Garnier. <https://classiques-garnier.com/synonymie-et-marqueurs-de-haut-degre-sens-conceptuel-sens-associatif-polysemie.html>
- Gaetone, D. (2008). Un ensemble hétéroclite : Les adverbes de degré en français. *Congrès mondial de linguistique française*, 216. <https://www.linguistiquefrancaise.org/articles/cmlf/abs/2008/01/cmlf08125/cmlf08125.html>
- Gaetone, D. (2013). Esquisse d'un guide des perplexes : Problèmes de définition et de classification des adverbes de degré en français. *Langue française*, 1, 37-50.
- Zribi-Hertz, A. (2015). De l'évolution des propriétés du mot grave en français européen moderne. *La dia-variation en français actuel. Études sur corpus, approches croisées et ouvrages de référence*, 63-98.



## La recherche sur le français québécois et la mutualisation des ressources

Le français québécois a fait l'objet d'innombrables travaux depuis l'institutionnalisation de la linguistique au Québec dans les années 1960 et 1970 (pour un aperçu, voir Dostie 2020 et Martineau, Remysen et Thibault 2022). Ces travaux ont été réalisés en grande partie à partir de corpus oraux et textuels dont certains sont devenus au fil du temps des références importantes, voire incontournables (voir Boisvert et Laurendeau 1988). Que l'on pense par exemple au corpus sociolinguistique Montréal 1971 (à l'origine d'une série de trois corpus réalisés en 1971, 1984 et 1995 auprès des mêmes personnes, ouvrant ainsi la porte à des analyses longitudinales ; voir Sankoff 2018) ou encore à la Banque de données textuelles de Sherbrooke (réalisé entre 1977 et 2005, ce corpus est à l'origine d'une série de travaux en statistique lexicale avant d'avoir servi de base à la rédaction du dictionnaire *Usito* ; voir Martel et Cajolet-Laganière 2004).

Constitués pour répondre à des objectifs de recherche précis, ces corpus présentent un potentiel d'exploitation important qui va au-delà des intentions initiales de leurs créatrices et créateurs. C'est sans compter que certains ensembles de données, notamment (mais pas exclusivement) ceux qui ont été constitués avant l'ère numérique, se trouvent parfois à l'heure actuelle dans un état fragile qui compromet leur survie ou sont de manière générale difficiles d'accès. Mis sur pied dès 2020, le Fonds de données linguistiques du Québec a pour objectif de mutualiser certaines des ressources constituées pour l'étude du français au Québec et de faciliter leur accès et leur consultation (voir Remysen et Saint-Amant Lamy à paraître). Plus qu'un simple dépôt de corpus, le Fonds répond plus particulièrement à une triple vocation, soit une vocation scientifique, une vocation patrimoniale et culturelle ainsi qu'une vocation pédagogique et sociale. Il tire profit des avancées que permettent les approches les plus récentes dans le domaine des humanités numériques (voir par exemple Citton 2015 ; Dacos et Mounier 2015) et s'inscrit dans la mouvance des données ouvertes et du partage de contenus numériques suivant les principes FAIR (Wilkinson *et al.* 2016).

Dans le cadre de cette communication, nous reviendrons sur les principales motivations qui animent ce projet et sur les possibilités d'analyse qu'il permet dans le domaine de la variation diatopique du français. Nous présenterons par ailleurs les principales orientations théoriques, méthodologiques et techniques qui sous-tendent la constitution du Fonds. Nous prendrons à témoin deux types de corpus, oraux et écrits, pour illustrer certaines des fonctionnalités déjà disponibles sur la plateforme du projet pour consulter et interroger les données. Nous reviendrons enfin sur les principaux défis que soulève la réalisation du Fonds (notamment sur le plan du traitement des données et de l'interopérabilité), sur son rôle en tant qu'outil de découvrabilité et sur les futurs développements projetés.

## Références

- Boisvert, Lionel et Paul Laurendeau (1988), « Répertoire des corpus québécois de langue orale », *Revue québécoise de linguistique*, vol. 17, n° 2 (*Problèmes de lexicographie en Amérique*), p. 241-259.
- Citton, Yves (2015), « Humanités numériques : une médiapolitique des savoirs encore à inventer », *Multitudes*, n° 59, p. 169-180.
- Dacos, Marin et Pierre Mounier (2015), *Humanités numériques : état des lieux et positionnement de la recherche française dans le contexte international*, Paris, Institut français.
- Dostie, Gaétane (dir.) (2020), *Combats pour la linguistique au Québec (1960-2000) : courants, théories, domaines*, Bruxelles, P.I.E. Peter Lang.
- Martel, Pierre et Hélène Cajolet-Laganière (2004), « L'apport de la Banque de données textuelles de Sherbrooke : des nomenclatures enrichies », dans Louis Mercier (dir.), *Français du Canada – Français de France VI. Actes du sixième Colloque international d'Orford, Québec, du 26 au 29 septembre 2000*, Tübingen, Max Niemeyer, p. 263-277.
- Martineau, France, Wim Remysen et André Thibault (2022), *Le français au Québec et en Amérique du Nord*, Paris, Éditions Ophrys.
- Remysen, Wim et Hugo Saint-Amant Lamy (à paraître), « Les corpus oraux montréalais à l'ère numérique : des bobines magnétiques à la diffusion sur la plateforme du Fonds de données linguistiques du Québec », dans Hélène Blondeau, Marty Laforest et Wim Remysen (dir.), *(Re)donner la parole aux corpus montréalais (1971-2021) : regards rétrospectif et prospectif*, Montréal, Presses de l'Université de Montréal.
- Sankoff, Gillian (2018), « Before there were corpora : the evolution of the Montreal French Project as a longitudinal study », dans Suzanne Evans Wagner et Isabelle Buchstaller (dir.), *Panel studies of variation and change*, New York, Routledge, p. 21-51.
- Wilkinson, Mark. D., Michel Dumontier, IJsbrand Jan Aalbersberg, et al. (2016), « The FAIR Guiding Principles for scientific data management and stewardship », *Scientific Data* 2016, n° 3. <https://doi.org/10.1038/sdata.2016.18>

## À la recherche de la syntaxe du breton contemporain : enquête auprès des locuteurs natifs

Parle-t-on le même breton d'une génération à l'autre et d'un territoire à l'autre ? C'est à cette question que l'enquête cherche à répondre, s'inscrivant de ce fait dans la lignée des différents atlas linguistiques, notamment l'*Atlas linguistique de la Basse-Bretagne*, publié en 1927, par Pierre le Roux, *Le Nouvel Atlas linguistique de la Basse-Bretagne*, publié en 2001, sous la direction de Jean Le Dû, ainsi que de l'*Atlas linguistique de la région de Pontivy*, publié en 2004 par Pascal Rolland. Mais à l'inverse de ces précédents volumes qui dressèrent un portrait détaillé du lexique breton, la présente étude s'attache, elle, à saisir l'évolution de la syntaxe du breton tel qu'il est parlé aujourd'hui, en mettant en perspective l'ensemble des territoires de la Basse-Bretagne, ainsi qu'en comparant les locuteurs issus de deux générations différentes : ceux nés avant 1960 et ceux nés après. Chaque locuteur participant à l'enquête a été élevé en breton, et ce de manière au moins équivalente au français, ou à une autre langue. Aussi, le breton est-il sa ou une de ses premières langues de communication. L'enquête se fonde sur une collecte de terrain, directement auprès des locuteurs natifs, individuellement soumis à un protocole constitué de différentes séries de questions et de stimuli visuels destinés à capter la syntaxe bretonne telle qu'ils l'ont intériorisée. Leurs réponses, enregistrées et conservées dans une base de données sonores, sont ensuite étudiées grâce à une grille d'analyse fondée sur les principales caractéristiques de la syntaxe bretonne. Les résultats sont notamment accessibles par le biais d'une carte interactive permettant de naviguer à travers les différentes zones géographiques et les deux générations représentées. Cette enquête se révèle des plus éclairantes pour comprendre les évolutions et changements affectant le breton du début du XXI<sup>e</sup> siècle à une époque où, contrairement à celles au cours desquelles ont été publiés certains des précédents atlas, plus aucun locuteur natif du breton n'a pu connaître un environnement et une Bretagne majoritairement brittophones. Assiste-t-on réellement à une unification de la syntaxe bretonne et à une influence grandissante du français et d'autres langues éventuelles ? Les différences générationnelles et territoriales sont-elles de plus en plus marquées ou, au contraire, difficilement perceptibles ? Les premiers résultats de cette enquête en cours permettent déjà d'apporter des éléments de réponse.

## Differential Object Marking in Southwestern French: Effects of Local factors

Anne Dagnac<sup>12</sup>

Juliette Thuilier<sup>12</sup>

Iona Brédif<sup>3</sup>

<sup>1</sup> Université Toulouse Jean Jaurès

<sup>2</sup> CLLE (CNRS/UT2J)

<sup>3</sup> LNLP (UT2J)

Previous works on French DOM present it either as inexistant or highly marginal (Binder & Stark 2018, 2023), rare (Fagard & Mardale 2014), common (Blanche-Benveniste 1997) or fully standard in Toulouse (Séguy, 1973). Our general aim is to sort out and document this phenomenon and its fine-grained properties in Southwestern French, in order to replace it in the typological and theoretical debates (for Romance languages, see e.g. Irimia & Mardale 2023, and references herein).

This paper specifically assesses the relative weight of animacy and syntactic structure in the acceptability of Southwestern French Differential Object Marking such as (1).

- (1) Je I' aime pas à lui  
 I PRO.3SG.ACC like NEG to PRO.3SG  
 'I don't like him'

Fagard and Mardale (2014), building on previous typological work, list factors influencing its acceptability: "local factors" like the animacy of the referent and the form of the *à*-PP, and "global factors" like the type of lexical verb involved and discourse structure (topicalisation, contrast).

Experimental work to assess the respective impact of animacy *vs.* dislocation (Dagnac & Thuilier, 2020) has shown that animacy impacts acceptability, though in a lesser way than discourse structure. This paper further focuses on the local factors, i.e., animacy and the type of structure following *à*, and adapts Dagnac & Thuilier's (2020) experimental design that take into account the non-normative and regional nature of such structures. To this end, we set up a 7-point Likert scale judgement task. The first variable, Animacy, has two values ([+animate], [-animate]); the second one, *à*-PP\_form, has 3 values (Personal Pronoun, Demonstrative Pronoun, Demonstrative NP). The other factors are controlled. The 6 conditions are illustrated in (2), where the target is B's answer:

(2) Item types

A: On a décidé de l'accueillir dans la troupe, **le poète argentin** ; tu sais celui qui s'inspire de l'Illiade.

B: Ah bon ? Moi, je l'apprécie moyen **à lui.** Condition A

B: " " **à celui-là.** Condition B

B: " " **à ce poète.** Condition C

A: On a décidé de **le** réciter dans la pièce, **le poème argentin** ; tu sais celui qui s'inspire de l'Illiade.

B: Ah bon ? Moi, je l'apprécie moyen **à lui.** Condition D

B: " " **à celui-là.** Condition E

B: " " **à ce poème.** Condition F

We hypothesize that both variables impact the judgements, *à*-PP\_form more strongly than Animacy, yielding the following hierarchy: Condition A [+animate, PersPRO]> Condition B [+animate, DemPRO]> condition E [-animate, DemPRO]> Condition C [+animate, DemNP]> condition [-animate, DemNP]. Condition D [-animate, PersPRO] is expected to be

independently degraded due to the animacy constraint on P3 personal strong pronouns (Abeillé & Godard 2021: 1061-1062).

The experiment was conducted with 38 participants (17-30 y.o., from Occitanie and Nouvelle Aquitaine), each judging 24 experimental items (latin square repartition) and 36 fillers. Fig. 1 gives the distribution of their judgements. We built a mixed-effects ordinal regression to model the participants' judgements, with Animacy and  $\grave{a}$ -PP\_form as fixed effects and Participants and Items as random effects. The main results confirm the hypothesis: inanimate referents lower the probability of having a high rate (coef. -0.56,  $p < .0001$ ), whereas demonstrative (coef. 0.52,  $p < .0001$ ) and personal (coef. 0.65,  $p < .0001$ ) pronouns increase it. The variable effects for each rate are plotted in Fig. 2. However, our prediction on condition D [-animate, PersPRO] is not borne out: participants don't strongly rule out [-animate] personal pronouns – as strongly as they rule out the 8 agrammatical fillers (mean = 2.14) – a fact that may be consistent with empirical observations of on-going diachronic change.

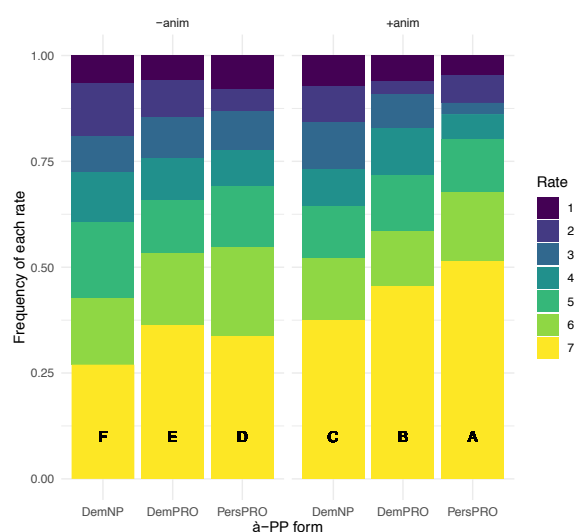


Fig. 1: Distribution by condition



Fig 2. Predictors effects in ordinal regression model

## References

- Abeillé, A. & Danièle G. (Eds.). (2021). *Grande Grammaire du Français*. Arles: Actes Sud, Paris: Imprimerie nationale éditions, 2 volumes, 2 628 p.
- Binder L. & Stark E. (2023), Differential object marking in French: Myth or reality?. In M.A. Irimia & A. Mardale (Eds.), *Differential Object Marking in Romance : Towards microvariation* (pp. 56-84). John Benjamins.
- Binder L. & Stark E. (2018), Differential object marking in French: Myth or reality?, présentation au colloque *Micro- variation dans le marquage différentiel de l'objet roman*, INaLCO, Paris, 9-10 novembre 2018.
- Blanche-Benveniste, C. 1997). *Approches de la langue parlée en français*. Paris:Ophrys.
- Dagnac, A. & Thuilier, J. (2020), Marquage différentiel de l'objet en Français du Sud-Ouest : une approche expérimentale, *L'Information grammaticale*, 166, 29-35.
- Fagard, B. & Alexandru Mardale. (2014). 'Non, mais tu l'as vu à lui ?' Analyse(s) du marquage différentiel de l'objet en français, *Verbum* 26 (1),145-170.
- Irimia, M. A. & Mardale, A. (Eds.). (2023). *Differential Object Marking in Romance: Towards microvariation*. Amsterdam/Philadelphie: John Benjamins.
- Rohlf, G. (1971). Autour de l'Accusatif prépositionnel dans les langues romanes (concordances et discordances). *Revue de linguistique romane*, 35, 312-334.
- Séguy, J. (1973). L'accusatif prépositionnel en gascon et dans le français du sud-ouest, *Travaux de linguistique et de littérature*, 11(1),. 429-433.

**Conférence LLcD 2024 : Langues, variation, changements : approches synchroniques, diachroniques, typologiques et comparées (session générale)**

**Langage pictural et langues des signes :  
étude de degrés d'iconicité et d'invariants translangagiers**

Thomas Sähn

Saghie Sharifzadeh

*Sorbonne Université – CeLiSo*

*Sorbonne Université – CeLiSo*

Grâce à un dialogue entre morphosyntaxe, sociolinguistique et sémiotique visuelle, nous proposons de mettre en regard le langage visuel propre aux représentations picturales avec les langues visuo-gestuelles que sont les langues des signes.

Selon la tripartition peircienne, il existe trois grandes familles de signes se distinguant par le lien qu'ils entretiennent avec leur référent : les indices, les icônes et les symboles (cf. Peirce 2017/1931 : 2.243<sup>ssq.</sup>). En reproduisant certains traits du référent, l'icône, ou signe iconique, entretient un lien « motivé » avec son référent mais la qualité et le choix des traits sélectionnés est nécessairement arbitraire (cf. Eco 1978 : 153 ; Groupe  $\mu$  1992 : 138). Grâce à notre expérience du monde, nous avons en effet stabilisé, pour chaque objet réel ou imaginaire, des traits le caractérisant et le distinguant des autres objets. Plus un signe iconique partage de tels traits (proto)typiques avec son référent – pensons par exemple à un dessin dit 'réaliste', une photo ou une vidéo –, plus son degré d'iconicité est élevé. À l'inverse, plus nous réduisons le nombre de ces traits communément partagés, plus les signes nécessitent l'apprentissage d'un code spécifique et se rapprochent du symbole, dont fait partie la majorité des signes linguistiques des langues vocales. Le degré d'iconicité peut donc varier d'un signe iconique à l'autre, sans que les icônes soient l'apanage des énoncés picturaux. Ce lien particulier caractérise des formes d'énoncés très diverses, du bruitage dans le domaine audio-visuel aux livres tactiles pour enfants ou même aux onomatopées des langues vocales (cf. Klinkenberg 1996 : 377-378). Si ces dernières constituent le mode d'expression symbolique par excellence, il en est tout autrement des langues des signes, langues naturelles, intrinsèquement liées à un processus d'iconicisation de l'expérience (cf. Cuxac 1998 : 86-87).

La littérature a mis en évidence l'importance de l'iconicité dans la formation du lexique des langues des signes (cf. Taub 2000 ; Perniss *et al.* 2010). Afin de mieux de comprendre l'articulation des unités signifiantes dans ces dernières, nous proposons une analyse comparée des degrés d'iconicité des langues des signes et du langage pictural. Nous appuyant sur le lexique de la langue des signes française (LSF), nous montrerons dans un premier temps que les langues singées se situent au centre d'un continuum allant du symbolique pur à un degré d'iconicité maximal, et qu'elles partagent en cela des caractéristiques communes aussi bien avec les langues scripturo-vocales qu'avec le langage pictural. Nous retrouvons cette particularité à tous les niveaux d'articulation des unités de la LSF, permettant de passer d'unités minimales constitutives de ses lexèmes à des syntagmes puis des énoncés plus complexes. Aussi proposerons-nous dans un second temps un classement de ces unités signifiantes en fonction des liens qu'elles entretiennent avec leur référent, dans les langues des signes et dans le langage pictural, mettant ainsi en lumière des phénomènes de variation lexicale dans l'espace

et dans le temps. Si l'intégration de ces unités au sein de récits en langues des signes est régie par des propriétés communes à la fois avec le langage pictural et avec les autres langues, nous montrerons dans un dernier temps que les langues signées se démarquent de ces derniers par l'existence d'invariants translinguistiques propres, reflet de mécanismes d'iconicisation indépendants de tout ancrage culturel et/ou linguistique.

**Mots clés :** Langues des signes, iconicité, langage pictural

### **Références bibliographiques citées**

Cuxac, Christian. (1998). Construction de références en Langue des Signes Française. *Sémiotiques*, 15, 85–105.

Eco, Umberto. (1978). Pour une reformulation du concept de signe iconique. *Communications*, 29, *Image(s) et culture(s)*, 141–191.

Groupe  $\mu$ . 1992. *Traité du signe visuel. Pour une rhétorique de l'image/*. Paris : Le Seuil.

Klinkenberg, Jean-Marie. 1996. *Précis de sémiotique générale*. Bruxelles : De Boeck.

Peirce, Charles Sanders. 2017/1931. *Écrits sur le signe* [traduit de l'anglais par Gérard Deledalle]. Paris : Le Seuil.

Perniss, Pamela, Robin L. Thompson, & Gabriella Vigliocco. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, 1, 1–15.

Taub, Sarah. 2001. *Language from the body: Iconicity and metaphor in American sign language*. Cambridge : Cambridge University Press.



## Prototypicalité des arguments et animéité dans le domaine postverbal du français

Juliette Thuilier<sup>12</sup>  
Cunha<sup>4</sup>

Pegah Faghiri<sup>3</sup>

Léa Pineau<sup>2</sup>

Yanis da

<sup>1</sup> Université Toulouse Jean Jaurès

<sup>2</sup> CLLE (CNRS/UT2J)

<sup>3</sup> University of Amsterdam

<sup>4</sup> LLF (CNRS/UPC)

Les travaux en psycholinguistique et typologie ont mis en avant le rôle des hiérarchies de proéminence dans l'organisation de la phrase (Lamers & de Swart, 2012 ; Haspelmath 2020). Selon la hiérarchie d'animéité, les référents humains ou animés tendent à être placés dans les positions syntaxiques les plus saillantes (Bock & Warren 1985, Bresnan et al 2007 pour l'anglais) ou linéairement avant les référents inanimés (Branigan et al 2008 pour une revue). Les récents travaux sur le français montrent que la fonction sujet attire effectivement les référents animés et humains (da Cunha & Abeillé 2022 sur l'alternance passive), mais que l'organisation du domaine postverbal ne suit pas la tendance attendue : Thuilier et al (2021) observent un effet d'anti-animéité selon lequel l'objet direct (OD) inanimé tend à précéder l'objet indirect (OI) animé et formulent l'hypothèse selon laquelle la structure canonique (ici l'ordre OD OI) est renforcée par des arguments prototypiques. Nous présentons les résultats d'une expérience de rappel de phrase (Lombardi & Potter 1992) permettant de tester cette hypothèse et plus largement l'effet du caractère animé sur l'ordre des compléments postverbaux. Le rappel de phrases consiste à présenter une série de phrases au participant et à lui demander de les rappeler après une tâche de distraction, à partir d'une amorce qui correspond au début de chaque phrase (cf. schéma du déroulement en Figure 1). Les 24 items expérimentaux varient selon l'animéité du SN-OD [concret/abstrait/animé] et l'ordre [SNSP/SPSN] et comprend 12 verbes de transfert avec SP-destinataire humain (type 'confier'), 12 verbes locatifs avec SP-destination concret (type 'placer') (1). La longueur et la définitude des constituants ont été fixées pour favoriser l'ordre non canonique SPSN. Les réponses de 80 participants analysées permettent de montrer que le taux d'erreur (nombre de phrases correctement rappelées) est dépendant de la prototypicalité des phrases stimuli : le nombre d'erreur est plus élevé quand l'ordre est SPSN et quand le SN n'est pas concret (Figure 2). Parmi, les phrases correctement rappelées, on observe que la tendance majoritaire est de rappeler les phrases SPSN dans l'ordre canonique SNSP, mais que cette tendance est moins forte pour les verbes de transfert (Figure 4) que pour les verbes locatifs (Figure 3), ce qui correspond à une tendance à avoir l'ordre SP-humain en premier. De plus, conformément aux résultats de Thuilier *et al.* (2021), les SN concrets sont plus fréquemment replacés dans l'ordre SNSP que les SN abstraits et animés (Figures 3 & 4). Ces résultats sont confirmés par les analyses statistiques (régression logistique à effets mixtes). Ils mettent au jour un effet d'animéité dans le domaine postverbal du français, jusque-là non identifié, et confirment le renforcement de la structure SNSP par la présence d'un OD prototypique (ici concret).

