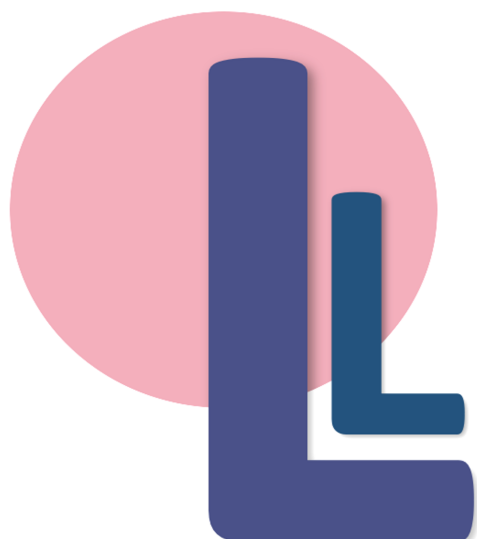




LLcD 2025

2ème rencontre du réseau

LANGUES & LANGAGE
À LA CROISÉE DES DISCIPLINES

BOOK OF ABSTRACTS

LIVRET DES RÉSUMÉS

Université de Lille • 1-5 septembre
ESJ Lille, 50, rue Gauthier-de-Châtillon
Sciences Po Lille, 9 Rue Auguste Angellier
Colloque, Table ronde et Ecole d'été

Bienvenue !

2^e Rencontre internationale LLcD 2025

Colloque : 1-3 septembre 2025

Ecole d'été : 3-5 septembre 2025

L'Université de Lille accueille la 2^e rencontre annuelle du réseau thématique *Langues & Langage à la croisée des Disciplines*.

Langues et Langage à la croisée des Disciplines (LLcD) est un réseau de recherche soutenu par le CNRS. Il organise une conférence internationale en France chaque année dans le but de faire progresser la compréhension scientifique du langage humain et des systèmes linguistiques, et de servir de forum favorisant la création de nouvelles collaborations et la fertilisation croisée entre différentes disciplines de recherche et approches.

Les Rencontres du Réseau LLcD incluent un colloque, une table ronde et une école d'été. Une thématique est mise en avant chaque année, notamment à travers une table ronde. Cette année, la Table Ronde, intitulée *Appropriation et réappropriation des langues : approches théoriques, expérimentales et de corpus en acquisition et en réhabilitation du langage*, aura lieu le 3 septembre.

L'école d'été (du 3 au 5 septembre) est consacrée au thème suivant :

Théorie et données en Linguistique

LLcD 2025 se tiendra dans les locaux de l'Ecole Supérieure de Journalisme (ESJ) et de Sciences Po de Lille. Le cocktail d'initiation du 1^{er} septembre au soir aura lieu à l'hôtel Carlton. La station de Métro la plus proche aux lieux de la conférence est : "République Beaux-Arts".

Adresses :

- ESJ Lille : 50, rue Gauthier-de-Châtillon, Lille
- Sciences Po Lille, 9 Rue Auguste Angellier, Lille

Soyez les bienvenus.e.s !

Contact : llcd2025@sciencesconf.org

Comités d'organisation

Comité porteur



Anne Carlier
Sorbonne Université, STIH



Sonia Cristofaro
Sorbonne Université, STIH



Pascal Denis
Inria, CRISAL



Maia Duguine
CNRS, IKER



Eva Soroli
Université de Lille, STL

Comité Local



Eva Soroli
Université de Lille, STL



Stéphanie Caët
Université de Lille, STL



Bert Cappelle
Université de Lille, STL



Kate Brantley
Université de Lille, STL



Adélie Buclier
Université de Lille, STL



Emmanuelle Canut
Université de Lille, STL



Lieven Danckaert
Université de Lille, STL



Pascal Denis
Inria, CRISTAL



Axelle Domingues
Sorbonne Université,STIH



Mahamadou Doumbia
Université de Lille, STL



Mikaela Keller
Université de Lille, CRISTAL



Benoît Leclercq
Université de Lille, STL



Bastien Liétard
Université de Lille, INRIA



Montserrat Rangel Vicente
Université de Lille, STL



Nicolas Ruytenbeek
Université de Leuven;
Université de Lille, STL



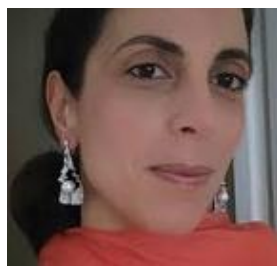
Damien Sileo
Université de Lille, INRIA



Alina Tsikulina
Université de Lille, STL



Ana Laura Vega Umaña
Université de Lille, STL



Eleni Valma
Université de Lille, STL

***Merci à Adélie Buclier, Montserrat Rangel Vicente, Nicolas Ruytenbeek et Eleni Valma
pour la traduction des abstracts et la réalisation de ce livret !***

Comité scientifique

Anne Abeillé
Université Paris Cité, LLF

Marta Abrusan
Institut Jean Nicod, ENS, Paris &
Université de Genève

Giovanni Agresti
Université Bordeaux Montaigne,
IKER - CNRS

Maxime Amblard
Université de Lorraine, LORIA -
CNRS

Peter Arkadiev
Université de Fribourg

Varun DC Arrazola
Université de Leyde

Michel Aurnague
Sorbonne Nouvelle, LATTICE

Mathieu Avanzi
Université de Neuchâtel

Lena Baunaz
Université Côte d'Azur, Bases,
Corpus, Langage

Sabrina Bendjaballah
Nantes Université, LLING

Claire Beyssade
Université Paris 8, SLF

Caroline Boglioti
Université Paris Nanterre, MoDyCo

Gemma Boleda
Université Pompeu Fabra

Olivier Bonami
Université Paris Cité, LLF

Chloé Braud
IRIT Toulouse

Heather Burnett
Université Paris Cité, LLF

Jenny Doetjes
Université de Leyde

Andreas Dufter
Université Munich

Margherita Farina
UMR Histoire des théories
linguistiques - CNRS

Sebastian Fedden
Université Sorbonne Nouvelle,
LACITO - CNRS

Abdellah Fourtassi
Aix-Marseille Université, TALEP

Valantis Fyndanis
Cyprus University of Technology

Luca Gasparri
Université de Lille, STL

Antoine Gautier
Université Paris-Sorbonne, STIH

Cyrille Granget
Université Toulouse-Jean Jaurès,
LNPL

Maximilien Guérin
Université de Lille, STL

Céline Guillot
ENS Lyon, IHRIM

Fatima Hamlaoui
University of Toronto

Nabil Hathout
Université Toulouse-Jean Jaurès,
CLLE

Martin Hilpert
Université de Neuchâtel

Nikolaus Himmelmann
Université de Cologne

Jill Hohenstein
King's College London

Pernelle Lorette
Université de Mannheim

Paolo Mairano
Université de Lille, STL

Fabienne Martin
Université Toulouse 2, LISST

Philippe Monneret
Sorbonne Université, STIH

Franck Neveu
Sorbonne Université, STIH

Christophe Parisse
Université Paris Nanterre, MODYCO

Cédric Patin
Université de Lille, STL

Alain Peyraube
EHESS, CRLAO

Gilles Philippe
Université de Lausanne

Paola Pietrandrea
Université de Lille, STL

Serge Pinto
Aix-Marseille Université, LPL

Sophie Prévost
Sorbonne Nouvelle, LATTICE

Laurent Prévot
Aix-Marseille Université, LPL

Thomas Rainsford
Université de Stuttgart

Christophe Rey
CY Cergy Paris Université, LT2D

Marie-Anne Sallandre
Université Paris 8 - Vincennes -
Saint-Denis, SFL

Patricia Cabredo Hofherr Université
Paris 8 - Vincennes - Saint-Denis,
SFL

Bert Cappelle
Université de Lille, STL

Elisabetta Carpitelli
Université Grenoble Alpes, GIPSA

Soonja Choi
San Diego State University

Benoît Crabbé
Université Paris Cité, LLF CNRS

Denis Creissels
DDL CNRS

Lieven Danckaert
Université de Lille, STL

Marie-Catherine De Marneffe
Université catholique de Louvain,
CENTAL

Hamida Demirdache
Nantes Université, LLING CNRS

Ilse Depraetere
Université de Lille, STL

Guillaume Desagulier
Université Bordeaux Montaigne,
CLIMAS

Holger Diessel
University of Jena

Sascha Diwersy
Université de Paul Valéry, Praxiling
CNRS

Huiyun Hu
Université de Lille, STL

Richard Huyghe
Université Paris Cité, LLF

Adèle Jatteau
Université de Lille, STL

Sophie Kern
Université Lyon 2, DDL

Anetta Kopecka
Université Lyon 2, DDL

Laurence Labrune
Université Toulouse-Jean Jaurès,
CLLE

Aimée Lahaussais
HTL - CNRS

Nicola Lampitelli
Université Paris Nanterre, LLL

Peter Lauwers
Université de Gand

Pascale Leclercq
Université Paul Valéry, EMMA

Gudrun Ledegen
Université Rennes 2, PREFICS

Gaël Lejeune
Sorbonne Université, STIH

Maarten Lemmens
Université de Lille, STL

Lene Schoesler
Université de Copenhagen

Michelle Sheehan
Newcastle University

Alexandra Simonenko
Université de Gand

Achim Stein
Université de Stuttgart

Dejan Stosic
Université Toulouse-Jean Jaurès,
CLLE

Fayssal Tayalati
Université de Lille, STL

Tim Van de Cruys
Université de Leuven

Freek Van de Velde
Université de Leuven

Lieven Vandelanotte
Université de Namur

Jean-Christophe Verstraete
Université de Leuven

Nigel Vincent
University of Manchester

Jacqueline Visconti
University of Birmingham

Marzena Watorek
Université Paris 8 - Vincennes -
Saint-Denis, SLF

Nos partenaires



CNRS

Centre national de la recherche scientifique - Sciences humaines & sociales



Université de Lille

Humanités



ESJ Lille

Ecole supérieure de journalisme de Lille



Sciences Po Lille

Sciences Po - Grande école de Sciences humaines & sociales

**INRIA**

Institut national de recherche en sciences et technologies du numérique

**Sorbonne Université**

Faculté de Lettres

**CRISTAL-UMR 9189**

Centre de Recherche en Informatique, Signal et Automatique de Lille

**STL-UMR 8163**

Laboratoire Savoirs, Textes, Langage

**MESHS Nord**

Maison européenne des sciences de l'Homme et de la Société

**STIH**

Laboratoire Sens, Texte, Informatique, Histoire

**IKER-UMR 5478**

Centre de recherche sur la langue et les textes basques

**Ecole Doctorale**

Université de Lille - SHS

Nos invitées à la Table Ronde / Round Table guests



Anahita Basirat
ScaLab & Univ. de Lille



Jill Hohenstein
King's College London, UK

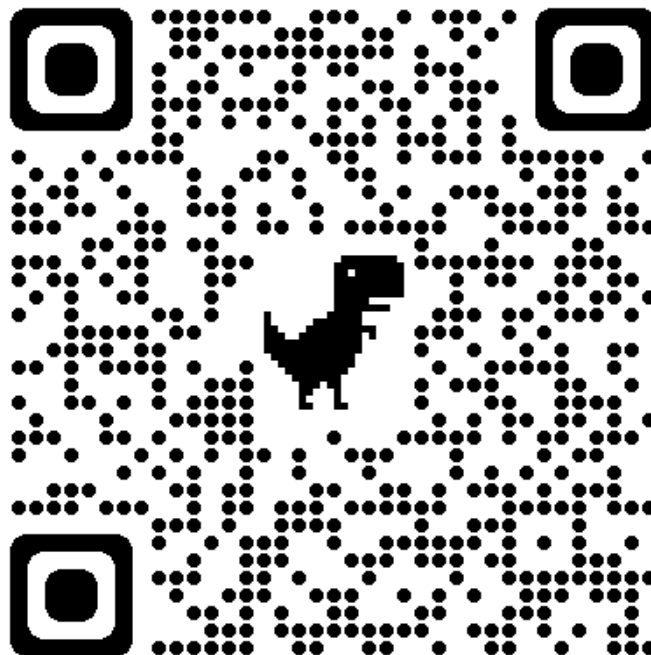


Sophie Kern
CNRS DDL & Univ. de Lyon

Plénière / Plenary speaker

Jill Hohenstein

Lien vers le programme / Program



Liste des résumés / List of abstracts

Marco Angster, Mia Batinic Angster, Jakov Proroković (University of Zadar), Gordana Hrzica (University of Zagreb) & Marijana Kresić (University of Zadar)	14
Silvia Ballarè (University of Bologna) & Paolo Della Putta (University of Turin).....	17
Silvia Ballarè, Caterina Mauri, Ludovica Pannitto & Eleonora Zucchini (University of Bologna)	21
Katia Benaiges-Saura, María-Teresa Moret-Oliver & Iraide Ibarretxe-Antuñano (University of Zaragoza).....	24
Ditte Boeg Thomsen (University of Copenhagen).....	27
Ramón Boldt (FAU Erlangen-Nürnberg)	30
Julie Bordes (Université Paris 8).....	33
Isabelle Briotet (Laboratoire ELLIADD).....	35
Gioia Cacchioli (Université de Genève)	38
Hugo Carbonnelle (Sorbonne Université - STIH) & Guillaume Kurz (ENS – AOROC)	42
Samuel Chakmakjian (Inalco-ERTIM).....	45
Gilles Col (Université de Poitiers, UR 15076 FORELLIS)	49
Rebecca Connor-Evans (Sheffield Hallam University)	52
Mireille Copin & Inès Saddour (University of Toulouse Jean Jaurès).....	54
Gauthier Delaby, Timothy Coleman & Robert Hartsuiker (Ghent University).....	58
Ruyue Deng (Laboratoire EDA, EA 4071, Université Paris Cité), Huiyun Hu (STL, Université de Lille) & Yanjing Bi (Pékin Chine Capital University of Economics and Business)	61
Marie Dewulf (Ghent University).....	65
Dong & Nicolas Audibert (Laboratoire de Phonétique et Phonologie, Université Sorbonne Nouvelle & CNRS).....	68
Valentina Fedchenko (INALCO)	74
Océane Foubert (University of Lille/STL)	77
Birgit Füreder (Mannheim Universität), Cyrille Granget (Université Toulouse Jean Jaurès), Martin Haiden (Université de Nantes) & Isabel Repiso (Université de Haute Alsace)	80
Cristina Garcia Holgado (University of Poitiers, University of Corsica).....	83
Johannes Gerwien (Heidelberg University).....	86
Matthias Gille Levenson (Aix-Marseille Université), Lucence Ing (Inria) & Carolina Macedo (Ecole nationale des chartes)	87
Patrizia Giuliano (Università degli Studi di Napoli 'Federico II', Dipartimento Studi Umanistici) & Simona Anastasio (Université de Toulouse Jean Jaurès, LNPL).....	90

Matilde Gonçalves (NOVA FCSH)	93
Daria Grigoreva (University of the Basque Country)	96
Nicolas Guillot (Université Bordeaux-Montaigne)	100
Walid Hafsi (Laboratoire de Linguistique de Nantes, LLING UMR 6310)	105
Antoine Heintzmann (UMR 7023 SFL (CNRS/Paris VIII University) - Lyon III University)	110
Eugen Hill (University of Cologne)	113
Henrik Hornecker (Humboldt-Universität zu Berlin)	116
Nigar Ali Hussein (Université Paul Valéry, Montpellier)	119
Oksana Ivchenko & Natalia Grabar (University of Lille)	122
Maria Karali (Aristotle University of Thessaloniki)	127
Hadil Karawani (University of Konstanz)	130
Franziska Keller (Paris Lodron Universität Salzburg)	131
Madeline Ladore (Universität Göttingen)	135
Pierre Larrivée (Université de Caen)	139
Alexander Letuchiy (Université HSE, Moscou / Institut de la langue russe Vinogradov de l'Académie des sciences de Russie, Moscou)	141
Svetlana Leuranguer (Université Toulouse Jean Jaurès), Juliette Thuilier (Université Toulouse Jean Jaurès, CLLE) & Cyrille Granget (Université Toulouse Jean Jaurès, LNPL)	144
Jinyu Li & Adèle Jatteau (Laboratoire Savoirs, Textes, Langage (Université de Lille))	148
Elsa Liste Lamas (ZHAW, Switzerland)	153
Sylvain Loiseau (Université Sorbonne Paris Nord) & Maximilien Guérin (Université de Lille)	156
Karin Madlener-Charpentier (Justus-Liebig-Universität Gießen)	160
Beth Malory (University College London)	164
Michail Marinis (Harvard University & The Ohio State University)	167
Clayton Marr (Ohio State University)	170
Andrei Munteanu (McGill University)	173
Elizaveta Onufrieva (INALCO, SeDyL)	177
Elena Nikishina (HSE University, Vinogradov Russian Language Institute of the Russian Academy of Sciences)	181
Yuanlong Peng (Sorbonne Nouvelle) & Xin Li (Sorbonne Université, STIH)	185
Quentin Peyras (Université Paul Valéry Montpellier 3 - Laboratoire ReSO)	188
Adam Pospíšil (Charles University)	192

Elena Riekhakaynen (Saint-Petersburg State University) and Lada Shumakova (Saint-Petersburg State University; HSE University)	195
Jean Rohleder (LACITO-CNRS).....	198
Catherine Ruchon (Université de Lille) & Nada Jonchere (Université de Montpellier Paul-Valéry).....	202
Anindita Sahoo (Indian Institute of Technology Madras), Celia Shiyu He (University of Bristol), Krishna Raj S. R. (Indian Institute of Technology Madras) & Dagmar Divjak (University of Birmingham)	207
Maguelone Sauzet (Université de Neuchâtel).....	210
Haykanush Sazhumyan & Alice Blumenthal-Dramé (Albert Ludwig University of Freiburg) .	214
Nadia Scanavino (Università di Torino) & Guglielmo Inglese (Università di Torino)	218
Tobias Scheer (Université Côte d'Azur).....	222
Jian Shen Shih (National Chengchi University Graduate Institute of Linguistics)	224
Denys Teptiuk (University of Tartu) & Tatiana Nikitina (CNRS-LaCiTO)	227
Alina Tsikulina & Eva Soroli (University of Lille, STL)	230
Sarah Turner (Coventry University) & Jeannette Littlemore (University of Birmingham)	234
Tobias Ungerer (University of Toronto), Theresa Matzinger (University of Vienna), Michael Pleyer (Nicolaus Copernicus University) & Stefan Hartmann (Heinrich Heine University Düsseldorf)	237
Agathe Wallet, Johanna Cordova, Samuel Chakmakjian, Damien Nouvel, Ilaine Wang & Victoria Khurshudyan (Inalco).....	241
Marinus Wiedner (University of Freiburg)	246
Hing Yuet Fung (The University of Hong Kong).....	249
QiuJun Zhang (Umeå University).....	254
Maria Zielenbach (Vrije Universiteit Amsterdam)	257

Abstracts (par ordre alphabétique)

**Marco Angster, Mia Batinic Angster, Jakov Proroković (University of Zadar),
Gordana Hrzica (University of Zagreb) & Marijana Kresić (University of Zadar)**

Code-switching and borrowing in bilingual first language acquisition

Keywords: code-switching, word internal mixing, Croatian-Italian bilingualism, bilingual first language acquisition, borrowing

Abstract

Bilingual first language acquisition is a domain that can be approached from a variety of perspectives, employing theoretical concepts and methods drawn from first, second language acquisition and contact linguistics. Some of the concepts typical of the mentioned disciplines can overlap from the point of view of their empirical basis, but rest on different theoretical assumptions. For example, the insertion of material from one language in an utterance of another language can be interpreted as borrowing or as code-switching/-mixing mostly depending on the perspective taken by the researcher (see Gardani 2022 for a survey). However, as Gardner-Chloros (2009) points out, some scholars argue against a clear-cut boundary between borrowing and code-switching, whose distinction might be better conceived as a continuum on a diachronic dimension where code-switching is the starting point leading to the establishment of more stable loans. This continuum is nonetheless characterised by inter-speaker variation (Aikhenvald 2002: 197) which stretches between contact-induced speech behaviour (code-switching) and contact-induced language change (borrowing; Haspelmath 2009).

In this paper, we will consider word internal mixing in the spontaneous productions of two Croatian-Italian bilingual children. The sample, collected by the parents in the form of diary notes, covers the period between 3 and 7 years for the older and between 15 months and 4 years and 9 months for the younger child. The children live in Croatia but, apart from the family, they are not exposed to a bilingual environment.

The aim of this paper is to classify the word internal transfer of items (phonological strings, inflectional or derivational morphemes, lexical roots) from one language to the other taking a theory neutral stance leaving aside the dichotomy between code-switch and loan. Despite this neutrality, we will make use of the terminology introduced by Matras and Sakel (2007) and Sakel (2007) further elaborated by Gardani (2018) for classifying the different types of transfer.

Given that the bilingualism of the two children is balanced, we expect the instances of transfer to be bidirectional and that the two children display the same types of transfer from Italian into Croatian and vice versa. On the other hand, some types of transfer fade out as the children grow, while other types are continuously attested throughout the sample – therefore showing a greater degree of stability. We will discuss the different types of transfer in the light of the borrowability hierarchies found in the literature (e.g. Thomason & Kaufman 1988: 74-76; Matras 2007: 61).

Mots-clés: changement de code, mélange interne des mots, bilinguisme croate-italien, acquisition de la première langue bilingue, emprunts linguistiques

Résumé

L'acquisition bilingue de la langue maternelle est un domaine qui peut être abordé sous différents angles, en utilisant des concepts théoriques et des méthodes issus de l'acquisition de la langue maternelle, de l'acquisition d'une deuxième langue et de la linguistique de contact. Certains concepts typiques des disciplines mentionnées peuvent se recouper du point de vue de leur base empirique, mais reposent sur des hypothèses théoriques différentes. Par exemple, l'insertion d'éléments d'une langue dans un énoncé d'une autre langue peut être interprétée comme un emprunt ou comme un changement/mélange de code, principalement en fonction du point de vue adopté par le chercheur (voir Gardani 2022 pour une étude). Cependant, comme le souligne Gardner-Chloros (2009), certains chercheurs s'opposent à une distinction claire entre emprunt et changement de code, dont la distinction pourrait être mieux conçue comme un continuum sur une dimension diachronique où le changement de code est le point de départ menant à l'établissement d'emprunts plus stables. Ce continuum est néanmoins caractérisé par des variations entre locuteurs (Aikhenvald 2002 : 197) qui s'étendent entre le comportement linguistique induit par le contact (changement de code) et le changement linguistique induit par le contact (emprunt ; Haspelmath 2009).

Dans cet article, nous nous intéresserons au mélange interne des mots dans les productions spontanées de deux enfants bilingues croato-italiens. L'échantillon, recueilli par les parents sous forme de notes dans un cahier de bord, couvre la période comprise entre 3 et 7 ans pour l'aîné et entre 15 mois et 4 ans et 9 mois pour le cadet. Les enfants vivent en Croatie mais, en dehors de leur famille, ils ne sont pas exposés à un environnement bilingue.

L'objectif de cet article est de classer le transfert interne de mots (chaînes phonologiques, morphèmes flexionnels ou dérivationnels, racines lexicales) d'une langue à l'autre en adoptant une position théoriquement neutre et en laissant de côté la dichotomie entre changement de code et emprunt. Malgré cette neutralité, nous utiliserons la terminologie introduite par Matras et Sakel (2007) et Sakel (2007), puis développée par Gardani (2018) pour classer les différents types de transfert.

Étant donné que le bilinguisme des deux enfants est équilibré, nous nous attendons à ce que les cas de transfert soient bidirectionnels et que les deux enfants présentent les mêmes types de transfert de l'italien vers le croate et vice versa. D'autre part, certains types de transfert s'estompent à mesure que les enfants grandissent, tandis que d'autres types sont continuellement attestés tout au long de l'échantillon, montrant ainsi un plus grand degré de stabilité. Nous discuterons des différents types de transfert à la lumière des hiérarchies d'emprunt trouvées dans la littérature (par exemple Thomason & Kaufman 1988 : 74-76 ; Matras 2007 : 61).

Références

- Aikhenvald, A. Y. (2002). *Language Contact in Amazonia*. Oxford: Oxford University Press.
- Gardani, F. (2018). On morphological borrowing, *Language and Linguistics Compass* 12, e12302 [1-17].
- Gardani, F. (2022). Contact and borrowing. In: Ledgeway, A. & Maiden, M. (eds.), *The Cambridge Handbook of Romance Linguistics*. Cambridge: Cambridge University Press, pp. 845-869.
- Gardner-Chloros, P. (2009). *Code-switching*. Cambridge: Cambridge University Press.
- Haspelmath, M. (2009). 'Lexical borrowing: concepts and issues'. In Haspelmath, M. and Tadmor, U. (eds), *Loanwords in the World's Languages: A Comparative Handbook*. Berlin: Mouton de Gruyter, pp. 35–54.
- Matras, Y. (2007). The borrowability of structural categories. In: Matras, Y. & Sakel, J. (eds.), *Grammatical borrowing in cross-linguistic perspective*. Berlin/New York: De Gruyter, pp. 31-73.
- Matras, Y. & Sakel, J. (2007). Introduction. In: Matras, Y. & Sakel, J. (eds.), *Grammatical borrowing in cross-linguistic perspective*. Berlin/New York: De Gruyter, pp. 1-13.
- Sakel, J. (2007). Types of loan: Matter and pattern. In: Matras, Y. & Sakel, J. (eds.), *Grammatical borrowing in cross-linguistic perspective*. Berlin/New York: De Gruyter, pp. 15-29.
- Thomason, S. G. & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Silvia Ballarè (University of Bologna) & Paolo Della Putta (University of Turin)

Popular Italian and learner varieties: A comparative analysis of existential constructions

Keywords: Popular Italian; L2 Italian; sub-standard varieties; linguistic simplification; language variation.

Abstract

Popular Italian (PI) is a variety spoken by the lower strata of the population and has been described by several authors (Berruto 1983) as a fossilized learner variety (LV) of Italian. Like LV (Saldana et al. 2021; Chini 2005: 108; Guerini 2024), PI is characterized by variants resulting from acquisitional strategies, such as paradigmatic simplification. Currently, PI is considered obsolescent due to the increased literacy levels of Italian speakers (Cerruti 2018). However, traces of PI remain, particularly among older speakers (Berruto 2012, §3; Guerini 2024). Conversely, immigrants who have been in Italy for many years often speak fossilized LV at basic stages of acquisition.

This study aims to compare PI and LV to determine whether they exhibit similarities, focusing specifically on the behavior of a single linguistic variable: verb agreement in existential constructions (EC, Chambers 2004, Gelso 2009). ECs fit within the model of sociolinguistic variables (and thus can be studied using a solid methodology), and extensive literature on the topic has identified various linguistic factors that may correlate with the lack of agreement.

The study analyzes the behavior of this variable in low- and medium-formality speech among speakers with low educational levels and L2-Italian speakers. These uses are then compared with the productions of highly educated native speakers in the same interactional context.

The data comes from the KIParla corpus (Mauri et al. 2019), specifically from modules dedicated to speakers born in Italy (I) and speakers born and raised abroad (A). In the (I) category, all instances of EC with plural subjects are extracted. Two distinct datasets are created: (I_L), which includes all occurrences produced in PI, and (I_H), which includes constructions produced by educated speakers. The geographic origin of the speakers is considered as an additional metadata.

For the (A) category, the occurrences of the construction are analyzed within two modules, approximately 50 hours of which have been transcribed so far. These modules involve speakers from Bangladesh, China, Morocco, Peru, Romania, and Ukraine residing in Bologna or Turin. In this case, L1 and the length of stay in Italy are taken into account.

The data is annotated according to the following linguistic parameters (Berruto & Cerruti 2015): NP structure, adjacency between the verb and NP, and adjacency between the verb and the head of the NP. To determine whether EC behaves similarly in PI and LV, three statistical analyses are performed and then compared.

First, we examine whether the distribution of the variable is statistically significant by comparing (I_L) and (A). Then, the weight of various linguistic and extra-linguistic factors is ranked within the

three sub-corpora using conditional inference trees and random forests (Tagliamonte & Baayen 2012), initially comparing (I_L) and (A), and subsequently comparing these with (I_H).

If the subgroups (I_L) and (A) exhibit similar behavior, it would suggest that, at least for EC with plural subjects, PI resembles LV. If this result is confirmed in future studies for other linguistic variables, it would support the idea that features of PI are still present in the Italian diasystem, albeit represented by a different population: foreign speakers using fossilized LV.

Mots-clés : italien populaire ; italien L2 ; variétés sub-standard ; simplification linguistique ; variation linguistique.

Résumé

L'italien populaire (IP) est une variété parlée par les couches sociales les plus basses de la population et a été décrit par plusieurs auteurs (Berruto 1983) comme une variété d'apprenants (VA) fossilisée de l'italien. Comme la VA (Saldana et al. 2021 ; Chini 2005 : 108 ; Guerini 2024), l'IP est caractérisé par des variantes résultant de stratégies acquisitionnelles, telles que la simplification paradigmatique. Actuellement, l'IP est considéré comme obsolète en raison de l'élévation du niveau d'alphabétisation des locuteurs italiens (Cerruti 2018). Toutefois, des traces d'IP subsistent, en particulier chez les locuteurs plus âgés (Berruto 2012, §3 ; Guerini 2024). À l'inverse, les immigrants installés en Italie depuis de nombreuses années parlent souvent une VA fossilisée aux stades de base de l'acquisition.

Cette étude vise à comparer l'IP et la VA afin de déterminer s'ils présentent des similarités, en se concentrant spécifiquement sur le comportement d'une seule variable linguistique : l'accord verbal dans les constructions existentielles (CE, Chambers 2004, Gelso 2009). Les CE s'inscrivent dans le modèle des variables sociolinguistiques (et peuvent donc être étudiées par une méthodologie solide), et une vaste littérature sur le sujet a identifié divers facteurs linguistiques susceptibles de corrélérer avec l'absence d'accord.

L'étude analyse le comportement de cette variable dans des contextes de parole peu ou moyennement formels chez des locuteurs faiblement scolarisés et des locuteurs d'italien L2. Ces usages sont ensuite comparés aux productions de locuteurs natifs très scolarisés dans le même contexte interactionnel.

Les données proviennent du corpus KIParla (Mauri et al. 2019), plus précisément des modules dédiés aux locuteurs nés en Italie (I) et aux locuteurs nés et élevés à l'étranger (A). Pour la catégorie (I), toutes les occurrences de CE avec des sujets pluriels sont extraites. Deux ensembles de données distincts sont créés : (I_L), qui inclut toutes les occurrences produites en IP, et (I_H), qui inclut les constructions produites par des locuteurs instruits. L'origine géographique des locuteurs est également prise en compte comme métadonnée supplémentaire.

Pour la catégorie (A), les occurrences de la construction sont analysées dans deux modules dont environ 50 heures ont été transcrites jusqu'à présent. Ces modules impliquent des locuteurs originaires du Bangladesh, de Chine, du Maroc, du Pérou, de Roumanie et d'Ukraine résidant à Bologne ou à Turin. Dans ce cas, la L1 et la durée de séjour en Italie sont prises en considération.

Les données sont annotées selon les paramètres linguistiques suivants (Berruto & Cerruti 2015) : structure du GN, adjacency entre le verbe et le GN, et adjacency entre le verbe et le noyau du GN. Afin de déterminer si les CE se comportent de manière similaire en IP et en VA, trois analyses statistiques sont réalisées puis comparées.

Nous examinons d'abord si la distribution de la variable est statistiquement significative en comparant (I_L) et (A). Ensuite, le poids de divers facteurs linguistiques et extra-linguistiques est classé au sein des trois sous-corpus au moyen d'arbres d'inférence conditionnelle et de forêts aléatoires (Tagliamonte & Baayen 2012), en comparant d'abord (I_L) et (A), puis en comparant ces résultats avec (I_H).

Si les sous-groupes (I_L) et (A) présentent un comportement similaire, cela suggérerait qu'au moins pour les CE avec des sujets pluriels, l'IP ressemble à la VA. Si ce résultat est confirmé dans de futures études pour d'autres variables linguistiques, cela appuierait l'idée que les traits de l'IP sont encore présents dans le diasystème italien, mais représentés par une population différente : des locuteurs étrangers utilisant une VA fossilisée.

Références

Berruto, Gaetano. 1983. "L'italiano popolare e la semplificazione linguistica". *Vox Romanica*. 42, 38-79.

Berruto, Gaetano. 2012. *Sociolinguistica dell'italiano contemporaneo*. Seconda edizione. Roma, Carocci.

Berruto, Gaetano / Cerruti, Massimo. 2015. "Un esercizio di analisi variazionista: l'accordo verbale nel costrutto locativo-esistenziale-presentativo". In Busà, Maria Grazia / Gesuato, Sara (Eds.). *Lingue e contesti. Studi in onore di Alberto M. Mioni*. Padova, CLEUP, 609-620.

Cerruti, Massimo. 2018. "La formazione di varietà intermedie tra 'dialetto' e 'standard' in situazioni europee: alcune considerazioni dall'angolatura italiana". *Rivista Italiana di Dialettologia* 42, 79-99.

Chambers, Jack K. 2004. "Dynamic typology and vernacular universals". In Kormann, Bernd (Ed.), *Dialectology meets typology: Dialect grammar from a cross-linguistic perspective*. Berlino – New York: de Gruyter. 128–145.

Chini, Marina. 2005. *Che cos'è la linguistica acquisizionale*. Roma, Carocci.

Cortelazzo, Manlio. 1972. *Avviamento critico allo studio della dialettologia italiana*. In *Lineamenti di italiano popolare*. Pisa, Pacini.

Gelso, Daniela. 2009. "Creoli e interlingue: un confronto". *Italiano LinguaDue* 1(2). 68-97.

Guerini, Federica. 2024, *L'italiano Popolare*, in Ballarè, Silcia / Fiorentini, Ilaria / Miola, Emanuele, (a cura di), *Le varietà dell'italiano contemporaneo*. Roma, Carocci. 67-80.

Mauri, Caterina / Ballarè, Silvia / Gorla, Eugenio / Cerruti, Massimo / Suriano, Francesco. 2019. *KIParla corpus: a new resource for spoken Italian*. In Raffaella Bernardi, Roberto Navigli & Giovanni Semeraro (Eds.). *Proceedings of the 6th Italian conference on Computational Linguistics CLiC-it*, Torino, Accademia University Press.

Saldana, Carmen / Smith, Kenny / Kirby, Simon / Culbertson, Jenny. 2021. "Is regularisation uniform across linguistic levels? Comparing learning and production of unconditioned probabilistic variation in morphology and word order". *Language Learning and Development* 17(2). 158-188.

Tagliamonte, Sali A. / Baayen, R. Harald. 2012. "Models, forests and trees of York English: Was/were variation as case study for statistical practice". *Language Variation and Change* 24(2). 135-178.

Silvia Ballarè, Caterina Mauri, Ludovica Pannitto & Eleonora Zucchini (University of Bologna)

Transcribing internal diversity: challenges and solutions from the KIParla Corpus

Keywords: spoken corpus; diversity; non-standard; transcription practices

Abstract

While transcription is a foundational practice in the creation of spoken language corpora, it becomes especially challenging when aiming to represent linguistic diversity in an accurate and reusable way. This contribution presents insights from the KIParla corpus (www.kiparla.it, Mauri et al. 2019), a modular and evolving corpus of spoken Italian designed to document internal sociolinguistic variation. It contains spoken interactions recorded ecologically in different contexts, involving speakers with diverse socio-geographic characterization. For this reason, the corpus captures non-standard linguistic features and several cases of code-switching with Italo-romance dialects and other languages, such as Arabic, Spanish, and Chinese.