(1) Pendant la répétition, elle a mis

une machine extraordinaire		[concret/SNSP]
un enthousiasme exceptionnel	sur la scène	[abstrait/SNSP]
un figurant intimidant		[animé/SNSP]
sur la scène	une machine extraordinaire	[concret/SPSN]
	un enthousiasme exceptionnel	[abstrait/SPSN]
	un figurant intimidant	[animé/SPSN]

## Bibliographie

- Bock, J. K. & Warren R.K. (1985) Conceptual accessibility and syntactic structure in sentence formulation. *Cognition* 21, 47–67.
- Branigan H. P., Pickering M. J. & Tanaka M. (2008) Contributions of animacy to grammatical function assignment and word order during production. *Lingua*, 118, 172–189.
- Bresnan J., Cueni A., Nikitina T. & Baayen. R. H. (2007) Predicting the dative alternation. In G. Boume, I. Kraemer & J. Zwarts (Eds.), *Cognitive Foundations of Interpretation*. Amsterdam. Royal Netherlands Academy of Science.
- da Cunha, Y. & Abeillé, A. (2022). L’alternance actif/passif en français parlé : un modèle statistique. *Langue française*, 216, 63-80.
- Haspelmath, M. (2020). Role-reference associations and the explanation of argument coding splits. *Linguistics*, 59(1), 123-174.
- Lamers, M. & de Swart, P. (éds) (2012). *Case, word order and prominence*. Studies in Theoretical Psycholinguistics. Dordrecht. Springer Netherlands, 2012.
- Lombardi, L. & Potter M. C. (1992) The regeneration of syntax in short term memory. *Journal of Memory and Language* 6(31). 713–733.
- Thuilier, J., Grant, M., Crabbé, B. & Abeillé, A. (2021) Word order in French: the role of animacy, *Glossa: a journal of general linguistics* 6(1): 55.

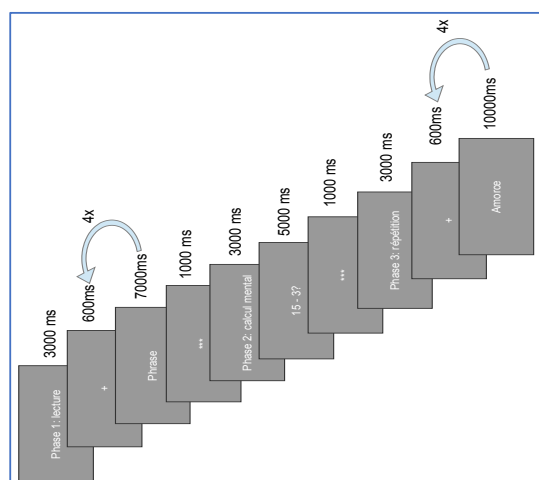


Fig. 1 : Déroulement de l'expérience

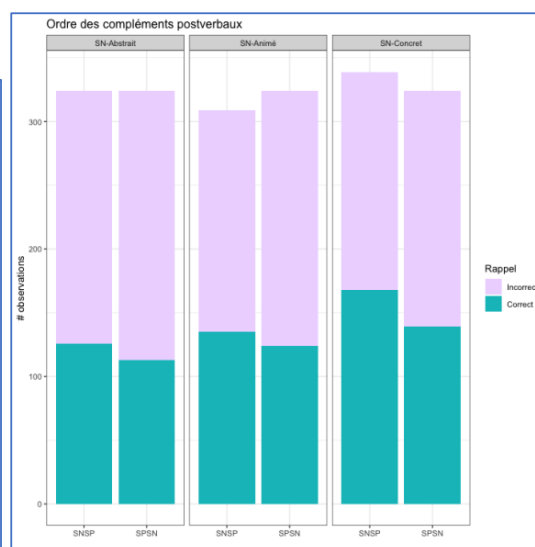
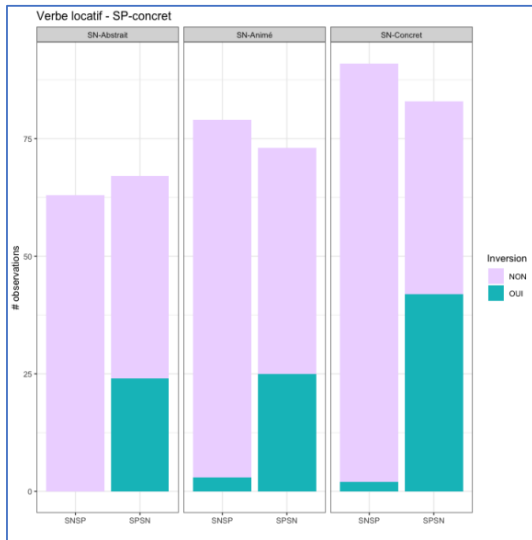
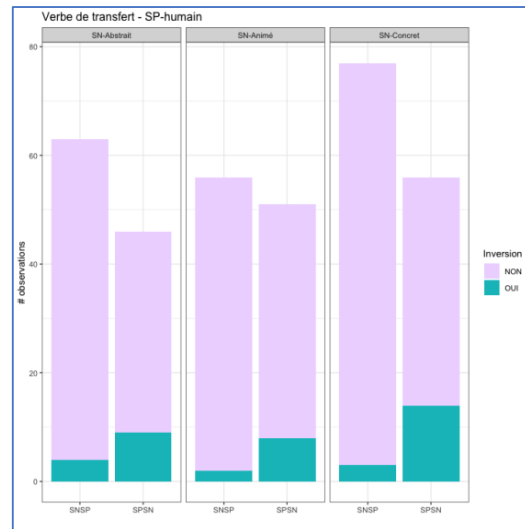


Fig. 2 : Nombre d'erreurs et de rappels corrects



**Fig. 3** : Rappels avec ou sans inversion (Locatif)



**Fig. 4** : Rappels avec ou sans inversion (Transfert)

# Sharp as a tack: assessing the role of iconicity and non-arbitrariness in word production processes

Giulio Massari<sup>1</sup>, Fanny Meunier<sup>1</sup>, Raphaël Fargier<sup>1</sup>

<sup>1</sup>Université Côte d’Azur, CNRS, BCL, France

## ABSTRACT

Despite standard assumptions about the arbitrariness of sound-meaning correspondences, empirical studies show the existence of non-arbitrary mappings in natural language (Perniss & Vigliocco, 2014). For instance, crosslinguistically, speakers are biased to label round shapes via phonemes like /a, b, m/, and spiky shapes via others like /i, k, t/: a phenomenon known as the *bouba-kiki* effect (Ramachadran & Hubbard, 2001). The effect has been largely observed experimentally with pseudowords (Fort et al., 2015), as well as statistically within natural language lexica (Sidhu et al., 2021). However, less attention has been dedicated to its experimental replication with existing words. Consequently, little is known of whether iconicity should make a difference in the cognitive representations of words, and therefore matter for the sake of lexical access in production – i.e., the cognitive flow going from conceptual activation to word form encoding (Levelt et al., 1999). To investigate this issue, we tested whether lexical retrieval in language production would be facilitated when there is iconic congruence between conceptual and phonological representations of words.

We developed a picture naming task in French, consisting of 80 images (Alario & Ferrand, 1999; Bonin et al., 2003). Half of the stimuli were selected for being examples of iconic congruence (e.g., *ballon*, *cactus*): such images have round or spiky shapes and, according to the sound symbolic model in Sidhu et al. (2021), their corresponding words’ phonology holds congruent round or sharp associations. The other half of the stimuli were selected for being neutral in sound symbolic terms, i.e., non-iconic (e.g., *aile*, *dauphin*). Given empirical correlations between iconicity and age of acquisition (Monaghan et al., 2014), as well as theoretically similar conceptions of the role of the two in lexical representation (Elsherif et al., 2023), stimuli were also chosen according to an AoA divide: each condition consisted in half of the words being acquired early (i.e., before 4.5 years), and the other half being acquired later. Items across conditions were further balanced on characteristics known to be influential upon naming (e.g., word frequency). Reaction times and accuracy measures were collected on a sample of 24 French native speakers, ranging 18-34 years of age. Results revealed that iconic items, compared to non-iconic ones, showed significantly higher percentages of correct responses, as well as faster response times. A significant interaction was also found between iconicity and age of acquisition on both measures, such that the effect of iconicity was significant on both accuracy (78% vs 59%) and RTs (1140 vs 1220 milliseconds) only for lately acquired words. Such results support the idea that iconicity impacts lexical retrieval by facilitating the production of words featuring non-arbitrary links between phonological and

visuo-conceptual features. Moreover, they are compatible with a partially shared interpretation for iconicity and age of acquisition effects, based on their role over the access for distributed lexical representations (Elsherif et al., 2023). Therefore, these results have potential implications for our understanding of language production processes, suggesting that some degree of interactivity between conceptual and phonological levels of representation must be recognized in word retrieval (Strijkers et al., 2017).

**Keywords:** iconicity; arbitrariness; *bouba-kiki* effect; language production; lexical access; word retrieval; cognitive representation; picture naming.

## BIBLIOGRAPHY

- Alario, F.X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers*, *31*(3), 531-552.
- Bonin, P., Peereman, R., Malardier, N., Méot, A., & Chalard, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 158-167.
- Elsherif, M.M., Preece, E., & Catling, J.C. (2023). Age-of-acquisition effects: A literature review. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(5), 812-847.
- Fort, M., Martin, A., & Peperkamp, S. (2015). Consonants are more important than vowels in the Bouba-Kiki effect. *Language and Speech*, *58*(4), 247-266.
- Levelt, W.J.M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-75.
- Monaghan, P., Shillcock, R.C., Christiansen, M.H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B*, *369*: 20130299.
- Perniss, P., & Vigliocco, G. (2014). *Philosophical Transactions of the Royal Society B*, *369*: 20130300.
- Ramachandran, V.S., & Hubbard, E.M. (2001). Synaesthesia – a window into perception, thought and language. *Journal of Consciousness Studies*, *8*(12), 3–34.
- Sidhu, D.M., Westbury, C., Hollis, G., & Pexman, P.M. (2021). Sound symbolism shapes the English language: The maluma/takete effect in English nouns. *Psychonomic Bulletin & Review*, *28*, 1390-1398.
- Strijkers, K., Costa, A., & Pulvermüller, F. (2017). *NeuroImage*, *163*, 206-219.

## Le traitement de la parole sifflée par des auditeurs non entraînés

Meunier, F. <sup>(1)</sup>, Tran Ngoc, A. <sup>(1)</sup>, Meyer J. <sup>(2)</sup>

*(1) Université Côte d'Azur, CNRS, BCL, Nice, France*

*(2) Université Grenoble Alpes, CNRS, GIPSA-Lab, Grenoble, France*

Les expériences comportementales sur la parole sifflée présentées ici ont en commun de s'intéresser à la façon dont des auditeurs naïfs (qui ne connaissent pas ce type de parole) parviennent à traiter un signal transposant la parole usuelle en sifflements modulés en fréquence et amplitude. Ce type de parole sert principalement en milieu rural à adapter la phonétique des langues à la communication à distance. Il permet aux experts de s'échanger des phrases longues et non attendues avec des taux d'intelligibilité relativement élevés (de 65 à 80% de reconnaissance des mots suivant le type de phrase ; Moles, 1970, Meyer et al., 2024) alors qu'elles demeurent incompréhensibles aux néophytes même lorsqu'ils sont informés qu'il s'agit d'un signal langagier. Pourtant, la parole sifflée s'appuie sur des caractéristiques phonologiques et lexicales très similaires à la parole usuelle. Elle existe dans une grande diversité de langues, parfois même en reflétant des variations dialectales phonétiques (comme sur les îles Canaries, Diaz, 2017) ; C'est ce qui nous permet d'utiliser cette forme de parole naturellement simplifiée pour comprendre certains processus de perception de la parole. En testant des participants non entraînés nous nous assurons qu'ils utiliseront leurs connaissances issues de la forme parlée usuelle.

Les expériences que nous avons menées portent principalement sur des signaux issus de langues non tonales pour lesquelles le sifflement transforme le timbre parlé des voyelles en différentes hauteurs et transpose donc des indices acoustiques issus des formants : ainsi, /i/ est généralement sifflé le plus aigu, /o/ le plus grave, et /e/ et /a/ sont intermédiaires (Meyer, 2015). Les consonnes, comme en parole usuelle, modifient les fréquences des voyelles par l'effet des mouvements articulatoires. Nous avons réalisé une série d'expériences sur la catégorisation de voyelles et de consonnes sifflées qui utilisent une méthodologie expérimentale commune. Nous faisons entendre des phonèmes sifflés (voyelles isolées ou en fin de séquence, et consonnes en contexte VCV) et les participants devaient identifier le phonème entendu parmi quatre proposés. Les résultats montrent que les voyelles comme les consonnes sont toujours catégorisées bien mieux que le hasard. Les participants non entraînés sont donc capables de catégoriser le signal en s'appuyant sur les traitements engagés par la parole usuelle. En testant des participants de langues natives variées, nous avons observé des profils de catégorisation différents et hiérarchisés en fonction de la proximité avec la langue qui est sifflée (Meyer, Dentel, Meunier, 2017). Par ailleurs, nous avons également mis en évidence des effets de l'expertise musicale et de l'instrument de prédilection. Globalement, les musiciens experts sont les meilleurs pour catégoriser les phonèmes sifflés, et au sein de ce groupe, les flûtistes sont bien au-dessus des violonistes ou des pianistes, avec cependant parfois une variabilité importante en fonction des différents phonèmes (Tran Ngoc, Meyer, Meunier, 2020). Dans cette présentation, nous proposerons une synthèse des résultats obtenus et nous mettrons en évidence la manière dont différentes expériences individuelles (langue native, expertise musicale) modulent les traitements appliqués à la parole.

Mots-clés : catégorisation de phonèmes, parole sifflée, perception de la parole, psycholinguistique, musiciens

## Bibliographie :

Díaz Reyes, D. (2017). *El lenguaje silbado en la Isla de El Hierro* (segunda edición ampliada), Tenerife: Le Canarien ediciones, La Orotava.

Meyer, J., Rolland, V., Socas, T., Díaz Reyes, D. (2024). A Sentence Comprehension Test with Whistled Spanish Experts. In Botinis, A. (ed.), *Proc. 14th Exling Conference*. Athens, Greece.

Moles, A. (1970). Etude sociolinguistique de la langue sifflée de Kuskoy. *Revue de Phonétique Appliquée*, 14(15), 78-118.

Meyer, J., Dentel, L., Meunier, F. (2017). Categorization of Natural Whistled Vowels by Naive Listeners of Different Language Background. *Front. Psychol.* 8, 25.

Tran Ngoc, A., Meyer, J. and Meunier, F. (2020). Whistled vowel identification by French listeners. *Interspeech*. DOI: 10.21437/Interspeech.2020-2697

## L'émergence d'un nouveau suffixe : *-(i)ssimo* vs. *-issime* en français contemporain

Fabio Montermini

UMR 5263 CLLE, CNRS & Université de Toulouse Jean Jaurès

Dans ce travail j'analyse le comportement d'un suffixe de diffusion récente en français, *-(i)ssimo*, dans le but de le comparer avec le suffixe similaire, et plus commun, *-issime*. Les deux partagent tant leur origine (ils sont empruntés de l'italien) que leur sens, que l'on peut caractériser comme génériquement superlatif. Cependant, ils présentent également plusieurs différences, notamment concernant leurs propriétés syntaxiques et phonologiques. Observer ces différences permet non seulement de décrire leurs propriétés individuelles, mais aussi d'analyser les mécanismes selon lesquels des schémas morphologiques ayant une origine commune se structurent différemment dans la même langue.

Les premiers mots en *-(i)ssimo* sont attestés à partir des années 1960, alors que *-issime* a été introduit en français à partir du 15<sup>e</sup> siècle (Perko, 2010). Les propriétés de ce dernier sont proches de celles d'un suffixe canonique en ce qui concerne la sélection des bases et la forme que prennent ses dérivés. *-(i)ssimo* forme principalement des noms propres d'enseignes, produits, événements, etc. Cependant, des emplois récents de mots en *-(i)ssimo* en discours s'observent également. En particulier, le suffixe semble fonctionner comme une variante 'branchée' de *-issime*, et sa diffusion dans cet emploi est possiblement liée à la présence en français de *-o* en tant que terminaisons génériquement expressive (Antoine, 1993 ; Kilani-Schoch & Dressler 1999).

Plus en détail, les différences principales entre les deux suffixes sont les suivantes (les données sont tirées de Plénat, 2002; Bengtsson, 2019 et de la base de données utilisée dans l'étude) :

- *-issime* se lie principalement à des adjectifs et plus marginalement à des noms sans déterminer de changements de catégorie ; *-(i)ssimo* prend des bases de catégories variées, y compris des verbes (1) ;
- *-issime* sélectionne des allomorphes savants et peut entraîner des réductions phonologiques de ses bases ; *-(i)ssimo* entraîne des modifications des bases plus nombreuses et variées (2) ; de plus, ce dernier peut présenter une voyelle initiale différente de /i/ (3).

(1) rarissime généralissime	incroyablissimo Parfumissimo Courissimo
(2) vérissime catholissime	Acquissimo (← acquisition) Optissimo (← opticien)
(3)	Pizzissimo Vélossimo Fondussimo

Je propose une analyse à partir d'une base de données de 527 mots en *-issime* et 446 mots en *-(i)ssimo* extraits de corpus et de sources en ligne. En tant qu'affixe canonique *-issime* impose des contraintes catégorielles et sémantiques fortes à ses bases, qui sont principalement des adjectifs (ou des noms) susceptibles de recevoir une connotation superlative. Ces contraintes sont en revanche moins fortes pour *-(i)ssimo*, qui sélectionne ses bases principalement sur une base formelle, en privilégiant en particulier celles qui peuvent être réduites.



De plus, les dérivés en *-issime* sont soumis à des contraintes de fidélité avec la base, qui sont moins rigides dans le cas de *-(i)ssimo* (cf. les exemples en (2)).

Enfin, l'allomorphie de l'exposant observée avec *-(i)ssimo* suggère que celui-ci peut être vu comme une série de contraintes hiérarchisées sur la forme des dérivés (/isimo/ > /Vsimo/), dans lesquelles la voyelle correspond par défaut à /i/, sauf si les propriétés segmentales de la base déterminent l'émergence d'une voyelle différente.

## References

- Antoine F. (1993). Les apocopes en [o] dans le français actuel : élément de réflexion. *Le français moderne*, LXI.1, 28-36.
- Bengtsson A. (2019). L'évolution du suffixe *-issime* : un inventaire et une fréquence des formes attestées dans Frantext. *Bergen Language and Linguistics Studies*, 10.1.
- Kilani-Schoch M. & Dressler W.U. (1999). Perspective morphopragmatique sur les formations en *-o* du français branché. In: I. Mel'čuk (Ed), *Dictionnaire explicatif et combinatoire du français contemporain, Recherches lexico-sémantiques IV*, (pp. 55-66). Montréal: Les Presses de l'Université de Montréal.
- Perko G. (2010). Le suffixe *-issime* dans le paysage dérivationnel du nom propre en français. In: M. Iliescu, P. Danler & H. Siller-Runggaldier (Eds.), *Actes du XXV<sup>e</sup> Congrès International de Linguistique et Philologie Romanes, Innsbruck 2007*, tome I (pp. 465-470). Berlin: de Gruyter.
- Plénat M. (2002). Jean-Louis Fossat : Fossatissime. Note sur la morphophonologie des dérivés en *-issime*, in: L. Rabassa (Ed.), *Mélanges offerts à Jean-Louis Fossat. Cahiers d'Etudes Romanes (CERCLiD)*, 11/12, 229-248.

## Les marqueurs typographiques et linguistiques associés aux néologismes : études en production et en réception.

Léa Pineau<sup>1</sup>, Fabio Montermini<sup>1</sup> & Juliette Thuilier<sup>2,1</sup>

<sup>1</sup> CLLE (Université de Toulouse, CNRS, Toulouse, France).

<sup>2</sup> Université Toulouse Jean Jaurès.

Le sentiment néologique (Ben Hariz Ouenniche, 2009; Gardin et al., 1974; Sablayrolles, 2003) ou intuition néologique (Lombard et al., 2021) est l'habileté métalinguistique qui permet de juger l'appartenance ou la nouveauté d'un mot vis-à-vis du lexique conventionnel. Selon Dal & Namer (2018), Sablayrolles (2019) et Winter-Froemel (2020), les néologismes peuvent être associés à l'écrit à des marqueurs, tels que les guillemets (cf. (1)), les commentaires (cf. (2)), les motifs syntaxiques (cf. (3)).