Italian is considered a well-documented language, but its internal variation poses challenges similar to those encountered when working with under-resourced languages. In our contribution, we will describe the workflow adopted in the creation of the KIParla corpus, also considering the problems arising from the type of data and the measures adopted to face them.

Due to the nature of the data, KIParla employs a flexible system that incorporates both orthographic and conversational conventions (Jefferson 2004) to represent prosodic and interactional features (overlaps, pauses, hesitations...), that include symbols for marking code-switching. Moreover, when transcribing deviations from the standard, we always aspire to balancing faithfulness and usability; however, normalization comes with challenges: on one hand, it is necessary to support querying and NLP integration, on the other, it may conceal diversity and have theoretical implications, which is imperative to avoid.

The system flexibility comes with the necessity to consider each case individually, therefore recordings are transcribed manually. Given the size of the corpus (2.326.171 tokens, 220 hours), manual transcription can only be performed as a shared task, involving multiple people. More specifically, data are transcribed by interns (BA and MA students), properly trained and periodically monitored. This may hinder transcription homogeneity; so to limit the subjectivity of the different transcribers, transcriptions are revised and harmonized by a junior researcher.

Finally, we are exploring whether it is viable to reduce transcription time with ASR, without compromising quality. We will illustrate the preliminary results of an experiment carried out using Whisper, OpenAI's speech-to-text model, which has shown that its output may not be suitable yet to replace manual annotation, since it requires a significant amount of post-processing to meet the transcription standards of the KIParla corpus (eg. re-introducing non-standard realisations, fixing audio-transcript alignment and segmentation, assigning units to speakers, annotating conversational and prosodic features). Moreover, we are experimenting a script to detect inconsistencies and, when possible, automatically correct them.

The goal of this presentation is to critically examine the methodology developed and refined over the years in transcribing authentic spoken language. We will highlight how, despite being a well-documented language, Italian still presents significant challenges. By reflecting on our experience, we aim to showcase the adjustments made to improve accuracy and consistency, offering insights into the complexities of speech transcription and the solutions adopted to address them.

Mots-clés: corpus oral ; diversité ; non standard ; pratiques de transcription

Résumé

Bien que la transcription soit une pratique fondamentale dans la création de corpus de langue parlée, elle devient particulièrement difficile lorsqu'il s'agit de représenter la diversité linguistique de manière précise et réutilisable. Cette contribution présente les conclusions tirées du corpus KIParla (www.kiparla.it, Mauri et al. 2019), un corpus modulaire et évolutif de l'italien parlé conçu pour documenter les variations sociolinguistiques internes. Il contient des interactions orales enregistrées de manière écologique dans différents contextes, impliquant des locuteurs présentant des caractéristiques sociogéographiques diverses. Pour cette raison, le corpus capture des caractéristiques linguistiques non standard et plusieurs cas de changement de code avec des dialectes italo-romans et d'autres langues, telles que l'arabe, l'espagnol et le chinois.

L'italien est considéré comme une langue bien documentée, mais ses variations internes posent des défis similaires à ceux rencontrés lors du travail avec des langues sous-documentées. Dans notre contribution, nous décrivons le processus adopté pour la création du corpus KIParla, en tenant également compte des problèmes découlant du type de données et des mesures adoptées pour y faire face.

En raison de la nature des données, KIParla utilise un système flexible qui intègre à la fois les conventions orthographiques et conversationnelles (Jefferson 2004) pour représenter les caractéristiques prosodiques et interactionnelles (chevauchements, pauses, hésitations...), y compris les symboles permettant de marquer les changements de code. De plus, lorsque nous transcrivons des écarts par rapport à la norme, nous aspirons toujours à trouver un équilibre entre fidélité et facilité d'utilisation ; cependant, la normalisation pose des défis : d'une part, il est nécessaire de prendre en charge les requêtes et l'intégration du TALN, d'autre part, elle peut masquer la diversité et avoir des implications théoriques, ce qu'il est impératif d'éviter.

La flexibilité du système implique la nécessité d'examiner chaque cas individuellement, c'est pourquoi les enregistrements sont transcrits manuellement. Compte tenu de la taille du corpus (2 326 171 tokens, 220 heures), la transcription manuelle ne peut être effectuée que sous forme de tâche partagée, impliquant plusieurs personnes. Plus précisément, les données sont transcrites par des stagiaires (étudiants en licence et en master), correctement formés et régulièrement contrôlés. Cela peut nuire à l'homogénéité de la transcription ; afin de limiter la subjectivité des différents transcripteurs, les transcriptions sont révisées et harmonisées par un jeune chercheur.

Enfin, nous étudions la possibilité de réduire le temps de transcription grâce à la reconnaissance vocale automatique (ASR), sans compromettre la qualité. Nous présenterons les résultats préliminaires d'une expérience menée à l'aide de Whisper, le modèle de reconnaissance vocale d'OpenAI, qui a montré que ses résultats ne sont peut-être pas encore adaptés pour remplacer

l'annotation manuelle, car ils nécessitent un post-traitement important pour répondre aux normes de transcription du corpus KIParla (par exemple, réintroduire les réalisations non standard, corriger l'alignement et la segmentation audio-transcription, attribuer des unités aux locuteurs, annoter les caractéristiques conversationnelles et prosodiques). De plus, nous expérimentons un script pour détecter les incohérences et, lorsque cela est possible, les corriger automatiquement.

L'objectif de cette présentation est d'examiner de manière critique la méthodologie développée et affinée au fil des années pour transcrire la langue parlée authentique. Nous mettrons en évidence comment, malgré le fait qu'il s'agisse d'une langue bien documentée, l'italien présente encore des défis importants. En nous appuyant sur notre expérience, nous souhaitons présenter les ajustements apportés pour améliorer la précision et la cohérence, en offrant un aperçu des complexités de la transcription de la parole et des solutions adoptées pour y remédier.

Références

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In: Lerner, G. H. (a cura di), *Conversation Analysis: studies from the first generation*. Amsterdam: John Benjamins, pp. 13-31.

Mauri, C. & Ballarè S. & Gorla, E. Cerruti M. & Suriano F. (2019). KIParla corpus: a new resource for spoken Italian. In: Bernardi, R., R. Navigli & G. Semeraro (eds.), *Proceedings of the 6th Italian Conference on Computational Linguistics CLiC-it*.

Katia Benaiges-Saura, María-Teresa Moret-Oliver & Iraide Ibarretxe-Antuñano
(University of Zaragoza)

Deixis matters: Catalan and Aragonese speakers prioritize deixis over Spanish in motion events

Keywords: Intra-typological variation; Deixis; Path; Language production task

Abstract

This study examines the interaction between Path (sub)components and Manner in several linguistic systems: Catalan, Aragonese, and their frontier varieties. It evaluates their status and usage, as well as whether their speakers exhibit the same preferences as Spanish speakers, who prefer encoding Path in the main verb when describing complex events (Ibarretxe-Antuñano, 2025). Additionally, the research focuses on how Romance languages with seemingly divergent systems encode Deixis, exploring potential intra-typological variation. Specifically, the study analyzes which information is described, where it is mentioned (inside or outside main verb) and whether scenes properties influences speakers' choices.

Path and Deixis are notions related to the trajectory of motion. Deixis refers to how the motion is directed relative to the space of the Speaker. Conceptually, it is the combination of Path schema and a special space, namely that of the Speaker (Matsumoto, Akita & Takahashi, 2017). There are three typological groups in relation to Deixis (Gathercole, 1977, 1978): (i) languages that allow both interlocutors (Speaker and Addressee) as deictic center, (ii) languages that only allow Speaker, and (iii) languages in which this category is not relevant. This affects the use of GO and COME and it is one of the main differences between Romance languages. In Spanish and Western Aragonese, venitive encoding is only allowed when the Speaker is the deictic center. In contrast, Catalan (except for Valencian) and Eastern Aragonese extend this encoding to include both the Speaker and the Addressee as possible deictic centers (Ibarretxe-Antuñano, 2009; Ibarretxe-Antuñano, Hijazo-Gascón & Moret-Oliver, 2017; Hijazo-Gascón & Ibarretxe-Antuñano, 2010).

To investigate these differences, an experimental study was conducted with sixty native speakers. Participants were divided into five groups based on their language: Spanish monolinguals, Western Aragonese, Eastern Aragonese, Catalan and speakers of frontier varieties. Among the Catalan and Aragonese speakers (n=48), all were bilingual in Spanish with early and systematic exposure to both languages.

The study employed a production task using 63 video clips depicting self-motion events, adapted from Matsumoto (2025). Clips were organized around three variables: Path, Manner and Deixis. All components competed for their selection as they appeared simultaneously in scenes. Participants had to locate themselves in the scene and verbally describe what was happening, based on available resources in their language. The study aimed to identify language-specific patterns and examine possible intra-typological and dialectal variation in motion encoding and in salience of motion event components.

Results show that all speakers tend to encode Path in the main verb. However, Catalan, Aragonese and frontier variety speakers exhibit a stronger preference for encoding Deixis in main verb compared to Spanish Speakers. Furthermore, they encode Deixis in the main verb in contexts where Spanish speakers choose Path. Regarding deictic verbs, Catalan and Eastern Aragonese speakers use the verb COME to express motion toward the Speaker and toward the Addressee, a pattern that diverges from that observed in Spanish. Results suggest that Deixis holds significant cognitive salience for them (see Koga, 2017).

Mots-clés : Variation intra-typologique ; Deixis ; Trajectoire ; Tâche de production langagière

Résumé

Cette étude examine l'interaction entre les (sous-)composants de Trajectoire et de Manière dans plusieurs systèmes linguistiques catalans, aragonais et leurs variétés frontalières. Elle évalue le statut et l'usage de ces langues, ainsi que la question de savoir si leurs locuteurs présentent les mêmes préférences que les locuteurs espagnols, qui privilégient l'encodage de la Trajectoire dans le verbe principal lors de la description d'événements complexes (Ibarretxe-Antuñano, 2025). De plus, la recherche se concentre sur la façon dont les langues romanes aux systèmes apparemment divergents encodent la Deixis, explorant la variation intra-typologique potentielle. Spécifiquement, l'étude analyse quelles informations sont décrites, où elles sont mentionnées (à l'intérieur ou à l'extérieur du verbe principal) et si les propriétés des scènes influencent les choix des locuteurs.

La Trajectoire et la Deixis sont des notions liées à la trajectoire du mouvement. La Deixis fait référence à la façon dont le mouvement est dirigé par rapport à l'espace du Locuteur. Conceptuellement, c'est la combinaison du schéma de Trajectoire et d'un espace spécial, notamment celui du Locuteur (Matsumoto, Akita & Takahashi, 2017). Il existe trois groupes typologiques en relation avec la Deixis (Gathercole, 1977, 1978) : (i) les langues qui permettent à la fois aux interlocuteurs (Locuteur et Destinataire) d'être le centre déictique, (ii) les langues qui ne permettent qu'au Locuteur, et (iii) les langues dans lesquelles cette catégorie n'est pas pertinente. Cela affecte l'usage de ALLER et VENIR et constitue l'une des principales différences entre les langues romanes. En espagnol et en aragonais occidental, l'encodage ventitif n'est autorisé que lorsque le Locuteur est le centre déictique. En revanche, le catalan (excepté le valencien) et l'aragonais oriental étendent cet encodage pour inclure à la fois le Locuteur et le Destinataire comme centres déictiques possibles (Ibarretxe-Antuñano, 2009 ; Ibarretxe-Antuñano, Hijazo-Gascón & Moret-Oliver, 2017 ; Hijazo-Gascón & Ibarretxe-Antuñano, 2010).

Pour étudier ces différences, une étude expérimentale a été menée avec soixante locuteurs natifs. Les participants ont été répartis en cinq groupes selon leur langue : espagnol monolingue, aragonais occidental, aragonais oriental, catalan et locuteurs de variétés frontalières. Parmi les locuteurs catalans et aragonais (n=48), tous étaient bilingues en espagnol avec une exposition précoce et systématique aux deux langues.

L'étude a employé une tâche de production utilisant 63 clips vidéo représentant des événements d'auto-mouvement, adaptés de Matsumoto (2025). Les clips étaient organisés autour de trois variables : Trajectoire, Manière et Deixis. Tous les composants ont concouru pour leur sélection car ils apparaissaient simultanément dans les scènes. Les participants devaient se situer dans la scène et décrire verbalement ce qui se passait, en se basant sur les ressources disponibles dans

leur langue. L'étude visait à identifier les patterns spécifiques aux langues et à examiner la variation intra-typologique et dialectale possible dans l'encodage du mouvement et dans la saillance des composants d'événements de mouvement.

Les résultats montrent que tous les locuteurs tendent à encoder la Trajectoire dans le verbe principal. Cependant, les locuteurs catalans, aragonais et de variétés frontalières présentent une préférence plus forte pour l'encodage de la Deixis dans le verbe principal comparativement aux locuteurs espagnols. De plus, ils encodent la Deixis dans le verbe principal dans des contextes où les locuteurs espagnols choisissent la Trajectoire. Concernant les verbes déictiques, les locuteurs catalans et aragonais oriental utilisent le verbe VENIR pour exprimer le mouvement vers le Locuteur et vers le Destinataire, un pattern qui diverge de celui observé en espagnol. Les résultats suggèrent que la Deixis détient une saillance cognitive significative pour eux (voir Koga, 2017).

Références

- Hijazo-Gascón, A. & Ibarretxe-Antuñano, I. (2010). Tipología, lexicalización y dialectología aragonesa. *Archivo de Filología Aragonesa*, 66, pp. 245-279.
- Ibarretxe-Antuñano, I. (2009). Path salience in motion events. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura and S. Ösçaliskan (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Oz*. New York: Routledge.
- Ibarretxe-Antuñano, I. (2025). The interaction of motion event (sub)components in Spanish motion verbs. In Yo Matsumoto (ed.), *Motion Event Descriptions Across Languages*. Mouton de Gruyter.
- Ibarretxe-Antuñano, I., Hijazo-Gascón, A. & Moret-Oliver, M. (2017). The importance of minority languages in semantic typology: The case of Aragonese and Catalan. In I. Ibarretxe-Antuñano (Ed.), *Motion and space across languages: Theory and applications*. Amsterdam, John Benjamins.
- Koga, H. (2017). Self-motion expressions in Japanese, English, German, and Russian: A comparative study using translation corpora. In Yo Matsumoto (ed.), *The typology of motion expressions*. Tokyo, Kurosio.
- Matsumoto, Y., K. Akita & K. Takahashi. (2017). The functional nature of deictic verbs and the coding patterns of deixis: An experimental study in English, Japanese, and Thai. In I. Ibarretxe-Antuñano (Ed.), *Motion and space across languages*. Amsterdam: John Benjamins.
- Matsumoto, Y. (2025). *Motion Event Descriptions Across Languages*. Berlin, Mouton de Gruyter.

Ditte Boeg Thomsen (University of Copenhagen)

Does Associated Motion grammar affect attention and memory? Visual processing and change detection in bilingual speakers of Nahuatl

Keywords: Associated Motion; thinking for speaking; change detection; eye-tracking; Nahuatl

Abstract

Psycholinguistic experiments have provided evidence that crosslinguistic variation in the coding of motion events is associated with differences in motion cognition (e.g., Soroli, Hickmann & Hendriks, 2019). While there is a strong tradition of investigating motion events that are lexically coded as main events, little is known about language-cognition relations for motion events that are grammatically expressed, i.e., coded as the *background* for other events. Grammatically expressed motion, called Associated Motion (AM), is typologically widespread, but still underdescribed, having only recently been recognized as a comparative concept (Guillaume & Koch, 2021). AM offers speakers a tool for habitually expressing meanings like ‘(go and) eat’ or ‘talk (while coming)’, as in (1) from Nahuatl (Uto-Aztecan, Mexico), where the suffix *-ti* adds a simultaneous going event to the primary scattering event designated by the verb:

(1) Tequila Nahuatl

Kin-xihxinih-ti

3PLO-scatter-AM.CONC.IT

‘She scatters them (going).’

According to the *thinking-for-speaking* hypothesis (Slobin, 2003), speakers develop cognitive routines for monitoring and storing those aspects of events that their grammar encourages them to communicate about routinely. Speakers who routinely use AM can thus be expected to pay more attention to and have enhanced memory for background motion compared to speakers who do not. Languages with AM are now often spoken by bilingual speakers in language-contact contexts where the dominating language does not possess AM, and *if* the use of AM affects cognition, this effect is likely to be weaker in bilingual speakers who use AM less routinely.

In a field experiment with 16 bilingual speakers of Nahuatl (Uto-Aztecan, Mexico) and Spanish (age range: 44-87 years, 11 female), the current study investigated whether variation in the use of AM predicted variation in motion memory and attention. Nahuatl AM systems are highly elaborate, with morphological paradigms contrasting coming/going motion and three temporal relations: prior, concurrent and subsequent motion. Using change detection, the experiment tested whether AM supports the ability to detect changes in motion direction in videos presenting motion as the background for other events. Using eye-tracking, the experiment further measured speakers’ gaze trajectories while viewing these videos in preparation for speaking.

In the change-detection experiment, variation in the bilingual speakers’ use of AM predicted variation in their ability to detect direction changes one day later ($\beta = 1.44$, $p < 0.05$). As for visual processing, all speakers fixated most on scene components relevant for processing the primary events (hands and objects), but the more speakers used AM, the more they fixated on scene components relevant for processing motion (legs and paths, $\beta = 0.06$, $p < 0.05$).

These results indicate an effect of AM on attention and memory, thus extending the findings from research on lexically expressed motion to the underdescribed, but typologically widespread phenomenon of grammatically expressed motion. Further, as these results depend on differences between bilingual speakers in the same speech communities, they suggest that linguistic variability extends to cognitive variability, with differential patterns of attention and memory associated with differences in the use of grammatical patterns present in only one of the speakers' two languages.

AM: Associated Motion, CONC: concurrent motion, IT: itive, away from the deictic centre, 3PLO: third-person plural object

Mots-clés : Mouvement Associé ; thinking for speaking ; détection de changement ; suivi oculaire ; nahuatl

Résumé

Les expériences psycholinguistiques ont fourni des preuves que la variation translinguistique dans le codage des événements de mouvement est associée à des différences dans la cognition du mouvement (par exemple, Soroli, Hickmann & Hendriks, 2019). Bien qu'il existe une forte tradition d'investigation des événements de mouvement qui sont lexicalement codés comme événements principaux, peu de choses sont connues sur les relations langage-cognition pour les événements de mouvement qui sont grammaticalement exprimés, c'est-à-dire codés à l'arrière-plan d'autres événements. Le mouvement grammaticalement exprimé, appelé Mouvement Associé (MA), est typologiquement répandu, mais encore sous-décrit, n'ayant été reconnu que récemment comme un concept comparatif (Guillaume & Koch, 2021). Le MA offre aux locuteurs un outil pour exprimer habituellement des significations comme « (aller et) manger » ou « parler (en venant) », comme en (1) du nahuatl (uto-aztèque, Mexique), où le suffixe -ti ajoute un événement de déplacement simultané à l'événement principal de dispersion désigné par le verbe :

(1) Tequila Nahuatl

kin-xihxinih-ti

3PLO-scatter-AM.CONC.IT

« Elle les disperse (en allant). »

Selon l'hypothèse du *thinking-for-speaking* (Slobin, 2003), les locuteurs développent des routines cognitives pour surveiller et stocker ces aspects des événements que leur grammaire les encourage à communiquer de manière routinière. Les locuteurs qui utilisent couramment le MA peuvent donc être attendus à porter plus d'attention et à avoir une mémoire renforcée pour le mouvement d'arrière-plan comparé aux locuteurs qui ne le font pas. Les langues avec MA sont maintenant souvent parlées par des locuteurs bilingues dans des contextes de contact linguistique où la langue dominante ne possède pas de MA, et si l'utilisation du MA affecte la cognition, cet effet est susceptible d'être plus faible chez les locuteurs bilingues qui utilisent le MA moins couramment.

Dans une expérience de terrain avec 16 locuteurs bilingues du nahuatl (uto-aztèque, Mexique) et de l'espagnol (tranche d'âge 44-87 ans, 11 femmes), l'étude actuelle a investigué si la variation dans l'utilisation du MA prédit la variation dans la mémoire du mouvement et l'attention. Les systèmes de MA nahuans sont hautement élaborés, avec des paradigmes morphologiques contrastant mouvement venant/allant et trois relations temporelles : antérieur, concurrent et subséquent. En utilisant la détection de changement, l'expérience a testé si le MA soutient la capacité à détecter des changements dans la direction du mouvement dans des vidéos présentant le mouvement comme arrière-plan d'autres événements. En utilisant le suivi oculaire, l'expérience a en outre mesuré les trajectoires du regard des locuteurs pendant qu'ils visionnaient ces vidéos en préparation pour parler.

Dans l'expérience de détection de changement, la variation dans l'utilisation du MA par les locuteurs bilingues a prédit la variation dans leur capacité à détecter les changements de direction un jour plus tard ($\beta = 1.44$, $p < 0.05$). Quant au traitement visuel, tous les locuteurs ont fixé le plus sur les composants de scène pertinents pour traiter les événements principaux (mains et objets), mais plus les locuteurs utilisaient le MA, plus ils fixaient sur les composants de scène pertinents pour traiter le mouvement (jambes et pieds, $\beta = 1.06$, $p < 0.05$).

Ces résultats indiquent un effet du MA sur l'attention et la mémoire, étendant ainsi les découvertes de la recherche sur le mouvement lexicalement exprimé au phénomène sous-décrit, mais typologiquement répandu du mouvement grammaticalement exprimé. De plus, comme ces résultats dépendent des différences entre locuteurs bilingues dans les mêmes communautés linguistiques, ils suggèrent que la variabilité linguistique s'étend à la variabilité cognitive, avec des patterns différentiels d'attention et de mémoire associés aux différences dans l'utilisation des patterns grammaticaux présents dans seulement une des deux langues des locuteurs.

MA : Mouvement Associé, CONC : mouvement concurrent, IT : itif, loin du centre déictique, 3PLO : objet pluriel de troisième personne.

Références

Guillaume, A. & Koch, H. (2021). Introduction: Associated Motion as a grammatical category in linguistic typology. In: Guillaume, A. & Koch, H. (eds.), *Associated Motion*. Berlin/Boston: De Gruyter Mouton, pp. 3–30.

Slobin, D.I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In: Gentner, D. & Goldin-Meadow, S. (eds.), *Language in Mind*. Cambridge, Massachusetts: The MIT Press, pp. 157–191.

Soroli, E., Hickmann, M. & Hendriks, H. (2019). Casting an eye on motion events: Eye tracking and its implications for typology. In: Aurnague, M. & Stosic, D., *The Semantics of Dynamic Space in French: Descriptive, Experimental and Formal Studies on Motion Expression*. Amsterdam: John Benjamins, pp. 250–288.

Ramón Boldt (FAU Erlangen-Nürnberg)

*Northwest Germanic *ā < *ē₁ in time and space*

Abstract

In my talk, I am aiming at answering the question if it is possible to narrow down the change of Northwest Germanic *ā from earlier *ē₁ in time and space. Special emphasis will be placed on methodological problems.

Every survey work and every (comparative) historical grammar of the Germanic languages mentions in more or less detail the fact that Gothic long ē, where it continues Germanic stressed *ē₁, corresponds to or presupposes ā in the Northwest Germanic languages (in selection: AhdGr: 57 f.; AsGr: 61 ff.; Fulk 2018: 60 f.; Grønvik 1998: 87; Ringe & Taylor 2014: 10 ff.).

It is striking that although it is often emphasised that the earliest attestations of the North Germanic languages already show ā in the expected position¹, hardly any concrete textual evidence is used to further determine the spread of the change within the remainder of the Northwest Germanic dialect continuum in more detail. Accordingly, Fulk (2018: 61 note 4) rightly notes that the dialect-geographical observations concerning the change *ē₁ > *ā made by Bremer almost 150 years ago (1886: 12–29) have hardly been taken into account in recent research.

Additionally, in the recent literature, problematic attestations with <e> instead of expected <a> like the late PN *Theudomērus* or *Dagedērus* from the 8th century are usually interpreted as dialectal variations or as expressing scribal traditions (AhdGr: 58: “traditionelle Schreibung ... Schreibtradition”). This happens tacitly, methodological reflections on these interpretations are never made explicit.

This is remarkable because a) the change *ā < *ē₁ is usually taken not only as a central feature of Northwest Germanic but introduced when making the case for a unified Proto-Northwest-Germanic, and b) recent reflections in variational linguistics and dialectology suggest that variation in written sources by no means must reflect dialectal variation but can, among other things, also reflect social and situational differences of different kinds (Elementaler & Voeste 2019: 61).

As it seems, in light of this, premature and heuristically unjustified to build elaborate models of the dialectal variation of Northwest Germanic upon few problematic instances of proposed changes like the one given above, I will argue for more caution in dealing with written sources in historical dialectology.

¹ See Proto-Norse *wajemāriR* ‘the not ill-famed’ (Goth. *merjan* ‘announce’, Burg. PN *Sigismērem*) on the Thorsberg chape (ca. 200); SuR (#99).

Résumé

Dans ma communication, je chercherai à déterminer s'il est possible de cerner l'évolution temporelle et spatiale du germanique du Nord-Ouest **ā* par rapport à l'ancien **ē*₁. Je mettrai l'accent sur les problèmes méthodologiques.

Tous les travaux de synthèse et toutes les grammaires historiques (comparatives) des langues germaniques mentionnent, plus ou moins en détail, que le *ē* long gothique, lorsqu'il continue le **ē*₁ germanique accentué, correspond à *ā* ou le présuppose dans les langues germaniques du nord-ouest (voir sélection : AhdGr : 57 ss. ; AsGr : 61 ss. ; Fulk 2018 : 60 ss. ; Grønvik 1998 : 87 ; Ringe & Taylor 2014 : 10 ss.).

Il est frappant de constater que, bien qu'il soit souvent souligné que les plus anciennes attestations des langues germaniques du Nord présentent déjà *ā* à la position attendue², peu de preuves textuelles concrètes sont utilisées pour déterminer plus en détail la propagation du changement dans le reste du continuum dialectal germanique du Nord-Ouest. Par conséquent, Fulk (2018 : 61, note 4) note à juste titre que les observations dialecto-géographiques concernant le changement **ē*₁ > **ā* faites par Bremer il y a près de 150 ans (1886 : 12–29) n'ont guère été prises en compte dans les recherches récentes.

De plus, dans la littérature récente, les attestations problématiques comportant un <e> au lieu du <a> attendu, comme le PN tardif *Theudomērus* ou le *Dagedērus* du VIII^e siècle, sont généralement interprétées comme des variations dialectales ou comme l'expression de traditions scripturales (AhdGr : 58 : « traditionnelle Schreibung ...Schreibtradition »). Cela se produit tacitement, les réflexions méthodologiques sur ces interprétations n'étant jamais explicitées.

Ceci est remarquable car a) le changement **ā* < **ē*₁ est généralement considéré non seulement comme une caractéristique centrale du germanique du Nord-Ouest, mais introduit dans la thèse d'un proto-germanique du Nord-Ouest unifié, et b) des réflexions récentes en linguistique variationnelle et en dialectologie suggèrent que la variation dans les sources écrites ne doit en aucun cas refléter une variation dialectale, mais peut, entre autres, refléter également des différences sociales et situationnelles de différentes natures (Elmentaler & Voeste 2019 : 61).

Étant donné qu'il semble prématuré et heuristiquement injustifié de construire des modèles élaborés de la variation dialectale du germanique du Nord-Ouest à partir de quelques exemples problématiques de changements proposés, comme celui présenté ci-dessus, je préconiserai une plus grande prudence dans l'utilisation des sources écrites en dialectologie historique.

² Voir le proto-norrois *wajemāriR* « celui qui n'a pas mauvaise réputation » (gothique *merjan* « annoncer », burgonien PN *Sigismērem*) sur la chape de Thorsberg (vers 200) ; SuR (#99).

Références

AhdGr = Braune, Wilhelm; Heidermanns, Frank, 2023: *Althochdeutsche Grammatik I, Phonologie und Morphologie*. Berlin, Boston.

AsGr = Gallée, Johan Hendrik, 1993: *Altsächsische Grammatik*. Berlin, New York.

Bremer, Otto, 1886: *Germanisches ē*. in: PBB 11, 1–76.

Elmentaler, Michael; Voeste, Anja, 2019: Areale Variation Im Deutschen Historisch: Mittelalter Und Frühe Neuzeit. In: Schmidt, Jürgen Erich (Ed.): *Language and Space. An International Handbook of Linguistic Variation* (= HSK 30), Berlin, Boston, 61–100.

Fulk, R. D., 2018: *A Comparative Grammar of the Early Germanic Languages*. Amsterdam, Philadelphia.

Grønvik, Ottar, 1998: Untersuchungen zur älteren nordischen und germanischen Sprachgeschichte. Frankfurt am Main.

Krahe-Meid = Krahe, Hans; Meid, Wolfgang, 1969: *Germanische Sprachwissenschaft*. Band I. Berlin.

Ringe, Don; Taylor, Ann, 2014: *The development of Old English*. Oxford.

SuR = Krause, Wolfgang, 1971: *Die Sprache der urnordischen Inschriften*. Heidelberg.

Julie Bordes (Université Paris 8)

Préférences d'encodage de l'aspect progressif au présent en anglais L2 chez les francophones

Résumé

L'encodage de l'action en déroulement au présent en français est réalisé via des moyens lexicaux non obligatoires tels que « être en train de » ou au Présent, contrairement à l'anglais qui le grammaticalise via la forme en Be+ING (De Wit et Brisard, 2014). Cette différence d'encodage rend l'acquisition de cette forme difficile pour les francophones, utilisant souvent des informations sur la télicité du prédicat pour l'accorder, là où les anglophones préfèrent la borne à droite visuelle de l'action (Leclercq, 2008, Demagny, 2013).

Dans le cadre d'une recherche sur cette opposition entre la forme en Be+ING au présent et le présent simple, nous avons sélectionné 40 participants universitaires et étudiant l'anglais, de niveau intermédiaire à avancé. Ils ont réalisé 3 tâches différentes, dont les items variaient en la télicité/atélicité du prédicat, la présence/absence d'une borne à droite dans le clip. Les trois tâches étaient des tâches de préférences où les participants choisissaient entre :

1. Deux phrases vis-à-vis d'une vidéo,
2. Deux vidéos vis-à-vis d'une phrase,
3. Deux verbes pour remplir un vide dans une phrase vis-à-vis d'une vidéo.

Nous avons ensuite comparé ces résultats à ceux d'anglophones natifs. Les tâches ont également été réalisées en français par des francophones, en gardant les mêmes items visuels, en cours d'analyse.

Nous relevons deux stratégies de réponses : accord visuel (simple avec action bornée, Be+ING avec action non-bornée), et accord linguistique (simple avec verbe télique, Be+ING avec verbe atélique). Tous les participants ont préféré un accord visuel (FL1ANGL2 : 60,5% visuel ; 55% linguistique / ANGL1 : visuel = 96% ; linguistique = 54%).

Lorsque les items combinaient un accord à la fois visuel et linguistique, le temps de réponse et la similarité avec les natifs augmentait chez les FRL1ANGL2. Chez les anglophones, l'accord télique est généralement limité aux situations où l'accord visuel est possible. Une ANOVA a montré que les différentes tâches ont un impact sur les FRL1ANGL2, mais pas sur les natifs. Dans la tâche 3, les apprenants ont beaucoup plus utilisé Be+ING, imputable au fait qu'il s'agissait d'une tâche de production ciblée, contrairement à T1 et T2.

Cette recherche montre l'impact des structures de la L1 sur l'acquisition des L2 (Leclercq et Benazzo, 2021), et en particulier dans l'expression aspectuelle. Cependant, notre étude apporte une nuance sur l'impact de la tâche dans l'expression et le traitement de la progressivité.

Abstract

The encoding of ongoing action in the present tense in French is achieved through non-mandatory lexical means such as *être en train de* or the simple Present, unlike English, which grammaticalizes it via the Be+ING form (De Wit & Brisard, 2014). This difference in encoding makes the acquisition of this form difficult for French speakers, who often rely on information about the telicity of the predicate to choose, whereas English speakers prefer the visual right boundary of the action (Leclercq, 2008; Demagny, 2013).

As part of a study on this opposition between the Be+ING form in the present and the simple present, we selected 40 university students learning English at an intermediate to advanced level. They carried out three different tasks, in which the items varied according to the telic/atelic nature of the predicate and the presence/absence of a right boundary in the clip. The three tasks were preference-based tasks where participants chose between:

1. Two sentences in relation to a video;
2. Two videos in relation to a sentence;
3. Two verbs to fill a blank in a sentence in relation to a video.

We then compared these results with those of native English speakers. The tasks were also carried out in French by French speakers, using the same visual items; this part is currently being analyzed.

We observed two response strategies: visual agreement (simple present with bounded action, Be+ING with unbounded action) and linguistic agreement (simple present with a telic verb, Be+ING with an atelic verb). All participants preferred a visual agreement (FR L1 – EN L2: 60.5% visual; 55% linguistic / EN L1: visual = 96%; linguistic = 54%).

When the items combined both visual and linguistic agreement, response time and similarity with native speakers increased for FR L1 – EN L2 participants. Among native English speakers, telic agreement is generally limited to situations where visual agreement is possible. An ANOVA showed that the different tasks had an effect on FR L1 – EN L2 participants, but not on native speakers. In Task 3, learners used Be+ING much more, which can be attributed to the fact that this was a targeted production task, unlike T1 and T2.

This research highlights the impact of L1 structures on L2 acquisition (Leclercq & Benazzo, 2021), particularly in aspectual expression. However, our study nuances this by revealing an impact of task type on the expression and processing of progressivity.

Références

- De Wit, A., & Brisard, F. (2014). A Cognitive Grammar account of the semantics of the English present progressive. *Journal of Linguistics*, 50(1), 49-90.
- Leclercq, P. (2008). L'influence de la langue maternelle chez les apprenants adultes quasi-bilingues dans une tâche contrainte de verbalisation : Étude de l'expression du déroulement en français et en anglais. *Acquisition et interaction en langue étrangère*, 26.

Isabelle Briotet (Laboratoire ELLIADD)

Etude de la parole atypique : Quelle méthodologie? Quelles contraintes?

Mots-clés : parole atypique ; trouble développemental du langage ; sémantique interprétative ; interaction ; fonctionnalité

Résumé

L'étude de la parole atypique, ou devrait-on dire « des paroles atypiques » tant elle revêt, notamment chez l'enfant, un caractère idiolectal, pourrait s'apparenter à l'étude de langues peu dotées. A l'instar de Devevey, nous « considér[ons] la parole atypique comme la matérialisation d'un acte de langage. Ainsi, en la plaçant à ce niveau nous la considérons comme un objet d'étude à part entière, nous proposons d'en décrire les caractéristiques » (Devevey, 2024). Dans le cas de la parole atypique, comme dans le cas des langues peu dotées, différents types d'enjeux se posent : des enjeux méthodologiques, d'une part, concernant le traitement des données orales (recueil d'un échantillon représentatif, conventions et formats de transcription à adopter, etc.) ; et d'autre part, des enjeux linguistiques (étude de structures linguistiques rares voire jamais décrites, étude des marques de l'oralité, prise en compte de la multimodalité de la communication, etc.). Souvent considérée comme marginale, la parole atypique chez l'enfant est cependant répandue (7,5 % des enfants sont concernés par un Trouble Développemental du Langage (Maillart, 2023; Norbury et al., 2016)), aussi, son étude mérite attention et ressources, tout comme d'autres langues.

Cette communication vise à exposer les défis relatifs à la transcription et à l'analyse de la parole atypique, soulevés dans le cadre d'un projet de recherche sur la fonctionnalité dans l'interaction d'enfants présentant une parole atypique. Souhaitant proposer une vision différente de la pathologie du langage chez l'enfant, nous postulons que certains enfants présentant une parole atypique ont, dans certains contextes de vie quotidienne, une « fonctionnalité interactionnelle », que nous définissons en nous basant sur le paradigme de l'énaction. Ce dernier nous permet d'envisager le langage comme une activité collaborative, incarnée et située ; et d'admettre que le sens émerge, au cours du processus interactionnel, par des processus de co-construction. En modélisant l'interaction dans un cadre énatif, nous proposons une méthodologie permettant de mettre en évidence ces processus. Cette démarche nous confronte inéluctablement aux contraintes inhérentes à l'analyse de corpus de parole atypique en linguistique conversationnelle, se rapprochant, nous semble-t-il, de celles pesant sur l'analyse de données orales de langues peu dotées (Comment tendre à l'exhaustivité et à la précision sans se perdre dans des informations non pertinentes ? Comment retranscrire le plus fidèlement et objectivement possible les productions, en essayant de conserver leur caractère naturel ? Quelles conventions de transcription choisir ?, etc.). Elle implique également d'adopter une approche méthodologique et des paradigmes linguistiques traditionnellement moins mobilisés en la matière, comme la sémantique interprétative, qui postule que le sens n'est pas prédonné, mais qu'il est le fruit de parcours interprétatifs (Rastier, 2009). Cette théorie nous semble être un cadre particulièrement intéressant pour résoudre des problématiques liées à l'étude de la parole atypique (Comment transcrire des productions non canoniques, les mots et les structures syntaxiques idiolectales ?, etc.). Ces considérations pourraient être étendues à l'étude des langues peu dotées.

Keywords: atypical speech; developmental language disorder; interpretive semantics; interaction; functionality

Abstract

The study of atypical speech – perhaps we should say *atypical utterances* given that, especially in children, it often takes on an idiolectal character – could be related to the study of under-documented languages. Following Devevey, we “[consider] *atypical speech as the materialization of a speech act. By placing it at this level, we regard it as an object of study in its own right, and we propose to describe its characteristics.*” (Devevey, 2024). In the case of atypical speech, as with under-documented languages, different types of issues arise: on the one hand, methodological issues concerning the treatment of oral data (gathering a representative sample, deciding on transcription conventions and formats, etc.); and, on the other hand, linguistic issues (study of rare or previously undescribed linguistic structures, study of markers of orality, taking into account the multimodality of communication, etc.). Often seen as marginal, atypical speech in children is in fact widespread (7.5% of children are affected by a Developmental Language Disorder (Maillart, 2023; Norbury et al., 2016)); its study therefore deserves attention and resources, just like other languages.

This paper aims to present the challenges related to the transcription and analysis of atypical speech, raised within the framework of a research project on functionality in the interaction of children exhibiting atypical speech. Wishing to offer a different perspective on language pathology in children, we posit that some children with atypical speech do, in certain contexts of everyday life, display an “interactional functionality,” which we define based on the paradigm of enaction. This paradigm allows us to consider language as a collaborative, embodied, and situated activity, and to acknowledge that meaning emerges, during the interaction, through processes of co-construction. By modeling interaction within an enactive framework, we propose a methodology that makes it possible to highlight these processes. This approach inevitably confronts us with the constraints inherent to the analysis of atypical speech corpora in conversational linguistics, constraints which, we feel, resemble those that weigh on the analysis of oral data from under-documented languages (How can one aim for exhaustivity and precision without being overwhelmed by irrelevant information? How can one transcribe productions as faithfully and objectively as possible while trying to preserve their natural character? Which transcription conventions should be chosen?, etc.). It also involves adopting a methodological approach and linguistic paradigms that are traditionally less mobilized in this field, such as interpretive semantics, which posits that meaning is not pre-given, but rather the product of interpretive pathways (Rastier 2009). This theory is, according to us, a particularly relevant framework for addressing issues linked to the study of atypical speech (How should we transcribe non-canonical productions, idiolectal words and syntactic structures?, etc.). These considerations could also be extended to the study of under-documented languages.

Références

Devevey, A. (2024). Quels paradigmes pour modéliser la parole atypique? In : *Langage oral—Nouvelles approches incarnées de la parole atypique* (De Boeck Supérieur), pp. 17-59.

Maillart, C. (2023). *Le trouble développemental du langage : Un handicap invisible*. <https://orbi.uliege.be/bitstream/2268/307131/1/maillart-handicap-invisible.pdf>

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder : Evidence from a population study. *Journal of Child Psychology and Psychiatry*, 57(11): 1247-1257. <https://doi.org/10.1111/jcpp.12573>

Rastier, F. (2009). *Sémantique interprétative* (3e éd). Presses universitaires de France.

Keywords: Complement Clauses ; Relative Clauses ; Tigrinya ; Ethio-semitic ; Syntax

Abstract

1. Overview. Head-final languages are expected to have clause-final complementizers and are not expected to display prefixes. However, in Tigrinya (Ethio-semitic, Eritrea/Ethiopia, SOV) the complementizer *kəmzi-* is a prefix that precedes the subordinate verb. It occurs in complement clauses introduced by verbs of factivity (p.ex. ‘know’), non-factivity (p.ex. ‘think’), utterance (p.ex. ‘say’) and perception (p.ex. ‘see’, as in (1)).

(1) *ʔit-om ʔawədat dəbdabə kəmzi-ts'əhaf-u rəʔij-ε-jom*
DEM-MP boy.MP letter COMP-write.PFV-3MP see.GER-1S-OBJ.3MP

‘I saw that the boys wrote a letter.’

I claim that *kəmzi* is bimorphemic, consisting of the preposition *kəm* ‘like’, whose use is shown in (2), and the prefix *zi-* ‘that’, whose use is shown in (3). This prefix introduces relative clauses and comparative clauses.

(2) *kəm-ay təmaharit ʔi-jja*
like-OBJ.1S student.FS COP.PRES-3FS

‘She is a student like me.’

(3) a. *ʔit-a nsxa zi-hab-ka-ni mətʂ'əhaf ʔatfiʔ-i-ja*
DEM-FS you.MS ZI-give.PFV-2MS-OBJ.1S book.FS lose.GER-1S-OBJ.2FS

‘I lost the book that you gave me.’

b. *Solomon kab-t-l ʔanə zi-ħasibə-xu-wo ji-ʕabi*
Solomon than-DEM-MS I ZI-think.PFV-1S-OBJ.3MS 3MS-be_tall.IPFV

‘Solomon is taller than I thought.’

Several other Ethio-Semitic languages (Tigre, Amharic, Argobba and Zay a.o.) make use of similative morphemes, with the meaning of ‘like’, to introduce some types of complement clauses and purpose clauses (Treis 2017). Note that *kəmzi-* is similar to the complementizer *kaʔanna* ‘as if’ in Standard Arabic, that is claimed to consist of the particles *ka-* ‘like’ and *ʔanna* ‘that’ (Ahmed 2015). Note also that, similarly to *zi-*, in Tunisian Arabic the element that introduces relative clauses, *elli*, also introduces complement clauses (Hewett 2023). The phenomenon presented above is thus found across Semitic languages.

2. Analysis. Because *zi-* appears in both relative clauses and comparative clauses, it has been proposed to analyze it as a marker of $\bar{A}$ -movement (Cacchioli 2023). Given that complement clauses present this prefix, I claim that they also involve $\bar{A}$ -movement. I provide a syntactic account for complement clauses in Tigrinya by following works on the semantics of complementation: I propose that what raises in the structure is an Operator coindexed with a phonologically null nominal (Kratzer 2006, Moulton 2009). This null nominal is contextually dependent on the matrix verb: it could be CLAIM (Arsenijević 2009), FACT (Kayne 2010), EVENT (Poletto and Sanfelici 2021) and so forth. More technically, *kəm* heads a functional projection that

dominates vP and in whose specifier an operator is base-generated. This operator is coindexed with a phonologically null nominal that heads a functional projection that dominates CP. This functional projection is the complement of the matrix verb and modifies the null nominal (Stowell 1981, Moulton 2009). The operator raises to Spec,CP, while the embedded verb moves to C°. The Spec-head configuration between the operator and the verb triggers the appearance of *zi-* that is prefixed onto the verb. The syntactic derivation is schematized and simplified in the tree below.

(4)

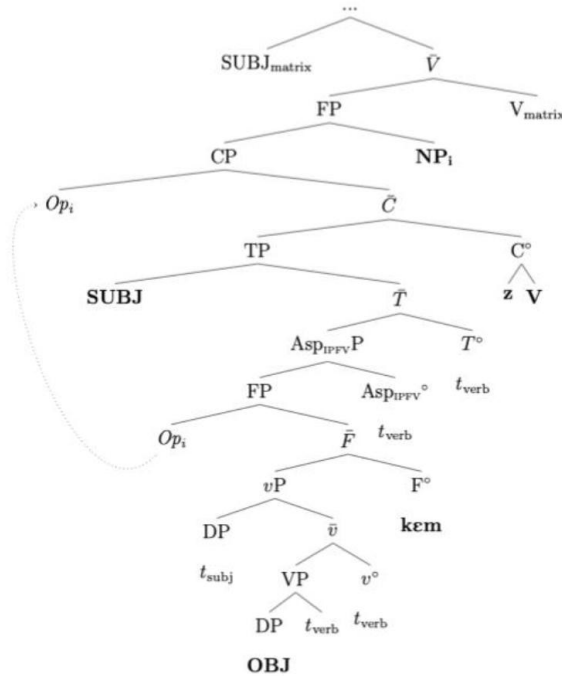


Fig. 1: Syntactic derivation of complement clauses in Tigrinya

The word order that emerges from this series of operations is SUBJ-OBJ-*kem-zi*-V which is exactly what we see in the complement clause in (1).

Mots-clés : Propositions complétives ; Propositions relatives ; Tigrinya ; Éthio-sémitique ; Syntaxe

1. Vue d'ensemble. Les langues à tête finale sont censées avoir des complémenteurs en position finale de proposition et ne sont pas censées présenter de préfixes. Cependant, en tigrinya (éthio-sémitique, Érythrée/Éthiopie, SOV), le complémenteur *kemzi-* est un préfixe qui précède le verbe subordonné. Il apparaît dans les propositions complétives introduites par des verbes de factivité (p.ex. « savoir »), de non-fectivité (p.ex. « penser »), d'énonciation (p.ex. « dire ») et de perception (p.ex. « voir »), comme en (1).

(1) ?it-om	?awɛdat	dɛɓdabɛ	kemzi- ts'ɛhaf-u	rɛʔij-ɛ-jom
DEM-MP	garçon.MP	lettre	COMP-écrire.PFV-3MP	voir.GER-1S-OBJ.3MP

'J'ai vu que les garçons ont écrit une lettre.'

Je soutiens que *kemzi* est bimorphémique, consistant en la préposition *kem* « comme », dont l'usage est montré en (2), et le préfixe *zi-* « que », dont l'usage est montré en (3). Ce préfixe introduit les propositions relatives et comparatives.

(2) *kɛm-ay* tɛmaharit ʔi-ja
 comme-OBJ.1S étudiant.FS COP.PRES-3FS

'Elle est étudiante comme moi.'

(3) a. ʔit-a nsxa *zi*-hab-ka-ni mɛts'ħaf ʔatfiʔ-i-ja
 DEM-FS tu.MS Zt-donner.PFV-2MS-OBJ.1S livre.FS perdre.GER-1S-OBJ.2FS

'J'ai perdu le livre que tu m'as donné.'

b. Solomon kab-t-l ʔane *zi*-ħasibɛ-xu-wo ji-ʕabi
 Solomon que-DEM-MS I Zt-penser.PFV-1S-OBJ.3MS 3MS-être_grand.IPFV

'Solomon est plus grand que je ne le pensais.'

Plusieurs autres langues éthio-sémitiques (tigré, amharique, argobba et zay entre autres) font usage de morphèmes similaires, avec le sens de « comme », pour introduire certains types de propositions complétives et de propositions de but (Treis 2017). Notez que *kemzi-* est similaire au complémenteur *ka/ʔanna* « que » de l'arabe standard, dont il est affirmé qu'il consiste en les particules *ka-* « comme » et *ʔanna* « que » (Ahmed 2015). Notez également que, de manière similaire à *zi-*, en arabe tunisien l'élément qui introduit les propositions relatives, *elli*, introduit aussi les propositions complétives (Hewett 2023). Le phénomène présenté ci-dessus se trouve donc dans les langues sémitiques.

2. Analyse. Parce que *zi-* apparaît à la fois dans les propositions relatives et comparatives, il a été proposé de l'analyser comme un marqueur de mouvement-A (Cacchioli 2023). Étant donné que les propositions complétives présentent ce préfixe, je soutiens qu'elles impliquent également un mouvement-A. Je fournis un compte rendu syntaxique des propositions complétives en tigrinya en suivant les travaux sur la sémantique de la complémentation : je propose que ce qui monte dans la structure est un Opérateur indexé avec un nominal nul phonologiquement (Kratzer 2006, Moulton 2009). Ce nominal nul est contextuellement dépendant du verbe matrice : il pourrait être CLAIM (Arsenijević 2009), FACT (Bayne et Poletto et Sanfelici 2021) et ainsi de suite. Plus techniquement, *kem* dirige une projection fonctionnelle qui domine VP et dans le spécifieur de laquelle un opérateur est généré de base. Cet opérateur est co-indexé avec un nominal nul phonologiquement qui dirige une projection fonctionnelle qui domine CP. Cette projection fonctionnelle est le complément du verbe matrice et modifie le nominal nul (Stowell 1981, Moulton 2009). L'opérateur monte vers Spec,CP, tandis que le verbe enchâssé monte vers C°. La configuration Spec-tête entre l'opérateur et le verbe déclenche l'apparition de *zi-* qui est préfixé sur le verbe. La dérivation syntaxique est schématisée et simplifiée dans l'arbre ci-dessous.

(4)

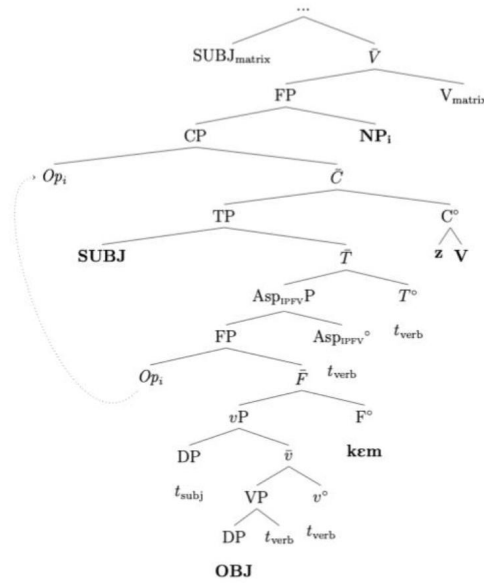


Fig. 1 : Dérivation syntaxique des propositions complétives en tigrinya

L'ordre des mots qui émerge de cette série d'opérations est SUBJ-OBJ-kem-zi-V qui est exactement ce que nous voyons dans la proposition complétive en (1).

Références

- Ahmed, H. E. I. (2015). *Verbal Complementizers in Arabic* [Ph.D. Thesis]. The University of Utah.
- Arsenijević, B. (2009). Clausal complementation as relativization. *Lingua*, 119, 39–50.
- Cacchioli, Gioia. 2023. "The Tigrinya Zi- Prefix: A Morphological Reflex of Successive-Cyclic Movement." *Brill's Journal of Afroasiatic Languages and Linguistics* 15(1): 232–61.
- Hewett, M. (2023). *Types of $\bar{A}$ -dependencies* [Ph.D. Thesis]. The University of Chicago.
- Kayne, R. S. (2010). Why Isn't This A Complementizer? In *Comparisons and Contrasts*. Oxford University Press.
- Kratzer, A. (2006). Decomposing Attitude Verbs. Talk given in honor of Anita Mittwoch, The Hebrew University Jerusalem.
- Moulton, K. (2009). *Natural Selection and the Syntax of Clausal Complementation* [Ph.D. Thesis]. University of Massachusetts Amherst.
- Poletto, C., & Sanfelici, E. (2021). Against Complementizers. In S. Wolfe & C. Meklenborg (Eds.), *Continuity and Variation in Germanic and Romance* (pp. 370–403). Oxford University Press.
- Stowell, T. (1981). *Origins of Phrase Structure* [Ph.D. Thesis]. MIT.
- Treis, Y. (2017). Similitive morphemes as purpose clause markers in Ethiopia and beyond. In Y. Treis & M. Vanhove (Eds.), *Similitive and Equative Constructions: A cross-linguistic perspective* (pp. 91–142). John Benjamins Publishing Company.

*When “majority loses” is a better etymology than “majority wins”: the case of the Proto-Romance suffix */-kl-/*

Keywords: Romance languages; Proto-Romance; comparative method; derivational morphology; marginalized linguistic data

Abstract

When analyzing the so-called diminutive formations that were inherited in the Romance languages, the diversity of etymologies is strikingly evident (Dębowiak, 2013, pp. 156–164): some forms are thought to derive from Latin suffixal derivatives in *-iculus*, *-a*, *-um*, others are regarded as reflexes of derivatives in *-uculus*, *-a*, *-um*, while still others are thought to belong to both types. The decision to trace a given form back to one type, the other, or both, appears to be guided by a well-known principle in comparative linguistics: majority wins (Campbell, 2020, p. 148). Thus, the conventional etymology of the words for ‘louse’ in the Romance languages is *pēduc(u)lus*, not *pēdic(u)lus*, because, aside from a few marginal Ladin forms (Kramer et al., 1993, pp. 275–277), most inherited forms derive from this type. Along the same lines, when Meyer-Lübke (1935, pp. 39, 56, 207, 319, 772) identifies nouns tracing back to both variants, he systematically cites the etymon in *-iculus* before that in *-uculus*, without questioning this variation or the formation of one type before the other.

Such fluctuation is already documented in Latin and has already been the subject of several attempts to explain it within the Latin system: according to Meyer-Lübke (1894, pp. 466–469) and Leumann (1977, pp. 306–307), the suffix was originally determined by the theme to which it was added, but this initial distribution became blurred over time; in contrast, Ward (1951, p. 481), considering the written sources, offers a purely sociolinguistic explanation: variants in *-uculus* would be linked to basilectal, overwhelmingly oral varieties.

A shift in perspective, from the past of Latin to its future, brings forth new insights: the comparative analysis of the Romance languages has emerged as an exceptionally productive approach, significantly expanding our understanding of their common ancestor (Hall, 1983; Buchi & Schweickard, 2008; Dworkin, 2016). A pan- and proto-Romance framework allows for a finer analysis of derivation. Turning once again to the so-called diminutive formations, the study of Proto-Romance etyma in */-kl-/ reveals a recurrent pattern of two etymological types (Hall, 1983, p. 125; Heidemeier, 2014, pp. 19–21): one in */-’ɪkl-/, the other in */-’ʊkl-/. Yet, the reconstruction of one of these variants often rests on marginalized data that are scarcely represented in traditional sources. A broader perspective, rather than a case-by-case investigation of these etyma, quickly suggests that minority variants may, in some cases, represent the most ancient forms from an etymological standpoint.

This analysis, applicable to */pe’dɪkl-u/ ~ */pe’dʊkl-u/ ‘louse’, */ar’tɪkl-u/ ~ */ar’tʊkl-u/ ‘joint’ and many other nouns, serves as a reminder of the crucial importance of paying scrupulous attention to dialectal data, especially marginal ones, in the field of linguistic reconstruction and of accepting that, sometimes, *majority loses*.

Mots-clés: langues romanes ; proto-roman ; méthode comparative ; morphologie dérivationnelle ; données linguistiques marginalisées

Résumé

Lorsqu'on analyse les formations dites diminutives héritées dans les langues romanes, la diversité des étymologies est frappante (Dębowiak, 2013, p. 156-164) : certaines formes dériveraient de dérivés suffixaux latins en *-iculus*, *-a*, *-um*, d'autres seraient des reflets de dérivés en *-uculus*, *-a*, *-um*, tandis que d'autres encore appartiendraient aux deux types. Le choix de rattacher une forme donnée à l'un ou l'autre type, ou aux deux, semble être guidé par un principe bien connu en linguistique comparée : *la majorité l'emporte* (Campbell, 2020, p. 148). Ainsi, l'étymologie conventionnelle des mots pour « pou » dans les langues romanes est *pēduc(u)lus*, et non *pēdic(u)lus*, car, hormis quelques formes ladines marginales (Kramer et al., 1993, p. 275-277), la plupart des formes héritées dérivent de ce type. De même, lorsque Meyer-Lübke (1935, p. 39, 56, 207, 319, 772) identifie des noms remontant aux deux variantes, il cite systématiquement l'étymon en *-iculus* avant celui en *-uculus*, sans remettre en question cette variation ni la formation d'un type avant l'autre.

Une telle fluctuation est déjà documentée en latin et a fait l'objet de plusieurs tentatives d'explication au sein du système latin : selon Meyer-Lübke (1894, p. 466-469) et Leumann (1977, p. 306-307), le suffixe était initialement déterminé par le thème auquel il était ajouté, mais cette répartition initiale s'est estompée au fil du temps ; à l'inverse, Ward (1951, p. 481), considérant les sources écrites, propose une explication purement sociolinguistique : les variantes en *-uculus* seraient liées à des variétés basilectales, majoritairement orales.

Un changement de perspective, du passé du latin à son avenir, apporte de nouvelles perspectives : l'analyse comparative des langues romanes s'est révélée une approche exceptionnellement productive, élargissant considérablement notre compréhension de leur ancêtre commun (Hall, 1983 ; Buchi & Schweickard, 2008– ; Dworkin, 2016). Un cadre pan- et proto-roman permet une analyse plus fine de la dérivation. Si l'on revient aux formations diminutives, l'étude de l'étyme proto-roman en **/-kl-/* révèle une récurrence de deux types étymologiques (Hall, 1983, p. 125 ; Heidemeier, 2014, p. 19–21) : l'un en **/-ʰkl-/*, l'autre en **/-ʷkl-/*. Cependant, la reconstruction de l'une de ces variantes repose souvent sur des données marginales, peu représentées dans les sources traditionnelles. Une perspective plus large, plutôt qu'une étude au cas par cas de ces étymologies, suggère rapidement que les variantes minoritaires pourraient, dans certains cas, représenter les formes les plus anciennes d'un point de vue étymologique.

Cette analyse, applicable à **/pe'dɪkl-u/ ~ */pe'dʊkl-u/* « pou », **/ar'tɪkl-u/ ~ */ar'tʊkl-u/* « joint » et à de nombreux autres noms, rappelle l'importance cruciale d'accorder une attention scrupuleuse aux données dialectales, notamment marginales, dans le domaine de la reconstruction linguistique et d'accepter que, parfois, *la majorité soit perdante*.

Références

- Buchi, É. & Schweickard, W. (Eds.) (2008–). *Dictionnaire Étymologique Roman*. <http://www.atilf.fr/DERom> (date de consultation : 08 juin 2025).
- Campbell, L. (2020). *Historical Linguistics. An Introduction* (4th ed.). Edinburgh: Edinburgh University Press.
- Dębowiak, P. (2013). *La formation diminutive dans les langues romanes*. PhD's Thesis. University of Cracow.
- Dworkin, S. (2016). *Do Romanists need to reconstruct Proto-Romance?: The case of the Dictionnaire Étymologique Roman (DÉRom) project*. *Zeitschrift für romanische Philologie*, 132(1), 1–19.
- Hall, R. A., Jr. (1983). *Comparative Romance Grammar*. Vol. 3: *Proto-Romance Morphology*. Amsterdam: Benjamins.
- Heidemeier, U. (2014). Reconstruction dérivationnelle. In É. Buchi & W. Schweickard (Eds.), *Dictionnaire Étymologique Roman (DÉRom)*. Vol. 1: *Genèse, méthodes et résultats* (pp. 211–246). Berlin: De Gruyter.
- Kramer, J. (Ed.), Mehren, U., Fiacre, K.-J., Schlösser, R. & Thybussek, E.-M. (1993). *Etymologisches Wörterbuch des Dolomitenladinischen* (Vol. 5). Hamburg: Buske.
- Leumann, M. (1977). *Lateinische Laut- und Formenlehre* (5th ed.). Munich: Beck.
- Meyer-Lübke, W. (1894). *Grammatik der romanischen Sprachen*. Vol. 2: *Formenlehre*. Leipzig: Reisland.
- Meyer-Lübke, W. (1935). *Romanisches Etymologisches Wörterbuch* (3rd ed.). Heidelberg: Winter.
- Ward, R. L. (1951). Stop Plus Liquid and the Position of the Latin Accent. *Language*, 27(4), 477–484.

Transducing Diaphonemes and Archigraphemes: Armenian as an example of multi-variant alignment

Keywords: diaphoneme; archigrapheme; phonological variation; transducers; lexical alignment

Abstract

Armenian is an Indo-European isolate with two standardized variants, its own alphabet, and two distinct orthographic norms. Eastern and Western Armenian differ both phonologically (Vaux, 1998) and orthographically, and while digital resources for both are expanding, they remain limited. A core challenge in developing NLP tools for Armenian lies in aligning data across these variants. This is an essential step toward supporting the language's pluricentric structure and data. This paper proposes a finite-state approach to address this challenge by leveraging the concepts of diaphonemes and archigraphemes (abstract units that resolve phonological and graphemic mismatches, respectively). We present a method for aligning two large-scale Armenian lexicons, along with preliminary results.

Armenian exhibits regular correspondences between variant specific phonemes, which we model using *diaphonemes* (abstract units representing said corresponding but different phonemes). In many cases, these are already represented by the same grapheme in Armenian script. Conversely, some phonemes shared by both variants are written using different graphemes due to the two orthographic norms. We address these cases through *archigraphemes*, abstract graphemic units that neutralize orthographic divergence to support more accurate alignment.

	Western Armenian	Eastern Armenian	Archigrapheme
(1) 'wall'	պատ, <i>pat</i> /bad/, [bad]	պատ, <i>pat</i> /pat/, [pat]	պատ
(2) 'chicken'	հաւ, <i>haw</i> /hav/, [hav]	հաւ, <i>hav</i> /hav/, [hav]	HwV
(3) 'floor'	յարկ, <i>yark</i> /harg/, [harg]	հարկ, <i>hark</i> /hark/, [hark]	Hարկ

Table 1: Examples of phonologic and orthographically divergent cognate pairs

Our two lexicons were compiled from different sources: the Eastern Armenian lexicon (59,430 lemmas) draws primarily from the EANC online corpus (Khurshudyan, 2022), while the Western Armenian lexicon (33,951 lemmas) was developed using resources from the DALiH³ project. We used the foma toolkit (Hulden, 2009) to construct three finite-state transducers. Two generate phonetic transcriptions for each variant using their standard orthography as input. A third transducer implements the archigraphemic transformation, replacing divergent graphemes with neutral abstract symbols.

As a baseline, we extracted all lexeme pairs with identical orthographic forms across the two lists, yielding 15,258 matches. Applying the archigraphemic transformation to the remaining data uncovered an additional 7,132 matched pairs. This resulted in a total of 22,390 alignments,

³ <https://digitalarmenian.huma-num.fr>

approximately 66% of the smaller (Western Armenian) lexicon. These results demonstrate that script-level normalization significantly improves lexical matching across dialects, even before applying phonological abstraction.

Armenian's pluricentric nature, unified script but different orthographic norms may be particular, but we propose that a transductive model based on either diaphonemes, archigraphemes or both could be effectively adapted to other language pairs which exhibit systematic variation across dialects or writing systems, providing a foundation for future work in lexical alignment, and multivariant language modelling.

	W. Armenian	E. Armenian
<i>Number of words in the lexicon</i>	33,951	59,430
<i>Orthographically identical lexemes</i>	15,258	15,258
<i>Lexemes extracted using archigraphemes</i>	7,132	7,132
Total aligned	22,390	22,390

Table 2: Preliminary results of lexicon alignments

Mots-clés: diaphonème, archigraphème, variation phonogique, transducteurs, alignement lexical

Résumé

L'arménien est une langue indo-européenne isolée qui comporte deux variantes standardisées, son propre alphabet et deux normes orthographiques distinctes. L'arménien oriental et l'arménien occidental diffèrent tant sur le plan phonologique (Vaux, 1998) que sur le plan orthographique, et bien que les ressources numériques pour ces deux variantes soient en expansion, elles restent limitées. L'un des principaux défis dans le développement d'outils de TALN pour l'arménien réside dans l'alignement des données entre ces variantes. Il s'agit d'une étape essentielle pour prendre en charge la structure pluricentrique et les données de la langue. Cet article propose une approche à états finis pour relever ce défi en s'appuyant sur les concepts de diaphonèmes et archigraphèmes (unités abstraites qui résolvent respectivement les incohérences phonologiques et graphémiques). Nous présentons une méthode d'alignement de deux lexiques arméniens à grande échelle, ainsi que des résultats préliminaires.

L'arménien présente des correspondances régulières entre des phonèmes spécifiques à chaque variante, que nous modélisons à l'aide de diaphonèmes (unités abstraites représentant lesdits phonèmes correspondants mais différents). Dans de nombreux cas, ceux-ci sont déjà représentés par le même graphème dans l'écriture arménienne. À l'inverse, certains phonèmes communs aux deux variantes sont écrits à l'aide de graphèmes différents en raison des deux normes orthographiques. Nous traitons ces cas à l'aide d'archigraphèmes, des unités graphémiques abstraites qui neutralisent les divergences orthographiques afin de permettre un alignement plus précis.

	Arménien occidental	Arménien oriental	Archigraphème
(1) « mur »	պաւոն, <i>pat</i> /bad/, [bad]	պաւոն, <i>pat</i> /pat/, [pat]	պաւոն
(2) « poulet »	հաւԼ, <i>haw</i> /hav/, [hav]	հաւ՝, <i>hav</i> /hav/, [hav]	HwV
(3) « sol »	յարկ, <i>yark</i> /harg/, [harg]	հարկ, <i>hark</i> /hark/, [hark]	Hարկ

Tableau 1 : Exemples de paires de mots apparentés divergents sur le plan phonologique et orthographique

Nos deux lexiques ont été compilés à partir de différentes sources : le lexique arménien oriental (59 430 lemmes) provient principalement du corpus en ligne EANC (Khurshudyan, 2022), tandis que le lexique arménien occidental (33 951 lemmes) a été développé à partir des ressources du projet DALiH⁴. Nous avons utilisé la boîte à outils foma (Hulden, 2009) pour construire trois transducteurs à états finis. Deux d'entre eux génèrent des transcriptions phonétiques pour chaque variante en utilisant leur orthographe standard comme entrée. Un troisième transducteur met en œuvre la transformation archigraphémique, en remplaçant les graphèmes divergents par des symboles abstraits neutres.

Comme base de référence, nous avons extrait toutes les paires de lexèmes ayant des formes orthographiques identiques dans les deux listes, ce qui a donné 15 258 correspondances. L'application de la transformation archigraphémique aux données restantes a permis de découvrir 7 132 paires correspondantes supplémentaires. Cela a donné un total de 22 390 alignements, soit environ 66 % du lexique le plus petit (arménien occidental). Ces résultats démontrent que la normalisation au niveau de l'écriture améliore considérablement la correspondance lexicale entre les dialectes, même avant d'appliquer l'abstraction phonologique.

La nature pluricentrique de l'arménien, avec son écriture unifiée mais ses normes orthographiques différentes, peut sembler particulière, mais nous proposons qu'un modèle transductif basé soit sur les diaphonèmes, soit sur les archigraphèmes, soit sur les deux, pourrait être efficacement adapté à d'autres paires de langues qui présentent des variations systématiques entre les dialectes ou les systèmes d'écriture, fournissant ainsi une base pour les travaux futurs sur l'alignement lexical et la modélisation multivariée des langues.

⁴ <https://digitalarmenian.huma-num.fr>

	Arménien occidental	Arménien oriental
<i>Nombre de mots dans le lexique</i>	33 951	59 430
<i>Lexèmes orthographiquement identiques</i>	15 258	15 258
<i>Lexèmes extraits à l'aide d'archigraphèmes</i>	7 132	7 132
Total aligné	22 390	22 390

Tableau 2 : Résultats préliminaires des alignements de lexiques

Références

Hulden, M. (2009). Foma: a Finite-State Compiler and Library. Dans J. Kreutel (Éd.), *Proceedings of the Demonstrations Session at {EACL} 2009* (pp. 29-32). Athens, Greece: Association for Computational Linguistics.

Khurshudyan, V. a. (2022). Eastern Armenian National Corpus: State of the Art and Perspectives. Dans V. a.-G. Khurshudyan (Éd.), *Proceedings of the Workshop on Processing Language Variation: Digital Armenian (DigitAm) within the 13th Language Resources and Evaluation Conference* (pp. 28--37). Marseille, France: European Language Resources Association.

Vaux, B. (1998). *The Phonology of Armenian*. Oxford University Press.

Pourquoi « pas de souci » ?

Mots-clés : marqueurs discursifs, navigation dans le dialogue, corpus oral, entretiens d'embauche, vocaux

Résumé

« Pas de souci » est apparemment en train de devenir une expression assez invasive dans nos interactions langagières quotidiennes sans que l'ombre d'un « souci » particulier ne soit réellement visible, comme par exemple à la boulangerie quand vous demandez un pain de campagne et que l'on vous réponde « pas de souci ! ». Cette communication a ainsi pour objectif de se demander les raisons réelles de l'emploi de cette expression, qui se comporte dorénavant comme un « marqueur de discours » : pourquoi dit-on « pas de souci » et pas « oui », ou « je vous en prie », ou « ok » etc ? Est-ce qu'il y a un vrai « souci » dans le dialogue ou dans le discours, en dehors de la difficulté de dialoguer... ? Dans quels types de contexte est employé « pas de souci », et par qui ?

Afin de proposer des réponses à ces questions, nous avons observé deux jeux de données de taille fort différente mais proposant une présence du marqueur à peu près équivalente : un corpus d'entretiens d'embauche recueilli à Neuchâtel par J. Brosy (135 occ. pour 592.799 mots ; Brosy 2019), et le corpus « Les Vocaux » (53 occ. pour 272.266 mots, Glikman et al 2025). Ces données ont été annotées manuellement avec l'outil URS de la plate-forme TXM et nous focalisant sur les contextes gauche et droit, la présence de marqueurs discursifs immédiatement à gauche et à droite de « pas de souci », le changement ou non de tour de parole et/ou de locuteur, la place de « pas de souci » dans le tour de parole, et enfin le type et le rôle des locuteurs (uniquement pour les entretiens d'embauche).

Les résultats de ces observations nous conduisent à mettre en évidence deux motifs linguistiques différents, illustrant deux aspects de l'expression « pas de souci », et correspondant aux deux rôles locutifs dans l'interaction des entretiens d'embauche. Ainsi l'usage de « pas de souci » par le candidat se rapproche du rôle de « ok » et de son rôle de marquage de transition (Bangerter et al 2024) tandis que celui des recruteurs, chez qui on relève la présence de différents marqueurs discursifs indiquant la conclusion/déduction, l'usage de « pas de souci » est plutôt lié à une forme d'aboutissement avant de passer à autre chose.

Les données issues des Vocaux semblent confirmer ce qui a été observé pour les recruteurs, c'est-à-dire un discours se caractérisant essentiellement par des marques indiquant des formes d'aboutissement avant de procéder à des transitions.