- (1) une envahissante cravate « *écrouélique* », souvent de couleur verte, qui monte jusqu'au menton
- (2) En discutant avec les *nuit-deboutistes* (ça se dit, ça ?), on se rend rapidement compte que, comme nous, ils ne portent pas particulièrement les GAFAM [...] dans leurs cœurs.
- (3) il y a un XIXe siècle *corbinesque* comme il y a un XIXe siècle *perrotien*

L'objectif de notre étude est d'observer la présence, la proportion et le type de marqueurs typographiques ou linguistiques associés aux néologismes par les locuteurs-énonciateurs à l'écrit, puis de mesurer l'influence de ceux-ci sur le sentiment néologique des locuteurs-destinataires.

En production, nous avons réalisé une étude de corpus. Nous avons extrait 1093 mots de fréquence 1, suffixés en *-ien*, *-ique* ou *-esque* du corpus FrenchWeb2017 (TenTen). Ces hapax sont absents du Petit Robert de la Langue Française et du TLFi. Nous avons annoté la présence et le type de marqueurs associés. Nous constatons que 22,9 % des 1093 néologismes sont associés à au moins un marqueur typographique ou linguistique (cf. Figure 1). Les locuteurs-énonciateurs semblent privilégier les guillemets qui représentent 64,0 % de tous les marqueurs recensés, devant les motifs syntaxiques (22,1 %) et les commentaires (14,0 %) (cf. Figure 2).

En réception, nous avons réalisé une étude expérimentale visant à mesurer l'influence des marqueurs sur le sentiment néologique des locuteurs-destinataires vis-à-vis de 42 néologismes dans leur contexte authentique, issus de notre première étude. Les items expérimentaux sont déclinés en 6 conditions ([+néologisme, -marqueur], [+néologisme, +guillemets], [+néologisme, +commentaire], [-néologisme, -marqueur], [-néologisme, +guillemets], [-néologisme, +commentaire]) et repartis en carré latin. 635 participants francophones natifs ont rempli le questionnaire en ligne, en répondant à la question « Selon vous, y a-t-il un mot

nouveau dans la phrase suivante ? », et le cas échéant, en indiquant le mot précis considéré comme nouveau parmi tous ceux de la phrase. Nous constatons que la présence de guillemets ou de commentaires augmente de façon significative le taux de qualification comme « mot nouveau » des néologismes (85,7% avec marqueur versus 82,1 % sans marqueur ;  $\chi^2(1)=28.4, p<0.001$ ) et des mots attestés (4,5 % avec marqueur versus 3,3 % sans marqueur ;  $\chi^2(1)=9.93, p=0.002$ ) (cf. Figure 3). L'influence exercée par les guillemets (marqueur implicite) ou les commentaires (marqueur explicite) semble identique.

En conclusion, la prise en compte du contexte énonciatif s'avère importante pour les études en néologie (Gardin et al., 1974, p.48 ; Lavale Ortiz, 2019, p.216 ; Tardy, 1974, p.102). Enfin, l'influence d'autres facteurs sur le sentiment néologique est suggérée par nos résultats : le suffixe du néologisme et l'âge des locuteurs.

Mots-clés : néologismes – intuition néologique – marqueurs linguistiques – guillemets – commentaires – motifs syntaxiques.

#### ANNEXE : Graphiques

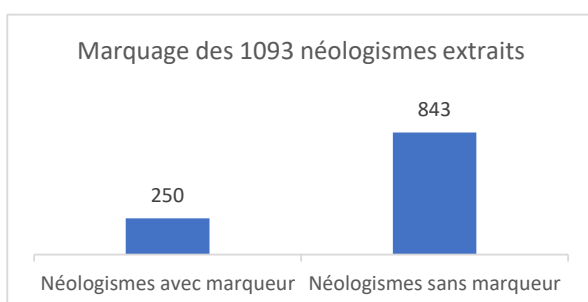


Figure 1 : PRODUCTION\_Marquage des 1093 néologismes

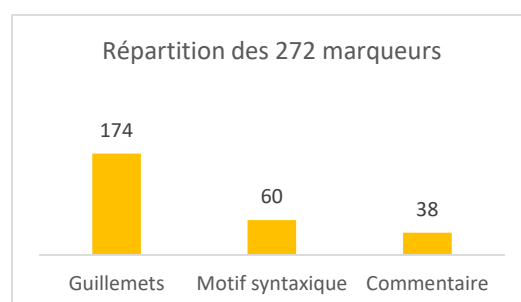


Figure 2 : PRODUCTION\_Répartition des 272 marqueurs

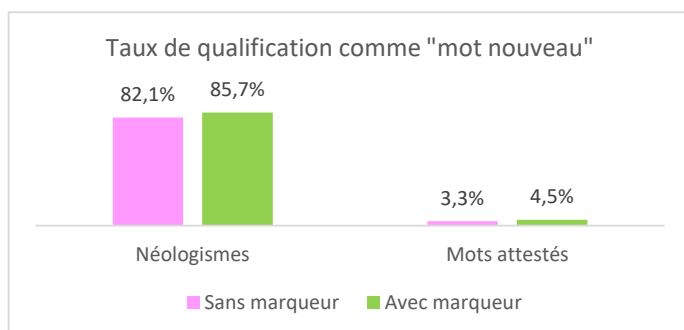


Figure 3 : RECEPTION\_Taux de qualification "mot nouveau": influence des marqueurs

## **BIBLIOGRAPHIE :**

Ben Hariz Ouenniche, S. (2009). Diminuer les fluctuations du sentiment néologique. *Neologica*, 3, 41–55.

<https://doi.org/10.15122/isbn.978-2-8124-4229-2.p.0041>

Dal, G., & Namer, F. (2018). Les occasionnalismes et la question de la productivité Le locuteur à l'œuvre. Pourquoi ? Comment ? *Neologica*, 12, 71–89.

Gardin, B., Lefèvre, G., Marcellesi, C., & Mortureux, M. Fr. (1974). A propos du « sentiment néologique ». *Langages*, 36, 45–52.

Lavale Ortiz, R. M. (2019). Bases para la fundamentación teórica de la neología y el neologismo : La memoria, la atención y la categorización. *Círculo de lingüística aplicada a la comunicación*, 80, 201–226.

Lombard, A., Huyghe, R., & Gygax, P. (2021). Neological intuition in French : A study of formal novelty and lexical regularity as predictors. *Lingua*, 254, 103055. <https://doi.org/10.1016/j.lingua.2021.103055>

Sablayrolles, J. F. (2003). Le sentiment néologique. In J.F. Sablayrolles (Eds.), *L'innovation lexicale* , (pp. 279-295). Honoré Champion.

Sablayrolles, J. F. (2019). *Comprendre la néologie. Conceptions, analyses, emplois*. Limoges : Lambert-Lucas.

Tardy, M. (1974). Néologie et fonctions du langage. *Langages*, 8(36), 95–102. <https://doi.org/10.3406/lgge.1974.2278>

Winter-Froemel, E. (2020). Les néologismes, un sous-type de l'innovation lexicale Réflexions à partir d'études de corpus généraux dans trois langues romanes. *Neologica*, 14. Perception, 25–46.

<sup>1</sup> Université Toulouse Jean Jaurès

<sup>2</sup> CLLE (CNRS/UT2J)

<sup>3</sup> University of Amsterdam

<sup>4</sup> LLF (CNRS/UPC)

Les travaux de Blanche-Benveniste (2000) et Hamma *et al.* (2017) montrent que certains verbes psychologiques, comme *choquer* ou *étonner*, sont plus fréquents à la voix passive qu'à la voix active dans les corpus oraux. Étant donné la structure argumentale de ces verbes (Arg1=*stimulus* ; Arg2=*expérienceur*), l'utilisation du passif permet de mettre le rôle sémantique le plus proéminent (*expérienceur*) en position sujet (Ferreira 1994). D'autres travaux ont montré l'influence du caractère animé des référents sur le choix de la voix passive (McLaughlin 2010, da Cunha & Abeillé 2020 2022, Thuilier *et al.* 2021 pour le français) : en cas d'asymétrie d'animéité entre Arg1 et Arg2, l'argument animé tend à occuper la fonction syntaxique sujet. Nous explorons l'interaction entre le type de verbe et l'animéité à l'aide d'un protocole de rappel phrases (Tanaka *et al.* 2011), dans lequel nous manipulons l'animéité du *stimulus* et la voix, ainsi que le rôle sémantique de l'Arg2, en faisant varier le type de verbe comme variable inter-item : 12 items avec verbe à Arg2 *expérienceur* (1) et 12 items à Arg2 *stimulus* (2). Les autres variables (longueur des constituants, définitude) ont été contrôlées afin de favoriser la production de la voix passive, réputée rare.

(1) Arg2-Exp : **Pendant la nuit,**

- un {tir de carabine/voleur de tracteur} a épouventé le villageois
- le villageois a été épouventé par un {tir de carabine/voleur de tracteur}

(2) Arg2-Stim : **Tout au long du procès,**

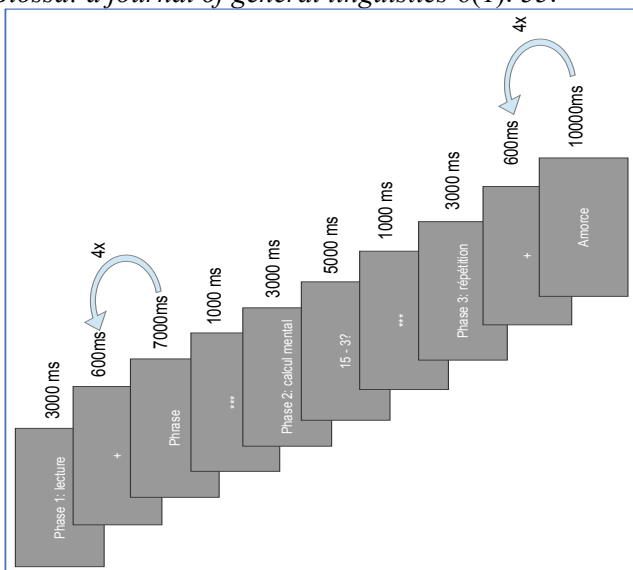
- un témoin de la défense a apprécié {l'atmosphère/la magistrate}
- {l'atmosphère/la magistrate} a été appréciée par un témoin de la défense

La passation s'organise en 24 blocs de 4 phrases, contenant chacun un item expérimental, avec répartition en carré latin. Pour chaque bloc, 4 phrases s'affichent successivement à l'écran et le participant a pour consigne de les mémoriser. Après une tâche de distraction, les amorces (en gras dans les exemples) des 4 phrases s'affichent successivement et le participant est invité à rappeler la phrase visée à voix haute à partir de l'amorce (cf. schéma du déroulement en Fig. 1). Les réponses de 80 participants ont été enregistrées, transcrites et triées selon qu'elles étaient correctement rappelées et qu'elles présentaient ou non une inversion de voix. Les taux d'inversion sont présentés en Fig. 2 : les participants tendent à changer la voix pour mettre l'*expérienceur* en position sujet, en particulier dans le cas des verbes mentionnés dans la littérature. Le fait que le stimulus soit inanimé renforce clairement cette tendance. L'analyse statistique des résultats (régression logistique à effets mixtes avec participants et items comme effets aléatoires) confirme ces observations : un stimulus animé défavorise l'inversion (coef. -0,71 ;  $p < 0,0001$ ) et l'interaction entre la structure active et l'Arg2 *expérienceur* favorise fortement l'inversion (coef. 2,01 ;  $p < 0,0001$ ). Nos résultats confirment les observations faites dans la littérature et permettent de documenter précisément l'interaction entre les rôles sémantiques *stimulus* et *expérienceur*, et le caractère animé des référents, dans le cas des verbes psychologiques du français.

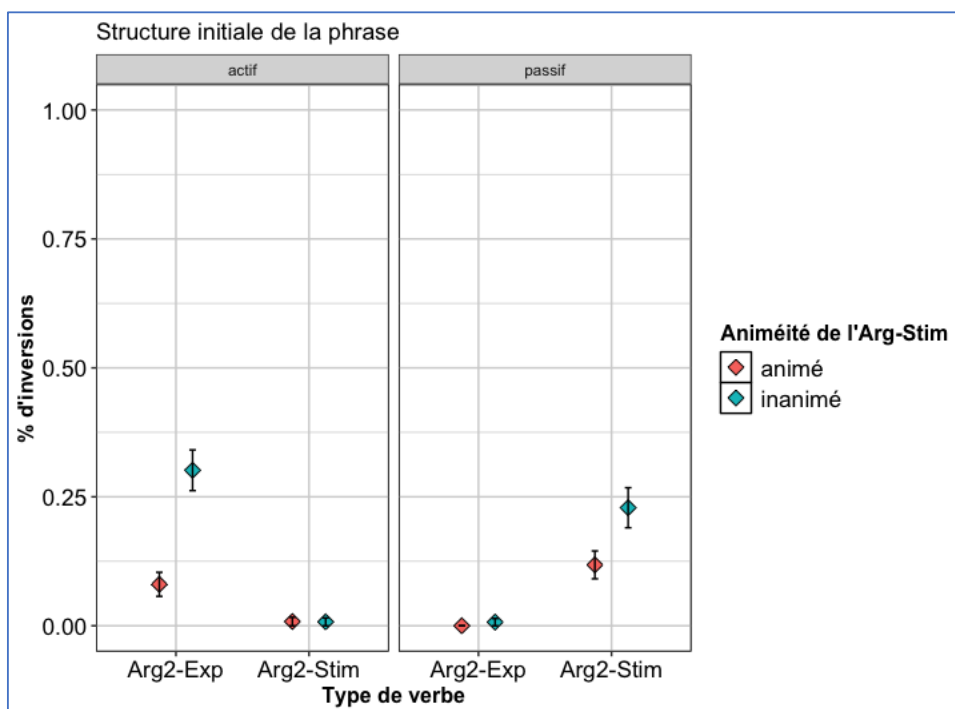
## Références

- Blanche-Benveniste, C. (2000). Analyse de deux types de passifs dans les productions de français parlé. L. Schøsler (éd.). *Le passif, Études romanes*, 45, 303-319.
- da Cunha, Y. & Abeillé, A. (2020). L'alternance actif/passif en français : une étude statistique sur corpus écrit. *Discours : revue de linguistique, psycholinguistique et informatique*, (27), 25.
- da Cunha, Y. & Abeillé, A. (2022). L'alternance actif/passif en français parlé : un modèle statistique. *Langue française*, 216, 63-80.

- Ferreira, F. (1994) Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language* 33(6). 715–736.
- Hamma, B. & Tardif, A. & Badin, F. (2017). Le passif à l’oral. P. Larrivée & F. Lefevre, (éds). *Français contemporain vernaculaire* (FRACOV).
- McLaughlin, M. (2010). L’influence de l’anglais sur la syntaxe du français : une étude de cas concernant la voix passive. *2ème Congrès Mondial de Linguistique Française*, 129.
- Tanaka, M. N., Branigan, H. P., McLean, J. F., & Pickering, M. J. (2011). Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language*, 65:318–330.
- Thuilier, J., Grant, M., Crabbé, B. & Abeillé, A., (2021) Word order in French: the role of animacy, *Glossa: a journal of general linguistics* 6(1): 55.



**Fig 1 : Déroulement de l’expérience**



**Fig 2 : Taux d’inversion (actif vers passif à gauche ; passif vers actif à droite)**

## Who fulfills a wish? On supernatural ergatives in Nakh-Daghestanian languages

Blessings and curses are peculiar in that, like imperatives, they imply performing a wish, but, unlike imperatives, the intended performer is not the addressee but a supernatural power. Impersonal constructions are typical for wish-formulas, since, as there is often only one supernatural power, there is no need to refer to it by linguistic means:

Russian

(1) čtob    teb'a                  razorva-l-o!  
 COMPL    you.SG.ACC    tear-PST-SG.N  
 'May you be torn apart!'

Another option is to express the performer explicitly: *Que dieu vous garde!* The name of the superficial power in blessings and curses can be grammaticalized in the languages of the world, e.g. Wolof (Guerin 2020). In Jóola Fóoñi (Atlantic), the biclausal construction with the matrix clause *Emitey ekaan* 'Let God make (that)', was reanalysed as a monoclausal construction where the sequence *Emitey ekaan* acts as an optative particle (Creissels & Bassène 2022):

(2)    Emitey            ε-kaan            Emitey            ε-bəncət-əə  
 God(E)            sI.clE-make    God(E)            sI.clE-forgive-I.clA  
 Lit. 'Let God make that God will forgive him / her!'

In this talk, I will show how the performer of blessings and curses is expressed in Nakh-Daghestanian languages (Northern Caucasus).

Most languages of Daghestan have a morphological optative, a category dedicated to the expression of blessings and curses. The implicit performer of most wishes is commonly understood to be Allah. Native speakers often translate optative utterances into Russian using the word 'Allah', even if it was not present in the original sentence; in other cases, however it is explicitly expressed as an address.

Notably, in some languages, eg. Rutul (Lezxic), the word 'Allah' may be expressed as an ergative subject in intransitive sentences which normally do not allow ergative arguments (3):

(3) Rutul  
 allah-a-ra                  saɾ                  jiš-ij                  žu-di                  duɣ  
 Allah-OBL-ERG            healthy 1.become-OPT            REFL.M-ATTR            son  
 May his son be healthy!

Unlike Rutul, Mehweb (Dargwa) intransitive verbs cannot have the ergative of *Allah* in the optative construction, but regularly add a causative marker; through valency increase, they add a position for the supernatural ergative. The optative of the verb thus causativized has exactly the same meaning as its non-causative counterpart (4), except that the former may introduce Allah as an ergative subject (5).

(4) Mehweb  
 hum-be ʔa<sup>h</sup>χ d-uh-a-b!  
 road-pl good npl-become:pfv-irr-opt

‘May you have a good trip!’

(5) Mehweb

allah-li hum-be ʔaʕ d-uh-aq-a-b!

Allah-obl(erg) road-pl good npl-become:pfv-caus-irr-opt

‘May you have a good trip with the help of God!’

To sum up, some Daghestanian languages explicitly express the performer of the blessings and curses as an agent even in the constructions which otherwise do not suppose the presence of agent. The nearest analogy are meteorological constructions, which often have dummy (expletive) subjects, a cross-linguistically common phenomenon also attested in Daghestan (‘world’, ‘weather’ or such). Cross-linguistically, meteorological predicates are also attested with so-called mythological subjects (see examples from Japanese and Uralic languages in Malchukov & Ogawa 2011, Salo 2011). While the subject of meteorological predicates is an S, the subject of blessings and curses is an Agent, and is thus coded by the ergative case.

#### References

Creissels, Denis, and Alain Christian Bassène. ‘The emergence of an optative marker in Jóola Fóõñi (Atlantic)’. *Syntax and Semantics* 22. 101-119.

Guérin, Maximilien. Optatif et prohibitif en wolof : description et hypothèses de grammaticalisation

Malchukov, Andrej, and Akio Ogawa. 2011. Towards a typology of impersonal constructions. In: Malchukov, Andrej, and Anna Siewierska, eds. Impersonal constructions: A cross-linguistic perspective. Vol. 124. John Benjamins Publishing.

Salo, Merja. 2011. Meteorological verbs in Uralic languages – are there any impersonal structures to be found. In: Malchukov, Andrej, and Anna Siewierska, eds. Impersonal constructions: A cross-linguistic perspective. Vol. 124. John Benjamins Publishing.



### Nominative-Genitive variation in Russian close appositional constructions

This corpus-based study deals with the Nominative-Genitive variation found in the encoding of close appositional constructions in Russian, which is illustrated in (1):

- (1) *Dlja* *nee* *ponjati-e* *čest-i* / *čest'* *ne* *by-l-o*  
 for she.GEN **notion-NOM** **honor-GEN** **honor.NOM** NEG be-PST-N  
*pustym* *zvuk-om*  
 empty-INS.SG.M sound-INS.SG  
 ‘For her, the concept of honor was not an empty phrase’.

As can be seen from (1), the noun *čest'* ‘honor’ can have either Nominative or Genitive encoding after the noun *ponjatie* ‘notion’ that both form a close appositional construction together. Both variants are completely valid for native speakers and in most of the contexts are equivalent. The corpus frequency of both variants is comparable with a slightly higher frequency of Nominative forms. Unlike well-studied concord variability found in Slavic close apposition (Logvinova 2022, 2024), Nominative-Genitive variation, has not received any attention until now. Given the overall poor typological awareness of the phenomenon of close apposition (however, see Logvinova, in press), this research adds a new parameter that contributes to the overall picture of constructional space of apposition.

In this study, we focus on the variation in constructions with the word *ponjatie*, although the same phenomenon is observed in constructions with other nouns as well. The noun that comes after *ponjatie* is hereinafter called *term*.

The study is based on the sample from the ruTenTen17 corpus, which totals to 2 000 manually annotated examples.

Based on the pilot corpus study, we propose the following factors to have an effect on the observed Nominative-Genitive variation (all factors were proven to have a statistically significant effect on the choice of encoding when considered separately). In all cases below > corresponds to “is associated with more Genitive than Nominative forms”:

- 1) Animacy of the term (inanimate > animate).
- 2) Presence of anaphoric relation between the term and a relative or personal pronoun (yes > no)
- 3) Usage of the term in the pre-text or later in the same text (yes > no)
- 4) Distance in words between the first mention of the term and its usage in the construction (shorter distance > larger distance)
- 5) Presence of adjectival or genitive dependents of the term (no > yes)
- 6) Semantics of the term (abstract noun > concrete noun)
- 7) Frequency of the term (frequent terms > infrequent terms).