Enfin, l'analyse des contextes situés à gauche de « pas de souci » révèle bien la présence de formes de difficultés, et ce autant dans les entretiens d'embauche que dans les vocaux.

Keywords: discourse markers; dialogue navigation; spoken corpus; job interviews; voice messages

Abstract

“*Pas de souci*” is visibly becoming a rather invasive expression in our everyday linguistic interactions, even when there is no actual “concern” to be seen—for instance, at the bakery when you ask for a loaf of country bread and the response is “*pas de souci !*”. This paper aims to examine the real reasons behind the use of this expression, which now behaves like a “discourse marker”: why do we say “*pas de souci*” rather than “*oui*”, “*je vous en prie*”, or “*ok*”? Is there really a “concern” in the dialogue or the discourse, aside from the difficulty of dialoguing itself? In which types of contexts is “*pas de souci*” used, and by whom?

To address these questions, we analyzed two datasets of very different sizes but with a roughly equivalent frequency of the marker: a corpus of job interviews collected in Neuchâtel by J. Brosy (135 occurrences for 592,799 words; Brosy 2019), and the “*Vocaux*” corpus (53 occurrences for 272,266 words; Glikman et al. 2025). These data were manually annotated using the URS tool on the TXM platform, with a focus on the left and right contexts: the presence of discourse markers immediately before and after “*pas de souci*”, whether or not a change of turn and/or speaker occurred, the position of “*pas de souci*” within the turn, and finally, the type and role of speakers (only for the job interviews).

The results of these observations highlight two different linguistic patterns, illustrating two facets of the expression “*pas de souci*” and corresponding to the two participant roles in job interview interactions. The candidates’ use of “*pas de souci*” is closer to the role of “*ok*”, marking a transition (Bangerter et al. 2024), while recruiters’ use – where we find co-occurring discourse markers of conclusion or deduction – is more often linked to a form of closure before moving on.

The *Vocaux* data appear to confirm what was observed for recruiters, namely a discourse characterized primarily by markers signaling closure before proceeding to a transition.

Finally, the analysis of contexts to the left of “*pas de souci*” clearly shows the presence of forms of difficulty, both in the job interview corpus and in the voice message corpus.

Références

Bangerter, A., Knutsen, D., Germanier, E., Col, G. et Brosy, J. (2024). Okay as a marker for coordinating transitions in joint actions: Effects of participant role and age in Swiss German and Swiss French interviews, *Journal of Pragmatics*, 230 : 166-186, <https://doi.org/10.1016/j.pragma.2024.07.006>.

(<https://www.sciencedirect.com/science/article/pii/S0378216624001371>)

Delahaie, J. et Col, G. (2024). Politeness and Discourse Markers, in: Hansen, M.-B. M. et Visconti, J. (eds), *Manual of Discourse Markers in Romance* Berlin, Boston: De Gruyter, pp. 319-350; <https://doi.org/10.1515/9783110711202-011>

Doshe, G. et Pusch, Claus D. (eds) (2007). *Les marqueurs discursifs. Sens et variation*, *Langue française* 154.

Fischer, K. (ed.) (2006). *Approaches to discourse particles*, Oxford : Elsevier.

Fraser, B. (1996). Pragmatic markers, *Pragmatics* 6 : 167–190.

Glikman, J., Mazziotta, N., Benzitoun, C. et al. (2025). Les Vaux [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage) - www.ortolang.fr, v0.0.2, <https://hdl.handle.net/11403/lesvoux/v0.0.2>.

Hansen, M.-B. M. (2006). A dynamic polysemy approach to the lexical semantics of discourse markers (with an exemplary analysis of French “*toujours*”), in: Kerstin Fischer (ed.), *Approaches to discourse particles*, Leiden : Brill, pp. 21–41.

Schiffrin, D. (1987). *Discourse Markers*. Cambridge : Cambridge University Press

Rebecca Connor-Evans (Sheffield Hallam University)

When A Monster Calls: reading as catharsis for parentally bereaved adolescents

Keywords: Text World Theory; Cognitive Stylistics; Reader Response; bereavement; Young Adult Fiction; *A Monster Calls*

Abstract

This paper examines the potential for cathartic reading experiences in bereaved adolescents through an interdisciplinary analysis of Patrick Ness's *A Monster Calls* (2011). Drawing on bibliotherapy, cognitive stylistics, and adolescent health research, the study explores how fiction can support young people in navigating parental loss. The research addresses the following questions: (1) How does *A Monster Calls* construct emotional experiences that resonate with bereaved adolescents? (2) In what ways does the novel's narrative structure, including its use of parable and intrusive fantasy, create space for emotional reflection and catharsis? (3) How do naturalistic reader responses illustrate the novel's impact on readers who have experienced bereavement?

A Monster Calls follows 13-year-old Conor O'Malley, whose mother is terminally ill with cancer. Conor's struggle with denial manifests as a recurring nightmare in which he inadvertently lets his mother fall to her death. Overwhelmed by guilt, Conor summons a monster—an anthropomorphised yew tree—who tells him three nested parablesque stories that ultimately guide Conor to confront his grief.

By employing Text World Theory (Gavins, 2007; Werth, 1999) to analyse key passages triangulated with reader responses gathered from online communities (Peplow & Carter, 2014), this paper investigates how the novel's cognitive and narrative features facilitate a cathartic experience for bereaved adolescents.

The paper concludes by discussing the broader implications of this research for educators, counsellors, and mental health practitioners. It argues that fictional narratives like *A Monster Calls* can offer a meaningful way for bereaved adolescents to navigate their emotions, particularly when paired with supportive reading environments. This study contributes to ongoing discussions about the role of literature in bereavement support and how language and narrative structure shape emotional experiences in young readers.

Mots-clés : Text World Theory, Stylistique cognitive, Réponse du lecteur, deuil, Fiction pour jeunes adultes, *A Monster Calls*

Résumé

Cet article examine le potentiel des expériences de lecture cathartique chez les adolescents endeuillés à travers une analyse interdisciplinaire du roman *A Monster Calls* de Patrick Ness (2011). S'appuyant sur la bibliothérapie, la stylistique cognitive et la recherche sur la santé des adolescents, cette étude explore comment la fiction peut soutenir les jeunes confrontés à la perte parentale. La recherche aborde les questions suivantes : (1) Comment *A Monster Calls* construit-il des expériences émotionnelles qui résonnent avec les adolescents endeuillés ? (2) De quelle manière la structure narrative du roman, y compris son usage de la parabole et du fantastique intrusif, crée-t-elle un espace pour la réflexion émotionnelle et la catharsis ? (3) Comment les

réponses naturalistes des lecteurs illustrent-elles l'impact du roman sur les lecteurs ayant vécu un deuil ?

A Monster Calls suit Conor O'Malley, 13 ans, dont la mère est en phase terminale d'un cancer. La lutte de Conor contre le déni se manifeste par un cauchemar récurrent dans lequel il laisse involontairement sa mère tomber vers sa mort. Accablé par la culpabilité, Conor invoque un monstre — un if anthropomorphe — qui lui raconte trois histoires paraboliques emboîtées qui guident finalement Conor vers une confrontation avec son chagrin.

En employant la Text World Theory (Gavins, 2007 ; Werth, 1999) pour analyser des passages clés triangulés avec des réponses de lecteurs recueillies auprès de communautés en ligne (Peplow & Carter, 2014), cet article examine comment les caractéristiques cognitives et narratives du roman facilitent une expérience cathartique pour les adolescents endeuillés.

L'article conclut en discutant les implications plus larges de cette recherche pour les éducateurs, conseillers et praticiens de la santé mentale. Il soutient que les récits fictionnels comme *A Monster Calls* peuvent offrir un moyen significatif aux adolescents endeuillés de naviguer leurs émotions, particulièrement lorsqu'ils sont associés à des environnements de lecture de soutien. Cette étude contribue aux discussions en cours sur le rôle de la littérature dans l'accompagnement du deuil et sur la façon dont le langage et la structure narrative façonnent les expériences émotionnelles chez les jeunes lecteurs.

Références

Gavins, J. (2007). *Text World Theory*. Edinburgh: Edinburgh University Press.

Hillis, S., N'konzi, J. P. N., Msemburi, W., Cluver, L., Villaveces, A., Flaxman, S., & Unwin, H. J. T. (2022). Orphanhood and caregiver loss among children based on new global excess COVID-19 death estimates. *JAMA pediatrics*, 176(11), 1145–1148.

Ness, P. (2011). *A Monster Calls*. Illustrated ed., London: Walker.

Parsons, S. (2011). *Long-term Impact of Childhood Bereavement: Preliminary Analysis of the 1970 British Cohort Study (BCS70)*. London: Child Well-being Research Centre.

Peplow, D. & Carter, R. (2014). Stylistics and real readers: In: Burke, M. (ed.) *The Routledge Handbook of Stylistics*. Abingdon: Routledge, 440–454.

Werth, P. (1999). *Text Worlds: Representing conceptual space in discourse*. New York Pearson Education Limited.

Mireille Copin & Inès Saddour (University of Toulouse Jean Jaurès)

Revisiting linguistic relativity: a comparison of linguistic and non-linguistic data in L2 French

Keywords: Conceptualization ; French L2 ; Causal event ; Segmentation ; Verbalization

Abstract

To observe both parts of the “thinking for speaking” hypothesis (Slobin 1996), research on linguistic relativity needs to investigate both linguistic and non-linguistic data (Athanasopoulos & Bylund, 2013). One non-linguistic task used to compare the conceptualization and the verbalization of events is the segmentation task (Gerwien & von Stutterheim, 2018), which provides insights to the number of events perceived and whether participants’ L1 or sociocultural backgrounds influence segmentation decisions. This study aims to investigate the conceptualization and verbalization of causal events in L2 French using two experimental tasks: a non-linguistic segmentation task and a verbalization one; with dynamic stimuli featuring a causer hitting an affectee with an object. The situations vary according to two conditions: contiguity between causing and resulting subevents (manipulating the directness of causation) and cultural context (protagonists with Oriental vs Occidental appearance). Participants were an L2 French group of Syrian Arabic (SA) speakers living in France (n=20), and an L1 French control group (n=22).

It has been shown that directness of causation affects the number of events conceptualized (Bellingham et al., 2020), while cultural context and orientation might lead to different segmentation preferences (Swallow & Wang, 2020). Thus, we expect more events conceptualized in the less direct condition (non-contiguous) and in the Occidental context; but also in the control group (because an individualistic orientation leads to more analytical cognitive styles (Nisbett et al., 2001)). Moreover, if event conceptualization is mediated by linguistic knowledge, the existence of more synthetic linguistic constructions in SA might lead to less event conceptualized by the learners. To account for other variables influencing verbalization, French proficiency is self-assessed at two time points (mean score = 4.78/7 at T1, 5.13 at T2; mean interval = 9 months). Participants repeated the two tasks at both time points. We compared the number of subevents identified in the segmentation task (number of clicks) and the number of utterances produced in the verbalization task, to investigate the (evolving) relationship between conceptual and linguistic events.

Generalized logistic models reveal that contiguity (but not context) affected both the likelihood of producing multiple utterances (fig. 1) and identifying multiple events in the segmentation task (fig. 2), suggesting a relationship between verbalization and conceptualization. However, the number of utterances produced in L2 French correlates with self-assessed proficiency scores, suggesting that linguistic output may be limited by proficiency, and does not fully reflect event conceptualization. Learners produced more utterances at T2, and for some of them, clicked more in the segmentation task – which reflects the improvement of proficiency scores over time. However, no direct correlation was found between the number of clicks and the number of utterances, paving the way for further research to investigate which factors affect segmentation decisions and the complex links between conceptualization and verbalization.

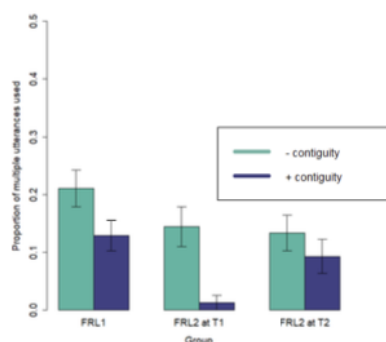


Figure 1: Proportion of descriptions containing multiple utterances to describe the collision, depending on the contiguity condition

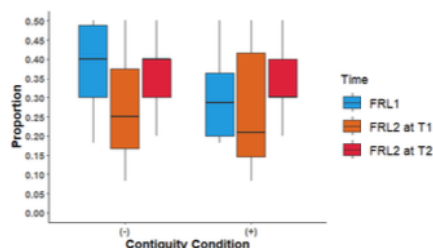


Figure 2: Proportion of videos with multiple clicks recorded during the segmentation tasks, depending on the contiguity condition

Mots-clés : Conceptualisation ; Français L2 ; Événement causal ; Segmentation ; Verbalisation

Résumé

Pour observer les deux volets de l'hypothèse « penser pour parler » (Slobin 1996), la recherche sur la relativité linguistique doit examiner à la fois les données linguistiques et non-linguistiques (Athanasopoulos & Bylund, 2013). Une tâche non-linguistique utilisée pour comparer la conceptualisation et la verbalisation d'événements est la tâche de segmentation (Gerwien & von Stutterheim, 2018), qui fournit des éclairages sur le nombre d'événements perçus et sur l'influence du L1 ou du contexte socioculturel des participants sur les décisions de segmentation. Cette étude vise à examiner la conceptualisation et la verbalisation d'événements causaux en français L2 en utilisant deux tâches expérimentales : une tâche non-linguistique de segmentation et une tâche de verbalisation à un seul essai, avec des stimuli dynamiques présentant un agent frappant un objet affecté. Les situations varient selon deux conditions : la contiguïté entre les sous-événements causants et résultants (manipulant la directivité de la causation) et le contexte culturel (protagonistes d'apparence orientale vs occidentale). Les participants étaient un groupe L2 français de locuteurs arabes syriens (AS) vivant en France (n=20), et un groupe contrôle L1 français (n=22).

Il a été montré que la directivité de la causation affecte le nombre d'événements conceptualisés (Bellingham et al., 2020), tandis que le contexte culturel et l'orientation peuvent conduire à différentes préférences de segmentation (Swallow & Wang, 2020). Ainsi, nous nous attendons à plus d'événements conceptualisés dans la condition moins directe (non-contiguë) et dans le

contexte occidental, mais aussi dans le groupe contrôle (car une orientation individualiste conduit à des styles cognitifs plus analytiques (Nisbett et al., 2001)). De plus, si la conceptualisation d'événements est médiée par la connaissance linguistique, l'existence de constructions plus synthétiques en AS pourrait conduire à moins d'événements conceptualisés par les apprenants. Pour tenir compte d'autres variables influençant la verbalisation, la compétence en français est auto-évaluée à deux moments (score moyen = 4,78/7 à T1, 5,13 à T2 ; intervalle moyen = 9 mois). Les participants ont répété les deux tâches aux deux moments. Nous avons comparé le nombre de sous-événements identifiés dans la tâche de segmentation (nombre de clics) et le nombre d'énoncés produits dans la tâche de verbalisation, pour examiner la relation (évolutive) entre événements conceptuels et linguistiques.

Les modèles logistiques généralisés révèlent que la contiguïté (mais pas le contexte) a affecté à la fois la probabilité de produire plusieurs énoncés (fig. 1) et d'identifier plusieurs événements dans la tâche de segmentation (fig. 2), suggérant une relation entre verbalisation et conceptualisation. Cependant, le nombre d'énoncés produits en français L2 corrèle avec les scores de compétence auto-évalués, suggérant que la production linguistique peut être limitée par la compétence, et ne reflète donc pas pleinement la conceptualisation d'événements. Les apprenants ont produit plus d'énoncés à T2, et pour certains d'entre eux, ont cliqué davantage dans la tâche de segmentation – ce qui reflète l'amélioration des scores de compétence au fil du temps. Cependant, aucune corrélation directe n'a été trouvée entre le nombre de clics et le nombre d'énoncés, ouvrant la voie à de futures recherches pour examiner quels facteurs affectent les décisions de segmentation et les liens complexes entre conceptualisation et verbalisation.

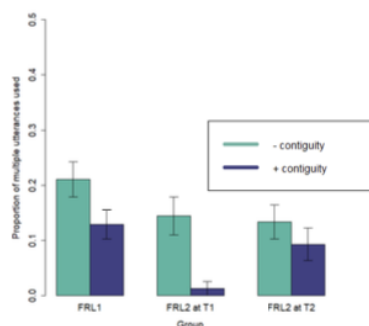


Figure 1: Proportion of descriptions containing multiple utterances to describe the collision, depending on the contiguity condition

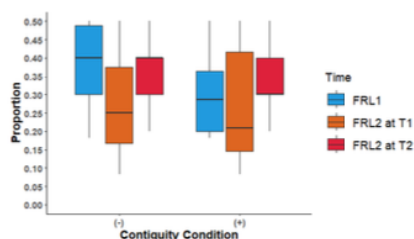


Figure 2: Proportion of videos with multiple clicks recorded during the segmentation tasks, depending on the contiguity condition

Références

- Athanasopoulos, P., & Bylund, E. (2013). The 'thinking' in thinking-for-speaking : Where is it? *Language, Interaction and Acquisition*, 4(1), 91-100. <https://doi.org/10.1075/lia.4.1.05ath>
- Gerwien, J., & von Stutterheim, C. (2018). Event segmentation : Cross-linguistic differences in verbal and non-verbal tasks. *Cognition*, 180, 225-237.
- Slobin, D. I. (1996). From « Thought and Language » to « Thinking for Speaking ». In J. Gumperz & S. Levinson (Éds.), *Rethinking Linguistic Relativity* (p. 70-96). Cambridge University Press.
- Bellingham, E., Evers, S., Kawachi, K., Mitchell, A., Park, S.-H., Stepanova, A., & Bohnemeyer, J. (2020). Exploring the Representation of Causality Across Languages : Integrating Production, Comprehension and Conceptualization Perspectives. In E. A. Bar-Asher Siegal & N. Boneh (Éds.), *Linguistic Perspectives on Causation* (p. 75-119). Springer. https://doi.org/10.1007/978-3-030-34308-8_3
- Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought : Holistic versus analytic cognition. *Psychological review*, 108(2), 291-310.
- Swallow, K. M., & Wang, Q. (2020). Culture influences how people divide continuous sensory experience into events. *Cognition*, 205, Article 104450. <https://doi.org/10.1016/j.cognition.2020.104450>

Gauthier Delaby, Timothy Coleman & Robert Hartsuiker
(Ghent University)

Contamination constructionnelle entre constructions formellement similaires

Preuves issues d'une expérience d'amorçage

Keywords : construction grammar; horizontal links; constructional contamination; structural priming; receptive construction; Dutch

Abstract

Pijpops & Van de Velde (2016) describe how an alternation within a construction can be influenced by neighbouring constructions with a similar form via a process they call constructional contamination. For example, in English passive constructions an adverb can either precede or follow the main verb (1). Hilpert & Flach (2022) show that this variation is partly determined by the frequency of an adverb-participle combination in a formally similar construction that only allows the order with the adverb first, viz. the noun phrase construction (2). The more an adverb and participle are used in the contaminating noun phrase construction, the more likely it is that they will appear in the order adverb-participle in the passive construction.

(1) *Stainless steels are [extensively used/ used extensively].*

(2) *an [extensively used/ *used extensively] technique.*

Such contamination effects have been attributed to longer-term effects of chunking (Pijpops & Van de Velde 2016): if a language user often encounters the string *extensively used* in the noun phrase construction, they will store it as a chunk, which can later be recycled in the target construction, thus creating a bias for the order adverb-participle. However, in earlier work, we have argued, on the basis of corpus data, that horizontal links contribute to constructional contamination, too: if a potentially contaminating construction is activated in the preceding context, this can bias language users producing an instance of the target construction towards choosing an alternant that is exclusively used in the contaminating construction (Delaby & Coleman accepted).

This paper presents the results of a priming experiment which tests the hypothesis that horizontal links can directly trigger constructional contamination. We focus on the Dutch receptive construction with *krijgen* 'to get', in which the agent prepositional phrase can appear either with the preposition *van* 'from/by' or *door* 'through/by' (3). 306 native speakers of Dutch were presented with sentences with a receptive construction which they had to complete, based on a accompanying picture. We tested whether the variation between *van* and *door* is influenced by prior activation of the regular passive construction with *worden/zijn* 'to be', in which the agent PP can only be expressed with the preposition *door* (4). Prime sentences were presented as maze tasks.

(3) *Hij krijgt boeken overhandigd [door/ van] zijn oma.*

'He is handed books by his grandmother.' (lit. 'got handed')

(4) *Boeken worden overhandigd aan hem [door/ *van] zijn oma.*

'Books are handed to him by his grandmother.'

Participants who were primed with a passive construction with an agent PP with the preposition *door* (4), were significantly more likely to use *door* for the agent PP in the receptive construction. Crucially, this effect was also found when participants were primed with a passive construction in which the agent PP was not expressed. This indicates that we are dealing with priming on the constructional level, and not with the mere repetition of a lexical item or abstract PP. Our results provide further evidence for the existence of horizontal links and for the possibility of such links directly triggering constructional contamination.

Mots-clés : grammaire constructionnelle ; liens horizontaux ; contamination constructionnelle ; amorçage structural ; construction réceptive ; néerlandais

Résumé

Pijpops & Van de Velde (2016) décrivent comment une alternance au sein d'une construction peut être influencée par des constructions voisines ayant une forme similaire via un processus qu'ils appellent *contamination constructionnelle*. Par exemple, dans les constructions passives anglaises, un adverbe peut soit précéder soit suivre le verbe principal (1). Hilpert & Flach (2022) montrent que cette variation est en partie déterminée par la fréquence d'une combinaison adverbe-participe dans une construction formellement similaire qui ne permet que l'ordre avec l'adverbe en premier, à savoir la construction syntagme nominal (2). Plus un adverbe et un participe sont utilisés dans la construction syntagme nominal contaminante, plus il est probable qu'ils apparaissent dans l'ordre adverbe-participe dans la construction passive.

(1) *Stainless steels are [extensively used/ used extensively].*

(2) *an [extensively used/ used extensively] technique

De tels effets de contamination ont été attribués aux effets à long terme du regroupement en blocs (*chunking*) (Pijpops & Van de Velde 2016) : si un utilisateur de la langue rencontre souvent la chaîne *extensively used* dans la construction syntagme nominal, il la stockera comme un bloc, qui peut ensuite être recyclé dans la construction cible, créant ainsi un biais pour l'ordre adverbe-participe. Cependant, dans des travaux antérieurs, nous avons soutenu, sur la base de données de corpus, que les liens horizontaux contribuent à la contamination constructionnelle, c'est-à-dire que si une construction potentiellement contaminante est activée dans le contexte précédent, cela peut biaiser les utilisateurs de la langue produisant une instance de la construction cible vers le choix d'un alternant qui est exclusivement utilisé dans la construction contaminante (Delaby & Coleman accepté).

Cet article présente les résultats d'une expérience d'amorçage qui teste l'hypothèse selon laquelle les liens horizontaux peuvent directement déclencher une contamination constructionnelle. Nous nous concentrons sur la construction réceptive néerlandaise avec *krijgen* « obtenir », dans laquelle le syntagme prépositionnel agent peut apparaître soit avec la préposition *van* « de/par » soit *door* « par/à travers » (3). 306 locuteurs natifs du néerlandais ont été présentés avec des phrases contenant une construction réceptive qu'ils devaient compléter, basée sur une image d'accompagnement. Nous avons testé si la variation entre *van* et *door* est influencée par l'activation préalable de la construction passive régulière avec *worden/zijn* « être », dans laquelle le SP agent ne peut être exprimé qu'avec la préposition *door* (4). Les phrases d'amorçage ont été présentées sous forme de tâches de labyrinthe.

(3) *Hij krijgt boeken overhandigd [door/ van] zijn oma.*

« Il se fait remettre des livres par sa grand-mère. » (litt. « il reçoit des livres remis »)

(4) **Boeken worden overhandigd aan hem [door/ van] zijn oma.*

« Des livres lui sont remis par sa grand-mère. »

Les participants qui ont été amorcés avec une construction passive avec un SP agent avec la préposition *door* (4), étaient significativement plus susceptibles d'utiliser *door* pour le SP agent dans la construction réceptive. De manière cruciale, cet effet a également été trouvé lorsque les participants ont été amorcés avec une construction passive dans laquelle le SP agent n'était pas exprimé. Cela indique que nous avons affaire à un amorçage au niveau constructionnel, et non à la simple répétition d'un élément lexical ou d'un SP abstrait. Nos résultats fournissent des preuves supplémentaires de l'existence de liens horizontaux et de la possibilité que de tels liens déclenchent directement une contamination constructionnelle.

Références

Delaby, G., & Coleman, T. (Accepted). Constructional contamination between two constructions with *krijgen* 'to get' in Dutch: national variation, bidirectionality and ambiguity avoidance. Accepted for publication in *Constructions and Frames*.

Hilpert, M., & Flach, S. (2022). A case of constructional contamination in English: Modified noun phrases influence adverb placement in the passive. In K. Krawczak, B. Lewandowska-Tomaszczyk, & M. Grygiel (Eds.), *Human Cognitive Processing* (Vol. 73, pp. 283–302). John Benjamins Publishing Company. <https://doi.org/10.1075/hcp.73.09hil>

Pijpops, D., & Van De Velde, F. (2016). Constructional contamination: How does it work and how do we measure it? *Folia Linguistica*, 50(2), 543–581. <https://doi.org/10.1515/flin-2016-0020>

Ruyue Deng (Laboratoire EDA, EA 4071, Université Paris Cité), Huiyun Hu (STL, Université de Lille) & Yanjing Bi (Pékin Chine Capital University of Economics and Business)

Enseignement des collocations en classe de FLE : difficultés, problèmes et perspectives

Mots-clés : collocation, pratique enseignante, didactique, FLE, phraséodidactique

Résumé

La collocation, unité phraséologique définie comme une « cooccurrence lexicale privilégiée de deux éléments linguistiques » (Tutin & Grossmann, 2002 : 2), représente plus de la moitié de nos discours (Lewis, 2000). Elle suscite un intérêt croissant parmi les didacticiens qui prennent progressivement conscience de son importance dans l'apprentissage des langues étrangères (Nesselhauf, 2003 ; González Rey, 2007 ; Cavalla, 2010, 2019). Par ailleurs, la collocation est étudiée sous différents angles par différents domaines. La phraséologie descriptive s'attache à décrire ses spécificités sémantico-syntaxiques et met en évidence des traits distinctifs tels que la binarité et la non-prédictibilité (Tutin & Grossmann, 2002), la dissymétrie sémantique (Hausmann, 1989) et la cooccurrence restreinte (Mel'čuk, 1993). Des études psycholinguistique (Wray, 2002 ; Edmonds, 2013) montrent, quant à elles, que cette unité phraséologique est reconnue et traitée plus rapidement que les constructions libres. Son acquisition serait influencée par divers facteurs, comme l'interférence de la L1 (Paquot & Granger, 2012), le degré de prise de conscience (Théophanous, 2021) et la fréquence d'exposition (Ellis, 2002).

Ces notions théoriques sont ensuite adaptées par les didacticiens afin d'optimiser leur enseignement en FLE (Cavalla et al., 2009 ; Cavalla & Legallois, 2020), ce qui a donné lieu à plusieurs dispositifs pédagogiques (González Rey, 2007, 2019 ; Cavalla, 2015 ; Yan et al., 2018). Cependant, malgré cette richesse théorique, la collocation demeure un objet marginalisé en classe de FLE, et les pratiques réelles d'enseignement restent peu documentées. Il devient dès lors indispensable de repenser l'articulation entre théorie et pratique, en interrogeant la transposabilité de ces savoirs en contexte didactique.

Notre étude vise à éclairer les pratiques d'enseignement des collocations en classe de FLE et à proposer des perspectives didactiques en tenant compte des défis rencontrés par les enseignants. L'analyse repose sur un corpus de 34 séances de FLE, observées durant deux mois, dispensées par cinq enseignants formés à l'enseignement des collocations à des apprenants allophones de niveaux A2 et B1.

Notre analyse s'articule en deux phases complémentaires. La première consiste à identifier les extraits de séances consacrés aux collocations, qu'ils soient abordés d'une approche explicite ou implicite, puis à les analyser afin d'évaluer dans quelle mesure les pratiques enseignantes intègrent les apports de la linguistique et de la psycholinguistique. Sur le plan linguistique, nous portons une attention particulière sur les explications métalinguistiques fournies par les enseignants, notamment celles relatives aux spécificités syntaxiques et sémantiques des collocations. D'un point de vue psycholinguistique, nous évaluons la prise en compte par les enseignants des facteurs favorisant l'acquisition des collocations, telle que la fréquence d'exposition ainsi que la gestion des interférences linguistiques. La seconde étape de l'analyse

visé à évaluer l'efficacité des pratiques enseignantes à partir des performances des apprenants, notamment par l'analyse de l'usage des collocations dans leurs productions écrites.

Les résultats révèlent que les apports des recherches en linguistique et en psycholinguistique sont encore peu intégrés dans l'enseignement des collocations. Environ 80% des pratiques observées présentent des limites récurrentes, parmi lesquelles une explication sémantique absente ou insuffisante, l'omission de certains traits syntaxiques, ou encore la non prise en compte de la langue maternelle des apprenants. De plus, les productions écrites de ces derniers témoignent d'un enseignement encore peu structuré et globalement peu efficace. Ces constats appellent à un renouvellement des approches didactiques, en faveur d'une meilleure articulation entre théorie et pratique, pour rendre l'enseignement plus structuré et plus rigoureux des collocations en classe de FLE.

Keywords: collocation; teaching practice; didactics; FLE; phraseodidactics

Abstract

Collocation, a phraseological unit defined as a *“privileged lexical co-occurrence of two linguistic elements”* (Tutin & Grossmann, 2002: 2), accounts for more than half of our discourse (Lewis, 2000). It has been attracting growing interest among didacticians, who are gradually becoming aware of its importance in foreign language learning (Nesselhauf, 2003; González Rey, 2007; Cavalla, 2010, 2019). Moreover, collocation is studied from different angles across various fields. Descriptive phraseology focuses on its semantic-syntactic specificities and highlights distinctive features such as binarity and unpredictability (Tutin & Grossmann, 2002), semantic asymmetry (Hausmann, 1989), and restricted co-occurrence (Mel'čuk, 1993). Psycholinguistic studies (Wray, 2002; Edmonds, 2013) show that this phraseological unit is recognized and processed more quickly than free combinations. Its acquisition appears to be influenced by several factors, such as L1 interference (Paquot & Granger, 2012), degree of awareness (Théophanous, 2021), and frequency of exposure (Ellis, 2002).

These theoretical notions are then adapted by didacticians in order to optimize their teaching in French as a Foreign Language (FLE) (Cavalla et al., 2009; Cavalla & Legallois, 2020), which has led to the development of several pedagogical approaches (González Rey, 2007, 2019; Cavalla, 2015; Yan et al., 2018). However, despite this theoretical richness, collocation remains marginalized in FLE classrooms, and actual teaching practices remain under-documented. It therefore becomes essential to rethink the relationship between theory and practice by questioning the transferability of this knowledge to teaching contexts.

Our study aims to shed light on the teaching practices of collocations in FLE classrooms and to propose didactic perspectives that take into account the challenges faced by teachers. The analysis is based on a corpus of 34 FLE lessons, observed over two months, delivered by five teachers trained in teaching collocations to allophone learners at A2 and B1 levels.

Our analysis unfolds in two complementary phases. The first consists in identifying lesson excerpts devoted to collocations, whether approached explicitly or implicitly, and then analyzing them to assess the extent to which teaching practices integrate contributions from linguistics and psycholinguistics. From a linguistic perspective, we focus on the metalinguistic explanations provided by teachers, particularly those related to the syntactic and semantic specificities of

collocations. From a psycholinguistic perspective, we evaluate teachers' consideration of factors that facilitate collocation acquisition, such as frequency of exposure and management of language interference. The second stage of the analysis aims to evaluate the effectiveness of teaching practices based on learners' performance, particularly through the analysis of collocation use in their written productions.

The results reveal that insights from linguistic and psycholinguistic research are still only minimally integrated into the teaching of collocations. About 80% of the observed practices show recurring shortcomings, including absent or insufficient semantic explanation, omission of certain syntactic features, and lack of consideration for learners' mother tongue. Moreover, learners' written productions reflect teaching that is still insufficiently structured and generally ineffective. These findings call for a renewal of didactic approaches, favoring a stronger connection between theory and practice, in order to make the teaching of collocations in FLE classrooms more structured and rigorous.

Références

- Cavalla, C. (2010). Propositions didactiques pour l'enseignement d'éléments phraséologiques en FLE. In Galatanu, O., Pierrard, M., Van Raemdonk, D. & Damar, M-E. (eds.), *Enseigner les structures langagières en FLE*. Peter Lang AG, pp.147-158.
- Cavalla, C. (2015). Collocations transdisciplinaires dans les écrits de doctorants FLS/FLE. *Linx*, 72, pp.95-110.
- Cavalla, C. (2019). Une entrée lexicale spécifiques : les collocations. In : Perassi, M., L., & Kwiecien, M., T. (eds.), *Palabras como puentes : estudios lexicológicos, lexicográficos, y terminológicos desde el Cono Sur*. Buena Vista Editora, pp.58-67.
- Cavalla, C., Crozier, E., Dumarest, D. & Richou, C. (2009). *Le vocabulaire en classe de langue*. Paris : CLE international.
- Cavalla, C., Legallois D. (2020). Caractériser et identifier les unités phraséologiques pour leur enseignement. *Action didactique*, 6, 12-30.
- Edmonds, E. (2013). Une approche psycholinguistique des phénomènes phraséologiques : le cas des expressions conventionnelles. *Langages*, n°189, pp. 121-138.
- 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, pp. 143-188.
- González Rey, I. (2007). *La didactique du français idiomatique*. Fernelmont : E.M.E.
- González Rey, I. (2019). *Le processus de conscientisation dans la phraséodidactique d'une L2*, Repères- DoRiF, n.18.
- Hausmann, F. J. (1989). Le dictionnaire de collocations. In : Hausmann F.J., Reichmann O., Wiegand H.E. & Zgusta L. (eds.), *Wörterbücher : ein internationales Handbuch zur Lexicographie. Dictionnaires*. Berlin, De Gruyter, pp. 1010-1019.
- Lewis, M. (2000). *Teaching collocation: Further developments in the lexical approach*. Hove: Language teaching publications LTP.

- Mel'čuk, I. (1993). La phraséologie et son rôle dans l'enseignement/apprentissage d'une langue étrangère. *Revue de Linguistique appliquée*, 92, pp. 82-113.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24/2, pp. 223-242.
- Paquot, M., Granger, S. (2012). Formulaic language in learner corpora. *Annual review of applied linguistics*, 32, 130-149.
- Théophanous, O. (2021). Apprendre les collocations : quels facteurs sont en jeu ? *Les Langues Modernes*, n° 3, pp. 64-72.
- Tutin, A., Grossmann, F. (2002). Collocation régulières et irrégulières : esquisse de typologie du phénomène collocatif, *Français de Linguistique Appliquée*, VIII-1, 7-26.
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge : CUP.
- Yan, R., Tran, T., T., H., Cavalla, C. (2018). Un corpus pour l'enseignement des collocations en FLE. *Etudes en didactique des langues, Les corpus / Corpora*, 31, pp.93-107.

Marie Dewulf (Ghent University)

Evaluating syntactic generalization in low-resource language models through targeted test suites

Keywords: syntactic evaluation; French; surprisal; language models; morphosyntax; test suites

Abstract

Low- and medium-resource language scenarios pose significant challenges for the development and evaluation of neural language models (LMs), particularly in capturing syntactic generalizations beyond surface-level regularities. Specifically, training corpora are predominantly in English, and existing evaluation methods may be biased towards English, potentially overlooking limitations in multilingual linguistic knowledge of models (cf. the works cited in this abstract). This work proposes an evaluation framework inspired by Warstadt et al. (2020) to systematically assess syntactic competence in French, a morphosyntactically rich and medium-resource language in the context of large-scale model pretraining. Building on targeted syntactic evaluation methodologies (Marvin & Linzen, 2018; Futrell et al., 2019), I develop diagnostic test suites that assess models' sensitivity to morphosyntactic and syntactic phenomena in French, extending existing English-focused resources.

My research aims to answer the following questions: (1) To what extent do multilingual and French-specific language models acquire abstract syntactic representations in a medium-resource training context? (2) How consistent are these representations across morphosyntactic and syntactic features? (3) Does morphosyntactic competence primarily depend on model size or training data?

I constructed a benchmark of 500 minimal pairs based on representative examples in grammar books specifically targeted towards English-speaking L2 learners of French. These minimal pairs contrast grammatical and ungrammatical sentences and cover multiple syntactic phenomena such as agreement (within verb and noun phrases), argument structure (e.g., direct vs. indirect object), tense and mood, question formation, negation and preposition use. For example, in the subject-verb agreement suite, verbs are varied to match or mismatch with the subject in person and number (e.g., *Tu travailles* vs. **Tu travaillent*). Each test item appears in controlled conditions that isolate syntactic dependencies. Surprisal values, calculated as the negative log probability of a token given the preceding context, are computed at the sentence-level. The syntactic sensitivity of the model is quantified by the extent to which predicted surprisal differences match theoretical expectations: grammatical sentences should be less surprising than ungrammatical ones.

Results show that CroissantLLMBase (Faysse et al., 2024), a smaller bilingual model trained on balanced English-French datasets, outperforms larger multilingual models such as Llama-2 (7B and 13B) (Touvron et al., 2023) on several syntactic dimensions, particularly argument structure, negation and preposition usage. This demonstrates that linguistically accurate behaviour in LLMs is not solely determined by model size or architecture but rather depends on the quantity and linguistic diversity of the training data.

This approach contributes to linguistic research in medium-resource settings by providing a principled framework for assessing the capture of grammatical knowledge in LMs beyond English. It enables model comparison across architectures and training regimes and offers insights into

the developmental trajectory of syntactic competence in models with constrained exposure. Furthermore, the test suite design is extensible to other typologically diverse languages, supporting broader efforts in cross-linguistic evaluation under data-scarce conditions. The methodology presented, which leverages grammar books to source representative sentences for language-specific phenomena and transforms them into a minimal-pair benchmark, can be adopted for other languages.

Mots-clés: évaluation syntaxique ; français ; surprise ; modèles linguistiques ; morphosyntaxe ; suites de tests

Résumé

Les scénarios linguistiques à ressources faibles et moyennes posent des défis importants pour le développement et l'évaluation des modèles linguistiques neuronaux (LM), en particulier pour saisir les généralisations syntaxiques au-delà des régularités superficielles. Plus précisément, les corpus d'entraînement sont principalement en anglais, et les méthodes d'évaluation existantes peuvent être biaisées en faveur de l'anglais, ce qui peut conduire à négliger les limites des connaissances linguistiques multilingues des modèles (cf. les travaux cités dans ce résumé). Ce travail propose un cadre d'évaluation inspiré de Warstadt et al. (2020) pour évaluer systématiquement la compétence syntaxique en français, une langue riche sur le plan morphosyntaxique et disposant de ressources moyennes dans le contexte du pré-entraînement de modèles à grande échelle. En m'appuyant sur des méthodologies d'évaluation syntaxique ciblées (Marvin & Linzen, 2018 ; Futrell et al., 2019), je développe des suites de tests diagnostiques qui évaluent la sensibilité des modèles aux phénomènes morphosyntaxiques et syntaxiques en français, élargissant ainsi les ressources existantes axées sur l'anglais.

Ma recherche vise à répondre aux questions suivantes : (1) Dans quelle mesure les modèles linguistiques multilingues et spécifiques au français acquièrent-ils des représentations syntaxiques abstraites dans un contexte d'entraînement à ressources moyennes? (2) Dans quelle mesure ces représentations sont-elles cohérentes entre les caractéristiques morphosyntaxiques et syntaxiques ? (3) La compétence morphosyntaxique dépend-elle principalement de la taille du modèle ou des données d'entraînement ?

J'ai construit un benchmark de 500 paires minimales à partir d'exemples représentatifs tirés de manuels de grammaire spécifiquement destinés aux apprenants anglophones de français en L2. Ces paires minimales opposent des phrases grammaticales et non grammaticales et couvrent plusieurs phénomènes syntaxiques tels que l'accord (au sein des groupes verbaux et nominaux), la structure argumentale (par exemple, complément d'objet direct vs indirect), le temps et le mode, la formation des questions, la négation et l'utilisation des prépositions. Par exemple, dans la série sur l'accord sujet-verbe, les verbes sont variés pour correspondre ou ne pas correspondre au sujet en personne et en nombre (par exemple, *Tu travailles* vs **Tu travaillent*). Chaque élément du test apparaît dans des conditions contrôlées qui isolent les dépendances syntaxiques. Les valeurs de surprise, calculées comme la probabilité logarithmique négative d'un token donné dans le contexte précédent, sont calculées au niveau de la phrase. La sensibilité syntaxique du modèle est quantifiée par la mesure dans laquelle les différences de surprise prédites correspondent aux attentes théoriques : les phrases grammaticales devraient être moins surprenantes que les phrases non grammaticales.

Les résultats montrent que CroissantLLMBase (Faysse et al., 2024), un modèle bilingue plus petit entraîné sur des ensembles de données équilibrés anglais-français, surpasse les modèles multilingues plus grands tels que Llama-2 (7B et 13B) (Touvron et al., 2023) sur plusieurs dimensions syntaxiques, en particulier la structure des arguments, la négation et l'utilisation des prépositions. Cela démontre que le comportement linguistique précis des LLM n'est pas uniquement déterminé par la taille ou l'architecture du modèle, mais dépend plutôt de la quantité et de la diversité linguistique des données d'entraînement.

Cette approche contribue à la recherche linguistique dans des environnements à ressources moyennes en fournissant un cadre théorique pour évaluer la capture des connaissances grammaticales dans les modèles linguistiques au-delà de l'anglais. Elle permet de comparer les modèles entre différentes architectures et différents régimes d'entraînement et offre des informations sur la trajectoire de développement de la compétence syntaxique dans les modèles à exposition limitée. De plus, la conception de la suite de tests est extensible à d'autres langues typologiquement diverses, ce qui soutient des efforts plus larges d'évaluation interlinguistique dans des conditions de données limitées. La méthodologie présentée, qui utilise des livres de grammaire pour trouver des phrases représentatives de phénomènes spécifiques à une langue et les transforme en une référence de paires minimales, peut être adoptée pour d'autres langues.

Références

- Marvin, R., & Linzen, T. (2018). Targeted Syntactic Evaluation of Language Models. In E. Riloff, D. Chiang, J. Hockenmaier, & J. Tsujii (Eds.), *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 1192–1202). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D18-1151>
- Faysse, M., Fernandes, P., Guerreiro, N. M., Loison, A., Alves, D. M., Corro, C., Boizard, N., Alves, J., Rei, R., Martins, P. H., Casademunt, A. B., Yvon, F., Martins, A. F. T., Viaud, G., Hudelot, C., & Colombo, P. (2024). *CroissantLLM: A Truly Bilingual French-English Language Model* (No. arXiv:2402.00786). arXiv. <https://doi.org/10.48550/arXiv.2402.00786>
- Futrell, R., Wilcox, E., Morita, T., Qian, P., Ballesteros, M., & Levy, R. (2019). Neural language models as psycholinguistic subjects: Representations of syntactic state. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 32–42). Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1004>
- Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C. C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., ... Scialom, T. (2023). *Llama 2: Open Foundation and Fine-Tuned Chat Models* (No. arXiv:2307.09288). arXiv. <https://doi.org/10.48550/arXiv.2307.09288>
- Warstadt, A., Parrish, A., Liu, H., Mohananey, A., Peng, W., Wang, S.-F., & Bowman, S. R. (2020). BLiMP: The Benchmark of Linguistic Minimal Pairs for English. *Transactions of the Association for Computational Linguistics*, 8, 377–392. https://doi.org/10.1162/tacl_a_00321

Dong & Nicolas Audibert (Laboratoire de Phonétique et Phonologie, Université Sorbonne Nouvelle & CNRS)

« Et [β]ah [v]as-y ! » Variation de /b/ et /v/ en français : Réduction ou lénition ?

Une analyse spectrographique des réalisations intervocaliques en français conversationnel

Mots-clés : obstruante labiale voisée ; approximante ; français conversationnel ; variation phonétique ; lénition

Résumé

Cette étude examine le phénomène d'affaiblissement des obstruantes labiales voisées /b/ et /v/ en position intervocalique en français conversationnel. La variation phonétique est généralement considérée comme une source du changement diachronique, issue des contraintes articulatoires et perceptives (par exemple, Ohala, 1990 ; Blevins, 2004 ; Lindblom, 1990 ; Beddor, 2023). La trajectoire de la lénition concernant les obstruantes voisées, *occlusive* > *fricative* > *approximante* > *zéro* (Trask, 2006, p.191 ; Crystal, 2008, p. 274), reflète un relâchement progressif de l'aperture en lien avec le mode d'articulation (Kirchner, 2004, p. 313 ; Broś et al., 2021). Ce phénomène d'affaiblissement bien documenté dans des langues romanes (par exemple, Simonet et al., 2012 ; Wayland et al., 2023), reste peu analysé en français malgré son occurrence relativement fréquente (par exemple, Wu & Adda-Decker, 2021).

L'étude présentée s'inscrit dans le cadre d'un projet visant à documenter les caractéristiques des variantes affaiblies de ces obstruantes labiales voisées en position intervocalique en français conversationnel, ainsi que les contextes dans lesquels elles surviennent, afin d'établir si ces variations relèvent d'un simple affaiblissement contextuel ou d'une amorce de changement sonore.

Dans le prolongement d'une première étude sur 5504 occurrences de /v/ (Dong & Audibert, 2024), produites par 10 femmes (n=2970) et 10 hommes (n=2534), nous appliquons le même protocole sur 2761 occurrences de /b/ produites par les mêmes locuteurs du corpus NCCFr (Torreira et al., 2010), incluant 1572 réalisations des femmes et 1189 des hommes. Le protocole d'annotation retenu repose sur l'inspection des spectrogrammes pour catégoriser les différents types de réalisation et identifier un ensemble d'indices acoustiques (Figure 1).

Les résultats montrent une variabilité importante des réalisations effectives de /b/ (n=2204) : 48% sont réalisés comme occlusives voisées, 20% comme approximantes sans indice spectrographique d'occlusion et avec une structure formantique stable, 7% comme intermédiaires avec des caractéristiques acoustiques des approximantes et un burst identifiable (voir l'illustration Figure 2 de ces trois types de réalisations), et 22% comme des réalisations nasalisées. En comparaison, 35% des /v/ ont été réalisés comme approximantes (n=1583) et 49% comme fricatives voisées (n=2197) (Dong & Audibert, 2024).

Comme pour le /v/, les variantes approximantes du /b/ se distinguent acoustiquement des réalisations canoniques, et présentent une durée plus courte avec un recouvrement important entre catégories indépendamment du locuteur. De plus, la mise en correspondance du taux d'occurrence des réalisations affaiblies du /v/ et /b/ par locuteur indique une corrélation positive

($r = .68$) notamment chez les femmes ($r = .93$), suggérant des dynamiques de variation partagées entre ces obstruantes au-delà des différences explicables par les variations de débit de parole.

L'inspection spectrographique des cas intermédiaires et des réalisations approximantes de /b/, partageant le même lieu d'articulation que leurs formes modales, est en accord avec les résultats de Duez (1995) et suggèrent que le burst pourrait devenir un indice perceptif secondaire comparativement à d'autres gestes articulatoires. Ces résultats préliminaires sont consistants avec l'hypothèse d'un changement sonore amorcé, au-delà d'une simple réduction contextuelle.

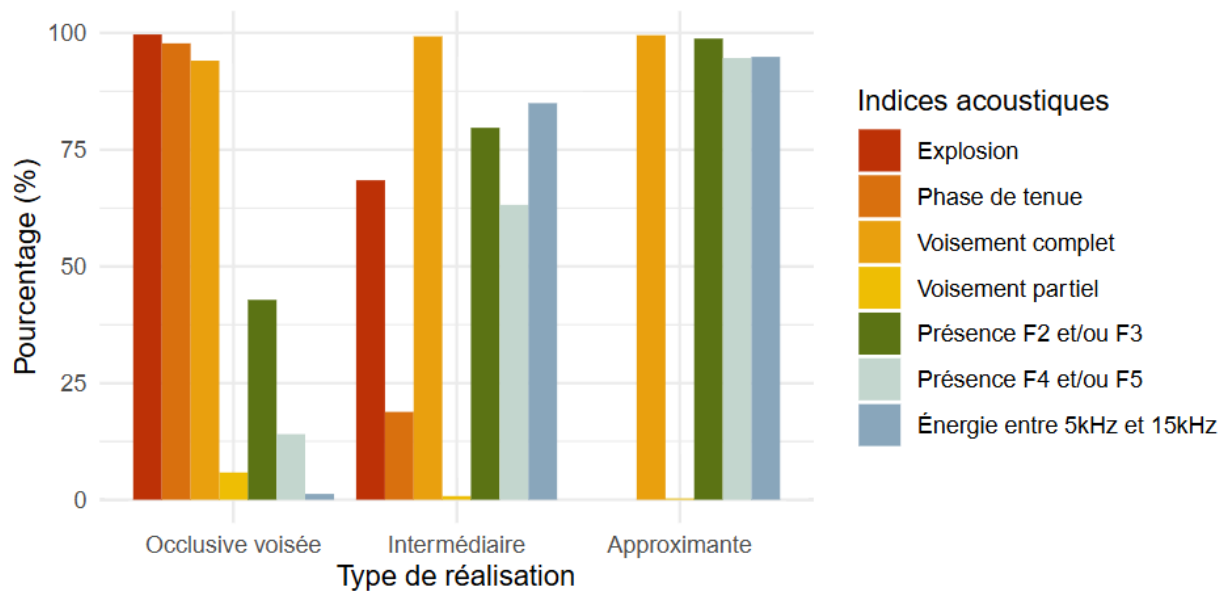


Fig. 1 : Répartition des indices acoustiques annotés par type de réalisation.

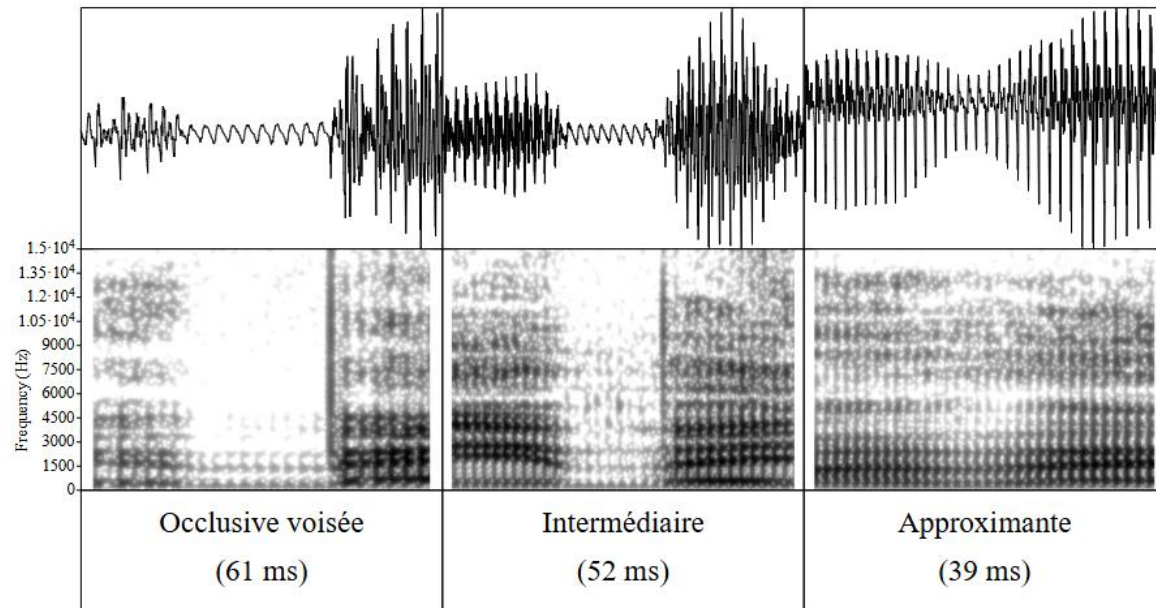


Fig. 2 : Exemples des réalisations du /b/ comme occlusive voisée, intermédiaire et approximante dans le contexte /ε_a/, produits par la locutrice LocF07. En complément de l'oscillogramme (haut), la distribution de l'énergie entre 0 et 15kHz est représentée dans le spectrogramme (bas). Les extraits de signal représentés incluent la réalisation des voyelles adjacentes, la durée de la consonne étant indiquée sous le spectrogramme.

Keywords: voiced labial obstruent; approximant; conversational French; phonetic variation; lenition

Abstract

This study examines the weakening phenomenon of the voiced labial stops /b/ and /v/ in intervocalic position in conversational French. Phonetic variation is generally considered a source of diachronic change, stemming from articulatory and perceptual constraints (e.g., Ohala, 1990; Blevins, 2004; Lindblom, 1990; Beddor, 2023). The lenition trajectory of voiced stops, *stop* > *fricative* > *approximant* > *zero* (Trask, 2006, p.191; Crystal, 2008, p.274), reflects a progressive relaxation of aperture in relation to manner of articulation (Kirchner, 2004, p.313; Broś et al., 2021). While this weakening phenomenon is well documented in Romance languages (e.g., Simonet et al., 2012; Wayland et al., 2023), it has received little attention in French despite its relatively frequent occurrence (e.g., Wu & Adda-Decker, 2021).

The present study is part of a project aiming to document the characteristics of weakened variants of these voiced labial stops in intervocalic position in conversational French, as well as the contexts in which they occur, in order to determine whether these variations represent simple contextual weakening or the onset of a sound change.

Building on a first study of 5,504 occurrences of /v/ (Dong & Audibert, 2024), produced by 10 female speakers (n=2970) and 10 male speakers (n=2534), we apply the same protocol to 2,761 occurrences of /b/ produced by the same speakers from the NCCFr corpus (Torreira et al., 2010), including 1,572 realizations from women and 1,189 from men. The annotation protocol was based

on spectrogram inspection to categorize the different types of realizations and identify a set of acoustic cues (Figure 1).

The results show significant variability in actual realizations of /b/ (n=2204): 48% were realized as voiced stops, 20% as approximants without spectrographic evidence of closure and with stable formant structure, 7% as intermediate forms with acoustic characteristics of approximants and an identifiable burst (see Figure 2 for illustrations of these three realization types), and 22% as nasalized realizations. By comparison, 35% of /v/ were realized as approximants (n=1583) and 49% as voiced fricatives (n=2197) (Dong & Audibert, 2024).

As with /v/, the approximant variants of /b/ are acoustically distinct from canonical realizations, showing shorter duration and considerable overlap between categories regardless of speaker. Moreover, matching the occurrence rates of weakened /v/ and /b/ realizations by speaker reveals a positive correlation ($r = .68$), particularly strong among women ($r = .93$), suggesting shared dynamics of variation between these stops beyond differences attributable to speech rate variation.

Spectrographic inspection of intermediate and approximant realizations of /b/, which share the same place of articulation as their modal forms, aligns with Duez (1995) and suggests that the burst may be becoming a secondary perceptual cue compared to other articulatory gestures. These preliminary results are consistent with the hypothesis of an incipient sound change, going beyond mere contextual reduction.

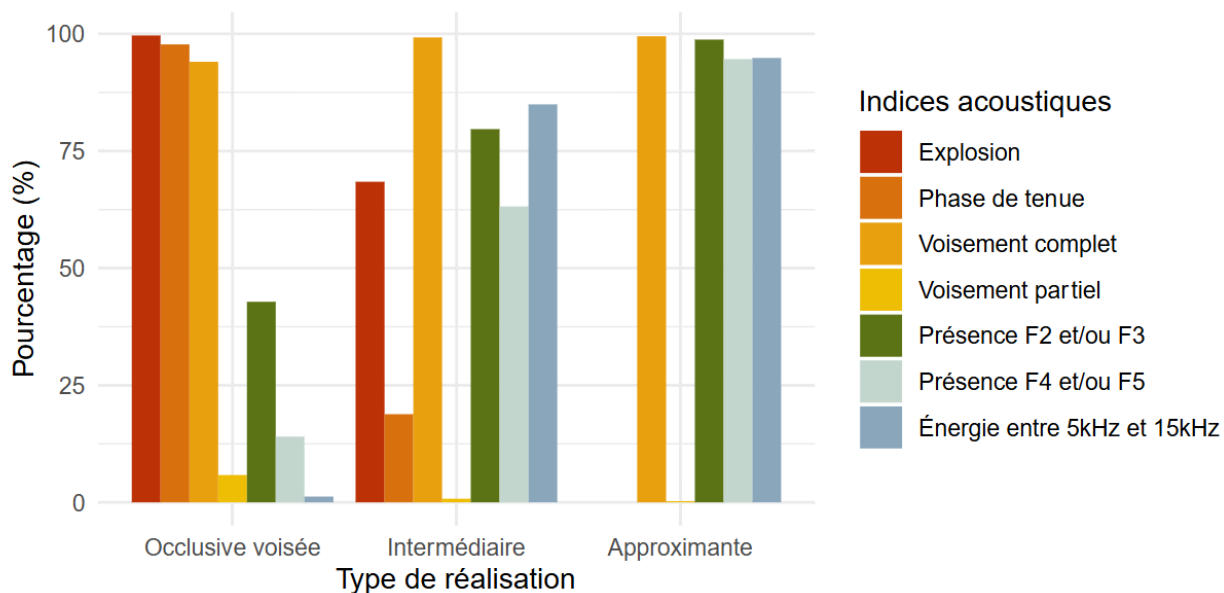


Fig. 1. Distribution of annotated acoustic cues according to the type of realization

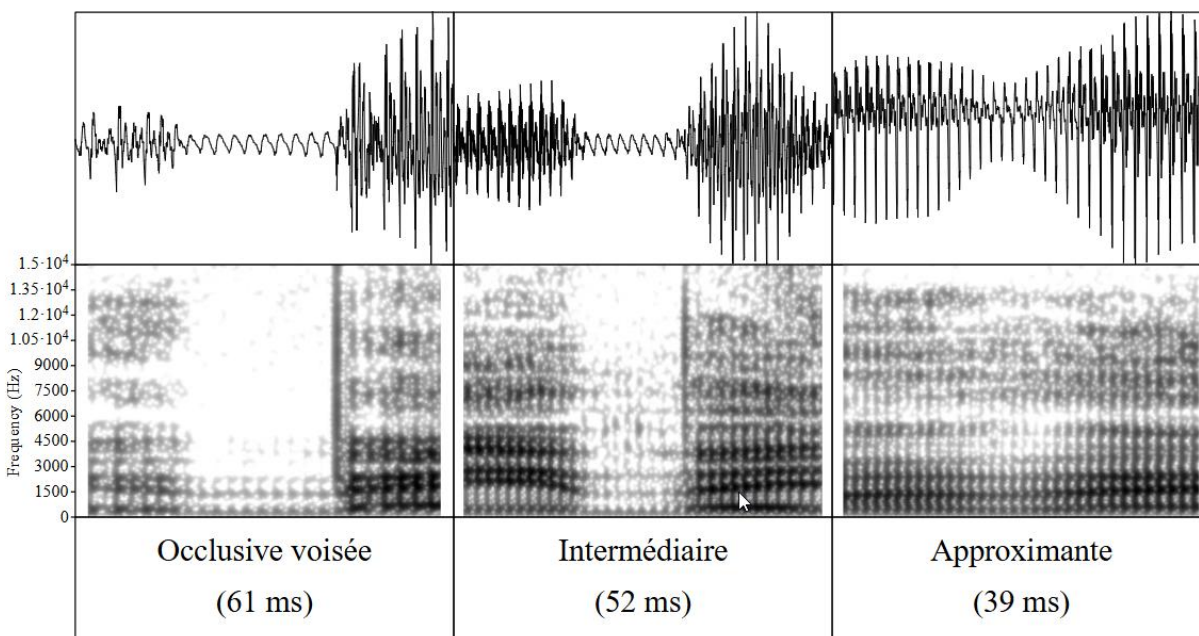


Fig. 2. Examples of /b/ produced as a voiced stop, an intermediate form, and an approximant in the /ε_a/ context by speaker LocF07. Waveform (top) and spectrogram (bottom, 0–15 kHz). Adjacent vowels are included; consonant duration is shown below the spectrogram

Références

- Beddor, P. S. (2023). Advancements of phonetics in the 21st century: Theoretical and empirical issues in the phonetics of sound change. *Journal of Phonetics*, 98, 101-228. <https://doi.org/10.1016/j.wocn.2023.101228>
- Blevins, J. (2004). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge University Press.
- Broś, K., Żygis, M., Sikorski, A. & Wołłejko, J. (2021). Phonological contrasts and gradient effects in ongoing lenition in the Spanish of Gran Canaria. *Phonology*, 38(1), 1-40. <https://doi.org/10.1017/S0952675721000038>
- Crystal, D. (2008). *A dictionary of linguistics and phonetics* (6th edition). Blackwell Pub.
- Dong, S. & Audibert N. (2024). Caractérisation acoustique des réalisations approximantes du /v/ intervocalique en français spontané. *35èmes Journées d'Études sur la Parole (JEP 2024)*, 13–22. <https://inria.hal.science/hal-04623055>
- Duez, D. (1995). On spontaneous French speech: aspects of the reduction and contextual assimilation of voiced stops. *Journal of Phonetics*, 23(4), 407-427. <https://doi.org/10.1006/jpho.1995.0031>
- Kirchner, R. (2004). Consonant lenition. In: Hayes, B., Kirchnern, R. & Steriade, D. (eds.), *Phonetically Based Phonology* (1st edition), pp. 313-345. Cambridge University Press.

- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In: Hardcastle, W. J. & Marchal, A. (eds.), *Speech Production and Speech Modelling*, pp. 403-439. Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_16
- Ohala, J. J. (1990b). The phonetics and phonology of aspects of assimilation. In Kingston, J. & Beckman, M. (eds.), *Papers in laboratory phonology, vol. 1: Between the grammar and the physics of speech*, pp. 258–75. Cambridge: Cambridge University Press.
- Simonet, M., Hualde, J. I. & Nadeu, M. (2012). Lenition of /d/ in spontaneous Spanish and Catalan. *13th Annual Conference of the International Speech Communication Association (Interspeech 2012)*, 1414-1417. <https://doi.org/10.21437/Interspeech.2012-50>
- Trask, R. L. (2006). *A dictionary of phonetics and phonology* (Reprinted). Routledge.
- Wayland, R., Tang, K., Wang, F., Vellozzi, S. & Sengupta, R. (2023). Quantitative acoustic versus deep learning metrics of Lenition. *Languages* 2023, 8(2), 98. <https://doi.org/10.3390/languages8020098>
- Wu, Y. & Adda-Decker, M. (2021). Réduction des segments en français spontané : apports des grands corpus et du traitement automatique de la parole. *Corpus* [En ligne], 22. <https://doi.org/10.4000/corpus.5812>

Valentina Fedchenko (INALCO)

Création d'un jeu de données pour l'entraînement d'un ASR sur les dialectes yiddish

Mots-clés : langues peu dotées ; jeu de données multidialectal ; reconnaissance vocale multidialectale ; dialectes yiddishs

Résumé

Cette présentation décrit la première étape d'un projet de création d'un système ASR pour l'ensemble des dialectes yiddish. Le yiddish, une langue moyennement dotée, commence à être intégré dans les grands modèles de langage (Kulick, 2023). Le problème des systèmes de reconnaissance vocale (ASR) pour le yiddish est en partie lié à sa proximité linguistique avec l'allemand : des modèles comme Whisper ou MMS identifient souvent le yiddish comme de l'allemand. En outre, le yiddish présente des difficultés typiques des langues minorées : une orthographe faiblement standardisée et une grande variabilité linguistique, notamment au niveau phonétique — le continuum dialectal yiddish est traversé par des isoglosses phonétiques.

Le système ASR pour les dialectes yiddish vise à détecter le dialecte et à produire une transcription du contenu en alphabet phonétique international (IPA) (cf. Wahee, 2023).

La première étape consiste à constituer un jeu de données multidialectal pour entraîner l'ASR aux différents dialectes du yiddish. Comme pour la plupart des langues peu dotées, on constate une absence de données annotées — c'est-à-dire des fragments audio alignés avec leurs transcriptions. Pour remédier à cette rareté, le projet s'appuie sur une riche collection de livres audio lus par des locuteurs natifs (cf. Yeroyan, 2024) : *Yiddish Book Center's Sami Rohr Library of Recorded Yiddish Books* et la *Yiddish Book Center Multimedia Library*, accessibles via www.archive.org. Les versions textuelles correspondantes de ces livres, obtenues par reconnaissance optique de caractères (OCR), fournissent l'alignement nécessaire pour l'entraînement ASR, en orthographe standardisée (Urieli, 2024).

Cette orthographe doit être convertie en transcription dialectale en IPA. La tâche de conversion est complexifiée par les variations dans le standard orthographique et par les ambiguïtés entre la phonétique dialectale et le canon orthographique, qui repose en partie sur le dialecte nord-est. L'orthographe standard ne reflète pas certaines variations phonétiques importantes selon les dialectes (par exemple, la voyelle notée *tsvey yudn* ("") peut être réalisée comme [aɪ] ou [ɛɪ] ([gaɪn] vs [gɛɪn] selon les régions). Le modèle doit donc apprendre à résoudre ces ambiguïtés en fonction du contexte dialectal.

Pour aborder cette tâche, nous avons sélectionné trois livres lus dans trois dialectes différents. Nous avons établi des règles de conversion phonologique et, dans un premier temps, effectué la conversion manuellement à l'aide de ces règles. À cela, nous avons ajouté des données sélectionnées issues du corpus REYD (Webber, 2022).

Ensuite, nous avons fine-tuné un modèle encodeur-décodeur (seq-to-seq) pré-entraîné, T5, sur des paires de phrases en orthographe standardisée (caractères hébraïques) et en transcription dialectale IPA, conditionnées par des identifiants de dialectes. Malgré des performances satisfaisantes sur les données de test, cette approche demande aux chercheurs d'être attentifs à

la spécificité dialectale des données. Les dialectes yiddish formant un continuum, l'étiquetage discret (ex. : *dialecte* : *nord-est*) s'avère réducteur. Le modèle risque de mal gérer les formes intermédiaires ou mixtes.

Pour l'instant, l'application de ce modèle à la production de données d'entraînement pour un système de reconnaissance vocale dialectal est envisageable, à condition de procéder à une présélection des livres lus par des locuteurs de dialectes distincts.

Keywords: low-resource languages; multidialectal dataset; multidialectal speech recognition; Yiddish dialects

Abstract

This presentation describes the first step of a project that aims to create an ASR system for all Yiddish dialects. Yiddish, a medium-resourced language, has only recently started being integrated into large language models (Kulick, 2023). The problem with ASR systems for Yiddish is partly due to its linguistic proximity to German: models such as Whisper or MMS often identify Yiddish as German. Moreover, Yiddish presents typical challenges of minoritized languages: weakly standardized orthography and significant linguistic variability, particularly at the phonetic level – the Yiddish dialect continuum is marked by multiple phonetic isoglosses.

The ASR system for Yiddish dialects aims to detect the dialect and produce a transcription of the content in the International Phonetic Alphabet (IPA) (cf. Wahee, 2023). The first step is to build a multidialectal dataset to train the ASR on the different Yiddish dialects. As with most low-resource languages, there is a lack of annotated data – i.e., audio fragments aligned with their transcriptions. To address this scarcity, the project relies on a rich collection of audiobooks read by native speakers (cf. Yeroyan, 2024): the Yiddish Book Center's *Sami Rohr Library of Recorded Yiddish Books* and the *Yiddish Book Center Multimedia Library*, both accessible via www.archive.org. The corresponding textual versions of these books, obtained through Optical Character Recognition (OCR), provide the alignment necessary for ASR training, in standardized orthography (Urieli, 2024).

This orthography must then be converted into dialectal transcription in IPA. The conversion task is complicated by variations in the standard orthography and by ambiguities between dialectal phonetics and the orthographic canon, which is partly based on the Northeastern dialect. Standard orthography does not capture certain important phonetic variations across dialects (for example, the vowel written *tsvey yudn* ("") may be realized as [aɪ] or [ɛɪ]—*gain* vs *gɛɪn* depending on the region). The model must therefore learn to resolve such ambiguities based on dialectal context.

To approach this task, we selected three books read in three different dialects. We established phonological conversion rules and, in the first stage, carried out the conversion manually using these rules. To this, we added selected data from the REYD corpus (Webber, 2022).

We then fine-tuned a pre-trained encoder-decoder (seq-to-seq) model, T5, on sentence pairs in standardized orthography (Hebrew characters) and dialectal IPA transcription, conditioned on dialect identifiers. Despite satisfactory performance on test data, this approach requires researchers to pay careful attention to the dialectal specificity of the data. Since Yiddish dialects

form a continuum, discrete labeling (e.g., dialect: Northeastern) proves reductive. The model risks handling intermediate or mixed forms poorly.

At this stage, applying this model to the production of training data for a dialectal speech recognition system is feasible, provided that audiobooks are preselected according to distinct dialects spoken by the narrators.

Références

Kulick, S., Ryant, N., Santorini, B., Wallenberg, J. & Urieli, A. (2023). A Part-of-Speech Tagger for Yiddish. <https://arxiv.org/abs/2204.01175> (date de consultation : 9 juin 2025).

Urieli, A., Clooney, A., Sigiel, M., Leyfer, G. (2024). Jochre 3 and the Yiddish OCR corpus. <https://arxiv.org/abs/2501.08442> (date de consultation : 3 avril 2025).

Wahee, A., Talařa, B., Sullivan, P., Elmadany, A., Abdul-Mageed, M. (2023). Vox Arabica: A Robust Dialect-Aware Arabic Speech Recognition System. In: Proceedings of Arabic NLP, pp. 441-449.

Webber, J. J., Lo, S. K., Bleaman, I. L. (2022). REYD – The First Yiddish Text-to-Speech Dataset and System. In: Proceedings of Interspeech, pp. 2363-2367.

Yeroyan, A., Karpov, N. (2024). Enabling ASR for Low-Resource Languages: A Comprehensive Dataset Creation Approach. <https://arxiv.org/abs/2406.01446> (date de consultation : 3 avril 2025).

Océane Foubert (University of Lille/STL)

Naming bereaved parents and their babies: a questionnaire-based study of the lexicon of perinatal bereavement

Keywords: lexicon; linguistic attitude; neologism; perinatal bereavement; questionnaire

Abstract

The suffering of parents experiencing perinatal bereavement is exacerbated by the subsequent silence surrounding the event (Lang et al. 2011). This lack of recognition of perinatal death is reflected in the absence of a specific term to refer to bereaved parents and their babies. This is one of the manifestations of what Fricker refers to as *hermeneutical injustice*, which she defines as “the injustice of having some significant area of one’s social experience obscured from collective understanding” (2007: 156).