In the talk we will present the results of applying the logistic regression model, taking into account most of the factors listed above (except for factors causing multicollinearity), to the data collected from the corpus. Using logistic regression will allow to disentangle the effect of seven parameters that were found to be relevant for the choice of encoding strategy in the pilot-study.

As a conclusion, we will explain the emergency of constructions with Nominative-Genitive variation in Russian and their place in the constructional space for apposition in Russian.

### References

- Logvinova N. Case Marking and Definiteness in Slavic Appositional Constructions // Slovene. 2022. Vol. 11. No. 1. P. 281-304.
- Logvinova N. Concord in Russian close appositional constructions: a quantitative study // Russian linguistics. 2024. Vol. 48. No. 1. Article 4.

# Comparing Headline Styles: Article Omission in Greek and English

Michail I. Marinis & Brian D. Joseph

*The Ohio State University*

For some languages, it is recorded as a peculiarity of headlines that the definite article (1a) may be absent (Borchmann, 2024; Mårdh, 1980) in contexts where, outside the headline language, omitting the article would be ungrammatical. Compare, for example, the structures in 1a and 1b.

(1) a. *Stocks collapse in 16,410,030-share day, but rally at close cheers brokers; bankers [are] optimistic, to continue aid.* New York Times, October 1929

b. \* *Stocks collapse in 16,410,030-share day, but rally at close cheers brokers; bankers optimistic, to continue aid.* Standard American English

More recently, in Marinis & Joseph (2024), some examples of headlines have been documented in the Greek press where the absence of an article is recorded in structures for which its absence is ungrammatical in Standard Modern Greek (Roussou & Tsimpli, 1994). For example, compare the structures in 2a and 2b.

(2) a. *epi 'soðion meta 'ksi ['ton] amerika 'non 'ce ['ton] cemali 'kon 'is ['tin] 'prijipon*  
Eleytheron Vima, 15/1/1923

b. \* *epi 'soðion meta 'ksi ['ton] amerika 'non 'ce ['ton] cemali 'kon 'is ['tin] 'prijipon*  
Standard Modern Greek  
'An incident between Americans and Kemalists in Prinkipos'

In order to investigate whether the absence of the definite article in Greek can be considered as part of a particular "grammar of the headlines," we examine the phenomenon through a corpus of titles that we created based on the Historical Archives of the newspapers To Vima and Ta Nea.

Our analysis of the material shows that the absence of an article should be considered a particular feature of the language of headlines for the English and Greek press. However, the phenomenon is much more limited for Greek than for English. The difference is mainly because the article's presence in Greek is much more marked than in English. This marking results from the status of the definite article, which is considered obligatory for the indication of definiteness. Compare, for example, structures 3a, 3b, 3c, and 3d.

(3) In Standard Modern Greek

a.	'ta	pe 'ðja	'pezun	'all the children play'
b.	*	pe 'ðja	'pezun	'all the children play'
c.		pe 'ðja	'pezun	'some children play'
d.	'kapja	pe 'ðja	'pezun	'some children play'

Thus, the fact that the absence of an article in Greek is marked as [-definite] makes it difficult to use the structure to express [+definite]. However, the existence of such structures, although limited in number compared to English, should be considered a peculiarity of the language of Greek press headlines.

**Keywords**

Headlines, Article, Article Omission, Modern Greek, English

**References**

Borchmann, S. (2024). Headlines as illocutionary subacts: The genre-specificity of headlines. *Journal of Pragmatics*, 220, 73-99.

Marinis, M. I. & Joseph, B. D. (2024). The structure of Greek Newspaper Headlines: The Case of 'Eleftheron Vima' in the 1920s. *Studies in Greek Linguistics* (43).

Mårdh, I. (1980). *Headlines: On the Grammar of English Front Page Headlines*. Lund: CWK Gleerup.

Roussou, A. & Tsimpli, I.-M. (1994). On the interaction of case and definiteness in Modern Greek. In Philippaki-Warbuton, I., Nicolaidis, K. & Sifianou, M. (eds.), *Themes in Greek linguistics*. Amsterdam: John Benjamins. 69–76.

**Les réalisations du *naʿt sababī* en français  
De l'arabe standard au français, il n'y a qu'un pas**

Fayssal Tayalati

UMR 8163 « STL » / Université de Lille

Notre communication explore la question de l'existence, dans une langue comme le français, du mode de qualification appelé *naʿt sababī* par les grammairiens classiques et modernes de l'arabe standard (désormais AS). Le nom *sabab* (dans *sababī*) signifie « cause » ou « lien » et fait référence à un mécanisme de qualification indirecte basé sur une relation qui unit deux entités. De manière générale, deux modes de qualification sont identifiés en AS (cf. entre autres Al-Ghalāyīnī (1966) ; Sylvestre de Sacy, (1831) ; Wright (1996 [1896-1898]) ; Ryding (2005) ; Larcher (2017) :

*naʿt ḥaqīqī* (attribut réel)

- |                               |                       |                          |
|-------------------------------|-----------------------|--------------------------|
| (1) zawǧ-u-n                  | wafīyy-u-n            | li-zawǧat-i=hi           |
| mari-M.SG-NOM-INDÉF           | fidèle.M.SG-NOM-INDÉF | à-femme.F.SG.GÉN=PR.M.SG |
| « un mari fidèle à sa femme » |                       |                          |

*naʿt sababī* (attribut lié)

- |   |                     |                          |
|---|---------------------|--------------------------|
| (2) a. fatāt-u-n  | ṭawīl-u-n           | šaʿr-u= <b>hā</b>        |
| filles.F.SG-NOM-INDÉF   | long.M.SG-NOM-INDÉF | cheveux.M.SG-NOM=PR.F.SG |
| « lit : une fille longs ses cheveux » « = une fille dont les cheveux sont longs » |                     |                          |
| b. fatāt-u-n  | ṭawīlat-u           | <b>š-šaʿr-i</b>          |
| filles.F.SG-NOM-INDÉF   | long.M.SG-NOM       | dét-cheveux.M.SG-GÉN     |
| « lit. une fille long les cheveux » « = une fille aux longs cheveux »             |                     |                          |

Dans le premier mode, la qualité est attribuée *exclusivement* et *directement* (i.e. sans que le complément y joue un rôle à un quelconque niveau) au *mari*. Le second mode se caractérise par l'implication du prédicat qualifiant dans deux relations prédicatives ordonnées et une contribution du complément au mécanisme de qualification globale du N1 : dans la première prédication, *ṭawīl* (long) qualifie *šaʿr-u=hā / š-šaʿr-i* (ses cheveux / les cheveux) formant un prédicat complexe, lequel est attribué (deuxième prédication) au N1. Le passage de la qualification du N2 au référent du N1 est assuré par la relation qui unit les deux noms, même si cette relation n'est pas de même nature. Sous (2a), c'est la relation de coréférence entre le pronom résomptif attaché au N2 qui assure le passage de la qualification au N1. Ce mécanisme opère même lorsque le N2 est un nom d'objet aliénable (3), preuve que c'est bien la relation de conférence qui opère et non le type lexical du N2 :

- |   |                |                         |
|---|----------------|-------------------------|
| (3) ǧārat-u-n   | ṣaǧīr-u-n      | manzil-u= <b>hā</b>     |
| voisine.F.SG-NOM-INDÉF  | petit-M.SG-NOM | maison.M.SG-NOM=PR.F.SG |
| « (lit.) une voisine petite sa maison » « = une voisine dont la maison est petite » |                |                         |

Sous (2b) le passage de la qualification du N2 au N1 est assuré lexicalement via la relation partie-tout qui unit les deux noms, ce que prouve l'impossibilité pour les noms d'objets aliénables de figurer dans ce deuxième contexte :

- |  |                |                         |
|--|----------------|-------------------------|
| (4) *ǧārat-u-n   | ṣaǧīrat-u      | <b>l-manzil-i</b>       |
| voisine.F.SG-NOM-INDÉF   | petit-M.SG-NOM | maison.M.SG-NOM=PR.F.SG |
| « (lit.) une voisine petite la maison » « = une voisine à la maison petite » |                |                         |

L'hypothèse que nous défendons est que le mode de qualification appelé *naʿt sababī* n'est pas l'apanage des langues sémitiques et s'instancie dans au moins deux constructions françaises, à savoir la construction Tough (désormais CT) et la construction restrictive partitive (désormais CRP) :

- (5) Le produit est facile à utiliser. (CT)

(6) une femme belle de visage (CRP)

Sans revenir sur les propriétés syntaxiques et distinctives des deux constructions (cf. pour la CT Rosenbaum 1967 ; Postal 1971 ; Giurgea & Soare (2010a,b) ; Hicks 2017 ; Danckaert et Tayalati (à paraître) ; pour la CRP Frei 1939 ; Hanon 1989 ; Salles 1998 ; Tayalati et Mostrov 2022), nous mettrons en exergue un ensemble de propriétés que ces dernières partagent avec celles sous (2), notamment la double prédication inhérente à ce mode de qualification ainsi que le rôle restrictif du complément pour la qualité exprimée. A cela s'ajoute l'existence d'un « lien » entre les deux noms – une condition nécessaire pour que la qualification transite du N2 au N1 – même s'il n'est pas de même nature. En effet, la CT instancie la stratégie syntaxique et exploite la relation de coréférence entre un pronom anaphorique sous-jacent et le N1. Ce pronom peut être explicite dans certaines variétés du français (ex. *Les chihuahuas<sub>i</sub> sont difficiles à les<sub>i</sub> faire manger*). Dans la CRP, le passage de la qualification du N2 au N1 exploite un lien lexical, particulièrement la relation partie-tout entre les référents des deux noms, et ce même lorsque le complément dans la CRP est une action (ex. *un logiciel facile d'utilisation*) : l'action est conçue comme une partie, ce que montre la possibilité d'employer le nom d'action en question dans les constructions de la possession inaliénables (cf. Mostrov 2010), que ce soit en avoir (ex. *Ce logiciel a une utilisation facile.*) ou dans le construction du SN inaliénable en à (ex. *un logiciel à l'utilisation facile*).

L'examen des propriétés des deux constructions sous le prisme de l'AS éclaire grandement leur sémantique commune et révèle d'autres propriétés qui leur sont communes, que ce soit au niveau des conditions de suppression du complément restrictif ou encore le statut des éléments à et de, lesquels ne sont pas des prépositions (*le livre est facile à lire*, \**le livre y est facile* ; *la voiture est agréable de conduite*, \**la voiture en est agréable*). L'ensemble des propriétés listées plaide, au-delà d'une syntaxe différente, en faveur de la catégorisation de la CRP et la CT comme deux constructions instanciant le même mode de qualification.

### Bibliographie sélective

- Al-Ghalāyīnī, M. (1966). *ḡāmi 'u d-durūsi l-'arabiyyati*. Vol 3. 10th ed. Beirut-Sidon: Al-maktaba l-'aṣriyya li-ṭibā'ati wa n-našri.
- Danckaert et Tayalati (à paraître), "The syntax and semantics of indirect predication in French" In Den Dikken Marcel, Kishimoto Hideki. *Formal Perspectives on Secondary Predication*, De Gruyter Mouton, 26 pages.
- Frei, H. (1939). « *Sylvie est jolie des yeux* ». In : Sechchaye A. et al. (éds.). *Mélanges de linguistique offerts à Charles Bally*. Genève : Georg, 185-192.
- Giurgea, I. & E. Soare. (2010a). "Modal non-finite relatives in Romance". In: Becker, M. & E.-M. Remberger (eds.). *Modality and mood in Romance*. Berlin: Mouton de Gruyter, 67-93.
- Giurgea, I. & E. Soare. (2010b). "Predication and the nature of non-finite relatives in Romance". In: Di Sciullo, A. M. & V. Hill (eds.). *Edges, heads, and projections: interface properties*. Amsterdam: Benjamins, 191-213.
- Hanon, S. (1988). « Qui a quoi ? Réflexions sur la possession inaliénable et le verbe *avoir* en français », *Revue romane* 23(2), 161-177.
- Hicks, G. (2017), "Tough-movement", in M. Everaert & H. van Riemsdijk (eds.), *The Wiley Blackwell Companion to Syntax*, Second Edition, Hoboken (NJ), John Wiley & Sons, Inc., 4534-4560.
- Larcher, P. (2017). *Syntaxe de l'arabe classique*. Université d'Aix-Marseille: Presses universitaires de Provence.
- Mostrov, V. (2010). Étude sémantique et syntaxique des compléments adnominaux en à et en de dénotant des parties. Thèse de doctorat. Université de Lille.
- Postal, P. (1971). *Cross-over Phenomena*. New York: Holt, Rinehart and Winston.
- Rosenbaum, P. (1967). *The Grammar of English Predicate Complement Constructions*. Cambridge (Mass.): The MIT Press.
- Ryding, K. C. (2005). *A reference grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.

- Salles, M. (1998). « Être un peu lent de la tête mais rapide des pieds ». *La Linguistique*, Vol. 34, fasc. 1 : 121-136.
- Sylvestre de Sacy, A.-I. (1831). *Grammaire arabe à l'usage des élèves de l'école spéciale des langues orientales vivantes, avec figures*, Seconde édition, corrigée et augmentée à laquelle on a joint un traité de la prosodie et de la métrique des Arabes, Paris, Imprimerie royale.
- Tayalati, F & V. Mostrov (2022). "The indirect attribute mode: a category that comes from elsewhere" In. Vassiliadou, H. & M. Lammert. *Crosslinguistic Perspective on Clear and Approximate Categorization*. Cambridge Scholars Publishing : 217-247.
- Wright , W. (1996 [1896-1898]). *A Grammar of the Arabic Language*. translated from the German of Caspari and edited with numerous additions and corrections. Third edition revised by W. Robertson Smith and M. J. de Goeje with a preface and addenda et corrigenda by Pierre Cachia. 2 vols. Librairie du Liban [reprint de Cambridge: Cambridge University Press].

## L'allomorphie à l'interface phonologie-syntaxe

### Amazigh BEDAR

Aix-Marseille Univ, CNRS, LPL, Aix-en-Provence, France

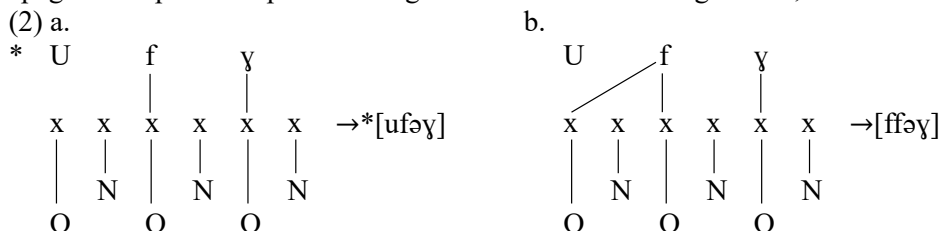
Mon objectif est de fournir une analyse d'un cas particulier d'allomorphie du radical qui ne peut être expliqué en s'échappant à la syntaxe. Les données proviennent du kabyle et contrastent des verbes trilitères  $C_1C_2C_3$  avec des verbes à consonne initiale geminée  $C_1C_1C_2$ . Les données en (1) contrastent les formes dérivées (causatives) avec les formes fléchies (accord avec le sujet de 3SG.M) des deux classes de verbes.

(1)	VERBE	DERIVATION CAUSATIF-VERBE	FLEXION 3SG.M-VERBE	GLOSE
a.	jzəm	ssə-jzəm	jə-jzəm	'couper'
	rzəf	ssə-rzəf	jə-rzəf	'visiter'
b.	ffəy	ss-ufəy	jə-ffəy	*ju-fəy
	ggəθ	ss-ujəθ	jə-ggəθ	*ju-gəθ

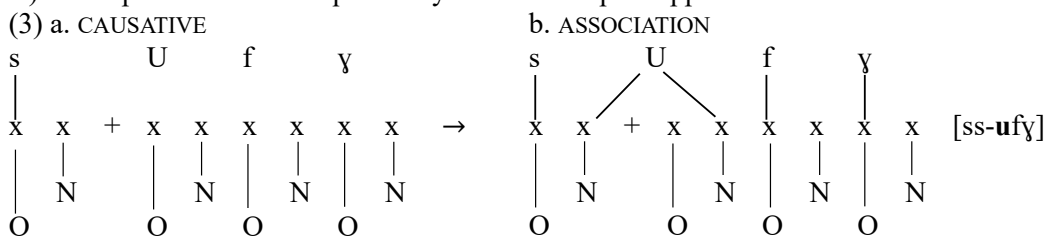
Les formes dérivées et fléchies des verbes trilitères en (1.a) sont régulières. Cependant, les verbes à initiale geminée en (1.b) présentent un cas d'allomorphie du radical : leurs formes causatives sont caractérisées par la présence d'une voyelle fantôme /U/, mais les formes fléchies ne contiennent pas cette voyelle.

**Question 1 :** Pourquoi les verbes en (1.b) présentent-ils la voyelle /U/ dans les formes causatives ?

Jebbour (1993) a démontré que ces verbes à initiale geminée de type  $C_1C_1C_2$  contiennent une voyelle /U/ dans leur forme sous-jacente : /UC<sub>1</sub>C<sub>2</sub>/. La voyelle /U/ ne peut être réalisée dans la forme simple en (2.a) car elle nécessite deux positions vocaliques au niveau du gabarit. Ceci est dû au fait que les voyelles sont phonologiquement longues en berbère, suivant l'analyse proposée par Lowenstamm (1991) pour le sémitique. Par conséquent, la consonne initiale /f/ se propage sur la première position du gabarit verbal et surface geminée, comme illustré en (2.b).



Lorsque le verbe est préfixé par le causatif /s-/ (3a.), la voyelle /U/ est réalisée sur la position vocalique du préfixe et sur la première position vocalique du gabarit verbal, comme illustré en (3.b). C'est pour cette raison que la voyelle fantôme peut apparaître dans la forme causative.



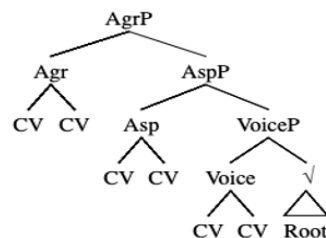
(Jebbour 1993 :13)

**Question 2 :** Si l'explication de cette allomorphie se limite à la phonologie, rien n'empêche l'apparition de la voyelle /U/ dans le cas de la flexion en (1.b), car le site qui héberge le préfixe de l'accord avec le sujet (3SG.M) possède aussi une position vocalique vide sur laquelle /U/ peut se réaliser. Comment, alors, pouvons-nous expliquer son absence dans cette forme ?

Pour répondre à cette question, je propose une analyse qui établit un lien *direct* entre la syntaxe et la phonologie afin d'expliquer ce contraste d'allomorphie du radical entre la flexion et la dérivation.

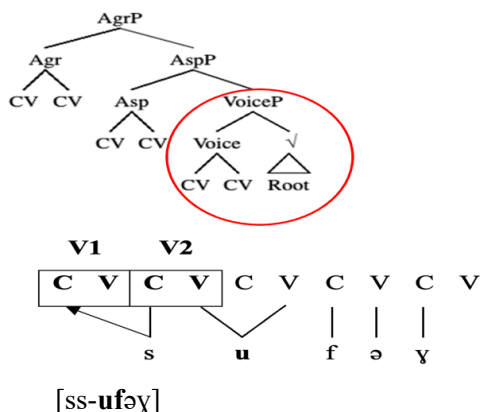
Les éléments nécessaires pour établir la structure du (4) domaine verbal sont les suivants : les accords avec le sujet précèdent les marqueurs d'aspect, qui eux-mêmes précèdent toujours les marqueurs de voix, ces derniers étant toujours proches de la racine. La mise en ensemble de ces éléments donne la structure suivante : [Agr [Asp [Voice [Root]]]].

Je propose que ces têtes syntaxiques soient complexes : chacune est bipartite, c'est-à-dire qu'elle est composée de deux unités CV, comme illustré en (4).

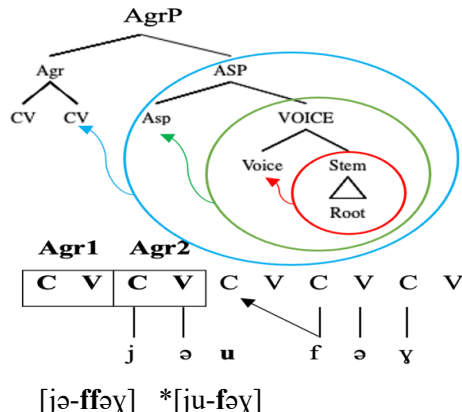


Je propose que, suivant Wood & Tylor (2024) et Workneh (2020), la racine et VoiceP constituent un seul domaine. Ainsi, la voyelle fantôme /U/ apparaît dans la forme causative en (5.a) car elle peut être réalisée sur la position V(ocalique) du site qui accueille les marqueurs de voix. Cependant, elle n'apparaît pas dans la forme fléchie en (5.b) parce que AgrP se situe au-dessus des têtes VoiceP et AspP. Dans ce cas, la racine se combine d'abord avec la voix, puis avec l'aspect et enfin avec l'accord avec le sujet. Comme le site des marqueurs de voix est vide, /U/ ne peut être réalisée, c'est la première consonne de la racine /f/ qui remplit le site initial du gabarit et surface geminée.

(5) a. CAUSATIF



b. 3SG.M



Ainsi, cette allomorphie du radical attestée entre la dérivation et la flexion n'est expliquée qu'à l'interface de la phonologie et de la syntaxe.

**Keywords :** allomorphie, phonologie, syntaxe, interface.

**Références :**

Jebbour, Abdelkrim. 1993. A note on blind roots in Berber. *Linguistica Communicatio* 1 (2): 211-224.

Lowenstamm, Jean. 1991. Vocalic length and syllabic structure in Semitic. *Semitic studies in honor of Wolf Leslau on the occasion of his 85th birthday*. Wiesbaden: Harrassowitz, 949-965.

Wood, Jim & Tyler, Matthew. 2024. Voice. *Lingbuzz*/007197.

Workneh, Desalegn belaynew. 2020. The voice system of Amharic. Doctoral dissertation, The Arctic University of Norway.



## **French as the Second Official Language in Nigeria: Symbolic or Reality**

**By**

**Abdulrahman Burour Ibrahim, Ph.D.**

**Department of Arabic and French, Kwara State University, Malete.**

**Corresponding email: [abdulrahman.ibrahim@kwasu.edu.ng](mailto:abdulrahman.ibrahim@kwasu.edu.ng)**

**Phone Number: +2347031343231.**

### **Abstract**

The proclamation of French as the second official language in Nigeria, alongside English by the military junta, late General Sanni Abacha in 1996 sparked debates and discussions among scholars, policymakers, and the public regarding its implementation, relevance, and impact on Nigeria's linguistic landscape. This paper aims to critically examine the status of French as the second official language in Nigeria, assessing whether this declaration has been effectively realized or remains merely symbolic. The methodology is a blend of qualitative and quantitative approaches, and focuses into the language policies, educational initiatives, and societal attitudes towards French language acquisition and usage in Nigeria. Preliminary findings suggest a mixed reception, with varying levels of implementation across different sectors and regions. While some initiatives have been put in place to promote French language learning and cultural exchange, challenges such as limited resources, inadequate infrastructure, and resistance to linguistic change have hindered its widespread adoption. This paper contributes to the ongoing discourse on language policy in multilingual societies and offers insights into the complexities of implementing a second official language in a diverse and linguistically rich country like Nigeria.

**Keywords: French; Second Official Language; Nigeria; Language Policy; Linguistic Diversity' National Identity.**

**“Vers l’Infini(tif) et Au-delà », Identifying Finiteness in Creoles: A Contrastive Study of the Complementation Systems of Martinican Creole and Haitian Creole.**

Finiteness, and its multifaceted features stemming from morphology, syntax, semantics, and pragmatics, presents unique challenges in the study of Creole languages. The latter, due to their isolating characteristics, challenge conventional linguistic typologies of finiteness therefore making them pivotal for comparative generative syntax. Indeed, the absence of traditional morphological inflections in isolating languages raises important questions about the legitimacy of a finite/nonfinite distinction for Creole languages, akin to debates in other isolating languages like Mandarin Chinese. By examining the behavior of Tense, Mood, and Aspect (TMA) markers, subject agreement/licensing (i.e. obviation, coreference), and negation scope/licensing in complement clauses, this study provides a comparative analysis of Martinican Creole (MC) and Haitian Creole (HC) to determine how applicable and relevant the concept of finiteness is to isolating languages like MC and HC.

This work explores the evolution of the complementation systems of MC and HC separately over a 40-year period (i.e. diachronically and intra-linguistically) and compares them with each other as they are attested nowadays (i.e. synchronically and inter-linguistically) by using data from my HC and MC corpora, and native speaker acceptability judgments. Additionally, I use a previously established framework (Bianchi 2003, Landau 2004, Adger 2007, Nikolaeva 2007, 2010; Wurmbrand 2014, 2015) to describe these systems in terms of their syntactic/semantic finite nature as well as the size of their syntactic projections. I examine causative verb HC/MC *fè* ‘make’, volitional verb HC *vle* and MC *lé* ‘want’, modals HC *ka/dwe* and HC *pé/dwet* ‘must/can’, raising verb HC *sanble* and MC *sanm* ‘seem’, and the impersonal expression HC/MC *fok* ‘one must’. These predicates host complement clauses with each their own syntactic properties with regards to subject licensing, temporal specification, the licensing of the HC/MC zero-marker, and the licensing of negation (as summarized in Figure 1). Diachronically, we observe featural differences for these predicates, some of which might be attributable to language contact. Synchronically, these predicates show properties that, at times, make MC and HC more similar to each other than previously claimed (e.g. with causative verbs), but at other times they appear more different from each other than expected (e.g. with modals).

In conclusion, the study shows that Creole languages, despite having little inflectional morphology, do in fact have clauses that should be considered finite and others nonfinite. Moreover, I show that finer-grained distinctions can be made based on the size of these complement clauses and that they compare cross-linguistically to languages outside of the Creole family.

References:

- Adger, D. (2007). Three Domains of Finiteness: a Minimalist Perspective. *Finiteness: Theoretical and Empirical Foundations*, 23–58.
- Bianchi, V. (2003). On Finiteness as Logophoric Anchoring. *Temps et Point de Vue/tense and Point Of View*, 213–246.
- Landau, I. (2004). The Scale of Finiteness and the Calculus of Control. *Natural Language & Linguistic Theory* 22 (4), 811–877.
- Nikolaeva, I. (2007). *Finiteness: Theoretical and Empirical Foundations*. OUP Oxford.
- Nikolaeva, I. (2010). Typology of Finiteness. *Language and Linguistics Compass* 4 (12), 1176–1189.
- Wurmbrand, S. (2014). Tense and Aspect in English Infinitives. *Linguistic Inquiry* 45 (3), 403–447.
- Wurmbrand, S. (2015). Restructuring Cross-linguistically. In *Proceedings of the northeastern linguistics society annual meeting*, Volume 45, pp. 227–240

Overview of Findings:

Finite complement clauses					
Predicate types	Subject Type	Temporal Specification	Zero-marker Licensing	Negation Type	Complement Type
<i>Make</i>	Free	Independent	Yes	Sentential	Indicative
<i>Seem</i>	Free	Independent	Yes	Sentential	
<i>Fok</i>	Free	Future-oriented	No	Sentential	Subjunctive
<i>Want</i>	Free	Future-oriented	No	Sentential	

Nonfinite complement clauses					
<i>Want</i>	Non-coreferential	Future-oriented	No	vP	Partial control
<i>Seem</i>	Anaphorical	Dependent	No	vP	Raising
<i>Epistemic Modals</i>	Anaphorical	Dependent	No	vP	Raising
<i>Deontic Modals</i>	Anaphorical	Anaphorical	No	vP	Exhaustive control
<i>Make</i>	Non-coreferential	Anaphorical	No	vP	

Figure 1 – Description of the complementation systems of MC and HC

*Note: The findings summarized in this figure suggest that the main properties that are absent in nonfinite complement clauses in HC and MC are 1) the free subject type, 2) the ability to license sentential negation, 3) the independent temporal specification on the complement clause and 4) its ability to license the zero-marker.*

## **Insécurité linguistique chez les jeunes créolophones de la Martinique : le jeu de rôle comme outil d'investigation de la compétence langagière dans la langue minorée.**

L'insécurité linguistique, ce sentiment de culpabilité ou d'illégitimité par rapport à sa façon de s'exprimer, est un phénomène qui survient dès lors qu'on prend conscience de l'existence et de la hiérarchisation des variations de langue (Remysen 2018). Chez les bilingues français-créole de la Martinique, l'insécurité linguistique est présente tant en français (français standard, français hexagonal, français régional martiniquais) qu'en créole (créole basilectal, créoles ordinaires) ; elle se manifeste plus fortement en créole, avec pour conséquence une auto-censure de l'expression dans cette langue, particulièrement chez les jeunes, qui éprouvent un malaise à parler cette langue qu'ils côtoient pourtant quotidiennement. Cette insécurité peut pénaliser leur usage de la langue et entraver le développement de leurs compétences langagières. Comment donc le linguiste et l'enseignant de langues peuvent-ils apprécier l'étendue des ressources langagières des jeunes dans une langue qu'ils parlent de façon restreinte? Depuis les années 70, de nombreuses actions de promotion de la langue créole ont été mises en place par les activistes martiniquais et, depuis les années 2000, par l'institution scolaire afin de promouvoir la légitimité de cette langue minorée, et d'en pérenniser l'usage et la transmission. Pourtant, le sentiment d'illégitimité persiste chez les jeunes créolophones.

Nous abordons cette problématique sous l'angle de l'approche communicative (Hymes 1984), et en particulier sous celui du jeu de rôle (Leblanc 2002). Ce dernier est une activité interactive qui s'effectue à deux ou à plusieurs participants et qui consiste en une simulation de la vie réelle. Le jeu de rôle est une activité sociale qui repose sur des cadres (notamment celui de la fiction) lesquels permettent de dépasser les contraintes de la vie réelle.

Dans cette communication, nous interrogeons les manifestations et les sources de l'insécurité linguistique des jeunes créolophones martiniquais. Nous nous appuyons pour cela sur des entretiens semi-directifs effectués en 2024 auprès de cinq enseignants de créole de niveau collège et lycée de l'Académie Martinique. Les questions portaient entre autres sur les difficultés rencontrées par les élèves en ce qui a trait au développement de leur compétence langagière et communicative en créole, ainsi que sur l'usage du jeu de rôle en classe comme outil d'enseignement et d'évaluation. L'analyse discursive des entretiens met en lumière le fait que l'usage répété de la langue dans un cadre bienveillant permet à ces jeunes locuteurs de « délier leur langue ». Il s'avère que le jeu de rôle en classe constitue un cadre favorisant le déploiement des ressources langagières. Suite à cette analyse, nous proposons que le jeu de rôle constitue un outil d'observation / évaluation de la compétence langagière et communicative dont le linguiste (comme l'enseignant de langues) peut se saisir afin de révéler l'attirail expressif des locuteurs de langues minorées. Finalement, nous montrons que le jeu de rôle, outil jusque-là quasiment inexploité par la recherche en créolistique, pourrait offrir un nouveau regard sur la compétence langagière en créole des jeunes générations, et donc sur l'avenir du créole.

Hymes, D. (1984). *Vers la compétence de communication*. Paris : Hatier-CRÉDIF.

Leblanc, Marie-Cécile (2002). *Jeu de rôle et engagement : évaluation de l'interaction dans les jeux de rôles de français langue étrangère*. Paris : L'Harmattan.

Remysen, W. (2018). « L'insécurité linguistique à l'école : un sujet d'étude et un champ d'intervention pour les sociolinguistes », *La linguistique et le dictionnaire au service de l'enseignement du français au Québec*, 25-59.

## Competition between personal and agent nouns with borrowed roots in contemporary Lithuanian

Lina Inčiuraitė-Noreikienė

Vilnius University, Faculty of Philology

Competition in inflection and derivation has attracted considerable attention in worldwide linguistics over the last few years (cf. Huyghe & Varvara, 2023; Stundžia & Inčiuraitė-Noreikienė, 2023; Stundžia & Inčiuraitė-Noreikienė, 2022; Inčiuraitė-Noreikienė, 2022; Vaskelienė, 2022; Huyghe, Wauquier, 2020; Rainer et al., 2019; Santana Lario et al., 2017; Inčiuraitė-Noreikienė & Stundžia, 2016). According to Štekauer (2017: 15), competition is “an inherent and universal feature of natural languages”. It is defined as “the coexistence of two or more affixes for the same base and for the expression of the same semantic category if restrictions (e.g., phonological, morphological) do not apply and no semantic or distributional differences are observed” (Fernández-Alcaina, 2017: 166). Debatable points revolve around “the same base” and “the same semantic category”. I hypothesize that competition may occur among nouns belonging to different but very close derivational (semantic) categories, such as those of personal and agent nouns, e.g. the derived denominal personal noun *eksportininkas*, -ė ‘exporter’ DI, DCL, DSL, DL, CCLL (from *eksportas* ‘an export’) competes with the derived deverbal agent noun *eksportuotojas*, -a ‘id.’ DSL, DL, CCLL (from *eksportuoti* ‘to export’) (Stundžia & Noreikienė, 2023: 59). It has been observed that only certain affixes are attached to borrowed bases, particularly in relation to personal and agent nouns. This observation prompts the immediate question: why do only certain nominal affixes attach to borrowed bases for personal and agent nouns, while others do not? The study adopts a synchronic approach to analyse the competition between hybrid suffixed and suffixed material borrowings (cf. Marchand, 1969; Urbutis, 2009). Descriptive, analytical, and comparative methods will be used to analyze the data.

The study also explores noun competition centered around clusters, which refer to a set of synonymous nouns belonging to different but closely related derivational (semantic) categories, such as personal and agent nouns. The data show that 519 clusters of competing personal and agent nouns (85%) consist of two members (e.g. *propagandininkas*, -ė ‘propagandist’ DCL, DL / *propagandistas*, -ė ‘id.’ DI). 73 clusters (12%) contain three members (e.g. *legionininkas* ‘legionary’ DCL, DL / *legionierius* ‘id.’ DI, DCL, DS, DL / *legionistas* ‘id.’ CCLL). 17 clusters (3%) have four members (e.g. *parfumerininkas*, -ė CCLL ‘parfumer’ / *parfumininkas*, -ė CCLL ‘id.’ / *parfumeristas*, -ė IS ‘id.’ / *parfumeris*, -ė CCLL ‘id.’). 3 clusters (0%) have five members (e.g. *strategininkas*, -ė ‘strategist’ IS / *strateguotojas* ‘id.’ CCLL / *strategistas* ‘id.’ IS / *strategikas* ‘id.’ DL / *strategas* ‘id.’ DI, DCL, DL) (cf. Stundžia & Inčiuraitė-Noreikienė, 2023; Stundžia & Inčiuraitė-Noreikienė, 2022; Inčiuraitė-Noreikienė, 2022). The competition between personal and agent nouns is influenced by a combination of semantic, stylistic, and genre-specific factors. The disparity in usage frequency between these competing personal and agent nouns will be demonstrated using normalized frequencies from the CCLL.

**Keywords:** competition, personal noun, agent noun, contemporary Lithuanian

## References

- Fernández-Alcaina, C. (2017). Availability and unavailability in English word-formation. In J. Santana-Lario, & S. Valera (Eds.), *Competing Patterns in English Affixation* (pp. 163–206). Peter Lang.
- Huyghe, R., & Varvara, R. (Eds.). (2023). Quantitative approaches to affix rivalry. *Thematic issue of Word Structure* 16(1), Edinburgh: Edinburgh University Press.
- Huyghe, R., & Wauquier, M. (2020). What's in an agent? *Morphology* 30(3). 185–218.
- Inčiuraitė-Noreikienė, L., & Stundžia, B. (2016). Searching for competing patterns in morphological derivation: The case of adjective borrowing, *Skase Journal of Theoretical Linguistics* 13(2), (pp. 189–214).
- Inčiuraitė-Noreikienė, L. (2022). Hibridinių vedinių ir skolinių konkurencija [Competition between hybrid derivatives and borrowings], *Baltistica* 57(1), (pp. 149–174). DOI: [10.15388/baltistica.57.1.2472](https://doi.org/10.15388/baltistica.57.1.2472)
- Marchand, Hans. (1969). *The Categories and Types of Present-day English Word Formation: A Synchronic Diachronic Approach*, 2nd edition. München: C. H. Beck'sche Verlagsbuchhandlung.
- Rainer, F., & Gardani, F., & Dressler, W. U., & Luschützky, H. C. (Eds.). (2019), *Competition in inflection and word-formation*. Cham: Springer.
- Santana-Lario, J., & Valera, S. (Eds.). 2017. *Competing patterns in English affixation*. Bern: Peter Lang.
- Štekauer, P. (2017). Competition in natural languages. In J. Santana-Lario, & S. Valera (Eds.), *Competing Patterns in English Affixation* (15–32). Peter Lang.
- Stundžia, B., & Inčiuraitė-Noreikienė, L. (2022). Vardažodinių hibridinių priesagų vedinių, pavadinančių asmenis, ir bendrašaknių skolinių konkurencija dabartinėje lietuvių kalboje [Competition between hybrid suffixed denominal personal nouns and common root borrowings in contemporary Lithuanian]. *Baltistica* 57(2), 239–261. DOI: [10.15388/baltistica.57.2.2486](https://doi.org/10.15388/baltistica.57.2.2486)
- Stundžia, B., & Inčiuraitė-Noreikienė, L. (2023). Exploring competing patterns in morphological derivation: the case of personal and agent nouns with borrowed roots in contemporary Lithuanian. *Skase Journal of Theoretical Linguistics* 20(2), 56–78.
- Urbutis, Vincas. (2009). *Žodžių darybos teorija* [Word Formation Theory]. Vilnius: Mokslo ir enciklopedijų leidybos institutas.
- Vaskelienė, J. (2022). Dėl bendrašaknių lietuvių kalbos daiktavardžių su priesagomis *-(i)acija* ir *-imas* [On common root nouns of the Lithuanian language with the suffixes *-(i)acija* and *-imas*], *Respectus Philologicus* 41(46), 94–106. DOI :[10.15388/RESPECTUS.2022.41.46.111](https://doi.org/10.15388/RESPECTUS.2022.41.46.111)

## Sources

CCLL – Corpus of Contemporary Lithuanian Language, Kaunas: Vytauto Didžiojo universitetas, 2011, <http://tekstynas.vdu.lt/tekstynas/> (Accessed 2024-04-20).

DCL – Stasys Keinys (ed.), Dabartinės lietuvių kalbos žodynas [Dictionary of Contemporary Lithuanian], 8th edition (revised and updated) Vilnius: Lietuvių kalbos institutas, 2021, elektroninis variantas, 2021 (<https://ekalba.lt/dabartines-lietuviu-kalbos-zodynas>) (Accessed 2024-04-20)

DI – Kompiuterinis tarptautinių žodžių žodynas [Computerised Dictionary of Internationalisms] ‘Interleksis’, Vilnius: Alma littera, Fotonija, 2008.

DL – Lietuvių kalbos žodynas [Dictionary of Lithuanian], 1–20 (1941–2002): electronic version, Gertrūda Naktinienė (ed.), Vilnius: Lietuvių kalbos institutas, 2005 (updated 2018) ([www.lkz.lt](http://www.lkz.lt)) (Accessed 2024-04-20).

DSL – Danutė Liutkevičienė (ed.) Bendrinės lietuvių kalbos žodynas [Dictionary of Standard Lithuanian] (<https://ekalba.lt/bendrines-lietuviu-kalbos-zodynas>) (Accessed 2024-04-20).

IS – The internet search engine Google.com was used to search for competing hybrid derivatives and common root borrowings.



## **Discursive Representation of Wildfires in the Mainstream Media:**

### **The Case of Anthropocentrism**

**Liudmila Arcimavičienė**

*Vilnius University, Faculty of Philology*

**Lina Marčiulionytė**

*Vilnius University, Faculty of Philology*

Given the unprecedented scale and intensity of wildfires observed across Europe during the summers of 2022 and 2023, their discursive representation in the mainstream news media is particularly valued for its framing patterns and implied ideological meanings. In this study, the discursive representation of wildfires in the news media is analysed from the perspective of anthropocentrism and its underlying linguistic features. For this purpose, the data sample of the Lithuanian and English news reporting from the state broadcasters BBC News and LRT News based on the following criteria: 1) the summers of 2022 and 2023, and 2) the key words wildfires and fires” in English, and “miško gaisrai, gaisrai” in Lithuanian. A total of 60 news media articles (with a total word count of 54,097) were examined, with 30 articles retrieved from each media outlet in both English and Lithuanian.

Procedurally, the collected data sample was analysed within the theoretical framework of Critical Discourse Studies (Fairclough, 1995/2013; Van Dijk, 2009, 2022; Angermuller, 2020), with a focus on thematic analysis. In the first step, the inductive thematic analysis (White, 2006; Van Leeuwen, 2008) was used to identify recurring themes or discourse topics (van Dijk, 2022: 183) in the data sample. Subsequently, deductive thematic analysis was applied to examine how the prevalent discourse topics align with human interests and environmental interests. It is hypothesised that, even though wildfires are given more discourse space in the news media narrative over time, their thematic construal will amplify and routinise anthropocentric attitudes. Finally, the relevance of time is tested by comparing the wildfire news reporting narratives from the summers of 2022 and 2023. The findings indicate that, despite certain attempts to create a sense of urgency regarding environmental interests, human-centered narrative is foregrounded in both media sources. This contributes to the routinisation of the climate change narrative, climate change skepticism (Sharman, 2014) and an underlying sense of uncertainty (Collins and Nerlich, 2016).

**Keywords:** news reporting, climate change discourse, wildfires, thematic analysis, narrative, cross-linguistic analysis, anthropocentrism, routinisation.

**References:**

- Angermuller, J. (2020). Poststructuralist Discourse Studies: from structure to practice. In A. de Fina, & A. Georgakopoulou (Eds.) *The Cambridge Handbook of Discourse Studies* (pp. 235–254). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108348195.012>
- Collins, L. C., & Nerlich, B. (2016). How certain is ‘certain’? Exploring how the English-language media reported the use of calibrated language in the Intergovernmental Panel on Climate Change’s Fifth Assessment Report. *Public Understanding of Science*, 25(6), 656-673.
- Fairclough, N. (1995/2013). *Critical Discourse Analysis: The Critical Study of Language*. Singapore: Routledge.
- Sharman, A. (2014). Mapping the climate sceptical blogosphere. *Global Environmental Change*, 26, 159-170. Elsevier. <https://doi.org/10.1016/j.gloenvcha.2014.03.003>
- Van Dijk, Teun A. (2009). News, discourse, and ideology. In K. Wahl-Jorgensen, & T. Hanitzsch (Eds.) *The Handbook of Journalism Studies* (pp. 191-204). New York and London: Routledge.
- Van Dijk, T. A. (2022). Ideology in cognition and discourse. S. K. Määttä, & K. M. Hall (Eds.). *Mapping Ideology in Discourse Studies* (pp. 118-137). Boston/Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9781501513602>
- Van Leeuwen, T. (2008). *Discourse and Practice: New Tools for Critical Discourse Analysis*. Oxford: Oxford University Press.
- White, P. R. R. (2006). Evaluative semantics and ideological positioning in journalistic discourse. In I. Lassen (Ed.) *Mediating Ideology in Text and Image: ten critical studies* (pp. 37-67). Amsterdam: John Benjamins. <https://doi.org/10.1075/dapsac.18.05whi>

# Why do objects have many names?

## A study on word informativeness in language use and lexical systems.

Anonymous submission

001 Human lexicons contain many different  
002 names that speakers can use to refer to the  
003 same object, e.g. speakers can call the same  
004 color *purple* or *magenta*, reflecting the fact  
005 that our semantic system allows for a soft  
006 mapping between referents and labels (Rosch  
007 and Mervis, 1975; Snodgrass and Vander-  
008 wart, 1980; Graf et al., 2016; Gualdoni et al.,  
009 2023). Previous work has either studied  
010 speaker adaptation to context (Graf et al.,  
011 2016; Monroe et al., 2017; Mädebach et al.,  
012 2022) or lexical systems without tackling in-  
013 context communication (Regier et al., 2015;  
014 Zaslavsky et al., 2018; Xu et al., 2020; Za-  
015 slavsky et al., 2021). We aim at bridging the  
016 gap between these traditions, exploring why a  
017 soft mapping between referents and names is  
018 optimal for communication.