To counter this lexical gap, French-speaking parents have coined lexical innovations to name themselves and their babies, particularly in online communities. For example, there have been several petitions to include the neologisms “*parange*” (a blend of “*parent*” and “*ange*” *angel*) and “*enfange*” (a blend of “*enfant*” *child* and “*ange*” *angel*) in dictionaries. Previous qualitative studies have gathered neologisms and attitudes towards these neologisms present in bereaved parents’ online discourse (Ruchon 2015; Caliendo & Ruchon 2020).

The aim of this study is to systematically record the attitudes of French-speaking bereaved parents towards self- and hetero-designations when referring to themselves and their babies. To do this, a questionnaire was set up for bereaved parents to gather their opinions on 13 lexical units for naming parents and 4 such units for naming babies. Two open-ended questions were also asked to allow parents to suggest other designations. To date, over 500 bereaved parents have responded to the questionnaires.

The results show that there is no consensus on a lexical innovation for naming parents and babies, and some even reject the neologisms *parange* and *enfange* put forward in the petitions. Indeed, many reject the euphemistic nature of angel-based lexical innovations. In addition, the responses to the open-ended questions revealed that some parents argue that what they need is not a different terminology but a different understanding of what it means to be a parent. Similar findings have been highlighted in a recent psychological study on bereaved mothers by Rossen et al. (2023: 16-17) who advocate for “a broader definition and understanding of motherhood, which encapsulates diverse experiences of motherhood and mothering”. Our results echo these findings and extend them to parenthood.

While the observation of the need to fill the lexical gap and the complexities involved in finding the right lexical items is a necessary step, it is also only the first. One of the following steps (currently underway) is to observe the attitudes of people who have not experienced perinatal bereavement towards these lexical items. Finally, the results of these questionnaire-based studies of the lexicon of perinatal bereavement will be used to provide input concerning care pathways to improve the communication experience of bereaved parents.

Mots-clés : lexique ; attitude linguistique ; néologisme ; deuil périnatal ; questionnaire

Résumé

La souffrance des parents vivant un deuil périnatal est exacerbée par le silence qui entoure par la suite l'événement (Lang et al. 2011). Ce manque de reconnaissance de la mort périnatale se reflète dans l'absence d'un terme spécifique pour désigner les parents endeuillés et leurs bébés. Ceci est l'une des manifestations de ce que Fricker désigne comme *injustice herméneutique*, qu'elle définit comme « l'injustice consistant à ce qu'un domaine significatif de l'expérience sociale d'une personne soit obscurci par une compréhension collective » (2007 : 156).

Pour contrer cette lacune lexicale, les parents francophones ont créé des innovations lexicales pour se nommer eux-mêmes et nommer leurs bébés, particulièrement dans les communautés en ligne. Par exemple, il y a eu plusieurs pétitions pour inclure les néologismes « *parange* » (mélange de « *parent* » et « *ange* ») et « *enfange* » (mélange d'« *enfant* » et d'« *ange* ») dans les dictionnaires. Des études qualitatives antérieures ont recueilli les néologismes et les attitudes envers ces néologismes présents dans le discours en ligne des parents endeuillés (Ruchon 2015 ; Caliendo & Ruchon 2020).

L'objectif de cette étude est d'enregistrer systématiquement les attitudes des parents endeuillés francophones envers les auto-désignations et hétéro-désignations lorsqu'ils font référence à eux-mêmes et à leurs bébés. Pour ce faire, un questionnaire a été mis en place pour que les parents endeuillés recueillent leurs opinions sur 13 unités lexicales pour nommer les parents et 4 de ces unités pour nommer les bébés. Deux questions ouvertes ont également été posées pour permettre aux parents de suggérer d'autres désignations. À ce jour, plus de 500 parents endeuillés ont répondu aux questionnaires.

Les résultats montrent qu'il n'y a pas de consensus sur une innovation lexicale pour nommer les parents et les bébés, et certains rejettent même les néologismes *parange* et *enfange* proposés dans les pétitions. En effet, beaucoup rejettent la nature euphémistique des innovations lexicales basées sur les anges. De plus, les réponses aux questions ouvertes révèlent que certains parents soutiennent que ce dont ils ont besoin n'est pas une terminologie différente mais une compréhension différente de ce que signifie être parent. Des résultats similaires ont été soulignés dans une étude psychologique récente sur les mères endeuillées par Rossen et al. (2023 : 16-17) qui plaident pour « une définition et une compréhension plus larges de la maternité, qui englobe les expériences diverses de la maternité et du maternage ». Nos résultats font écho à ces conclusions et les étendent à la parentalité.

Bien que l'observation du besoin de combler la lacune lexicale et les complexités impliquées dans la recherche des bons éléments lexicaux soit une étape nécessaire, c'est aussi seulement la première. L'une des étapes suivantes (actuellement en cours) consiste à observer les attitudes des personnes qui n'ont pas vécu de deuil périnatal envers ces éléments lexicaux. Enfin, les résultats de ces études basées sur questionnaire du lexique du deuil périnatal seront utilisés pour fournir des éléments concernant les parcours de soins afin d'améliorer l'expérience de communication des parents endeuillés.

Références

- Caliendo, G. & Ruchon, C. (2020). La nomination des enfants décédés en bas-âge et de leurs parents. D'une analyse du discours située à une linguistique d'intervention. *SHS Web of Conferences* 78, 01019. <https://doi.org/10.1051/shsconf/20207801019>
- Fricker, M. (2007). *Epistemic injustice: power and the ethics of knowing*. Oxford: Oxford University Press.
- Lang, A., Fleischer, A. R., Duhamel, F., Sword, W., Gilbert, K. R. & Corsini-Munt, S. (2011). Perinatal loss and parental grief: the challenge of ambiguity and disenfranchised grief. *Omega: Journal of Death and Dying*, 63, 183–196.
- Rossen, L., Opie, J. E., & O'Dea, G. (2023). A Mother's Voice: The Construction of Maternal Identity Following Perinatal Loss. *OMEGA - Journal of Death and Dying*. <https://doi.org/10.1177/00302228231209769>
- Ruchon, C. (2015). Des vertus antalgiques du discours ? L'expression de la douleur et de l'attachement dans les discours sur la maternité. PhD thesis. Université de Paris 13 Sorbonne Paris Cité.

Birgit Füreder (Mannheim Universität), Cyrille Granget (Université Toulouse Jean Jaurès), Martin Haiden (Université de Nantes) & Isabel Repiso (Université de Haute Alsace)

La primauté de l'agent en allemand et français à l'épreuve des 3 S :

Statut, syntaxe, saillance

Résumé

Le principe agent en premier est un principe universel de la cognition humaine qui conduit à accorder plus d'attention à l'entité X agissante dans un événement du type « X humain agit sur Y humain » et à exprimer cette scène au moyen d'une phrase dont le premier élément est un nom animé qui a le rôle d'agent (Griffin & Bock 2000). Plusieurs études linguistiques et/ou oculométriques ont montré ou suggéré que ce principe est une variable déterminante mais en interaction avec d'autres : le sens de lecture pratiqué (Essaulova et al. 2021), la flexibilité syntaxique de la langue utilisée (Cristante & Schimke 2020, Lee & Doherty 2019), la saillance visuelle de l'entité agissante (Hafri et al. 2023, Isasi-Isasmendi et al. 2018) et le statut de la langue utilisée dans le répertoire multilingue, première ou pas (Klein & Perdue 1997). Cependant ces résultats reposent sur quelques études dans un ensemble réduit de langues et de designs expérimentaux, d'autres études empiriques sont nécessaires pour mieux comprendre la cognition multilingue. Nous proposons une tâche de production avec suivi du regard. Partant des hypothèses selon lesquelles les phénomènes liés à 'agent-first' sont favorisés par l'usage d'une langue non première (Ln), d'une langue à sujet initial et par la localisation de l'entité agissante à gauche du participant, nous avons ainsi comparé le comportement verbal et oculaire de 44 adultes (18-26 ans) selon trois variables principales : le statut d'acquisition de la langue utilisée, opérationnalisé par un score de maîtrise (test ELAO), son statut syntaxique (français SVO vs allemand V2) et la saillance visuelle de l'entité agissante dans les 6 scènes cibles *coiffer, laver, masser, frapper, pincer, pousser* (à gauche et plus grand vs droite et plus petit). Les variables dépendantes utilisées sont la diathèse et la durée de fixation de l'agent. Les premiers résultats des descriptions orales des 22 adultes de L1 allemand montrent que lorsque l'entité agissante est à droite (condition défavorable), l'agent est en position initiale dans 80.5 % des énoncés descriptifs en L1 et dans 53.3 % des énoncés en français L2, ce qui va à l'encontre de l'hypothèse selon laquelle le principe agent en premier est plus fréquent en L2 qu'en L1. Ces résultats doivent être complétés par l'analyse, en cours, des descriptions du groupe français L1/allemand L2 et des mouvements oculaires mais ils montrent déjà que la mise en œuvre du principe « agent first » est tributaire du statut de la langue utilisée, dans un sens opposé à celui attendu. Ceci peut s'expliquer par l'inaccessibilité du lexique verbal en L2 et l'emploi fréquent de constructions alternatives présentatives (*c'est/il y a*).

Abstract

The *agent-first principle* is a universal principle of human cognition that leads people to pay more attention to the acting entity X in an event of the type “human X acts on human Y,” and to express this scene by means of a sentence whose first element is an animate noun with the role of agent (Griffin & Bock 2000). Several linguistic and/or eye-tracking studies have shown or suggested that this principle is a determining variable, but one that interacts with others: the reading direction practiced (Essaulova et al. 2021), the syntactic flexibility of the language used (Cristante & Schimke 2020; Lee & Doherty 2019), the visual salience of the acting entity (Hafri et al. 2023; Isasi-Isasmendi et al. 2018), and the status of the language used within the multilingual repertoire, whether first language or not (Klein & Perdue 1997). However, these results are based on only a few studies in a limited set of languages and experimental designs; further empirical research is needed to better understand multilingual cognition.

We propose a production task with eye-tracking. Based on the hypotheses that phenomena linked to *agent-first* are favored by the use of a non-first language (Ln), by a subject-initial language, and by the location of the acting entity to the left of the participant, we compared the verbal and ocular behavior of 44 adults (18-26 years) according to three main variables: the acquisition status of the language used, operationalized by a proficiency score (ELAO test); its syntactic status (French SVO vs. German V2); and the visual salience of the acting entity in six target scenes (*coiffer* “to comb”, *laver* “to wash”, *masser* “to massage”, *frapper* “to hit”, *pincer* “to pinch”, *pousser* “to push”), presented either on the left and larger or on the right and smaller. The dependent variables used were diathesis and fixation duration on the agent.

Preliminary results from the oral descriptions of the 22 German L1 adults show that when the acting entity is on the right (an unfavorable condition), the agent is in initial position in 80.5% of descriptive utterances in L1 and in 53.3% of utterances in French L2. This runs counter to the hypothesis that the *agent-first* principle is more frequent in L2 than in L1. These results must be complemented by the ongoing analysis of descriptions from the French L1/German L2 group and the eye movement data, but they already show that the implementation of the *agent-first* principle depends on the status of the language used, in a direction opposite to that expected. This may be explained by the inaccessibility of the verbal lexicon in L2 and the frequent use of alternative presentative constructions (*c’est / il y a*).

Références

- Cristante, V., & Schimke, S. (2020). The processing of passive sentences in German : Evidence from an eye-tracking study with seven- and ten-year-olds and adults. *Language, Interaction and Acquisition*, 11(2), 163-195. <https://doi.org/10.1075/lia.19013.cri>
- Essaulova, Y., Dolscheid, S., Reuters, S., & Penke, M. (2021). The Alignment of Agent-First Preferences with Visual Event Representations : Contrasting German and Arabic. *Journal of Psycholinguistic Research*, 50. <https://pub.uni-bielefeld.de/record/2952757>
- Griffin, Z. M., & Bock, K. (2000). What the Eyes Say About Speaking. *Psychological Science*, 11(4), 274-279. <https://doi.org/10.1111/1467-9280.00255>

Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition*, 175, 36-52. <https://doi.org/10.1016/j.cognition.2018.02.011>

Isasi-Isasmendi, A., Andrews, C., Flecken, M., Laka, I., Daum, M. M., Meyer, M., Bickel, B., & Sauppe, S. (2023). The Agent Preference in Visual Event Apprehension. *Open Mind: Discoveries in Cognitive Science*, 7, 240-282. https://doi.org/10.1162/opmi_a_00083

Klein, W., & Perdue, C. (1997). The Basic Variety (or: Couldn't natural languages be much simpler?). *Second Language Research* 13, 301-347.

Lee, J., & Doherty, S. (2019). Native and nonnative processing of active and passive sentences: The effects of processing instruction on the allocation of visual attention. *Studies in Second Language Acquisition*, 41(4), 853-879.

Cristina Garcia Holgado (University of Poitiers, University of Corsica)

Unsupervised Dialect Discrimination: An Analysis of Linguistic Sensitivity in Custom Pipelines and Pretrained LLMs in a Low-Resource Language

Keywords: dialect identification; linguistic variation; low-resource languages; clustering; unsupervised learning; CamemBERT; large language models

Abstract

This work investigates the automatic identification of two dialectal varieties of a low-resource (LR) regional language with insufficient linguistic documentation and spelling variation. Although language identification is widely considered a solved task in high-resource settings, performance remains limited in unsupervised scenarios and in the case of closely related language varieties (Goswami et al., 2020). Today, most approaches focus on supervised learning. However, for low-resource languages where dialects form a continuum with no clear-cut boundaries and available metadata is scarce, unsupervised methods are a suitable alternative as recent studies have demonstrated (Sato et al., 2020; Sung et al., 2024).

Here, we compare two unsupervised clustering pipelines. The first relies on surface-level features, using character-level TF-IDF vectors, reduced with t-SNE and clustered using KMeans. The second pipeline incorporates contextual embeddings from CamemBERT, a transformer model fine-tuned on a closely related language. These embeddings are combined with character n-gram features, reduced using PCA, and clustered using KMeans.

For evaluation, we assess both approaches in multiple ways. First, we use a corpus of dialectal texts with geolocated metadata to evaluate how well the resulting clusters align with the known geographic distribution of the two dialects. We observe that the clusters closely match the expected dialectal zones, confirming that the models discriminate between the two varieties in a way that reflects established linguistic descriptions. Secondly, external validation is performed using unseen texts with dialect labels assigned by linguists with limited familiarity with the language. Both pipelines produce comparable results, achieving F1 scores ranging from 0.80 to 0.90, depending on the split and feature set. In this context, we observe that the same texts tend to be misclassified across both setups, indicating errors are generally linked to linguistic-specific factors (such as French passages, diachronic variation and OCR-related noise) rather than to the clustering method itself. Moreover, one of the two dialects presented highly salient features that strongly influences the clustering behavior, particularly in the non-LLM model. This has shown that the other variety lacks equally distinctive linguistic traits. Thirdly, we analyzed the correlation between cluster confidence and predicted clusters with several text-level features. In the surface-based model, we observed that confidence decreases with higher lexical diversity, longer sentences, and French interference. In contrast, the LLM-based model was more stable across variation, showing higher confidence for longer and more French-similar texts. Although both models have shown to be capable of recovering dialectal zones that match linguistic geographic areas, the surface-based method has shown to be more sensitive to linguistic variation, while the contextual model appears more robust to noisy and heterogeneous text.

This work aims to highlight the strengths of surface-based and contextual representations for dialect discrimination in a low-resource language. It is integrated in an effort to explore how

computational methods can support and enrich fine dialectal descriptions in scenarios where supervised classification is not always feasible. Our long-term objective is to enable a finer-grained identification of sub-dialects in cases where no prior labels exist and available linguistic descriptions are scarce or incomplete.

Mots-clés: identification des dialectes ; variation linguistique ; langues à faibles ressources ; regroupement; apprentissage non supervisé ; CamemBERT ; grands modèles linguistiques

Résumé

Ce travail porte sur l'identification automatique de deux variétés dialectales d'une langue régionale à faibles ressources (LR) pour laquelle la documentation linguistique est insuffisante et l'orthographe variable. Bien que l'identification linguistique soit largement considérée comme une tâche résolue dans les environnements à ressources élevées, les performances restent limitées dans les scénarios non supervisés et dans le cas de variétés linguistiques étroitement apparentées (Goswami et al., 2020).

Aujourd'hui, la plupart des approches se concentrent sur l'apprentissage supervisé. Cependant, pour les langues à faibles ressources où les dialectes forment un continuum sans frontières clairement définies et où les métadonnées disponibles sont rares, les méthodes non supervisées constituent une alternative appropriée, comme l'ont démontré des études récentes (Sato et al., 2020 ; Sung et al., 2024).

Ici, nous comparons deux pipelines de clustering non supervisés. Le premier s'appuie sur des caractéristiques de surface, utilisant des vecteurs TF-IDF au niveau des caractères, réduits avec t-SNE et regroupés à l'aide de KMeans. Le deuxième pipeline intègre des intégrations contextuelles provenant de CamemBERT, un modèle de transformateur affiné sur une langue étroitement apparentée. Ces intégrations sont combinées avec des caractéristiques n-grammes de caractères, réduites à l'aide de PCA, et regroupées à l'aide de KMeans.

Pour l'évaluation, nous évaluons les deux approches de plusieurs manières. Tout d'abord, nous utilisons un corpus de textes dialectaux avec des métadonnées géolocalisées pour évaluer dans quelle mesure les clusters obtenus correspondent à la répartition géographique connue des deux dialectes. Nous observons que les clusters correspondent étroitement aux zones dialectales attendues, ce qui confirme que les modèles distinguent les deux variétés d'une manière qui reflète les descriptions linguistiques établies. Deuxièmement, une validation externe est effectuée à l'aide de textes inédits auxquels des linguistes ayant une connaissance limitée de la langue ont attribué des étiquettes dialectales. Les deux pipelines produisent des résultats comparables, avec des scores F1 compris entre 0,80 et 0,90, selon la répartition et l'ensemble de caractéristiques. Dans ce contexte, nous observons que les mêmes textes ont tendance à être mal classés dans les deux configurations, ce qui indique que les erreurs sont généralement liées à des facteurs linguistiques spécifiques (tels que les passages en français, les variations diachroniques et le bruit lié à l'OCR) plutôt qu'à la méthode de regroupement elle-même. De plus, l'un des deux dialectes présentait des caractéristiques très marquées qui influençaient fortement le comportement de regroupement, en particulier dans le modèle non LLM. Cela a montré que l'autre variété ne présentait pas de traits linguistiques aussi distinctifs. Troisièmement, nous avons analysé la corrélation entre la confiance des clusters et les clusters prédits avec plusieurs caractéristiques au niveau du texte. Dans le modèle basé sur la surface, nous avons observé que

la confiance diminuait avec une plus grande diversité lexicale, des phrases plus longues et l'interférence du français. En revanche, le modèle basé sur le LLM était plus stable malgré les variations, affichant une confiance plus élevée pour les textes plus longs et plus proches du français. Bien que les deux modèles se soient révélés capables de récupérer des zones dialectales correspondant aux zones linguistiques géographiques, la méthode basée sur la surface s'est avérée plus sensible aux variations linguistiques, tandis que le modèle contextuel semble plus robuste face à des textes bruités et hétérogènes.

Ce travail vise à mettre en évidence les atouts des représentations contextuelles et basées sur la surface pour la discrimination dialectale dans une langue à faibles ressources. Il s'inscrit dans le cadre d'une initiative visant à explorer comment les méthodes informatiques peuvent soutenir et enrichir les descriptions dialectales fines dans des scénarios où la classification supervisée n'est pas toujours possible. Notre objectif à long terme est de permettre une identification plus fine des sous-dialectes dans les cas où il n'existe aucune étiquette préalable et où les descriptions linguistiques disponibles sont rares ou incomplètes.

Références

Goswami, K., Sarkar, R., Chakravarthi, B. R., Fransen, T., & McCrae, J. P. (2020). Unsupervised Deep Language and Dialect Identification for Short Texts. Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain.

Sato, Y., & Heffernan, K. (2020). Dialect Clustering with Character-Based Metrics: In Search of the Boundary of Language and Dialect. Proceedings of the Twelfth Language Resources and Evaluation Conference, Marseille, France. European Language Resources Association.

Sung, H. W. M., & Prokić, J. (2024). Detecting Dialect Features Using Normalised Pointwise Information. Computational Linguistics in the Netherlands Journal, 13, 121–145. <https://clinjournal.org/clinj/article/view/177>

Keywords: Thinking-for-speaking, inner speech

Abstract

While drafting this abstract, my focus is on crafting a clear take-home message and arranging arguments to support my point. Occasionally, my thoughts drift to what will happen later today. However, thinking in that sense of the word—i.e., verbal thinking through inner speech—is not what modern-day researchers have in mind who engage in the "language-thought-debate." In this context, thought often refers to unconscious cognitive processing, a concept rooted in mid-20th century cognitive science and unavailable to earlier scholars like von Humboldt, Herder, or Boas. Modern-day research focuses on whether language systematically affects the mechanisms underlying conscious experience. However, what about verbal thinking via inner speech generation? Abundant evidence supports both Slobin's 'thinking for speaking' across languages and the crucial role of inner speech in cognitive tasks. Given that, from a psycholinguistic perspective, inner speech is generated very similarly to overt speech, I will demonstrate how 'thinking for speaking (with others)' naturally extends to 'thinking for speaking to yourself.' This extension invites new research into how language and typological differences shape higher cognitive processing, with implications for understanding both linguistic and cognitive diversity.

Mots-clés : thinking-for-speaking, discours intérieur

Résumé

En rédigeant cet abstract, je me concentre sur l'élaboration d'un message clair à retenir et sur l'organisation d'arguments pour soutenir mon propos. Occasionnellement, mes pensées dérivent vers ce qui se passera plus tard dans la journée. Cependant, penser dans ce sens du terme — c'est-à-dire, la pensée verbale à travers le discours intérieur — n'est pas ce que les chercheurs contemporains ont à l'esprit lorsqu'ils s'engagent dans le « débat langage-pensée ». Dans ce contexte, la pensée fait souvent référence aux processus cognitifs inconscients, un concept enraciné dans la science cognitive du milieu du XXe siècle et indisponible aux premiers érudits comme von Humboldt, Herder, ou Boas. La recherche contemporaine se concentre sur la question de savoir si le langage affecte systématiquement les mécanismes *sous-jacents* à l'expérience consciente. Cependant, qu'en est-il de la pensée verbale via la génération du discours intérieur ? De nombreuses preuves soutiennent à la fois le « thinking for speaking » de Slobin à travers les langues et le rôle crucial du discours intérieur dans les tâches cognitives. Ceci étant donné, d'un point de vue psycholinguistique, le discours intérieur est généré de manière très similaire au discours manifeste, je démontrerai comment le « thinking for speaking (with others) » s'étend naturellement au « thinking for speaking to yourself ». Cette extension invite à de nouvelles recherches sur la façon dont les différences linguistiques et typologiques façonnent les processus cognitifs supérieurs, avec des implications pour la compréhension de la diversité à la fois linguistique et cognitive.

Matthias Gille Levenson (Aix-Marseille Université), Lucence Ing (Inria) & Carolina Macedo (Ecole nationale des chartes)

Training sentence segmenters on Medieval languages

Keywords: Text segmentation; Multilingual alignment; Sentence embedding; BERT; Multilingual dataset; Medieval philology; Digital humanities

Abstract

Medieval languages, often categorized as low-resource languages due to their limited available data and lack of standardized orthography, present significant challenges to linguistic and philological research. Primary sources, predominantly manuscripts, are not only difficult to decipher (sometimes even with the most advanced Automatic Text Recognition [ATR] algorithms), but they also fail to conform to modern text editing standards. This research project explores the underexamined area of medieval multilingual text traditions, a field often avoided due to the complexity of analyzing multilingual traditions simultaneously, but which has been explored in recent years, thanks to the emergence of IT tools such as large language models [Meineke et al. 2020]. Our approach includes a comparative analysis of different text versions and translations produced through automatic alignment of sentences or clauses, employing edited texts and manuscript transcriptions to provide a comprehensive view of a given medieval textual tradition.

Actual alignment tools (such as Bertalign [Liu & Zhu, 2023]), which are designed for modern languages where the concept of the sentence is obvious, work with segments that are fixed upstream. Traditional segmentation methods based on punctuation fail to perform adequately with these medieval datasets, particularly with prose textual resources like ours. The lack of normalized punctuation in medieval texts poses significant segmentation challenges. Data from HTR and manuscript transcriptions are devoid of crucial punctuation, impeding effective segmentation. An example illustrating the absence of punctuation is shown in an excerpt from Gilles de Rome's, *De Regimine Principum* (Book 1, Part 2, Chapter 1) :

Cecondement les uertuz ne peust estre es sens corporelz car aussi que lespuissances naturelles ne obeissent pas a raison car ne peust croistre plus nemoins selon la reigle de raison ne plus ou moins digerez et aussi les senscorporelz ne obeissent pars a raison car il nest pas en nostre puissanceueoir plus cler ou moins cler

To address this, we developed a custom multilingual segmenter tailored for medieval languages, based on a BERT tokens classifier model, specifically trained on a variety of medieval languages, including Catalan, Castilian, English, French, Italian and Latin. The segmenter produces segmented text as follows, efficiently dividing text into segments, facilitating alignment with corresponding segments in translations and other texts :

*Cecondement les uertuz ne peust estre es sens corporelz
car aussi que les puissances naturelles ne obeissent pas a raison
car ne peust croistre plus ne moins selon la reigle de raisonne plus ou moins digerez
et aussi les sens corporelz ne obeissent pars a raison
car il nest pas en nostre puissance ueoir plus cler ou moins cler*

Our presentation at the “Developing Models for Linguistic Research: Training Data Usage in Low-Resource Scenarios” workshop will serve as a platform for discussing our methodology and the rationale behind our choices, such as the emphasis on precision over recall to enhance model efficiency. We will compare our methods for segmenting medieval data with several others, starting with a regex-based tokenization. We will also address challenges such as the discrepancy between the clean, edited datasets used for training and the raw, unedited data typically encountered in manuscript studies, on which we want to apply our model: we began our work on a mainly published corpus, before opening it up to sources acquired by ATR.

Our findings should advance medieval studies and provide insights applicable to similar challenges in other low-resource linguistic research areas.

Mots-clés: segmentation de texte ; alignement multilingue ; intégration de phrases ; BERT ; ensemble de données multilingues ; philologie médiévale ; humanités numériques

Résumé

Les langues médiévales, souvent classées parmi les langues à faibles ressources en raison du peu de données disponibles et de l'absence d'orthographe standardisée, posent des défis importants à la recherche linguistique et philologique. Les sources primaires, principalement des manuscrits, sont non seulement difficiles à déchiffrer (parfois même avec les algorithmes de reconnaissance automatique de texte [ATR] les plus avancés), mais elles ne sont pas non plus conformes aux normes modernes d'édition de texte. Ce projet de recherche explore le domaine peu étudié des traditions textuelles multilingues médiévales, un domaine souvent évité en raison de la complexité de l'analyse simultanée des traditions multilingues, mais qui a été exploré ces dernières années grâce à l'émergence d'outils informatiques tels que les grands modèles linguistiques [Meineke et al. 2020]. Notre approche comprend une analyse comparative de différentes versions de textes et traductions produites par l'alignement automatique de phrases ou de clauses, en utilisant des textes édités et des transcriptions de manuscrits afin de fournir une vue d'ensemble d'une tradition textuelle médiévale donnée.

Les outils d'alignement actuels (tels que Bertalign [Liu & Zhu, 2023]), conçus pour les langues modernes où le concept de phrase est évident, fonctionnent avec des segments qui sont fixés en amont. Les méthodes de segmentation traditionnelles basées sur la ponctuation ne fonctionnent pas correctement avec ces ensembles de données médiévaux, en particulier avec des ressources textuelles en prose comme les nôtres. L'absence de ponctuation normalisée dans les textes médiévaux pose des défis importants en matière de segmentation. Les données provenant de HTR et des transcriptions de manuscrits sont dépourvues de ponctuation essentielle, ce qui empêche une segmentation efficace. Un exemple illustrant l'absence de ponctuation est présenté dans un extrait de l'ouvrage *De Regimine Principum* (Livre 1, Partie 2, Chapitre 1) de Gilles de Rome :

Cecondement les uertuz ne peust estre es sens corporelz car aussi que les puissances naturelles ne obeissent pas a raison car ne peust croistre plus nemoins selon la reigle de raison ne plus ou moins digerez et aussi les senscorporelz ne obeissent pars a raison car il nest pas en nostre puissance ueoir plus cler ou moins cler.

Pour remédier à cela, nous avons développé un segmentateur multilingue personnalisé adapté aux langues médiévales, basé sur un modèle de classification de tokens BERT, spécialement entraîné sur diverses langues médiévales, notamment le catalan, le castillan, l'anglais, le français, l'italien et le latin. Le segmentateur produit un texte segmenté comme suit, divisant efficacement le texte en segments, facilitant l'alignement avec les segments correspondants dans les traductions et autres textes.

*Cecondement les uertuz ne peust estre es sens corporelz
car aussi que les puissances naturelles ne obeissent pas a raison
car ne peust croistre plus ne moins selon la reigle de raisonne plus ou moins digerez
et aussi les sens corporelz ne obeissent pas a raison
car il nest pas en nostre puissance ueoir plus cler ou moins cler.*

Notre présentation lors de l'atelier « Développer des modèles pour la recherche linguistique : utilisation des données d'entraînement dans des scénarios à faibles ressources » servira de plateforme pour discuter de notre méthodologie et de la logique qui sous-tend nos choix, tels que l'accent mis sur la précision plutôt que sur le rappel afin d'améliorer l'efficacité du modèle. Nous comparerons nos méthodes de segmentation des données médiévales avec plusieurs autres, en commençant par une tokenisation basée sur des expressions régulières. Nous aborderons également des défis tels que la divergence entre les ensembles de données propres et édités utilisés pour la formation et les données brutes et non éditées généralement rencontrées dans les études de manuscrits, auxquelles nous souhaitons appliquer notre modèle : nous avons commencé notre travail sur un corpus principalement publié, avant de l'ouvrir aux sources acquises par ATR.

Nos conclusions devraient faire progresser les études médiévales et fournir des informations applicables à des défis similaires dans d'autres domaines de recherche linguistique à faibles ressources.

Références

- M. Artetxe & H. Schwenk (2019). Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual Transfer and Beyond. Transactions of the Association for Computational Linguistics 7: 597–610. https://doi.org/10.1162/tacl_a_00288
- O. Kraif (2024). Adaptative Bilingual Aligning Using Multilingual Sentence Embedding. Preprint. arXiv. URL: <http://arxiv.org/abs/2403.11921>.
- L. Liu & M. Zhu (2023). Bertalign: Improved Word Embedding-Based Sentence Alignment for Chinese–English Parallel Corpora of Literary Texts. *Digital Scholarship in the Humanities* 38(2.2): 621–634. <https://doi.org/10.1093/lc/fqac089>
- C. Meinecke, D. J. Wrisley, & S. Jänicke (2020). Automated Alignment of Medieval Text Versions Based on Word Embeddings. Preprint. Open Science Framework. <https://doi.org/10.31219/osf.io/tah3y>

Patrizia Giuliano (Università degli Studi di Napoli 'Federico II', Dipartimento Studi Umanistici) & Simona Anastasio (Université de Toulouse Jean Jaurès, LNPL)

Italian and English perspectives in narratives: evidence from retellings of native speakers and L2 English learners

Keywords: perspective; thinking for speaking; narratives; morphosyntax; English L2

Abstract

Research on perspective in narratives has demonstrated that grammaticalized features engender crucial differences in the rhetorical stratagems speakers pick up when translating their thought into verbal language (cf. Carroll & von Stutterheim 2003, Carroll *et al.* 2008). For example, English natives tend to adopt a deictic perspective via the grammaticalized progressive aspect. Conversely, German speakers opt for an anaphoric perspective anchored to the completion of the previous event (Carroll *et al.* 2004). However, literature has focused much more on textual dissimilarities (in L1 and L2) as determined by grammars and much less on the enunciative and pragmatic variables (except for Gagliardelli 1999, Adamczewski 2002).

This study aims at illustrating some peculiarities of the Italian and English narrative perspectives in semi-spontaneous retellings produced by native speakers of both languages and by advanced (C1 level of the CEFRL) Italian learners of L2 English (16 informants per group). The data were collected by means of a short film titled *Quest* (T. Stellmach, 1997). The informants watched the film twice, the first time in its entirety, the second by three steps. This second time, each subject told the part of the film they had just watched to a naïve listener. We adopted a functional and discursive framework (cf. the *Quaestio* theory by Klein & von Stutterheim 1989, 1991 and the *Thinking for Speaking Hypothesis* by Slobin 1987, 1996, 2003) as well as an enunciative approach (Gagliardelli 1999, Adamczewski 2002), in order to consider the expectations of both the speaker and the listener while executing an oral task.

We examine a) the selection of verb morphology with a focus on the progressive vs non progressive aspect and its dependence on language properties and perspective and b) the morphological and syntactic strategies (main vs subordinate clauses; types of subordination) to make the transition from the foreground to the background of narration.

We aim at answering the following research questions (RQ):

RQ1: will Italian speakers ever select a deictic perspective such as the one picked up by English native speakers or will they prefer an alternative perspective?

RQ2: will they encode the alternation between the foreground and the background of narration by similar morphological and/or syntactic principles?

As to our L2 data, we finalize our exploratory analysis concerning the two basic RQ:

RQ3: have learners identified the thinking for speaking of the L2?

RQ4: is there any conceptual and linguistic transfer from L1 into L2?

In typological terms, results show that, even if English and Italian have similar tempo-aspectual markings (progressive), they can work very differently in the way information is 'packaged' and

organised in narrations, due to reasons going beyond grammar (Gagliardelli 1999). Indeed, perspective seems also to be shaped and moulded by interactive factors (objective vs subjective perspective) and textual genres.

As for L2 data, results show that learners cannot capture some of the peculiarities of the L2: this triggers either a transfer from their L1 narrative perspective in the L2 or a “mixed” narrative perspective combining both elements of the L1 and the L2.

Mots-clés : perspective ; thinking for speaking ; récits ; morphosyntaxe ; anglais L2

Résumé

Les recherches sur la perspective dans les récits ont démontré que les caractéristiques grammaticalisées engendrent des différences cruciales dans les stratégies rhétoriques que les locuteurs adoptent lorsqu'ils traduisent leur pensée verbale en langage (cf. Carroll & von Stutterheim 2003, Carroll et al. 2008). Par exemple, les locuteurs natifs anglais tendent à adopter une perspective déictique via l'aspect progressif grammaticalisé. À l'inverse, les locuteurs allemands optent pour une perspective anaphorique ancrée dans l'achèvement de l'événement précédent (Carroll et al. 2004). Cependant, la littérature s'est beaucoup plus concentrée sur les dissimilarités textuelles (en L1 et L2) telles que déterminées par les grammaires et bien moins sur les variables énonciatives et pragmatiques (sauf pour Gagliardelli 1999, Adamczewski 2002).

Cette étude vise à illustrer certaines particularités des perspectives narratives italienne et anglaise dans des récits semi-spontanés produits par des locuteurs natifs des deux langues et par des apprenants italiens avancés d'anglais L2 (niveau C1 du CECR) (16 informateurs par groupe). Les données ont été collectées au moyen d'un court film intitulé *Quest* (Bellmach, 1997). Les informateurs ont regardé le film dans son intégralité une première fois, puis une seconde fois par trois étapes. Cette seconde fois, chaque sujet a raconté la partie du film qu'il venait de regarder à un auditeur naïf. Nous avons adopté un cadre fonctionnel et discursif (cf. la théorie *Quaestio* de Klein & von Stutterheim 1989, 1991 et l'hypothèse *Thinking for Speaking* de Slobin 1987, 1996, 2003) ainsi qu'une approche énonciative (Gagliardelli 1999, Adamczewski 2002), afin de considérer les attentes de l'orateur et de l'auditeur tout en exécutant une tâche orale.