019 We analyze color naming data for English  
020 and Mandarin Chinese, gathered by (Monroe  
021 et al., 2017) —see Figure 1). We define the  
022 meaning of a word as the set of objects that the  
023 word denotes and, more concretely, the region  
024 in the visual space that these objects occupy  
025 (Dowty et al., 1981; Heim and Kratzer, 1998;  
026 Gärdenfors and Williams, 2001; Erk, 2009).  
027 We propose a new measure of word informa-  
028 tiveness ( $I_w$ ), namely the size of this region;  
029 see Figure 2a. With this measure, we show  
030 that speakers adapt to the context as expected  
031 given previous work, using more informative  
032 words in more challenging contexts (as illus-  
033 trated in Figure 1, and Figure 3).

034 Moreover, we examine the effects adapta-  
035 tion has on lexical systems, via simulations  
036 and empirical data collection. We define the  
037 informativeness of a lexical system  $L$  ( $I_L$ ) as  
038 the average of  $I_w$  over a number of interac-  
039 tions  $N$ . We compare actual systems to two  
040 types of alternative systems containing only  
041 high- or low-informativeness words. Actual

042 systems have *more* lexical items than these al-  
043 ternatives, thus are not the most compact in  
044 terms of number of words. We find these sys-  
045 tems are nevertheless more efficient than alter-  
046 natives: they achieve a good balance between  
047 communicative accuracy and communicative  
048 cost across interactions. We also identify a  
049 number of pragmatic effects that lead speak-  
050 ers to use over- and under-informative expres-  
051 sions, and discuss the limitations of a purely  
052 denotation-based notion of informativeness.

053 We conclude that optimal lexical systems  
054 are those where multiple words can apply to  
055 the same referent, conveying different amounts  
056 of information. Such systems allow speakers to  
057 simultaneously maximize communication ac-  
058 curacy and minimize the amount of informa-  
059 tion conveyed when communicating about ref-  
060 erents in contexts.



Figure 1: To allow successful identification of a target color chip (in the black frame), a general term like *purple* is sufficient when the context is not challenging (above), but not in a more challenging context target (bottom).

### References

- David R. Dowty, Robert Eugene Wall, and Stanley Peters. 1981. *Introduction to Montague Semantics*. Springer.
- Katrin Erk. 2009. Representing words as regions in vector space. In *Proceedings of the Thirteenth*

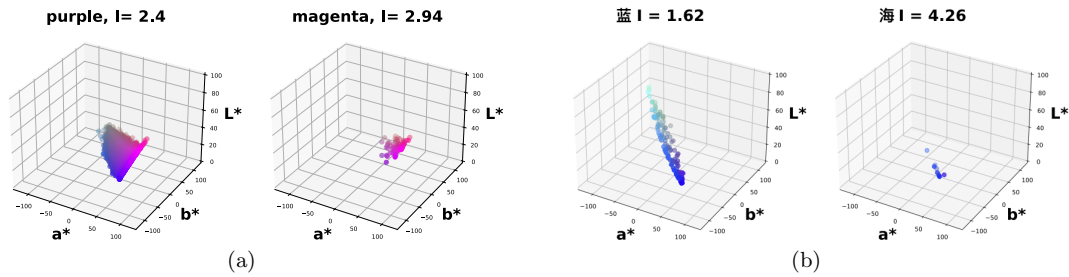


Figure 2: Denotation in color space of the words *purple* and *magenta* (a) and 蓝 “blue” and 海 “ocean” (b).

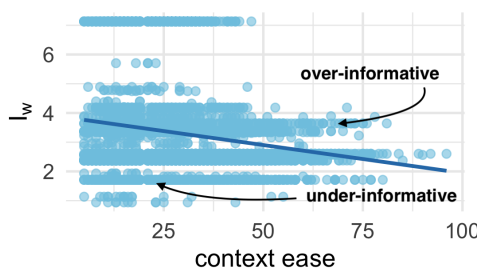


Figure 3: Harder contexts require higher  $I_w$ , English data of Monroe et al. (2017).

067  
068

*Conference on Computational Natural Language Learning (CoNLL-2009)*, pages 57–65.

069  
070  
071  
072  
073  
074  
075

Caroline Graf, Judith Degen, Robert X D Hawkins, and Noah D Goodman. 2016. Animal, dog, or dalmatian? level of abstraction in nominal referring expressions. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, pages 2261–2266, Austin, TX. Cognitive Science Society.

076  
077  
078  
079  
080  
081

Eleonora Gualdoni, Thomas Brochhagen, Andreas Mädebach, and Gemma Boleda. 2023. What’s in a name? A large-scale computational study on how competition between names affects naming variation. *Journal of Memory and Language*, 133:104459.

082  
083  
084  
085

Peter Gärdenfors and Mary-Anne Williams. 2001. Reasoning about categories in conceptual spaces. In *Proceedings of the IJCAI*, pages 385–392.

086  
087  
088

Irene Heim and Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Blackwell, Malden, MA.

089  
090  
091  
092  
093  
094

Will Monroe, Robert X.D. Hawkins, Noah D. Goodman, and Christopher Potts. 2017. Colors in context: A pragmatic neural model for grounded language understanding. *Transactions of the Association for Computational Linguistics*, 5:325–338.

A. Mädebach, Torubarova E., E. Gualdoni, and G. Boleda. 2022. Effects of task and visual context on referring expressions using natural scenes. In *Proceedings of the 44th Annual Conference of the Cognitive Science Society*. Cognitive Science Society. 095  
096  
097  
098  
099  
100

Terry Regier, Charles Kemp, and Paul Kay. 2015. *Word Meanings across Languages Support Efficient Communication*, pages 237–263. 101  
102  
103

Eleanor Rosch and Carolyn B Mervis. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4):573–605. 104  
105  
106  
107

Joan Gay Snodgrass and Mary Vanderwart. 1980. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology. Human learning and memory*, 6 2:174–215. 108  
109  
110  
111  
112  
113

Yang Xu, Emmy Liu, and Terry Regier. 2020. Numerical systems across languages support efficient communication. *Open Mind*, 4:1–14. 114  
115  
116

Noga Zaslavsky, Charles Kemp, Terry Regier, and Naftali Tishby. 2018. Efficient compression in color naming and its evolution. *PNAS*, 115(31):7937–7942. 117  
118  
119  
120

Noga Zaslavsky, Mora Maldonado, and Jennifer Culbertson. 2021. Let’s talk (efficiently) about us: Person systems achieve near-optimal compression. In *43rd Annual Meeting of the Cognitive Science Society*. 121  
122  
123  
124  
125

## Le traitement linguistique et cognitif de la composante de manière : le cas de l'interprétation simultanée français-anglais et anglais-français

Introduction. La manière comme composante sémantique des événements a été largement étudiée dans le domaine spatial, entre les langues à cadre verbal (LCV) et les langues à cadrage satellitaire (LCS). Elle est considérée plus saillante linguistiquement et cognitivement (Slobin, 1987, 2003, 2006; Talmy, 1985, 2000) en anglais (LCS) qu'en français (LCV) en raison de son encodage (lexical) plus fréquent (*he ran inside* vs *il entra en courant*). Cependant, la manière structure d'autres domaines (Fellbaum, 1999, 2002) comme la communication (*murmurer*), la création (*élaborer*) ou encore la mort (*noyer*).

Objectifs. Nous présenterons les résultats d'une étude antérieure (Combe & Stosic, 2024) dans laquelle nous testions l'hypothèse d'une saillance supérieure de la manière en anglais qu'en français *tous domaines confondus*, en comparant des discours originaux et des discours interprétés, afin d'observer si elle est plus transférée dans l'une des directions (*fr-ang* : FA ou *ang-fr* : AF), en fonction de la charge cognitive des interprètes, modulée via le contrôle du débit de parole source (Seeber, 2011).

Méthodologie. Nous avons sélectionné 8 discours politiques traitant de thèmes variés qui ont été interprétés simultanément (tableau 1) et avons relevé systématiquement toutes les expressions correspondant à la définition de la manière comme « Une valeur sémantique complexe (...) qui consiste en la diversification d'un procès, d'un état ou d'une qualité par une spécificité qualitative. » (Stosic, 2011, 2020).

Tableau 1. Présentation du corpus

Langue	Intervenants	Date	Taille (mots)	Débit de parole source (mots par minute)
ANG	Hilary Clinton	28.07.2016	1186	79
	Barack Obama	20.09.2016	1668	111
	Theresa May	20.01.2017	2217	148
	David Cameron	10.11.2015	2349	158
FR	François Mitterrand	24.09.1990	1394	93
	Jacques Chirac	26.01.2005	1565	104
	Dominique De Villepin	14.02.2003	1852	123
	Emmanuel Macron	22.09.2020	2236	149

Résultats. Notre premier résultat, présenté dans la figure 1, met en évidence la distribution différente de deux stratégies principales de spécification de la manière des événements entre orateurs politiques anglophones et francophones, les premiers privilégiant davantage l'encodage lexical, celui-ci étant significativement plus fréquent en anglais ( $Khi\text{-carré} = 4.4123$ ,  $p\text{-value} = 0.03568$ ) :

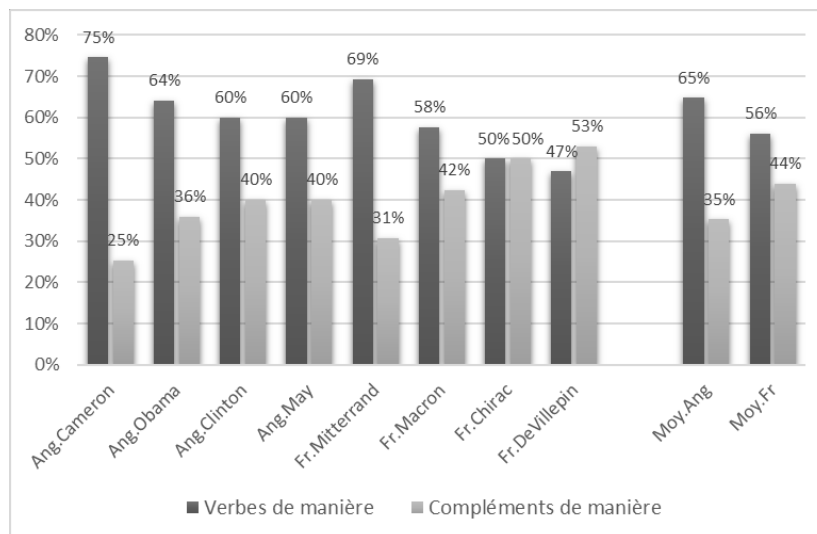


Figure 1. Encodage lexical vs syntactique de la manière

Deuxièmement, sur le plan interlinguistique (voir figure 2), la déperdition du sens de manière est nettement moins prononcée dans la direction FA (33 % en moyenne) que dans la direction AF (45 % en moyenne) en raison d'un transfert significativement plus élevé de la manière lexicale vers l'anglais ( $\text{Khi-carré} = 4.7311$ ,  $df = 1$ ,  $p = 0.02962$ ). D'autre part, les interprètes FA ajoutent davantage de manière, là où elle n'était pas codée dans les discours originaux. Enfin, alors que les taux de déperdition AF augmentent en fonction du débit de parole source, les interprètes FA ne semblent pas sensibles à cette variable.

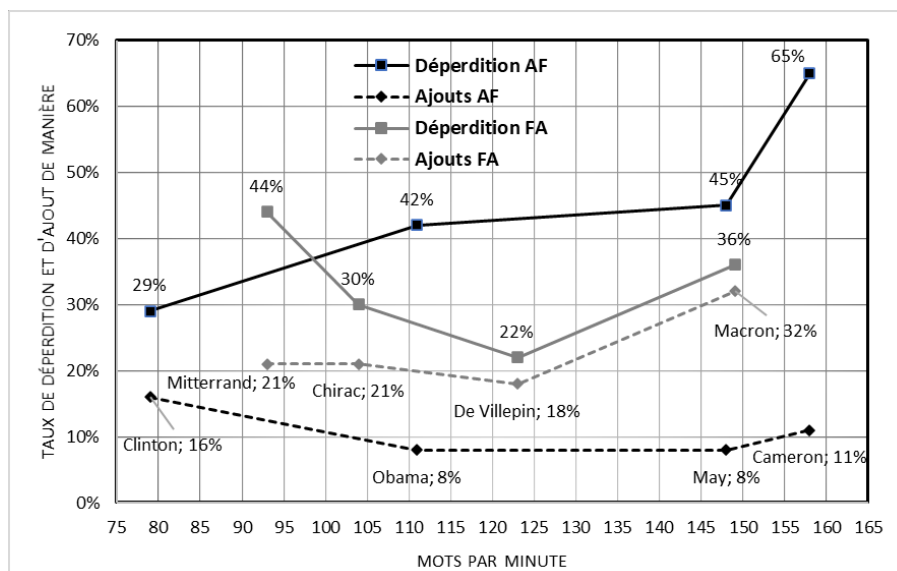


Figure 2. Taux de déperdition et d'ajout de manière

Discussion. De façon inattendue, l'anglais comme le français affichent une préférence pour le codage lexical de la manière, mais cette préférence est plus forte en anglais, ce qui était attendu dans l'hypothèse d'une généralisation des différences typologiques entre ces deux langues à plusieurs domaines sémantiques. De plus, le fait que les interprètes FA transfèrent davantage la manière indépendamment du débit de parole source suggère que cette composante sémantique bénéficie d'une codabilité plus forte en langue anglaise et que son traitement interlinguistique pourrait être cognitivement plus efficace quand le discours cible est produit en anglais. Notre étude fournit donc des indices supplémentaires que la manière est plus saillante en anglais (LCS) qu'en français (LCV), ce potentiellement quel que soit le domaine sémantique.

**Mots-clés :** *Sémantique de la manière ; Typologie de Talmy ; Interprétation simultanée ; Charge cognitive.*

### Références :

- Fellbaum, C. (1999). La représentation des verbes dans le réseau sémantique WordNet. *Langages*, 33(136), 27-40. <https://doi.org/10.3406/lgge.1999.2210>
- Fellbaum, C. (2002). On the semantics of troponomy. In R. Green, C. Bean, & S. H. Myaeng (Éds.), *The semantics of relationships : An interdisciplinary perspective* (p. 23-34). Kluwer Academic Publishers. [https://doi.org/10.1007/978-94-017-0073-3\\_2](https://doi.org/10.1007/978-94-017-0073-3_2)
- Seeber, K. G. (2011). Cognitive load in simultaneous interpreting : Existing theories — new models. *Interpreting. International Journal of Research and Practice in Interpreting*, 13(2), 176-204. <https://doi.org/10.1075/intp.13.2.02see>
- Slobin, D. I. (1987). Thinking for Speaking. *Annual Meeting of the Berkeley Linguistics Society*, 13, 435. <https://doi.org/10.3765/bls.v13i0.1826>
- Slobin, D. I. (2003). Language and Thought Online : Cognitive Consequences of Linguistic Relativity. In D. Gentner & S. Goldin-Meadow (Éds.), *Language in Mind : Advances in the study of language and thought* (p. 157-192). The MIT Press. <https://doi.org/10.7551/mitpress/4117.003.0013>
- Slobin, D. I. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse and cognition. In M. Hickmann & S. Robert (Éds.), *Space in languages : Cognitive systems and linguistic categories* (p. 59-81). John Benjamins. <https://psycnet.apa.org/doi/10.1075/tsl.66.05slo>
- Stosic, D. (2011). Le sens de manière comme critère de définition d'un paradigme. In *La construction d'un paradigme* (p. 117-142). Echinex. <https://shs.hal.science/halshs-00654939>
- Stosic, D. (2020). Defining the concept of manner : An attempt to order chaos. *Testi e Linguaggi*, 14, 127-150.
- Talmy, L. (1985). Lexicalization patterns : Semantic structure in lexical forms. In *Language typology and syntactic description: Vol. 3: Grammatical categories and the lexicon*. Cambridge University Press.
- Talmy, L. (2000). *Towards a cognitive semantics* (Vol. 2). MIT press.

## Resilience Predictors in International Students – the Role of Multilingualism

Advancements in technology and increased cross-border workforce mobility have led society to intricate challenges and meaningful transformations over recent decades. These developments have resulted in significant changes for societies and economies worldwide. Internationalisation became prevalent in the education field as well. Nowadays, international education has become one of the must-have experiences for aspiring academia members. However, some studies (Baier, 2013; Presbitero 2016; Zhou et al 2007) suggest that adapting to a new culture can be challenging and international students may experience culture shock and various adversities. Previous research has already associated unfavourable life events with numerous undesirable consequences. However, even when facing adversity, people who demonstrate significant resilience seem less affected by hardship and can maintain a positive outlook on life. Alvord and Grados (2005) note that in addition to understanding the factors that can place individuals at risk of negative outcomes, it is also important to develop an understanding of factors that can influence an individual's level of resilience. Therefore, conducting thorough research in this area is essential to find how students can build their resilience and shape effective policies and programs that cater to their diverse needs. As well as promote their successful integration into both academic and professional spheres.

To help fill this gap, this study investigates the experiences of international students in Germany as Germany is often named as one of the top destinations for students (UNESCOIS 2020). 30 international students coming from post-Soviet countries were selected to investigate international students' resilience predictors and the role, their multilingualism plays in this process. The reason for choosing this particular group is based on the fact that their motivation to seek higher education in Western Europe is still unclear in the scientific literature and not much is known about their ways of dealing with adversity. To assess the resilience level of International students from the post-Soviet countries the Connor-Davidson Resilience Scale (CD-RISC) was utilized. The CD-RISC (2003) is a 25-item scale which aims to determine the ability to deal effectively with stress and adversity. Afterwards, semi-structured interviews were conducted to elicit data regarding participants' internationalization experiences and their perceived role of multilingualism in their journey.

The study has shown that certain factors, such as good peer relationships, cultural ties and supporting school environments can positively influence resilience-building in students. The research found that bi and multilingualism can also play a role in strengthening students' self-confidence. Students were noted to see their multilingualism as a valuable asset in the German labour market. some even found ways of strengthening their financial situation using their multilingual skills (by tutoring or interpreting). Multilingualism also seems to often function as a coping mechanism to navigate difficult situations and uncertainties in new surroundings.

In summary, this is a multidisciplinary research, which emphasizes the importance of resilience predictors, including multilingualism and finding new instruments to promote international students' well-being throughout their study journey.



**References:**

- Alvord, M. K., & Grados, J. J. Enhancing Resilience in Children: A Proactive Approach. *Professional Psychology: Research and Practice*, (2005), 36(3), 238–245.  
<https://doi.org/10.1037/0735-7028.36.3.238>
- Connor KM, Davidson JR. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). *Depress Anxiety*. 2003;18(2):76-82. doi: 10.1002/da.10113. PMID: 12964174.
- Baier, Stefanie Theresia. "International students: Culture shock and adaptation to the U.S. culture." Eastern Michigan University, 2015. <https://commons.emich.edu/theses/523/>.
- Presbitero, Alfred. "Culture shock and reverse culture shock: The moderating role of cultural intelligence in international students' adaptation." *International Journal of Intercultural Relations* 54 (2016): 75-85.
- United Nations Educational, Scientific and Cultural Organization Institute of Statistics. "UIS Statistics." Accessed January 25, 2024. <https://uis.unesco.org/>.
- Zhou, Yuefang, Divya Jindal-Snape, Keith Topping, and John Todman. "Theoretical models of culture shock and adaptation in international students in higher education." *Studies in Higher Education* 32, no. 6 (2007): 751-768.  
<https://www.tandfonline.com/doi/full/10.1080/03075070701794833>

The aim of this presentation is to demonstrate that functions encoded in the grammatical system rather than events in the real world determine the forms of utterances. This demonstration is based on an analysis of a function in the grammatical system of Hdi (Central Chadic). This system encodes a category that directs the listener to interpret the proposition as denoting an event that actually happened or that will happen. Moreover, this proposition must be interpreted in connection with (a) some specific time; (b) with another proposition mentioned in discourse; or (c) in connection with a presupposition held by the speaker. This function is coded by an aspectual marker. The encoding of this function in the grammatical system of some languages, but not in others, demonstrates that the semantic structure of the language does not depend on activities in the real world or on the speakers' perception of, or experience with, those activities (Higginbotham et al. (eds.) 2000, Bohnemeyer and Pederson (eds.) 2011). The semantic structure encoded in the grammatical system thus represents an autonomous system.

Higginbotham, James, Fabio Pianesi, and Achille. C. Varzi (eds.). 2000. *Speaking of events*. New York-Oxford: Oxford University Press.

Bohnemeyer, Jürgen, and Eric Pederson (eds.). 2011. *Event representation in language and cognition*. Cambridge: Cambridge University Press.

## Visual and auditory perception in narrative discourse: a contrastive approach

### Abstract

Languages differ in the way they encode sensory perception and in how systematically they do it for particular senses (San Roque et al. 2015; Majid et al. 2018). Some encode perception mode through dedicated grammatical ‘evidentiality’ markers (Aikhenvald 2004), while others adhere to lexical means or employ ‘depictive’ devices such as onomatopoeias and ideophones associated with sensory meaning (Dingemanse 2011; Dingemanse & Majid 2012). One of the issues that remain unexplored is the interaction of the different strategies; for example, no attempt has been made to test whether the availability of grammatical evidentials or sensory ideophones corresponds to higher frequency of reference to the corresponding sense in discourse (consistent with Slobin’s 1987 ‘Thinking for Speaking’ hypothesis).

This study explores this issue based on data from seven understudied languages. First, we focus on four languages with different evidentiality systems: Udihe (Tungusic) with evidential particles marking visually perceived information (Nikolaeva & Tolskaya 2001), Nganasan (Samoyedic) with an auditory evidential (Gusev 2007), Selkup (Samoyedic) where the latensive mood is used for encoding different information sources (Urmančieva 2014), and Ngiyambaa (Pama-Nyungan), one of the few Australian languages with evidential morphology (Donaldson 1980). Second, we compare them to three languages without grammatical encoding of the sensory information source: Chuvash (Turkic, Russia), Wan (Mande, Côte d’Ivoire), and Ngarinyin (Worrorran, Australia).

We compare how two senses, SEEING and HEARING, are represented in traditional narratives across these languages. We look at whether they are encoded explicitly (through lexical or grammatical means, cf. [1]) or merely implied in context (e.g. the ideophone in [2] refers to the object’s visual characteristics).

(1) Nganasan (Brykina et al. 2018)

<i>Maat’aküə?</i>	<i>sojbumunut’ün</i>	<i>muŋka?</i>	<i>n’iməənɪ.</i>
what:EMPH:EMPH:NOM.PL	sound:AUD:OBL.3PL	tree:GEN.PL	ON:PROLADV

‘There’s some noise **heard** in the trees.’

(2) Wan (Nikitina et al. 2022)

<i>tē</i>	<i>bú</i>	<i>è</i>	<i>flà</i>	<i>á</i>	<i>téj-téj-téj</i>
fire	powder	DEF	be.white	ADJ.FOC	IDPH-IDPH-IDPH

‘The cinder is **sparkling** white!’

We find that several factors influence the encoding of information source in discourse. First, we observe overall prominence of visual perception in our data; furthermore, vision is more likely to be encoded lexically as compared to auditory perception (which is often left

uncoded). Second, we observe effects of rhetorical style: in Wan, systematic use of sound-related ideophones is characteristic of skilful narration, and contexts where hearing is implied often refer to the “title song” conventionally associated with the story. Third, environmental factors are reflected in the prominence of indices peculiar to a specific physical environment: tracing by footprints is characteristic of narratives in Siberia, whereas Ngarinyin narratives reflect the changing landscapes of the subtropics and the desert, which predominantly affect different senses throughout the year (e.g., the roaring floods of the rainy season are associated with auditory references, whereas the sweltering heat of the dry season is reflected in a prominence of visual expressions). We conclude that while there is some evidence for the 'Thinking for Speaking' hypothesis in this domain, visual and auditory expressions are equally influenced by non-cognitive and non-cultural factors, such as the physical environment.

## References

- Aikhenvald, A. Y. (2004), *Evidentiality*. Oxford: Oxford University Press.
- Brykina, M., et al. (2018), *Nganasan Spoken Language Corpus (NSLC)*. Version 0.2. <http://hdl.handle.net/11022/0000-0007-C6F2-8>
- Dingemanse, M. (2011), *The meaning and use of ideophones in Siwu*. Nijmegen: Max Planck Institute for Psycholinguistics.
- Dingemanse, M. & A. Majid. (2012), The semantic structure of sensory vocabulary in an African language. In: Miyake N., D. Peebles & R. P. Cooper (eds.), *CogSci 2012: Proceedings of the 34th Annual Conference of the Cognitive Science Society*, 300–305. Austin, TX: Cognitive Science Society.
- Donaldson, T. (1980), *Ngiyambaa: The language of the Wangaaybuwan*, Cambridge: Cambridge University Press.
- Gusev, V. Ju. (2007), Èvidencial'nost' v nganasanskom jazyke [Evidentiality in Nganasan]. In: V. S. Xrakovskij (ed), *Èvidencial'nost' v jazykax Evropy i Azii. Sbornik statej pamjati N.A. Kozincevoj*, Saint-Petersburg: Nauka, 415 – 444
- Majid, A., et al. (2018), Differential coding of perception in the world's languages. *Psychological and Cognitive Sciences* 45(115), 11369–11376.
- Nikitina, T., et al. (eds) (2022), *The SpeechReporting Corpus: Discourse Reporting in Storytelling*. CNRS-LLACAN & LACITO, <http://discoursereporting.huma-num.fr/index.html>
- Nikolaeva, I. & M. Tolskaya (2001), *A Grammar of Udihe*. Berlin: Mouton de Gruyter.
- San Roque L., et al. (2015), Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies. *Cognitive Linguistics* 26, 31–60.
- Slobin, D. I. (1987), Thinking for Speaking, *Annual Meeting of the Berkeley Linguistics Society* 13, 435–445.
- Urmančieva, A. Ju. (2014), Èvidencial'nye pokazateli sel'kupskogo jazyka: sootnošenie semantiki i pragmatiki v opisanii glagol'nyx grammem [Evidential Markers of Selkup: Correlation between Semantics and Pragmatics in the Description of Verbal Grams], *Voprosy jazykoznanija* 4, 66–86.

**Title :** (A)symétrie dans l’expression de la Source et du But dans quelques langues de Birmanie

**Abstract :**

Dans les événements spatiaux, le But et la Source ne sont pas des concepts égaux et les langues manifestent une asymétrie particulière à cet égard, comme l’ont mis en évidence Ikegami (1987), Bourdin (1997), Kopecka et Ishibashi (2011), ainsi que Vuillermet et Kopecka (2021). Tous ces travaux montrent que les langues n’encodent pas de la manière le But et la Source et affirment la primauté du But sur la Source, sur la base de preuves cognitives (Regier et Zheng 2007) et linguistiques.

L’exemple birman (1) illustre l’optionnalité d’un marqueur de But dans les expressions encodant le point final d’un déplacement (1a), à l’inverse des expressions encodant le point initial d’un déplacement (1b). Ainsi, l’absence de forme marque le complément locatif comme But, ce qui implique une asymétrie dans l’expression des deux concepts.

Birman :

- (1) a. သူရန်ကုန်(ကို) လာတယ်။  
 $\theta u^2$      $yaN^2gon^2$     =**Ko**<sup>2</sup>/**Ø**     $la^2$             =**Te**<sup>2</sup>  
 3SG    Yangon        =**ALL**/**Ø**    come            =**REAL**  
 ‘He came/comes (to) Yangon.’
- b. သူရန်ကုန်က လာတယ်။  
 $\theta u^2$      $yaN^2gon^2$     =**Ka**<sup>1</sup>/**\*Ø**     $la^2$             =**Te**<sup>2</sup>  
 3SG    Yangon        =**ABL**/**\*Ø**    come            =**REAL**  
 ‘He came/comes from Yangon.’

Dans cette présentation, nous étudierons l’expression du But et de la Source dans les événements de déplacement. Nous nous focaliserons sur les langues d’Asie-du Sud-Est et plus particulièrement de Birmanie, à la recherche d’une asymétrie dans l’expression des points initiaux et finaux d’une trajectoire. Ce travail s’inscrit dans le cadre du projet SpOTy (*Spatial Ontology and Typology*), qui vise à créer une typologie de l’expression de la Trajectoire et une ontologie (un réseau interconnecté de catégories conceptuelles) à l’aide de données concernant des langues géographiquement, génétiquement et typologiquement diverses.

Une partie des données utilisées dans cette étude a été collectée avec le stimulus vidéo Trajectoire (Ishibashi *et al.* 2006, Vuillermet et Kopecka 2019), d’autres à l’aide du stimuli *Frog Story* (Mayer, 1969).

Après avoir introduit le contexte théorique, notre méthodologie, ainsi que l’outil de requête SpOTy, nous présenterons les dispositifs morphosyntaxiques encodant de la Source et du But dans les descriptions de mouvements spontanés des différentes langues étudiées. Trois types d’outils linguistiques seront distingués : le premier type d’encodage relève du domaine nominal (nom nu, adpositions, noms relateurs), le deuxième du domaine verbal (constructions verbales sérielles). Le troisième type renvoie à une stratégie syntaxique (séquentialisation des événements) : cette construction particulière exprime les différentes parties d’une trajectoire, en divisant la phrase en deux propositions.

Enfin, nous nous focaliserons sur deux des langues présentées, *i.e.* le birman (sino-tibétain), et le shan (taï-kadaï). En utilisant l’outil de requête SpOTy (ainsi que le graphe de connaissance sous-jacent), nous montrerons que (1) l’asymétrie entre la source et le but existe dans ces

langues, bien qu'elle ne corresponde pas entièrement à l'hypothèse universaliste selon laquelle les langues favorisent systématiquement l'expression du But. Nous pointerons également (2) que l'utilisation des noms relateurs INTERIEUR et DESSUS est plus symétrique dans ces langues que dans d'autres langues comme le français, puisque ces noms relateurs peuvent être utilisés dans l'encodage du But et de la Source.

### Références:

- Bourdin, Philippe. 1997. On Goal-bias across languages: modal, configurational and orientational parameters. In Bohumil Palek (ed.), *Proceedings of LP'96*, 185–218. Prague, République tchèque : Charles University Press.
- Grinevald, Colette. 2011. On constructing a working typology of the expression of path. *Faits de Langues - Les cahiers* (3). 71–98.
- Ikegami, Yoshihiko. 1987. “Source” vs. “Goal”: a Case of linguistic dissymmetry. In René Dirven & Günter Radden (eds.), *Concepts of case*, 122–146. Tübingen, Allemagne : Gunter Narr Verlag.
- Ishibashi, Miyuki, Anetta Kopecka & Marine Vuillemermet. 2006. *Trajectoire : matériel visuel pour élicitation des données linguistiques*. DVD. Lyon, France: Fédération de Recherche “Typologie et Universaux Linguistiques” (TUL).
- Kopecka, Anetta & Miyuki Ishibashi. 2011. L'(a)-symétrie dans l'expression de la Source et du But : perspective translinguistique. *Faits de Langues - Les cahiers* (3). 131–150.
- Kopecka Anetta & Marine Vuillemermet (eds). 2021. *Source-Goal (a)symmetries across languages*. Amsterdam, Pays-Bas : John Benjamins.
- Mayer, Mercer. 1969. *Frog, Where Are You? by Mercer Mayer* (Children's Picture Books). Londre, Royaume-Uni : Penguin Random House.
- Regier, Terry & Mingyu Zheng. 2007. Attention to endpoints: A cross-linguistic Constraint on Spatial Meaning. *Cognitive Science* 31. 705–719.
- Vuillemermet Marine & Anetta Kopecka. 2019. Trajectoire: a methodological tool for eliciting Path of motion. *Methodological tools for linguistic description and typology, Special issue of Language, Documentation and Conservation* 16. 97-124.

### **Axial forms exhibit stronger cohesion with the ground nominal than directional markers**

Axial elements, such as ‘in front of,’ ‘above,’ and ‘under’ have been studied in various theoretical approaches (e.g. Levinson & Wilkins 2006; Cinque & Rizzi 2010; Svenonius 2006). Previous studies have noticed that these elements exhibit idiosyncratic morphosyntax across languages. For example, *wi* in (1a) takes a locative flag like nouns, but it does not require its ground nominal to be marked with a genitive flag. Similarly, *jelow* in (1b) can take a direct complement like adpositions, but it can also take an ezafe marker like nouns. However, the typological generalization concerning axial elements has rarely been reported.

- |                                |                                  |
|--------------------------------|----------------------------------|
| (1) a. Korean (Sohn 1994: 259) | b. Persian (Mahootian 1997: 170) |
| <i>san wi-eyse</i>             | <i>jelow(-ye) mæn be-šin</i>     |
| mountain top-LOC               | in front of (-EZ) me IMP-sit     |
| ‘at the top of the mountain’   | ‘Sit in front of me.’            |

In this presentation, based on a 60-language variety sample, I propose a generalization about the relationship to the ground nominal between *axial forms* and directional markers such as ‘from’ and ‘toward’:

- (2) Axial-directional marking generalization  
When languages express both axial region and direction, directional markers do not preferentially occur on ground nominals over axial forms.

(2) can encompass all patterns that simultaneously express the axial region and direction, regardless of the morphosyntactic features of axial elements and directional markers. This is exemplified by (3) and (4). (3a) shows the axial region denoted by an adposition and direction by a case, while (3b) expresses direction with a verb. In (4a), both are expressed by adpositions, whereas (4b) uses cases for both.

- |                         |                                   |
|-------------------------|-----------------------------------|
| (3) a. Finnish          | b. Iatmul (Jendraschek 2012: 271) |
| <i>sien-ten pää-lle</i> | <i>wun-a kadi'-ba a-yi</i>        |
| mushroom-PL.GEN on-ALL  | 1SG-GEN rear-LOC IMP-go           |
| ‘onto the mushrooms’    | ‘go behind me’                    |
- 
- |                        |                                |
|------------------------|--------------------------------|
| (4) a. Russian         | b. Bagvalal (Daniel 2019: 298) |
| <i>iz-pod stol-a</i>   | <i>L'er-L'i-s:</i>             |
| from-under table-GEN   | bridge-under-from              |
| ‘from under the table’ | ‘from under the bridge’        |

Lestrade et al. (2011) report a similar observation, but their observation applies only to languages where axial elements are considered adpositions. In contrast, (2) can be obtained by defining axial elements as *axial forms*:

- (5) Axial form  
An axial form is a linguistic form that denotes an axial region.

Axial elements have been analyzed as adpositions (Cinque 2010), as nouns (Matushansky & Zwarts 2019), and as “AxPart” (Svenonius 2006). However, each analysis falls short from a typological perspective because these word classes are language-specific.

Finally, I would suggest a diachronic explanation for (2). While axial forms are often grammaticalized from head nouns in genitive constructions (Svorou 1994), directional markers are often derived from verbs or adverbs (Dryer 2019; Vincent 1999). When expressing both axial region and direction, axial forms are grammaticalized from nouns in genitive constructions, and the possessed noun becomes a ground nominal, forming a stronger cohesion with the ground nominal. This is evident from the fact that axial forms tend to retain characteristics of nouns, as illustrated in (6).

- |                       |  |
|-----------------------|--|
| (6) a. Japanese       | b. Akebu (Shluinsky 2023: 95)          |
| <i>ki=no shita=ni</i> | <i>fī wì-cíí-kpá` tē lá kà-níh-kā`</i> |
| tree=GEN under=LOC    | LOC KP-tree-KP INDF POSS K-buttocks-K  |
| ‘under the tree’      | ‘under a tree’                         |

## References

- Cinque, Guglielmo. 2010. Mapping Spatial PPs: An Introduction. In Guglielmo Cinque & Luigi Rizzi (eds.), *Mapping Spatial PPs: The Cartography of Syntactic Structures, Volume 6*, 3–25. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195393675.003.0001>.
- Cinque, Guglielmo & Luigi Rizzi (eds.). 2010. *The Cartography of Syntactic Structures, Vol. 6*. Oxford: Oxford University Press.
- Daniel, Michael. 2019. Case study in part-of-speech typology. *STUF - Language Typology and Universals* 72(3). 297–311. <https://doi.org/doi:10.1515/stuf-2019-0012>.
- Dryer, Matthew. 2019. Grammaticalization accounts of word order correlations. In Karsten Schmidtke-Bode, Natalia Levshina, Susanne Maria Michaelis & Ilja Seržant (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*, 63–95. Berlin: Language Science Press.
- Jendraschek, Gerd. 2012. *A grammar of Iatmul*. Regensburg: Universität Regensburg PhD Thesis.
- Lestrade, Sander, Kees de Schepper & Joost Zwarts. 2011. The distribution of labor between adpositions and case within complex spatial PPs. *STUF - Language Typology and Universals* 64(3). 256–274. <https://doi.org/doi:10.1524/stuf.2011.0018>.
- Levinson, Stephen C. & David P. Wilkins (eds.). 2006. *Grammars of Space: Explorations in Cognitive Diversity* (Language Culture and Cognition). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486753>.
- Mahootian, Shahrzad. 1997. *Persian* (Descriptive Grammars). London: Routledge.
- Matushansky, Ora & Joost Zwarts. 2019. Tops and bottoms: axial nominals as weak definites. In Richard Stockwell, Maura O’Leary, Zhongshi Xu & Z.L. Zhou (eds.), *Proceedings of the 36th West Coast Conference on Formal Linguistics*, 270–280. Somerville, MA: Cascadilla Proceedings Project.
- Shluinsky, Andrey. 2023. Adpositions and adpositional relator nouns in Akebu. *Journal of African Languages and Linguistics* 44(1). 77–117.



- <https://doi.org/doi:10.1515/jall-2023-2006>.
- Sohn, Ho-min. 1994. *Korean* (Descriptive Grammars). London / New York: Routledge.
- Svenonius, Peter. 2006. The Emergence of Axial Parts. *Nordlyd* 33(1). 49–77.  
<https://doi.org/10.7557/12.85>.
- Svorou, Soteria. 1994. *The Grammar of Space*. Amsterdam: John Benjamins.  
<https://www.jbe-platform.com/content/books/9789027276575>.
- Vincent, Nigel. 1999. The evolution of c-structure: prepositions and PPs from Indo-European to Romance. *Linguistics* 37(6). 1111–1153.  
<https://doi.org/doi:10.1515/ling.37.6.1111>.

## FAIR database publication in typology: Language parameters and construction parameters in the CrossGram data registry

Replicability of research results minimally relies on data accessibility, but the data should ideally be FAIR: Findable, Accessible, Interoperable, and Reusable. For technical interoperability, the CLDF standard (Forkel et al. 2018) can be used, and here I discuss the design of the CLDF-compliant CrossGram data registry, which is specifically designed for typological datasets (it is currently in its beta version, available at <https://crossgram.clld.org/>, to be launched later in 2024). I describe how CrossGram enhances findability and reusability, and I highlight the two different data types that it supports (language parameters and construction parameters).

CrossGram makes typological data more findable in that it “brings to light” what is often “hidden away” in supplementary spreadsheet files (or even tables in PDF files, though this is becoming rare). Research papers typically limit themselves to summary tables or graphs and a few small maps, but ideally we want to access all typological datasets with the amenities known from CLLD applications such as *WALS Online* (Dryer & Haspelmath 2013, [wals.info](http://wals.info)) or Grambank (Skirgård et al. 2023, [grambank.clld.org](http://grambank.clld.org)). These provide not only easy exporting in CLDF format, but also easy searching and sorting in data tables, as well as map visualization, and links to references and Glottolog language information.

In addition, CrossGram provides glossed example sentences in tabular form, similar to the thousands of examples in the *APiCS* database (Michaelis et al. 2013). These are a particularly striking case of increased transparency, because it is not uncommon for example sentences to be hidden in PDF supplements (for example, Bugaeva 2022 has a supplement of 80 pages of annotated examples). Interlinear glossed text has a range of applications even independently of the typological claims that the examples illustrate, so this is another obvious improvement in reusability.

CrossGram supports two types of typological data: Language parameters that classify entire languages (i.e. parameters as in *WALS* and Grambank), and construction parameters that classify constructions. There are many grammatical meanings or functions that can be rendered by multiple constructions in a given language, and if we only consider language parameters, the language must be classified as “mixed” (or a minor construction must be ignored). For example, Kashmiri has both correlative relative clauses and postnominal relative clauses, so both of these strategies could be included in the database and their properties recorded. Stereotypically, typology consists in classifying languages into types, but in reality, languages often have multiple types coexisting with each other, so the addition of constructions and construction types as a data type makes typological databases more fine-grained.

Finally, CrossGram parameters (both language parameters and construction parameters) are not only explained succinctly and clearly, but there is also a sophisticated keyword annotation that allows users to easily find grammatical information on a wide range of topics, and integration with the envisaged “Grammaticon” reference catalogue is planned (see Haspelmath 2022). This will enhance findability and accessibility even further, and it will facilitate replication and (more generally) cumulative science.