Nous examinons a) la sélection de la morphologie verbale avec un focus sur l'aspect progressif vs non progressif et sa dépendance aux propriétés linguistiques et à la perspective et b) les stratégies morphologiques et syntaxiques (propositions principales vs subordonnées ; types de subordination) pour effectuer la transition du premier plan vers l'arrière-plan de la narration.

Nous visons à répondre aux questions de recherche suivantes (QR) : QR1 : les locuteurs italiens sélectionneront-ils jamais une perspective déictique telle que celle adoptée par les locuteurs natifs anglais ou préféreront-ils une perspective alternative ? QR2 : encoderont-ils l'alternance entre le premier plan et l'arrière-plan de la narration selon des principes morphologiques et/ou syntaxiques similaires ?

Concernant nos données L2, nous finalisons notre analyse exploratoire concernant les deux QR de base : QR3 : les apprenants ont-ils identifié le *thinking for speaking* de la L2 ? QR4 : y a-t-il un transfert conceptuel et linguistique de la L1 vers la L2 ?

En termes typologiques, les résultats montrent que, même si l'anglais et l'italien ont des marquages tempo-aspectuels similaires (progressif), ils peuvent fonctionner très différemment dans la façon dont l'information est « emballée » et organisée dans les récits, en raison de motifs allant au-delà de la grammaire (Gagliardelli 1999). En effet, la perspective semble aussi être façonnée et modelée par des facteurs interactifs (perspective objective vs subjective) et des genres textuels.

Concernant les données L2, les résultats montrent que les apprenants ne peuvent pas saisir certaines des particularités de la L2 : cela déclenche soit un transfert de leur perspective narrative L1 dans la L2, soit une perspective narrative « mixte » combinant à la fois des éléments de la L1 et de la L2.

Références

- Adamczewski, H. 2002. *The secret architecture of English grammar*. Prény sur Oise: EMA.
- Carroll, M. & von Stutterheim, C. 2003. Typology and information organisation: perspective taking and language specific effects in the construal of events. In: Giacalone Ramat, A. (ed.), *Typology and Second Language Acquisition*. Berlin: Mouton de Gruyter, pp. 365-402.
- Carroll, M., von Stutterheim, C. & Nüse, R. 2004. The thought and language debate: a psycholinguistic approach. In: Pechman, T. & Habel, C. (eds.), *Multidisciplinary Approaches to Language Production*. Berlin/New York: Mouton de Gruyter, pp. 184-218.
- Carroll, M., Lambert, M., von Stutterheim, C. & Rossdeutscher, A. 2008. Subordination in narratives and macrostructural planning: taking a comparative point of view. In: Fabricius Hansen C. & Ramm, M. (eds.), *'Subordination' versus 'Coordination' in Sentence and Text*. Amsterdam: John Benjamins, pp. 161-184.
- Gagliardelli, G. 1999. *Elementi di grammatica enunciativa della lingua inglese*. Bologna: CLUEB.
- Klein, W. & von Stutterheim, C. 1989. Referential movement in descriptive and narrative discourse. In: R. Dietrich, R. & Graumann, C. F. (eds.), *Language Processing in Social Context*. Amsterdam: Elsevier Science Publishers B.V, pp. 39-76.
- Klein, W. & von Stutterheim, C. 1991. Text structure and referential movement. *Sprache und Pragmatik*, 22: 1-32.
- Slobin, D. I. 1987. Learning to think for speaking. *Pragmatics* 1(1): 7-25.
- Slobin, D. I. 1996. From 'Thought and language' to 'thinking for speaking'. In: Gumperz, J. & Levinson, S. (eds.), *Rethinking Linguistic Relativity: Vol. Studies in the social and cultural foundations of language*. New York: Cambridge University Press, pp. 70-96.
- Slobin, D. I. 2003. Language and thought online: cognitive consequences of linguistic relativity. In: Gentner, D. & Goldin-Meadow, S. (eds.), *Advances in the Investigation of Language and Thought*. Cambridge MA: MIT Press, pp. 157-192.

Matilde Gonçalves (NOVA FCSH)

Cinq regards sur la mort : des représentations sociales et individuelles au développement de la personne au long de la vie

Mots-clés : mort/vie, représentations, développement de la personne, médiation, langage

Résumé

La présente communication s'intègre dans un projet de recherche, en phase initiale, à être développé au Portugal, qui vise à étudier, au niveau scientifique, d'une part comment la personne se dit et construit son identité narrative (Ricoeur, 1990 , 1991), et, d'autre part, les configurations linguistiques de l'ordre de l'exposer et du raconter (Bronckart, 1997). Au niveau social, ce projet prétend valoriser les histoires de vie de personnes âgées et/ou en fin de vie par leurs récits de vie et la création de leur héritage mémorial et social.

Deux présupposés sont sous-jacents à ce travail: c'est au niveau discursif et textuel que se situent les mécanismes de création de sens et les transactions incessantes entre représentations collectives et individuelles (Durkheim, 1898) ; le langage est médiation dans la prise de conscience du vécu (Vygotsky, 1997) et contribue à un apprentissage transformateur (Mezirow, 1997).

Dans cette communication, les objectifs à atteindre sont en lien avec les questions suivantes : Comment les personnes disent la vie et/ou la mort ? Comment les personnes se disent face à la vie et à la mort ? Quelles représentations sociales et individuelles sous-tendent les réponses à ces questions ?

Ce travail suit une approche qualitative et interprétative (Rastier, 2001), sans pour autant mettre de côté des données quantitatives. Méthodologiquement, des entretiens semi-directifs (Sauvayre, 2021) seront menés auprès de cinq personnes dans des situations distinctes face à la vie et à la mort (i. infirmier.e en soins palliatifs et doula de fin de vie ; ii : étudiant.e en médecine palliative ; iii. personne faisant du volontariat auprès de personnes en fin de vie ; iv. personne avec un cancer niveau 4 ; v : personne ayant un cancer et souffrant d'un burn-out). Le choix des entretiens se tient au fait de vouloir comprendre « le sens que les acteurs donnent à leurs pratiques, aux événements dont ils ont pu être les témoins actifs » Blanchet et Gotman (2007 : 24).

Partant des données obtenues par les entretiens, et en vue de répondre aux questions initiales, l'analyse se fera au niveau du contenu thématique et du champ lexical, des mécanismes de responsabilité énonciative et de l'organisation temporelle.

À cette étape de la recherche, il n'y a pas encore de résultats concrets. Toutefois, notre idée, à ce stade, est de mettre en place un dispositif afin d'étudier la construction d'identité face à des situations de grande vulnérabilité par l'utilisation des mécanismes linguistiques et textuels, comprendre les représentations individuelles et collectives ayant trait à la mort et à son rapport à la vie et contribuer ainsi à une meilleure compréhension du développement de la personne tout au long de sa vie.

Keywords: death/life; representations; personal development; mediation; language

Abstract

This presentation is part of an initial-phase research project, to be developed in Portugal, which scientifically aims to study, on the one hand, how individuals express themselves and construct their narrative identity (Ricoeur, 1990, 1991), and, on the other hand, the linguistic configurations of exposing and telling (Bronckart, 1997). At a social level, this project seeks to valorize the life stories of elderly people and/or people at the end of life through their life narratives and the creation of their memorial and social legacy.

Two premises underlie this work: it is at the discursive and textual level that the mechanisms of meaning-making and the ongoing transactions between collective and individual representations are situated (Durkheim, 1898); language functions as mediation in the awareness of lived experience (Vygotsky, 1997) and contributes to transformative learning (Mezirow, 1997).

The objectives of this presentation are linked to the following questions: How do people speak about life and/or death? How do people define themselves in relation to life and death? What social and individual representations underlie their responses to these questions?

This work adopts a qualitative and interpretive approach (Rastier, 2001), while also integrating quantitative data. Methodologically, semi-structured interviews (Sauvayre, 2021) will be conducted with five people in different situations regarding life and death: (i) a palliative care nurse and end-of-life doula; (ii) a student in palliative medicine; (iii) a person volunteering with individuals at the end of life; (iv) a person with stage 4 cancer; (v) a person with cancer also experiencing burnout. The choice of interviews is motivated by the desire to understand *“the meaning that actors give to their practices, to the events of which they may have been active witnesses”* (Blanchet & Gotman, 2007: 24).

Based on the interview data, and in order to answer the initial questions, the analysis will focus on thematic content and lexical fields, mechanisms of enunciative responsibility, and temporal organization.

At this stage of the research, there are no concrete results yet. However, our aim is to establish a framework for studying identity construction in situations of great vulnerability through linguistic and textual mechanisms, to understand the individual and collective representations relating to death and its connection to life, and thus to contribute to a better understanding of personal development throughout life.

Références

Blanchet A. et Gotman A. (2007). L'entretien. Paris : Armand Colin, coll. « 128 » (2e éd. 2015).

Bronckart, J.-P. (1997). Activité de langage, textes et discours. Pour un interactionnisme socio-discursif. Neuchâtel: Delachaux et Niestlé.

Durkheim, É. (1898). Représentations individuelles et représentations collectives. In: Revue de Métaphysique et de Morale, tome VI. In

https://classiques.uqam.ca/classiques/Durkheim_emile/Socio_et_philo/ch_1_representations/representations.html (date de consultation: 28 mars de 2025).

Mezirow, J. (1997). Transformative learning: Theory to practice. In P. Cranton (Ed.), Transformative learning in action: Insights from practice (pp. 5-12). Jossey-Bass.

Rastier, F. (2001). Arts et Sciences du Texte. Paris: PUF.

Ricoeur, P. (1990). Soi-même comme un autre. Paris : Éditions du Seuil.

Ricoeur, P. (1991). Temps et Récit I. Paris : Points, Essais.

Sauvayre, R. (2021). Initiation à l'entretien en sciences sociales - 2e éd. Méthodes, applications pratiques et QCM. Paris : Armand Colin.

Vygotsky, L. S. (2001). Linguagem e pensamento (3^a ed., M. da G. C. Matos, Trans.). Relógio D'Água.

Daria Grigoreva (University of the Basque Country)

From tenseless to tensed? On nominal tense markers in the Mataco-Guaicuru languages of South America

Keywords: nominal tense; grammaticalisation; demonstratives; language change; evidentiality

Abstract

Some Mataco-Guaicuru languages of South America are reported to possess so-called nominal tense - morphological tense encoded in non-verbal domain (e.g. Caminsky (2011) for Wichí, Vidal and Gutiérrez (2010) for Pilagá, Wichí and Nivacle). Nominal tense, a cross-linguistically rare phenomenon, was first introduced and studied by Nordlinger and Sadler (2004). The authors identified two types of nominal tense: independent nominal tense, where tense is encoded within the nominal itself and functions independently of clausal tense, and propositional nominal tense, where the tense marker affects the entire proposition, working solely or in conjunction with verbal tense. Campbell and Grondona (2012: p. 631) illustrate the propositional nominal tense in Nivacle using the following examples:

(1) Nivacle (Campbell and Grondona, 2012: p. 631)

- a. tsex **na** towak
 grow **DEM.CLOSE** river
 'the river is rising'
- b. tsex **xa** towak
 grow **DEM.ABSENT** river
 'the river rose'

Nivacle verbs, like those in other Mataco-Guaicuru languages, lack tense morphology. Examples (1a) and (1b) show that temporal distinctions rely on the demonstratives *na* and *xa*. However, this is a context-dependent interpretation, as these markers primarily indicate spatial deixis (proximity or absence) and evidentiality (visibility). Tense marking is a secondary feature they may encode. This study explores whether Nivacle and related languages are developing an unusual non-verbal tense-marking system and how this process might unfold.

The comparison with other languages within the family suggests that tense is not fully grammaticalised in these demonstratives but may be undergoing a process of grammaticalisation. The Mataco-Guaicuru languages exhibit strong similarities in their demonstrative systems, enabling comparative analysis to investigate how this phenomenon may have emerged. This study examines the meanings and functions of several cognate morphs within the language family. The main data sources used for the study are reference grammars and glossed texts.

Consider examples (2a), (2b), and (2c) from Toba, where demonstratives similarly trigger clausal temporal distinctions, as observed with the Nivacle cognates above. Messineo and Cúneo (2019) describe the temporal distinctions in these examples as arising purely from pragmatic inference. However, they might also be analysed as instances of propositional nominal tense, similar to the Nivacle examples (1a) and (1b).

(2) Toba (Messineo and Cúneo, 2019: p. 208)

- a. n-vi? **na** ʔad-qaya
 3-come **DEM.CLOSE** POS2-brother
 ‘Your brother is coming closer’.
- b. n-vi? **ka** ʔad-qaya
 3-come **DEM.ABSENT** POS2-brother
 ‘Your brother will come’.
- c. n-vi? **so** ʔad-qaya
 3-come **DEM.DISTAL** POS2-brother
 ‘Your brother (just) came’.

The comparative analysis provided in the study highlights ambiguous contexts where a shift in the scope of demonstratives might occur. Additionally, crosslinguistic parallels are drawn to illustrate similar developments in other languages. As a result of the analysis, a potential diachronic sequence is proposed. A comparative analysis of Mataco-Guaicuru cognates provides an initial exploration of how nominal tense may be developing, shedding light on the mechanisms and features that could drive its emergence. The study demonstrates how the interaction between spatial deixis and evidentiality in tenseless languages generates temporal implications and ambiguous contexts, ultimately leading to the potential development of propositional tense marking.

Mots-clés: temps nominal; grammaticalisation; démonstratifs; changement de langue; évidence

Résumé

Certaines langues mataco-guaicuru d'Amérique du Sud posséderaient un temps nominal, un temps morphologique encodé dans le domaine non verbal (par exemple, Caminsky (2011) pour le wichí, Vidal et Gutiérrez (2010) pour le pilagá, le wichí et le nivaclé). Le temps nominal, phénomène translinguistique rare, a été introduit et étudié pour la première fois par Nordlinger et Sadler (2004). Ces auteurs ont identifié deux types de temps nominaux : le temps nominal indépendant, où le temps est encodé dans le nominal lui-même et fonctionne indépendamment du temps de la proposition, et le temps nominal propositionnel, où le marqueur de temps affecte la proposition entière, fonctionnant seul ou en conjonction avec le temps verbal. Campbell et Grondona (2012 : p. 631) illustrent le temps nominal propositionnel en nivaclé à l'aide des exemples suivants :

(1) Nivaclé (Campbell and Grondona, 2012: p. 631)

- a. tsex **na** towak
 monte **DEM.CLOSE** rivière
 ‘la rivière monte’
- b. tsex **xa** towak
 monte **DEM.ABSENT** rivière
 ‘la rivière est montée’

Les verbes en nivacilé, comme ceux des autres langues mataco-guaicuru, sont dépourvus de morphologie des temps. Les exemples (1a) et (1b) montrent que les distinctions temporelles reposent sur les démonstratifs *na* et *xa*. Cependant, cette interprétation dépend du contexte, car ces marqueurs indiquent principalement la deixis spatiale (proximité ou absence) et l'évidentialité (visibilité). Le marquage des temps est une caractéristique secondaire qu'ils pourraient coder. Cette étude examine si le nivacilé et les langues apparentées développent un système de marquage des temps non verbal inhabituel et comment ce processus pourrait se dérouler.

La comparaison avec d'autres langues de la famille suggère que le temps n'est pas entièrement grammaticalisé dans ces démonstratifs, mais pourrait être en cours de grammaticalisation. Les langues mataco-guaicuru présentent de fortes similitudes dans leurs systèmes démonstratifs, ce qui permet une analyse comparative pour comprendre comment ce phénomène a pu émerger. Cette étude examine les significations et les fonctions de plusieurs formes apparentées au sein de la famille linguistique. Les principales sources de données utilisées pour cette étude sont les grammaires de référence et les textes glosés.

Considérons les exemples (2a), (2b) et (2c) du toba, où les démonstratifs déclenchent de la même manière des distinctions temporelles, comme dans le cas des mots apparentés du nivacilé ci-dessus. Messineo et Cúneo (2019) décrivent les distinctions temporelles dans ces exemples comme découlant purement de l'inférence pragmatique. Cependant, ils pourraient également être analysés comme des cas de temps nominal propositionnel, similaires aux exemples du nivacilé (1a) et (1b).

(2) Toba (Messineo and Cúneo, 2019: p. 208)

- a. n-vi? **na** ʔad-qaya
 3-venir **DEM.CLOSE** POS2-frère
 'Ton frère se rapproche'.
- b. n-vi? **ka** ʔad-qaya
 3-venir **DEM.ABSENT** POS2-frère
 'Ton frère viendra'.
- c. n-vi? **so** ʔad-qaya
 3-venir **DEM.DISTAL** POS2-frère
 'Ton frère viendra'.

L'analyse comparative présentée dans cette étude met en évidence des contextes ambigus où un changement dans la portée des démonstratifs pourrait se produire. De plus, des parallèles interlinguistiques sont établis pour illustrer des développements similaires dans d'autres langues. Suite à cette analyse, une séquence diachronique potentielle est proposée. Une analyse comparative des cognats du mataco-guaicuru offre une première exploration de l'évolution possible du temps nominal, éclairant les mécanismes et les caractéristiques qui pourraient favoriser son émergence. L'étude démontre comment l'interaction entre la deixis spatiale et l'évidentialité dans les langues sans temps génère des implications temporelles et des contextes ambigus, conduisant finalement au développement potentiel du marquage des temps propositionnels.

Références

- Caminsky, D. W. (2011). Temporal Markers in Wichí. *Santa Barbara Working Papers in Linguistics*, 22, 1–15.
- Campbell, L., & Grondona, V. (2012). Languages of the Chaco and Southern Cone. In *The Indigenous Languages of South America* (pp. 625–668). De Gruyter Mouton.
- Messineo, C., & Cúneo, P. (2019). Multifunctionality of deictic classifiers in the Toba language (Guaycuruan). In A. Y. Aikhenvald & E. I. Mihas (Eds.), *Genders and Classifiers: A Cross-Linguistic Typology* (pp. 197–221). Oxford University Press.
- Nordlinger, R., & Sadler, L. (2004). Nominal Tense in Crosslinguistic Perspective. *Language*, 80(4), 776–806.
- Vidal, A., & Gutiérrez, A. (2010). La categoría de ‘tiempo nominal’ en las lenguas chaqueñas. In *La renovación de la palabra en el bicentenario de la Argentina. Los colores de la mirada lingüística* (pp. 1347–1355). FFyL, UNCuyo.

Nicolas Guilliot (Université Bordeaux-Montaigne)

*Approche interdisciplinaire sur les langues régionales de Nouvelle-Aquitaine:
Variations, représentations sociales, cartographie, et transmission*

Mots-clés : syntaxe, variation, cartographie, sociolinguistique, didactique

Résumé

L'objectif de notre présentation sera de présenter les axes scientifiques, le matériel d'enquêtes et les premiers résultats du projet région PaRL (Patrimoine Régional Linguistique) sur les langues régionales en Nouvelle-Aquitaine. Le projet (<https://projetparl.hypotheses.org/>) vise à croiser trois niveaux d'étude, (i) description de la variation linguistique, (ii) représentations sociales, et (iii) enjeux didactiques, à partir d'un terrain propice puisque cette région fait apparaître des langues régionales aux statuts et usages très variés, dont l'occitan (gascon), le poitevin-saintongeais, et les parlers du croissant dans le domaine gallo-roman.

Un point fort du projet consiste à se focaliser sur la variation syntaxique, moins documentée que les autres formes de variation. Les objets intégrés dans l'étude, sensibles à la variation et pertinents pour les langues considérées, incluent la négation (cf. (1), et Guilliot & Becerra-Zita (2020) entre autres), le marquage différentiel de l'objet (cf. (3), et Dagnac & Thuillier (2020) entre autres), le temps (cf. (3), et Bras & Sibille (2021) entre autres), et les structures interrogatives (cf. (4), et Dagnac (2018) entre autres).

1) ?*l'aimé reque li et le diset (pouet) ren a prsouné. (Poitevin-Saintongeais, PaRL_85115_01)*

« Il n'aimait que lui-même, et ne disait rien à personne. »

(2) *Ce qui m'intéresse à moi, c'est pédagogiquement, d'essayer de répondre à ça. (Français régional, Fagard & Mardale (2014))*

(3) *L'ai vist de matin VS Lo vegèri ièr. (Occitan, Bras & Sibille (2021))*

« Je l'ai vu ce matin » VS « Je le vis hier. »

(4) *I sé trjou pa prqua (que) le m'avant atachae. (Poitevin-Saintongeais, PaRL_85115_01)*

« Je ne sais toujours pas pourquoi on m'a attaché. »

L'autre point fort réside dans l'interdisciplinarité de l'approche, permettant d'étudier la variation linguistique selon différents plans (description, représentations sociales des témoins, et approche pédagogique de la variation) en les corrélant.

La récolte de données combine enquêtes de terrain et en ligne, à travers la mise en place d'une plateforme (cf. Fig. 1 ; mise en ligne prochaine) permettant de fidéliser les témoins (création de comptes) sur les questionnaires mis en place durant le projet. L'étude linguistique se construit autour d'un conte à traduire (cf. Fig. 3), inspiré de celui développé dans le projet ANR SyMiLa, et incluant également le recensement des jugements des témoins sur les variations possibles. Ces jugements sont mis en regard avec une enquête sociolinguistique constituée d'un entretien semi-dirigé et d'assertions permettant d'évaluer le degré d'adhésion (cf. Fig. 4). Enfin, la variation linguistique et sa prise en compte sont également étudiées dans la pédagogie, avec une enquête ciblant les formateurs.rices.

Nous évoquerons enfin notre partenariat avec le projet de transcription de l'enquête Bourciez porté par Alexandre Génadot (cf. carte sur l'usage de la négation dans l'enquête Bourciez en Fig. 2), et le projet d'Atlas des Langues Régionales de France porté par Philippe Boula de Mareuil (cf. Fig. 1 avec intégration des textes supports des deux projets).

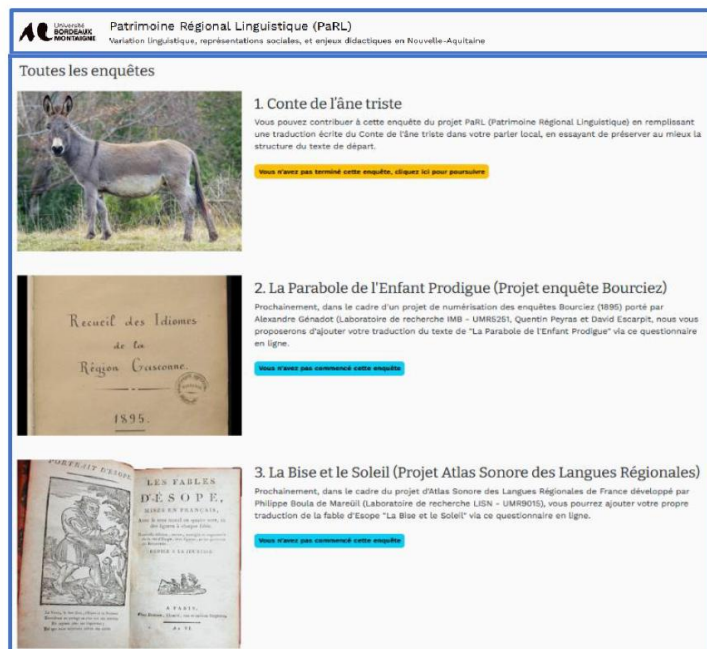


Fig. 1 : Aperçu de la plateforme en ligne PaRL

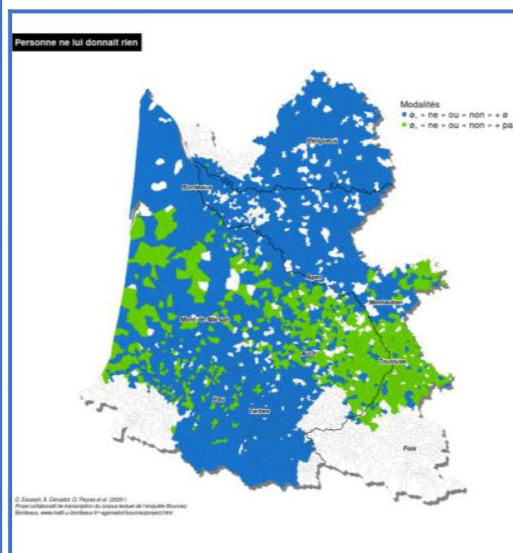


Fig. 2 : Carte linguistique enquête Bourciez

N°	Phrases à faire traduire		La phrase "modifiée" sur NEG/MDO/TPS/QU est:			
			Acceptabilité variations NEG-POL	Acceptabilité variations MDO	Acceptabilité variations TPS	Acceptabilité variations QU
			0-Spontané 1-Possible, et je l'utilise 2-Possible, mais je ne l'utiliserais pas 3-très étrange, impossible dans mon parler			
	Conte de l'âne triste	Réponse ou Traduction proposée				
1	C'est l'histoire d'un âne triste, pas très grand.	Réponse spontanée	0	0	0	0
		Réponse spontanée	0	0	0	0
2	Ce faisait très longtemps qu'il avait pas quitté son village.	Modifier avec/sans négation pas /point Modifier avec/sans MDO (cible: qu'il n'avait pas quitté à son village)	0	0	0	0
		Réponse spontanée	0	0	0	0
3	Il ne pouvait aller presque nulle part, parce qu'on l'avait attaché à un piquet.	Modifier avec/sans négation pas (cible: pouvait (pas) aller presque nulle part) Modifier avec/sans MDO (cible: on n'avait attaché à lui à un piquet)	0	0	0	0
		Réponse spontanée	0	0	0	0
4	Il n'aimait que lui-même, et ne disait rien à personne.	Modifier avec/sans négation pas (cible: n'aimait (pas) que lui-même) Modifier avec/sans négation pas (cible: ne disait (pas) rien à personne) Modifier avec/sans MDO (cible: il ne (s) disait qu'à lui-même)	0	0	0	0

Fig. 3 : Extrait du Conte de l'âne triste à traduire

18. Certaines variétés du [basque/poitevin-saintongeais/occitan] semblent plus 'correctes' que d'autres.

- ☐ Tout à fait d'accord
- ☐ D'accord
- ☐ Plutôt d'accord
- ☐ Indifférent
- ☐ Plutôt pas d'accord
- ☐ Pas du tout d'accord

Fig. 4 : Extrait de l'enquête sociolinguistique

Keywords: syntax, variation, mapping, sociolinguistics, didactics

Abstract

The aim of our presentation is to outline the scientific approaches, survey materials, and first results of the PaRL project (Patrimoine Régional Linguistique) about regional languages in Nouvelle-Aquitaine. The project (<https://projetparl.hypotheses.org/>) combines three levels of investigation: (i) the description of linguistic variation, (ii) social representations, and (iii) didactic challenges. This region provides a particularly rich field site, as it includes regional languages with highly diverse statuses and uses, such as Occitan (Gascon), Poitevin-Saintongeais, and the Croissant dialects within the Gallo-Romance area.

A major focus of the project is on syntactic variation, which has been much less documented than other types of variation. The phenomena examined – chosen for their sensitivity to variation and their relevance to the languages under study – include negation (cf. (1), and Guilliot & Becerra-Zita 2020, among others), differential object marking (cf. (2), and Dagnac & Thuillier 2020, among others), tense (cf. (3), and Bras & Sibille 2021, among others), and interrogative structures (cf. (4), and Dagnac 2018, among others).

1) *?l'aimé reque li et le diset (**pouet**) ren a prsouné. (Poitevin-Saintongeais, PaRL_85115_01)*

“He loved only himself and said nothing to anyone.”

(2) *Ce qui **m'intéresse à moi**, c'est pédagogiquement, d'essayer de répondre à ça. (Français régional, Fagard & Mardale (2014))*

“What interests me, pedagogically, is trying to respond to that.”

(3) *L'**ai vist** de matin VS Lo **vegèri** ièr. (Occitan, Bras & Sibille (2021))*

“I saw him this morning.” vs. “I saw him yesterday.”

(4) *I sé trjou pa prqua (**que**) le m'avant atachae. (Poitevin-Saintongeais, PaRL_85115_01)*

“I still don't know why they tied me up.”

Another key strength lies in the interdisciplinary nature of the approach, which makes it possible to study linguistic variation along different dimensions (description, speakers' social representations, and pedagogical approaches to variation) while correlating them.

Data collection combines fieldwork and online surveys, through the implementation of a platform (see Fig. 1; soon to be launched) designed to retain participants (via account creation) for the questionnaires deployed during the project. The linguistic study is built around a translation task of a folktale (see Fig. 3), inspired by the one developed in the ANR SyMiLa project, and also includes the collection of speakers' judgments about possible variants. These judgments are compared with a sociolinguistic survey consisting of a semi-structured interview and statements designed to assess the degree of agreement (see Fig. 4). Finally, linguistic variation and its treatment are also studied in pedagogy, through a survey targeting trainers.

We will also discuss our partnerships with the Bourciez survey transcription project led by Alexandre Génadot (see map of negation use in the Bourciez survey, Fig. 2), and with the Atlas of Regional Languages of France project led by Philippe Boula de Mareüil (see Fig. 1, which integrates supporting texts from both projects).

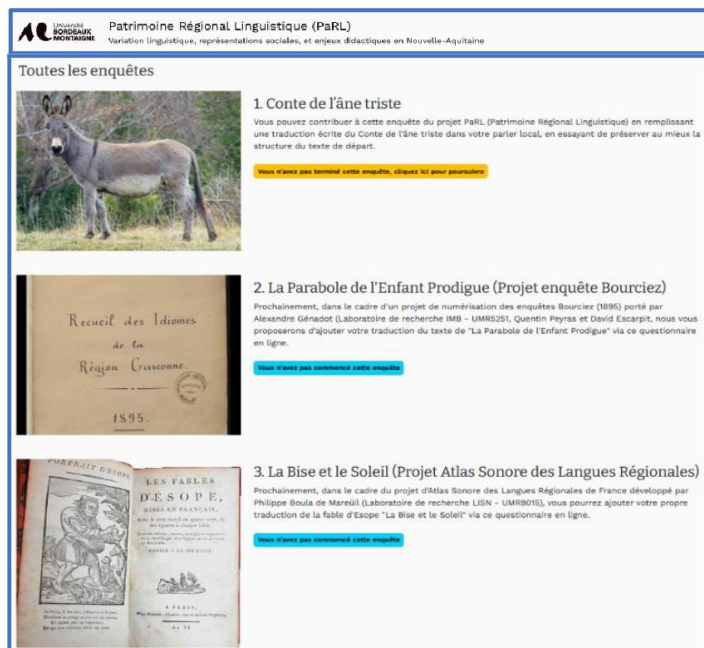


Fig. 1: Overview of the PaRL online platform

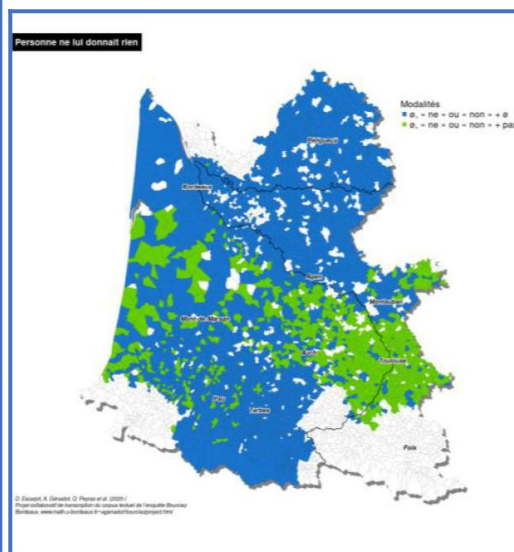


Fig. 2: Linguistic map from the Bourciez survey

N°	Phrases à faire traduire	Réponse ou Traduction proposée	La phrase "modifiée" sur NEG/MDO/TPS/QU est:			
			0 - Spontanée	1 - Possible, et je l'utilise	2 - Possible, mais je ne l'utiliserais pas	3 - Très étrange, impossible dans mon parler
	Conte de l'âne triste		Acceptabilité variations NEG-POL	Acceptabilité variations MDO	Acceptabilité variations TPS	Acceptabilité variations QU
1	C'est l'histoire d'un âne triste, pas très grand.	Réponse spontanée	0	0	0	0
2	Ca faisait très longtemps qu'il avait pas quitté son village.	Réponse spontanée	0	0	0	0
		Modifier avec/sans négation pas/point				
		Modifier avec/sans MDO (cible: qu'il n'avait pas quitté à son village)				
3	Il ne pouvait aller presque nulle part, parce qu'on l'avait attaché à un piquet.	Réponse spontanée	0	0	0	0
		Modifier avec/sans négation pas (cible: pouvait (pas) aller presque nulle part)				
		Modifier avec/sans MDO (cible: on l'avait attaché à un piquet)				
4	Il n'aimait que lui-même, et ne disait rien à personne.	Réponse spontanée	0	0	0	0
		Modifier avec/sans négation pas (cible: n'aimait (pas) que lui-même)				
		Modifier avec/sans négation pas (cible: ne disait (pas) rien à personne)				
		Modifier avec/sans MDO (cible: il ne s'aimait qu'à lui-même)				

Fig. 3: Excerpt from *The Sad Donkey* tale to be translated

18. Certaines variétés du [basque/poitevin-saintongeais/occitan] semblent plus 'correctes' que d'autres.

☐ Tout à fait d'accord

☐ D'accord

☐ Plutôt d'accord

☐ Indifférent

☐ Plutôt pas d'accord

☐ Pas du tout d'accord

Fig. 4: Excerpt from the sociolinguistic survey

Références

Bras, M. & Sibille, J. (2021). « Preterit and perfect in Romance: new insights from Occitan », dans Louis de Saussure et Laura Baranzini (eds.) *Aspects of Tenses, Modality and Evidentiality*, Leiden/Boston : Koninklijke Brill NV, pp. 136-161.

Dagnac, A. (2018). SyMiLa and the *Atlas linguistique de la France*: A tool for the study of Gallo-Romance syntax. *Glossa: a journal of general linguistics* 3(1): 85. 1–23, <https://doi.org/10.5334/gigl.543>

Dagnac, A. & Thuilier, J. (2020). « Français du Sud-Ouest et marquage différentiel de l'objet. » *L'information grammaticale*, 166, pp.29-35.

Fagard, B. et Mardale, A. (2014). « Non, mais tu l'as vu à lui ? Analyse(s) du marquage différentiel de l'objet en français », *Verbum* (Presses Universitaires de Nancy), Université de Nancy II, vol. 26 (1), pp.145-170.

Guilliot, N. & Becerra-Zita, S. (2020) « Concordance négative en Gallo : qui n'y comprend (pas) rien » dans *La Microvariation Syntaxique dans les Langues Romanes de France*, édité par J. Sibille, publié par Lambert-Lucas, Limoges.