## References

- Bugaeva, Anna & Nichols, Johanna & Bickel, Balthasar. 2022. Appositive possession in Ainu and around the Pacific. *Linguistic Typology* 26(1). 43–88. (doi:10.1515/lingty-2021-2079)
- Dryer, Matthew S. & Haspelmath, Martin (eds.). 2013. *WALS Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (<https://wals.info/>)
- Forkel, Robert & List, Johann-Mattis & Greenhill, Simon J & Rzymiski, Christoph & Bank, Sebastian & Cysouw, Michael & Hammarström, Harald & Haspelmath, Martin & Kaiping, Gereon A & Gray, Russell D. 2018. Cross-Linguistic Data Formats: Advancing data sharing and re-use in comparative linguistics. *Scientific Data* 5(1), 1–10.
- Haspelmath, Martin. 2022. Classificatory concepts for typology and the Grammaticon vision. Prague: Charles University. (Presented at the Workshop “The Central European languages in the World Atlas of Language Structures: Supplementing WALS data,” Prague, January 2022.) (<https://osf.io/skzut>)
- Michaelis, Susanne Maria & Maurer, Philippe & Haspelmath, Martin & Huber, Magnus (eds.) 2013. *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://apics-online.info>.)
- Skirgård, Hedvig & Haynie, Hannah J. & Blasi, Damián E. & Hammarström, Harald & Collins, Jeremy & Latarche, Jay J. & Lesage, Jakob et al. 2023. Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Science Advances* 9(16). eadg6175. (doi:10.1126/sciadv.adg6175)

# Estimer la prédictibilité dans les paradigmes dérivationnels à partir de scénarios générés par des LLM

Daniele Sanacore  
CLLE - Université de Toulouse

Nabil Hathout  
CLLE - Université de Toulouse

Fiammetta Namer  
ATILF - Université de Lorraine

La nature paradigmatique de la morphologie flexionnelle est aujourd’hui relativement consensuelle (Carstairs-McCarthy, 1994 ; Ackerman *et al.*, 2009), mais il n’en va pas de même pour la morphologie dérivationnelle, même si de nombreux auteurs ont proposé d’étendre les paradigmes à la dérivation (Bauer, 1997 ; Štekauer, 2014 ; Bonami et Strnadová, 2019 ; Hathout et Namer, 2022). Pour l’heure, nous ne savons pas encore à quoi ressemblent les paradigmes dérivationnels en grande partie parce que nous ne savons pas comment la prédictibilité se définit dans le lexique construit (Hathout et Namer, 2019). Paradigme et prédictibilité sont des notions étroitement liées en morphologie. En flexion, la présence d’une forme exprimant M.SG dans le paradigme d’un adjectif du français et celle d’une forme exprimant F.PL se prédisent mutuellement. Cette interprédictibilité définit une structure implicative dans les paradigmes flexionnels (Bonami, 2014).

Nous proposons dans cette étude de définir la prédictibilité en dérivation en utilisant des scénarios, i.e., des représentations d’une situation prototypique (Sanford et Garrod, 1998 ; Erk et Herbelot, 2024). Par exemple, un scénario de vente implique typiquement un vendeur, un acheteur, un bien vendu et de l’argent échangé. Plus précisément, il s’agit de produire des scénarios qui contiennent un couple de mots morphologiquement apparentés pour identifier les autres participants qu’ils contiennent. Si les deux mots appartiennent au même paradigme dérivationnel (ex. HERBE, DÉSHÉRBAGE), les scénarios qu’ils engendrent présentent une régularité importante qui se manifeste par le fait qu’ils parlent essentiellement des mêmes événements et qu’ils contiennent en grande partie les mêmes participants. À l’inverse, deux mots qui appartiennent à des paradigmes différents (ex. DÉSHÉRBAGE, HERBIER) n’imposent pas de réelles contraintes sur les scénarios qu’ils engendrent : ces derniers parlent de situations variées et impliquent des participants disparates.

Nous proposons de produire ces histoires à l’aide de *Large Language Models* (LLM) afin d’identifier les participants récurrents qu’elles contiennent et qui ne sont mentionnés dans le prompt. Par exemple, l’histoire en (1a) a été produite par *Mistral* en utilisant *Ollama* à partir du prompt : *Écris une histoire réaliste de 50 à 100 mots qui contient les mots “désherber” et “désherbant”. Donne-moi en plus la liste des participants.* (1b).

- (1) a. Après avoir testé plusieurs méthodes de **désherbage** dans son **jardin**, **Pierre** décida d’utiliser un **désherbant écologique** pour éliminer les **mauvaises herbes** qui envahissaient ses parterres de fleurs. Avec son **pulvérisateur**, il appliqua le produit pour **désherber** le jardin [...]
- b. **Pierre** : Jardinier amateur ;  
**Jardin** : Lieu de l’action ;  
**Mauvaises herbes** : Plantes indésirables ;

**Dés herbant écologique** : Produit anti-herbes ;  
**Pulvérisateur** : Outil qui applique le dés herbant.

Notre hypothèse est que les concepts fortement reliés sémantiquement vont coexister régulièrement dans les histoires. L'histoire (1a) parle d'herbes, du dés herbage, d'un jardin et d'un pulvérisateur. Ces participants apparaissent régulièrement dans les histoires construites pour le couple DÉSHÉRBAGE-DÉSHÉRBANT. Par ailleurs, ces participants prédisent ces deux lexèmes. Par exemple, dans les histoires produites pour HERBE et DÉSHÉRBAGE, on retrouve DÉSHÉRBER et DÉSHÉRBANT. Nous pouvons en déduire que HERBE, DÉSHÉRBER, DÉSHÉRBAGE et DÉSHÉRBANT font partie du même paradigme dérivationnel.

La méthode que nous proposons permet ainsi de découvrir les relations de prédictibilité sémantique pertinentes pour la morphologie dérivationnelle. Il devient ainsi possible de constituer de véritables paradigmes dérivationnels et de vérifier qu'ils ne correspondent pas à des familles de mots. Par exemple, les histoires comme celles en (1a) ne parlent pas généralement d'HERBORISTERIE, d'HERBORISTE ou d'HERBIER.

## Références

- ACKERMAN, F., BLEVINS, J. P. et MALOUF, R. (2009). Parts and wholes : Implicative patterns in inflectional paradigms. *Analogy in grammar : Form and acquisition*, 54 :81.
- BAUER, L. (1997). Derivational paradigms. *In Yearbook of morphology 1996*, pages 243–256. Springer.
- BONAMI, O. (2014). *La structure fine des paradigmes de flexion*. Thèse de doctorat, Université Paris 7-Denis Diderot.
- BONAMI, O. et STRNADOVÁ, J. (2019). Paradigm structure and predictability in derivational morphology. *Morphology*, 29(2) :167–197.
- CARSTAIRS-McCARTHY, A. (1994). Inflection classes, gender, and the principle of contrast. *Language*, pages 737–788.
- ERK, K. et HERBELOT, A. (2024). How to marry a star : Probabilistic constraints for meaning in context. *Journal of Semantics*, page fad016.
- HATHOUT, N. et NAMER, F. (2019). Paradigms in word formation : what are we up to? *Morphology*, 29(2) :153–165.
- HATHOUT, N. et NAMER, F. (2022). Paradis : a family and paradigm model. *Morphology*, pages 1–43.
- SANFORD, A. J. et GARROD, S. C. (1998). The role of scenario mapping in text comprehension. *Discourse processes*, 26(2-3) :159–190.
- ŠTEKAUER, P. (2014). Derivational paradigms. *The Oxford handbook of derivational morphology*, pages 354–369.

## An oblique correspondence between dominant word orders and corpus variations

Hing-Yuet Fung

*University of Hong Kong*

Type and token frequencies are important concepts in usage based theories. However, a mismatch between a language's dominant orders and the orders' corpus frequencies is often taken for granted, and it obscures our observation of potential shifts in usage. In particular, an asymmetry regarding the order of object (O), oblique (X), and verb (V), has been noted on WALS (Dryer & Gensler, 2013)—when O and X appear on the same side relative to the verb, OV languages are nearly as abundant in XOV and OXV orders, and for VO languages, VOX (1a) far outnumbers VXO; while variation data tend to show similar flexibility in both preverbal and postverbal domains (Liu, 2021).

(1) ENGLISH; dominant in VO, VOX (adopted from Zeman et al. 2023)

- a. I have been eating [ Cuban ]<sub>OBJ</sub> [ for a long time ]<sub>OBL</sub> .
- b. Can you please reconfirm [ with Mark ]<sub>OBL</sub> [ the price for the gas at Bondad ]<sub>OBJ</sub> ?

In light that the VXO order in English is motivated by a long O (1b), this study approaches the problem by building on the efficiency principles and methodologies in Hawkins (2014). Corpus variation is predicted to align with the typological asymmetry, by the interaction of two preferences: (A) preference for object and verb adjacency; (B) preference for the oblique to follow constituents which are necessary for event construal, notably the direct object. The productivity of the orders is predicted thus: the VOX order fulfills both preferences, hence the most productive, and VXO none of them; the XOV and OXV orders each fulfil one preference and are equally productive.

Instances are extracted from 111 eligible languages in UD, v2.12 (Zeman et al., 2023), which include languages with different word order types primarily categorised by WALS (Dryer & Haspelmath, 2013) and supplemented by Kalyan and Donohue (2023) and Skirgård et al. (2023). Lengths of co-occurring direct objects (O) and oblique nominals (X; non-core NP oblique arguments or adjuncts, following UD's coding) are calculated by tokens. The distributions of four ordering combinations are shown in Figure 1 by probability contours of the length pairs, with outliers removed. Each possible pair as found in UD contributes to a type count, and the contour lines enclose different percentages of data points by different transparencies. A reference line is drawn where both constituents are of equal length. In the preverbal domain of OV (Figure 1a), the two orders outline a similar space under the competition between the two word order preferences. For VO languages (Figure 1b), the VOX order yields to VXO when O grows to extreme lengths. Interestingly, this strong preference is not found in the postverbal domain for languages with no dominant order (Figure 1d), reflecting weak consistency in these languages. In conclusion, our observation confirms the asymmetry between OV and VO languages in corpus data as regards the position of an oblique phrase, which can be attributed to the interaction of functional principles.

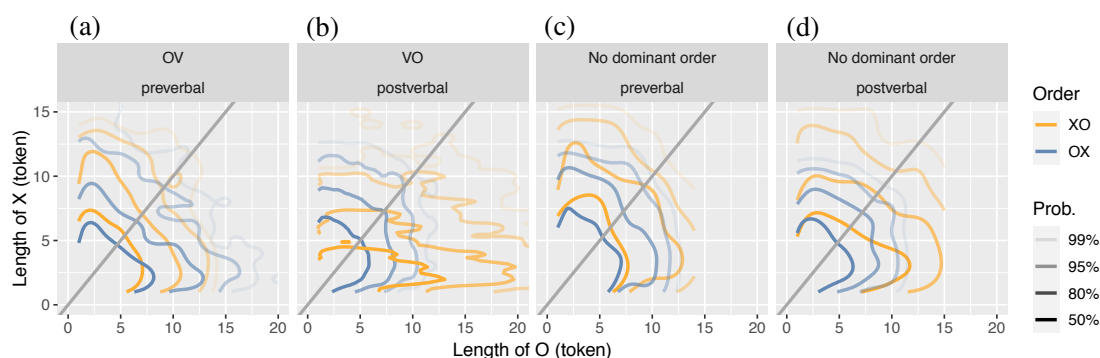


Figure 1: Distributions of length-pairs consisting of a direct object (O) and an oblique nominal (X)

**Keywords:** word order alternations, oblique phrase, asymmetry, efficiency principles

### References

- Dryer, M. S., & Gensler, O. D. (2013). Order of object, oblique, and verb (v2020.3) [Data set]. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Zenodo.
- Dryer, M. S., & Haspelmath, M. (Eds.). (2013). *WALS Online (v2020.3)*. Zenodo.
- Hawkins, J. A. (2014). *Cross-linguistic variation and efficiency*. Oxford University Press.
- Kalyan, S., & Donohue, M. (2023). The dimensions of morphosyntactic variation: Whorf, Greenberg and Nichols were right. *Linguistic Typology at the Crossroads*, 3(2), 132–190.
- Liu, Z. (2021). The crosslinguistic relationship between ordering flexibility and dependency length minimization: A data-driven approach. *Proceedings of the Society for Computation in Linguistics*, 4(1), 264–274.
- Skirgård, H., Haynie, H. J., Hammarström, H., Blasi, D. E., Collins, J., Latarche, J., . . . Gray, R. D. (2023, Mar). *Grambank v1.0*. Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.7740140> (Dataset)
- Zeman, D., Nivre, J., Abrams, M., Ackermann, E., Aepli, N., Aghaei, H., . . . Ziane, R. (2023). *Universal Dependencies 2.12*. Retrieved from <http://hdl.handle.net/11234/1-5150> (LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University)

## Classification phylogénétique des dialectes romans : à la croisée entre linguistique historique et apprentissage automatique

Marc Evrard<sup>1</sup>, Philippe Boula de Mareuil<sup>1</sup>, Alexandre François<sup>2</sup>, Antonio Romano<sup>3</sup>

<sup>1</sup> Université Paris-Saclay, CNRS, LISN, Orsay, France

<sup>2</sup> Lattice, CNRS, ENS-PSL, Université Sorbonne Nouvelle, Paris, France

<sup>3</sup> Università degli Studi di Torino, LFSAG, Turin, Italy

{philippe.boula.de.mareuil;marc.evrard}@lisn.fr,  
alexandre.francois@ens.fr; antonio.romano@unito.it

L'Europe est caractérisée par une grande diversité dialectale, qui reste encore à documenter. Pour l'Europe romane, différentes approches ont été proposées afin de rendre compte de la diversification du latin : les arbres phylogénétiques [1] et le modèle des vagues [2, 3]. Des méthodes computationnelles s'en inspirent, pour mettre en évidence divers phénomènes dialectaux [4, 5, 6, 7] ; cependant, les résultats, qui reposent sur des atlas linguistiques traditionnels, sont principalement fondés sur des mots isolés. Des informations plus riches, notamment grammaticales, peuvent être recueillies à partir de textes parallèles [8, 9, 10].

La présente étude s'appuie sur un corpus illustrant plusieurs dizaines de dialectes romans de France, Belgique, Suisse, Italie, Espagne et Portugal [11, 12], pour lesquels 145 innovations par rapport au latin ont été codées sous la forme de 1 (présence) ou 0 (absence). Sur la base d'enregistrements réalisés sur le terrain — des traductions de la fable d'Ésope « La bise et le soleil » (environ une minute de parole) et une liste d'une centaine de mots isolés, nous proposons de classer ces variétés romanes en suivant les principes de la dialectométrie, de la Méthode comparative [1] et en particulier de la glottométrie historique [13, 14], appliqués à l'aide d'outils informatiques. Les résultats de techniques d'apprentissage automatique (clustering hiérarchique, forêts d'arbres de décision...) confirment la robustesse d'une division Nord/Sud — la zone d'oïl étant, de loin, la plus innovante — et, secondairement, une opposition entre le Sud-Ouest (principalement ibéro-roman) et le Sud-Est (principalement italo-roman, plus conservateur). Parmi les caractéristiques les plus discriminantes figurent la chute des voyelles finales autres que *a*, qui caractérise les dialectes gallo-romans, et la simplification des géminées au nord de la ligne La Spezia–Rimini, en Italie. La plupart des innovations concernent des traits phonétiques ou phonologiques. Cependant, on trouve également des traits morphosyntaxiques et lexicaux, comme le sujet non-nul, dans les variétés gallo-romanes septentrionales, et la substitution de *CUM* par *APUD* 'près de' > *amb* « avec » dans les variétés occitano-romanes. En ne retenant que les innovations morphosyntaxiques, on retrouve une tripartition Nord vs Sud-Est vs Sud-Ouest.

En basant nos calculs sur l'ensemble des 145 traits (essentiellement phonétiques), la principale ligne de partage dans l'ensemble roman passe au milieu du domaine occitan, avec l'oïl et les dialectes intermédiaires regroupés au nord, le reste au sud. De fait, le nord-occitan partage nombre d'innovations avec le francoprovençal, le Croissant et la plupart des variétés d'oïl : par exemple, la palatalisation du latin *ca* et la chute des consonnes devenues finales en gallo-roman, notamment au participe passé (ex. auvergnat *eschaufat* [ejtsofã] < EXCALFATU(M) « réchauffé »). En ne retenant que les innovations morphosyntaxiques et lexicales, les variétés occitanes méridionales et septentrionales — dont celles du Croissant — ainsi que les variétés francoprovençales et catalanes se retrouvent regroupées avec les variétés ibéro-romanes. Des expériences avec 50 ou 60 innovations sélectionnées aléatoirement donnent également des



résultats cohérents. Nous comparerons ces résultats avec ceux que donne une sélection des innovations jugées les plus pertinentes — indépendamment — par trois linguistes.

## Références

- [1] Schleicher, August (1861[2010]), *Compendium der vergleichenden Grammatik der indogermanischen Sprachen*, Cambridge University Press, Cambridge.
- [2] Schmidt, Johannes (1872), *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*, Hermann Böhlau, Weimar.
- [3] Saussure, Ferdinand de (1916), *Cours de linguistique générale*, Payot, Paris.
- [4] Goebel, Hans (2003), « Regards dialectométriques sur les données de l'Atlas linguistique de la France (ALF) : relations quantitatives et structures de profondeur », *Estudis Romànics*, 25 : 59–121.
- [5] Nerbonne, John, Kleiweg, Peter, Heeringa, Wilbert & Manni, Franz (2007), « Projecting dialect differences to geography: bootstrap clustering vs. noisy clustering », in Preisach, Christine, Schmidt-Thieme, Lars, Burkhardt, Hans & Decker, Reinhold (Eds.), *Data analysis, machine learning, and applications*, Springer, Berlin (pp. 647–654).
- [6] Patriarca, Marco, Heinsalu, Els, Léonard, Jean-Léo (2020), *Language in Space and Time*, Cambridge University Press, Cambridge.
- [7] Scherrer, Yves (2021), « Les cartes dialectométriques interactives de dialektkarten.ch », in Thibault, André, Avanzi, Mathieu, Lo Vecchio, Nicholas & Millour, Alice (Eds.), *New Ways of Analyzing Dialectal Variation*, Éditions de linguistique et de philologie, Strasbourg (pp. 137–152).
- [8] Rohlfs, Gerhard (1937), *La struttura linguistica dell'Italia*, H. Keller, Leipzig.
- [9] Pellegrini, Giovan Battista (1973), « I cinque sistemi dell'italo-romanzo », *Revue roumaine de linguistique*, 18 : 105–129.
- [10] Cugno, Federica (2023), « Italian dialect classifications », *Dialectologia*, 10 : 197–230.
- [11] Boula de Mareüil, Philippe, Rilliard, Albert, Vernier, Frédéric (2018), « A Speaking Atlas of the Regional Languages of France », *11<sup>th</sup> International Conference on Language Resources and Evaluation*, Miyazaki (pp. 4133–4138).
- [12] Boula de Mareüil, Philippe, Bilinski, Éric, De Iacovo, Valentina, Romano, Antonio & Vernier, Frédéric (2021), « For a mapping of the languages/dialects of Italy and regional varieties of Italian », in Thibault, André, Avanzi, Mathieu, Lo Vecchio, Nicholas & Millour, Alice (Eds.), *New Ways of Analyzing Dialectal Variation*, Éditions de linguistique et de philologie, Strasbourg (pp. 267–288).
- [13] François, Alexandre (2014), « Trees, Waves and Linkages: Models of language diversification », in Bowerman, Claire & Evans, Bethwyn (Eds.), *The Routledge Handbook of Historical Linguistics*. Routledge, London (pp. 161–189).
- [14] Kalyan, Siva & François, Alexandre (2018), « Freeing the Comparative Method from the tree model: A framework for Historical Glottometry », in Kikusawa, Ritsuko & Reid, Lawrence (Eds.), *Let's talk about trees: Tackling problems in representing phylogenetic relationships among languages* (Senri Ethnological Studies, 98), National Museum of Ethnology, Osaka (pp. 59–89).

## Entre normes et variations : réflexions sur les usages contemporains des corpus en linguistique française

Christophe Benzitoun  
Université de Lorraine & CNRS ATILF  
[Christophe.Benzitoun@univ-lorraine.fr](mailto:Christophe.Benzitoun@univ-lorraine.fr)

Depuis un quart de siècle, les corpus se sont développés de manière très importante en linguistique française et ont permis d'approfondir nos connaissances notamment de phénomènes phonétiques, phonologiques, morphologiques, syntaxiques et sémantiques. Les corpus ont également rendu possible le renouvellement des recherches portant sur la variation et l'évolution en élargissant le champ d'investigation à des données comme le français parlé de conversation ou des écrits de « peu lettrés ». Cet élargissement a pour conséquence la nécessité d'apporter une attention particulière au statut des données et des exemples. Confrontés à des données de plus en plus hétérogènes, permettant de contraster des usages parfois fort éloignés, il est difficile d'interpréter de manière univoque les résultats.

Dans notre intervention, nous nous proposons de faire le point sur les pratiques contemporaines en linguistique de corpus dans le champ grammatical en mettant l'accent sur l'articulation entre normes et variations. Nous aborderons principalement deux phénomènes grammaticaux : les interrogatives partielles directes et l'accord du participe passé. Entre évolution de la langue et injonctions normatives, est-il possible de mesurer la part respective de ces deux dynamiques, l'une plutôt conservatrice et l'autre plus innovante ?

### Références bibliographiques

- CAPPEAU, P. (2005), Quand on ne trouve pas ce que l'on cherche..., *Faits de langue* 25, pp. 157-160.
- FARMER, K. L. (2015), *Sociopragmatic variation in yes/no and wh-interrogatives in hexagonal French: A real time study of French films from 1930 to 2009*, Phd Thesis, Indiana University.
- FOULET, L. (1921), Comment ont évolué les formes de l'interrogation, *Romania* tome 47 n°186-187, pp. 243-348.
- GAUCHER, D. (2015), Sémantique temporelle et accord du participe passé en français parlé : une analyse variationniste, *Journal of French Language Studies* 25, pp. 65-83.
- GURYEV, A. & LARRIVÉE, P. (2021), Routines discursives comme contextes d'émergence de l'interrogative partielle in situ en diachronie, *Langue française* 212, pp. 75-90.
- MARSAC, F. & PELLAT, J.-Ch. (dir.) (2018), *Le participe retrouvé : Grand corpus étiqueté sur le participe passé*, Paris : L'Harmattan.
- REINHARDT, J. (2019), *Regularity and variation in french direct interrogatives: The morphosyntax and intonation of question forms in reality TV shows, audio, books and teaching materials*, Phd thesis, Université de Bielefeld.
- STARK, E. & BINDER, L. (2021), L'inversion du sujet clitique en français oral : ultime apanage des interrogatives ?, *Langue française* 212, pp. 25-40.
- STARK, E. & RIEDEL, I. (2012), L'accord du participe passé dans les SMS francophones du corpus SMS suisse, *Romanistisches Jahrbuch*, 63(1), pp. 116-138.