Abstract

(4) **Question:** weš Klit-i Possible Answers : ħatta haja Or haja
 what ate-2pers.past even thing thing
 'what did you eat?' 'Nothing' 'something'

105

- (5) *Reda ma šaf š Sami*
 Reda NEG Saw NEG Sami
 i. $\rightarrow \text{see } (r, a)$
 ii. [TP Reda [T' ma-saf [NegP š [Neg' ~~ma-šaf~~ [VP Reda [V' ~~šaf~~ Ali]]]]]]]

The analysis. I propose a similar treatment of bare NPs and N-words as existential generalized quantifiers that merge within the VP domain. At this level, they compose with a covert negation operator (CN) via function composition, resulting in a negative quantifier. Following Chomsky's (1998) Agree operation, the formed negative quantifier, bearing interpretable negative features, undergoes a (wh-like) movement to the specifier of NegP (then to higher projections). This movement is driven by *ma*, located in NegP°, which carries uninterpretable negative features. Importantly, CN is not available in the presence of *š*, because the Agree operation that licenses it cannot take place when Spec-NegP is occupied. Also, CN is not available for other quantifiers like universals, since the lexicalization of negative quantifiers is highly constrained (see Horn (1973), NALL lexical gap).

Preverbal negative quantifiers (example (1-2)). The absence of *š* (example (1)) enables CN to adjoin to the phrase [(*ħatta*) *dajf*] in Spec-VP. Subsequently, the complex phrase [(*ħatta*) *dajf* CN] undergoes movement: first to Spec-NegP, attracted by *ma* located in NegP°, and then to Spec-TP, most likely for case, as shown in (6). At LF, CN combines with the existential quantifier contributed by [(*ħatta*) *dajf*] through function composition, as follows: $\llbracket \text{CN} \rrbracket \circ \llbracket ((\hbar\text{atta}) \text{dajf}) \rrbracket = \lambda P. \neg \exists x [\text{guest}(x) \ \& \ P(x)]$. The resulting negative quantifier undergoes Quantifier Raising (QR), leaving behind a trace, t_1 , which yields the structure [1 t_1 *kla*]. At the sentence level, the meaning of the composed negative quantifier combines with this structure via function application, resulting in the meaning in (1i).

- (6) [TP [(*ħatta*) *dajf* CN] [T' *ma-kla* [NegP [(*ħatta*) *dajf* CN] [Neg' *ma-kla* [VP [(*ħatta*) *dajf* CN] [V' *kla*]]]]]]]

In the presence of *š* (example (2)), CN is not available since it cannot be licensed via movement to Spec-NegP, which is occupied by *š*. Thus, the subject [(*ħatta*) *dajf*] merges in Spec-VP, then moves directly to Spec-TP, yielding the word order in (2). At Logical Form (LF), [(*ħatta*) *dajf*], which contributes only an existential quantifier (as CN is absent), undergoes QR and combines with [1 *š* t_1 *kla*] via function application, yielding the meaning shown in (2i), where the existential quantifier takes scope above negation. It's important to note that only bare NPs can produce this reading, as *ħatta* phrases are ungrammatical when they surface higher than *š*. I argue that this is because *ħatta* phrases, unlike bare NPs, are strong negative polarity items (NPIs) (Ladusaw, 1992; Giannakidou, 2000). Strong NPIs are those NPIs that are licensed exclusively in anti-additive environments, such as within the scope of sentential negation or without (Zwarts, 1998; Gajewski, 2011).

Postverbal existential quantifiers (example 3). As previously stated, the presence of the enclitic *š* blocks the occurrence of CN, preventing the formation of a negative quantifier. In cases like (3), the presence of *š* forces the existential quantifier to take scope below negation. This is expected because inverse scope in AA is not available. Thus, I assume that QR in Algerian Arabic is restricted to the first node of type *t* (arguably right above VP in example (3)). Notably, *ħatta* phrases are grammatical in this configuration due to their appearance within the scope of negation. As for the obligatory presence of *š* with postverbal (*ħatta*) phrases, it suggests that AA adheres to an economy principle that restricts the availability of CN to cases where a reading is impossible to derive in the presence of the enclitic *š*. That is, in the case of (3), CN would yield the same reading that we have with *š*.

Mots-clés : Négation, concordance négative, éléments de polarité négative, SN nus, quantificateurs existentiels.

Résumé

Les données. Cet article présente des données nouvelles de l'arabe algérien, démontrant que les SN nus préverbaux, tels que *kteb* « livre », et les mots-N, comme les syntagmes *ħatta* « même », présentent un comportement similaire au sein de la négation bipartite sentencielle *ma-V-š*. Spécifiquement, je montre que, contrairement au consensus dans la littérature affirmant que l'enclitique *š* ne peut être omis que lorsque la négation co-apparaît avec des mots-N comme les syntagmes *ħatta* en (1) (Benmamoun 1992, Ouhalla 2002, Soltan 2012, Al-Qassas 2015, entre autres), de nouvelles données montrent que : (i) l'enclitique *š* peut aussi être omis lorsque la négation co-apparaît avec des SN nus préverbaux, comme également montré en (1) ; (ii) dans les cas où l'enclitique est omis, la négation prend toujours une portée large sur ces SP contributeurs de quantificateurs existentiels ; et (iii) lorsque les syntagmes *ħatta* et les SN nus apparaissent en position postverbale, l'enclitique *š* retrouve son statut obligatoire, résultant également en une portée large de la négation, comme montré dans l'exemple (3). Ces similitudes frappantes sont surprenantes, considérant que les syntagmes *ħatta* et les SN nus présentent également deux différences clés : (i) les SN nus préverbaux, mais pas les syntagmes *ħatta*, sont grammaticaux avec la présence de l'enclitique *š*, donnant une lecture existentielle à portée large comme en (2), et (ii) les syntagmes *ħatta*, mais pas les SN nus, sont des mots-N qui donnent des réponses négatives fragmentaires comme montré en (4).

- (1) *(ħatta) dajf ma kla*
 even guest NEG ate
 i. $\neg \exists x$ [guest (*x*) & ate (*x*)]

- (2) *(*ħatta) dajf ma kla š*
 even guest NEG ate NEG
 i. $\exists x$ [guest (*x*) & $\neg$ ate (*x*)]

- (3) *ma kla (*š) (ħatta) dajf*
 NEG ate NEG even guest
 i. $\neg \exists x$ [guest (*x*) & ate (*x*)]

- | | | |
|--|--|---|
| (4) Question: <i>weš Klit-i</i>
what ate-2pers.past
'what did you eat?' | Possible Answers : <i>ħatta haja</i>
even thing
'Nothing' | Or <i>haja</i>
thing
'something' |
|--|--|---|

Hypothèses de base. Je suivrai la plupart des analyses et prendrai l'ordre des mots VSO dans les variétés d'arabe moderne pour constituer la configuration de phrase de base (Mohammad 1988, 1989, et travaux subséquents). Ainsi, les sujets préverbaux sont dérivés par mouvement — c'est-à-dire, ils fusionnent dans Spec-VP, puis se déplacent vers au moins Spec-TP, très probablement pour l'assignation de cas. Quant à la négation bipartite, le verbe fusionne en V° puis se déplace vers T° (pour le temps) via Neg° par mouvement de tête à tête. À Neg° , le proclitique *ma* se cliticise au verbe, formant *ma-V* avant de se déplacer ensemble vers T° . L'enclitique *š* est supposé rester dans sa position de fusion dans Spec-NegP. Ainsi, l'ordre des mots [S *ma-V-š* O] est dérivé, comme montré en (5ii) ci-dessous. Sémantiquement, je montre que *š*, et non *ma*, contribue à la négation sémantique en arabe algérien.

- (5) *Reda ma šaf š Sami*
 Reda NEG Saw NEG Sami
 i. $\neg \text{see}(r, a)$
 ii. [TP Reda [T' *ma-saf* [NegP *š* [Neg' ~~*ma-šaf*~~ [VP Reda-[V' ~~*šaf*~~ Ali]]]]]]]

L'analyse. Je propose un traitement similaire des SN nus et des mots-N comme quantificateurs généralisés existentiels qui fusionnent au sein du domaine VP. À ce niveau, ils se composent avec un opérateur de négation couvert (CN) via composition fonctionnelle, résultant en un quantificateur négatif. Suivant l'opération d'Accord de Chomsky (1998), le quantificateur négatif formé, portant des traits négatifs interprétables, subit un mouvement (de type wh) vers le spécifieur de NegP (puis vers des projections plus hautes). Ce mouvement est motivé par *ma*, situé en Neg° , qui porte des traits négatifs non-interprétables. Importamment, CN n'est pas disponible en présence de *š*, parce que l'opération d'Accord qui l'autorise ne peut avoir lieu lorsque Spec-NegP est occupé. De plus, CN n'est pas disponible pour d'autres quantificateurs comme les universaux, puisque la lexicalisation des quantificateurs négatifs est hautement contrainte (voir Horn (1973), lacune lexicale NALL).

Quantificateurs négatifs préverbaux (exemple 1-2). L'absence de *š* (exemple (1)) permet à CN de s'adjoindre au syntagme [(*ħatta*) *dajf*] dans Spec-VP. Subséquemment, le syntagme complexe [(*ħatta*) *dajf* CN] subit un mouvement : d'abord vers Spec-NegP, attiré par *ma* situé en Neg° , puis vers Spec-TP, très probablement pour le cas, comme montré en (6). À FL, CN se combine avec le quantificateur existentiel contribué par [(*ħatta*) *dajf*] par composition fonctionnelle, comme suit : $[[CN] \circ [(ħatta) dajf]] = \lambda P. \neg \exists x [\text{invité}(x) \ \& \ P(x)]$. Le quantificateur négatif résultant subit une Montée de Quantificateur (QR), laissant derrière une trace, *t_i*, qui donne la structure [1 *t_i* *kla*]. Au niveau de la phrase, le sens du quantificateur négatif composé se combine avec cette structure via application fonctionnelle, résultant dans le sens en (1i).

- (6) [TP [(*ħatta*) *dajf* CN] [T' *ma-kla* [NegP [(*ħatta*) *dajf* CN] [Neg' *ma-kla* [VP [(*ħatta*) *dajf* CN] [V' *kla*]]]]]]]

En présence de *š* (exemple (2)), CN n'est pas disponible puisqu'il ne peut être autorisé via mouvement vers Spec-NegP, qui est occupé par *š*. Ainsi, le sujet [(*ħatta*) *dajf*] fusionne dans Spec-VP, puis se déplace directement vers Spec-TP, donnant l'ordre des mots en (2). À la Forme Logique (FL), [(*ħatta*) *dajf*], qui ne contribue qu'un quantificateur existentiel (puisque CN est

absent), subit QR et se combine avec [1 š t_i kla] via application fonctionnelle, donnant le sens montré en (2i), où le quantificateur existentiel prend portée au-dessus de la négation. Il est important de noter que seuls les SN nus peuvent produire cette lecture, car les syntagmes *ħatta* sont agrammaticaux lorsqu'ils remontent plus haut que š. Je soutiens que c'est parce que les syntagmes *ħatta*, contrairement aux SN nus, sont des éléments de polarité négative forts (NPIs) (Ladusaw, 1992 ; Giannakidou, 2000). Les NPIs forts sont ces NPIs qui sont autorisés exclusivement dans des environnements anti-additifs, tels que dans la portée de la négation sentencielle ou *without* (Zwarts, 1998 ; Gajewski, 2011).

Quantificateurs existentiels postverbaux (exemple 3). Comme précédemment affirmé, la présence de l'enclitique š bloque l'apparition de CN, empêchant la formation d'un quantificateur négatif. Dans des cas comme (3), la présence de š force le quantificateur existentiel à prendre portée sous la négation. Ceci est attendu parce que la portée inverse en AA n'est pas disponible. Ainsi, j'assume que QR en arabe algérien est restreint au premier nœud de type *t* (vraisemblablement juste au-dessus de VP dans l'exemple (3)). Notamment, les syntagmes *ħatta* sont grammaticaux dans cette configuration en raison de leur apparition dans la portée de la négation. Quant à la présence obligatoire de š avec les syntagmes (*ħatta*) postverbaux, cela suggère que AA obéit à un **principe d'économie** qui restreint la disponibilité de CN aux cas où une lecture est impossible à dériver en présence de l'enclitique š. C'est-à-dire, dans le cas de (3), CN donnerait la même lecture que nous avons avec š.

Références

- Chomsky, N. (1998). Minimalist inquiries: The framework. *MIT Occasional Papers in Linguistics*, 15, 1-55.
- Gajewski, J. (2011). Licensing strong NPIs. *Natural Language Semantics*, 19(3), 309-348.
- Horn, L. R. (1973). Greek Grice: A brief survey of proto-conversational pragmatics. In *Proceedings of the Ninth Regional Meeting of the Chicago Linguistic Society* (pp. 205-214). Chicago Linguistic Society.
- Ladusaw, W. A. (1992). Expressing negation. In C. Barker & D. Dowty (Eds.), *Proceedings of the 2nd Conference on Semantics and Linguistic Theory (SALT II)* (pp. 237-253). Ohio State University Working Papers in Linguistics.
- Giannakidou, A. (2000). *Polarity sensitivity as (non) veridical dependency*. John Benjamins Publishing.
- Zwarts, F. (1998). Three types of polarity. In M. Harvey & S. Smallwood (Eds.), *Proceedings of the Sixth Conference on Semantics and Linguistic Theory (SALT VI)* (pp. 317-338). Cornell University Working Papers in Linguistics.

Assimilations et harmonies phonologiques en franco-provençal nord-dauphinois

Mots-clés : francoprovençal, phonologie, morphologie, assimilation, harmonie

Résumé

Le parler de Genas (Rhône) est un dialecte francoprovençal situé sur l'amphizone du Velin, établissant la transition entre les zones dialectales (sud-)lyonnaises et (nord-)dauphinoises. C'est une zone où le francoprovençal est particulièrement vulnérable, la ville de Genas, comptant pourtant 13698 habitants⁵, ne compte aujourd'hui plus que quelques locuteurs de la langue. L'un d'entre eux, monsieur Pierre Minjat, qui avait déjà réalisé pour la commune un audioguide bilingue nommé « Les Genadines » (voir Heintzmann, 2023), a réalisé pour cette étude des tâches d'élicitation lexicale et verbale, ainsi que des récitations de comtes et fables traduits et adaptés.

Comme bon nombre d'autres parlers romans, le système phonologique du genaillard est traversé par des phénomènes assimilatoires, y compris à distance. Les plus (typologiquement) communs d'entre eux sont les assimilations vocaliques. On observe en effet dans les paradigmes verbaux du genaillard, une assimilation nasale déclenchée par le suffixe de 3e personne pluriel, et ciblant la voyelle précédente de suffixes précédents (imparfait notamment):

		'chanter'	'lier'	'sortir'
1° p. singulier, présent	də	θāt- e	li- Ø	saʎ- e
3° p. pluriel, présent	e	θāt- ɔ̃	lij- ɔ̃	saʎ- ɔ̃
1° p. singulier, présent	də	θāt- av-e	lij- av-e	saʎ- av-e
3° p. pluriel, présent	e	θāt- ɔ̃v-ɔ̃	lij- ɔ̃v-ɔ̃	saʎ- āv-ɔ̃

Tableau 1 : Extrait de paradigmes verbaux du genaillard

On peut également rattacher à ces phénomènes le traitement palatalisant de la finale atone, un trait bien connu des francoprovençalistes (Durham, 1975; Stich, 1998), auparavant répandu dans l'ensemble de la zone franco-provençale, et pour lequel la zone du Velin fait office de « butte témoin » conservatrice (voir Tuaillon, 2007). En effet, si cette palatalisation peut être déclenchée par certaines consonnes coronales, dans d'autres cas, le traitement peut être amené à varier pour une même consonne, notamment en fonction de la voyelle tonique :

(1) /'teta/ 'tête' ≠ /'drɔti/ 'droite' /'sɛla/ 'chaise' ≠ /'goli/ 'faucille'
 /'pysa/ 'poussière' ≠ /'kasi/ 'casserole' /'fiʎa/ 'fille' ≠ /'oreʎi/ 'oreille'
 /'pyneza/ 'punaise' ≠ /'prazi/ 'prise' /'mira/ 'chatte' ≠ /'pjɛri/ 'pierre'
 /fa'rina/ 'farine' ≠ /θa'tani/ 'châtaigne' /'byja/ 'lessive' ≠ /'boji/ 'génisse, velle'

Enfin, de manière beaucoup plus marquée vis-à-vis du domaine roman, la consonne palatale /j/ peut également déclencher une harmonie consonantique sur les autres fricatives du mot, on compte notamment pour la première le suffixe harmonique déverbal /-jɔ̃/, dérivé du -TiŌ latin (ex: jitaʎɔ̃ 'citation', aʎoʎaʎɔ̃ 'association', aʎɔ̃pʎɔ̃ 'Assomption').

⁵ Recensement de population 2022 de l'INSEE.

Dans le cadre de la Théorie des Eléments (Backley, 2011), cadre de modélisation interne des segments s’inscrivant dans la continuité de la Phonologie du Gouvernement (Harris & Lindsey, 1995; Kaye et al., 1985), les phénomènes d’assimilation et d’harmonie reposent sur la propagation d’un élément actif, pouvant résulter de relations locales de licenciement ou d’une propagation harmonique généralisée de l’élément sur son rang au sein d’un domaine (Hulst, 2018). Ainsi, le dépouillement de la richesse et la complexité du traitement de la coronalité et de la palatalité par le diasystème francoprovençal constitue alors un apport bienvenu pour le parachèvement de cette approche récente et novatrice des traits constitutifs des unités langagières, au niveau le plus fondamental.

Keywords: Francoprovençal, phonology, morphology, assimilation, harmony

Abstract

The dialect of Genas (Rhône) is a Francoprovençal variety located in the Velin amphizone, which forms a transition between the (southern) Lyonnais and (northern) Dauphinois dialectal areas. This is a zone where Francoprovençal is particularly vulnerable: although the town of Genas has 13,698 inhabitants⁶, only a few speakers of the language remain today. One of them, Mr. Pierre Minjat – who had already produced for the municipality a bilingual audio guide entitled *Les Genadines* (see Heintzmann, 2023) – contributed to this study by carrying out lexical and verbal elicitation tasks, as well as recitations of tales and adapted fables in translation.

Like many other Romance varieties, the phonological system of *genaillard* is shaped by assimilatory processes, including long-distance ones. The most typologically common among these are vowel assimilations. In the verbal paradigms of *genaillard*, one can indeed observe nasal assimilation triggered by the 3rd person plural suffix, targeting the preceding vowel of earlier suffixes (notably in the imperfect):

		‘chanter’ (to sing)	‘lier’ (to bind)	‘sortir’ (to get out)
1st sg., present	də	θāt- ə	li- Ø	saʎ- ə
3rd pl., present	e	θāt- ɔ̃	lij- ɔ̃	saʎ- ɔ̃
1st sg., present	e	θāt- av-ə	lij- av-ə	saʎ- av-ə
3rd pl., present	də	θāt- ɔ̃v-ɔ̃	lij- ɔ̃v-ɔ̃	saʎ- āv-ɔ̃

Table 1: Extract from Genailard verb paradigms

These phenomena can also be linked to the palatalizing treatment of the unstressed final vowel, a well-known feature among Francoprovençal specialists (Durham, 1975; Stich, 1998). Formerly widespread throughout the entire Francoprovençal area, this trait now survives in the Velin zone, which functions as a “witness hill” (*butte témoin*) preserving it (see Tuailon, 2007). Indeed, while this palatalization can be triggered by certain coronal consonants, in other cases the outcome may vary for the same consonant, depending in particular on the stressed vowel:

⁶ 2022 INSEE Population Census.

(1) /'teta/ 'head' ≠ /'dratɪ/ 'right'	/'sɛla/ 'chair' ≠ /'goli/ 'sickle'
/'pysa/ 'dust' ≠ /'kasi/ 'saucepan'	/'fiɫa/ 'daughter' ≠ /'oreɫɪ/ 'ear'
/'pynɛza/ 'pin' ≠ /'prazi/ 'plug'	/'mira/ 'female cat' ≠ /'pjɛɾɪ/ 'stone'
/fa'rina/ 'flour' ≠ /θɑ'tani/ 'chestnut'	/'byja/ 'laundry' ≠ /'boji/ 'heifer'

Finally, and in a way that is much more distinctive with respect to the Romance domain, the palatal consonant /j/ can also trigger consonant harmony on the other fricatives of the word. A particularly clear case is found in the deverbal harmonic suffix /-ʃ̃/, derived from Latin *-TIŌ* (e.g., *ʃitaʃ̃* 'citation', *afoʃaʃ̃* 'association', *aʃ̃pʃ̃* 'Assumption').

Within the framework of Element Theory (Bakley, 2011) – an internal model of segments that builds on Government Phonology (Harris & Lindsey, 1995; Kaye et al., 1985) – phenomena of assimilation and harmony are analyzed as the spreading of an active element. This spreading may result from local licensing relations or from the generalized harmonic propagation of the element across its tier within a domain (Hulst, 2018). The examination of the richness and complexity of how coronality and palatality are treated across the Francoprovençal diasystem thus provides a valuable contribution to the ongoing refinement of this recent and innovative approach to the constituent features of linguistic units, at their most fundamental level.

Références

- Bakley, P. (2011). *An Introduction to Element Theory*. Edinburgh University Press.
<https://www.jstor.org/stable/10.3366/j.ctt1g0b2d7>
- Durham, S. P. (1975). *The Posttonic Vocalism of Proto-Franco-Provençal*. 91(3-4), 339-364.
<https://doi.org/10.1515/zrph.1975.91.3-4.339>
- Harris, J., & Lindsey, G. (1995). The elements of phonological representation. In *Frontiers of Phonology: Atoms, structures, derivations* (p. 34-79).
- Heintzmann, A. (2023, juin 29). *Edition and typological analysis of an oral corpus of “genaillard” (northern dauphinois) Francoprovençal : An input for the internal classification of a (severely) endangered gallo-romance language* [Présentation orale]. 53rd Linguistic Symposium on Romance Languages (LSRL 53), Paris, France.
- Hulst, H. van der. (2018). *Asymmetries in Vowel Harmony : A Representational Account*. Oxford University Press.
- Kaye, J., Lowenstamm, J., & Vergnaud, J.-R. (1985). The Internal Structure of Phonological Elements: A Theory of Charm and Government. *Phonology Yearbook*, 2, 305-328.
- Stich, D. (1998). *Parlons francoprovençal : Une langue méconnue*. Montréal l'Harmattan.
- Tuaillon, G. (2007). *Le francoprovençal. Tome premier, définition et délimitation Phénomènes remarquables / Gaston Tuaillon*. Musumeci.

Eugen Hill (University of Cologne)

Comparative method, linguistic conspiracies and the phylogenetic signal

Abstract

The Comparative Method has been developed for two tasks. The first is to establish which languages are genetically related. The second, traditionally called “reconstruction”, is to find out how exactly the systems of such languages relate to each other. However, often the application of Comparative Method still leaves open questions. This is particularly well known in case of so-called “Conspiracies” which are recurrently observed in language groups of different genealogical and areal affiliation.

What linguistic Conspiracies are, is best illustrated on Slavonic (Balto-Slavonic branch of Indo-European). That Slavonic languages constitute a group is well established due to numerous shared innovations in them. The historical relations between the individual Slavonic languages are equally well understood. What remains without explanation, is the strange inclination of the individual Slavonic languages towards sound changes transforming closed syllables into open ones (cf. Schenker 1993: 67). This drift, traditionally called the “Open Syllable Conspiracy”, is established for all of Slavonic.

In my talk, I will pursue the following two goals:

(a) I will suggest an explanation of the Slavonic Open Syllable Conspiracy in terms of language contact. I will argue that the decisive factor triggering conspiracies are not the structural properties of the contact languages but, rather, the dynamics of their interaction.

(b) I will discuss the consequences of such conspiracies for recent attempts to use phonotactic information for establishing genetic relatedness of languages (cf. Macklin-Cordes et al. 2021).

My core assumption is that Conspiracies emerge as by-products of goal-oriented sound changes which, at the same time, do not target the feature defining the relevant Conspiracy. Typically, Conspiracies result from a situation where a language goes through numerous different sound changes within a short period of time. Usually, this is caused by a higher-than-average intensity of language contact to a higher-than-average number of languages.

Most languages are subject to restrictions against particular consonant clusters (e.g., Latin against *dl*, Lithuanian against *ns*, etc.). When a language possessing such clusters is influenced by languages lacking them, these clusters may be eliminated by sound changes creating, epiphenomenally, open syllables (cf. Blevins & Garrett 1998; Blevins 2004), such as:

- metathesis: Proto-Balto-Slavic *gálvǫ (Lithuanian *galvà*) ‘head’ > Polish *głowa*,
- anaptyxis: Proto-Balto-Slavic *gálvǫ > Belorussian *galavá*,
- loss of tautosyllabic nasals: Proto-Balto-Slavic *rankǫ (cf. Lithuanian *rankà*) ‘hand’ > Bulgarian *rǎká*,
- palatalisation: Proto-Balto-Slavic *ariǫn (cf. Lithuanian *ariù*) ‘I plough’ > Old Church Slavonic *or’p*,
- etc.

By contact with a large number of different languages, changes of this type accumulate in that more and more closed syllables become open. Superficially, the data may suggest a targeted development toward a phonology with a preference for open syllables. Thus, Slavonic Open Syllable Conspiracy might be a mere function of the well-documented rapid spread of Slavonic out of a compact centre over a huge territory during 6th to 9th c.

Résumé

La méthode comparative a été développée pour deux objectifs. La première consiste à déterminer quelles langues sont génétiquement apparentées. La seconde, traditionnellement appelée « reconstruction », consiste à déterminer précisément les relations entre les systèmes de ces langues. Cependant, l'application de la méthode comparative laisse souvent des questions ouvertes. Ceci est particulièrement bien connu dans le cas des « conspirations », observées de manière récurrente dans des groupes linguistiques d'appartenance généalogique et spatiale différente.

La nature des conspirations linguistiques est parfaitement illustrée par le slave (branche balto-slave de l'indo-européen). Le fait que les langues slaves constituent un groupe est bien établi grâce à de nombreuses innovations communes. Les relations historiques entre les différentes langues slaves sont également bien comprises. Ce qui reste sans explication, c'est l'étrange tendance des langues slaves individuelles à modifier leur son, transformant les syllabes fermées en syllabes ouvertes (cf. Schenker 1993 : 67). Cette dérive, traditionnellement appelée « conspiration des syllabes ouvertes », est présente dans l'ensemble du slave.

Dans mon exposé, je poursuivrai les deux objectifs suivants :

(a) Je proposerai une explication de la conspiration des syllabes ouvertes slaves en termes de contact linguistique. Je soutiendrai que le facteur déterminant déclenchant ces conspirations ne réside pas dans les propriétés structurelles des langues de contact, mais plutôt dans la dynamique de leur interaction.

(b) J'analyserai les conséquences de ces conspirations sur les tentatives récentes d'utilisation de l'information phonotactique pour établir la parenté génétique des langues (cf. Macklin-Cordes et al. 2021).

Mon hypothèse principale est que les conspirations apparaissent comme des sous-produits de modifications phonétiques ciblées, qui, en même temps, ne ciblent pas la caractéristique définissant la conspiration concernée. Généralement, les conspirations résultent d'une situation où une langue subit de nombreuses modifications phonétiques différentes sur une courte période. Cela est généralement dû à une intensité de contact linguistique supérieure à la moyenne avec un nombre de langues supérieur à la moyenne.

La plupart des langues sont soumises à des restrictions concernant certains groupes de consonnes (par exemple, le latin contre *dl*, le lituanien contre *ns*, etc.). Lorsqu'une langue possédant de tels groupes est influencée par des langues qui en sont dépourvues, ces groupes peuvent être éliminés par des changements phonétiques créant, de manière épiphénoménale, des syllabes ouvertes (cf. Blevins & Garrett 1998 ; Blevins 2004), comme :

- métathèse : proto-balto-slave *gálvǫ́ (lituanien *galvą*) « tête » > polonais *głowa*,
- anaptyxie : proto-balto-slave *gálvǫ́ > biélorusse *galavá*,
- perte des nasales tautosyllabiques : proto-balto-slave *rankǫ́ (cf. lituanien *rankà*) « main » > bulgare *rǎká*,
- palatalisation : proto-balto-slave *ariǫ́n (cf. lituanien *ariù*) « Je laboure » > vieux slavon *or'q*,
- etc.

Au contact d'un grand nombre de langues différentes, ce type de changements s'accumule, de plus en plus de syllabes fermées devenant ouvertes. À première vue, les données suggèrent une évolution ciblée vers une phonologie privilégiant les syllabes ouvertes. Ainsi, la conspiration slave des syllabes ouvertes pourrait être simplement le résultat de la diffusion rapide et bien documentée du slave à partir d'un centre compact sur un vaste territoire, entre le VI^e et le IX^e siècle.

Références

- Blevins, Juliette. 2004. *Evolutionary Phonology. The emergence of sound patterns*. Cambridge: CUP.
- Blevins, Juliette & Garrett, Andrew. 1998. The origins of consonant-vowel metathesis. *Language* 74(4), 508-556.
- Schenker, Alexander M. 1993. Proto-Slavonic. In: Comrie, Bernard & Greville G. Corbett. *The Slavonic languages*. London: Routledge.
- Macklin-Cordes Jayden L., Claire Bower, Erich R. Round. 2021. Phylogenetic signal in phonotactics. *Diachronica* 38(2), 210-258.

Henrik Hornecker (Humboldt-Universität zu Berlin)

How to reconstruct verb valency for a proto-language? Two case studies for Proto-Germanic

Keywords: verb valency; Proto-Germanic; comparative method; morphosyntactic reconstruction; lexical reconstruction

Abstract

Verb valency is a complex phenomenon at the interface between different components of grammar. Research on it has primarily been conducted for contemporary languages under a synchronic perspective. But scholars as early as Dressler (1971) started to investigate verb valency topics in Historical Linguistics and to reconstruct verb valency (cf. also Panagl 1973; 1975; 1977; Krisch 1984). Recent examples include Barðdal & Eypórssón (2009; 2012), working on the reconstruction of oblique subjects for Proto-Germanic.

Although the term verb valency was originally coined in syntax as a verb's potential to take necessary complements as its dependents (cf. Tesnière 1959), verb valency is today conceptualized as a many-sided phenomenon on different grammatical layers or even as an aggregation of independent phenomena in different grammatical dimensions (Zifonun 2000: 364–375). Those are, e.g., a semantic, a pragmatic and a syntactic layer/dimension (which affects morphological forms like case, hence the term morphosyntax).

From a theoretical perspective, it is, thus, not trivial to reconstruct verb valency because it is necessary to do (morpho-)syntactic, semantic and pragmatic reconstruction to reconstruct verb valency as a whole for a proto-language. One has to bear in mind that reconstruction with the comparative method in those language components is different from phonological reconstruction. As the comparative method was developed for the comparison and reconstruction of phonemes, it must be adapted if it is applied to other lingual entities. If, e.g., valency patterns (i.e. patterns of verbal arguments in certain morphosyntactic forms) are to be reconstructed, one has to work out first, how reconstructing them is analogical to reconstructing phonemes and which differences there are. Second, one has to consider how the different layers/dimensions of verb valency interact in a synchronic as well as in a diachronic perspective. For the reconstruction of valency patterns, this means, one has to account for the number of arguments of a verb (a semantic/syntactic issue) and the morphosyntactic marking of those arguments and how both issues interact synchronically in general and in a proto-language in particular and diachronically in language change and in grammaticalization. Finally, it must be differentiated between the (lexical) reconstruction of the valency of single verbs (e.g., their valency patterns in a proto-language) and the reconstruction of the more abstract verb valency grammar of a proto-language (e.g., case usage; cf. Barðdal & Eypórssón 2012).

In my talk, I address these issues by presenting two case studies from the Germanic branch of the Indo-European language family. First, I reconstruct the valency of the Proto-Germanic verb **werpan* 'to throw' and discuss instances of case alternation. Second, I investigate the alternation between causative and anti-causative verbs in Proto-Germanic based on the research of Zanchi & Tarsi (2021) for Gothic. I focus on the alternation between *jan-* and *nan-*verbs like Goth. *distairan* 'to burst (tr.)' vs. *distaurnan* 'to burst (intr.)'.

Thus, my talk aims at demonstrating what is possible in the reconstruction of verb valency, which difficulties there are, how it can be done and where the limits of this approach are.

Mots-clés: valence verbale ; proto-germanique ; méthode comparative ; reconstruction morphosyntaxique ; reconstruction lexicale

Résumé

La valence verbale est un phénomène complexe à l'interface entre différentes composantes de la grammaire. Les recherches à ce sujet ont principalement porté sur les langues contemporaines dans une perspective synchronique. Cependant, dès Dressler (1971), des chercheurs ont commencé à étudier les thèmes de la valence verbale en linguistique historique et à reconstruire cette valence (voir également Panagl 1973 ; 1975 ; 1977 ; Krisch 1984). Parmi les exemples récents, citons Barðdal et Eypórsson (2009 ; 2012), qui ont travaillé sur la reconstruction des sujets obliques pour le proto-germanique.

Bien que le terme « valence verbale » ait été initialement inventé en syntaxe pour désigner la capacité d'un verbe à prendre des compléments nécessaires comme dépendants (cf. Tesnière 1959), la valence verbale est aujourd'hui conceptualisée comme un phénomène multiforme relevant de différentes strates grammaticales, voire comme une agrégation de phénomènes indépendants relevant de différentes dimensions grammaticales (Zifonun 2000 : 364-375). Il s'agit, par exemple, d'une dimension sémantique, pragmatique et syntaxique (qui affecte les formes morphologiques comme le cas, d'où le terme morphosyntaxe).

Du point de vue théorique, reconstruire la valence verbale n'est donc pas une mince affaire, car il est nécessaire de procéder à une reconstruction (morpho-)syntaxique, sémantique et pragmatique pour reconstruire la valence verbale dans son ensemble pour une proto-langue. Il faut garder à l'esprit que la reconstruction par la méthode comparative dans ces composantes linguistiques diffère de la reconstruction phonologique. Développée pour la comparaison et la reconstruction des phonèmes, la méthode comparative doit être adaptée si elle est appliquée à d'autres entités linguistiques. Par exemple, si l'on souhaite reconstruire des schémas de valence (c'est-à-dire des schémas d'arguments verbaux dans certaines formes morphosyntaxiques), il faut d'abord déterminer en quoi leur reconstruction est analogue à celle des phonèmes et quelles en sont les différences. Ensuite, il faut considérer comment les différentes couches/dimensions de la valence verbale interagissent dans une perspective synchronique et diachronique. Pour reconstruire les schémas de valence, il faut tenir compte du nombre d'arguments d'un verbe (question sémantique et syntaxique), de leur marquage morphosyntaxique et de la manière dont ces deux aspects interagissent de manière synchronique, en général et dans une proto-langue en particulier, et diachroniquement dans le changement linguistique et la grammaticalisation. Enfin, il faut distinguer la reconstruction (lexicale) de la valence de verbes isolés (par exemple, leurs schémas de valence dans une proto-langue) de la reconstruction de la grammaire plus abstraite de la valence verbale d'une proto-langue (par exemple, l'usage des cas ; cf. Barðdal et Eypórsson 2012).

Dans ma communication, j'aborde ces questions en présentant deux études de cas issues de la branche germanique de la famille des langues indo-européennes. Premièrement, je reconstruis la valence du verbe proto-germanique **werpan* « lancer » et j'analyse des cas d'alternance. Deuxièmement, j'étudie l'alternance entre verbes causatifs et anticausatifs en proto-germanique, en m'appuyant sur les recherches de Zanchi et Tarsi (2021) pour le gothique. Je me concentre sur l'alternance entre verbes en *jan-* et en *nan-*, comme le gothique *distairan* « éclater (tr.) » et *distaurnan* « éclater (intr.) ».

Mon exposé vise donc à démontrer les possibilités de reconstruction de la valence verbale, les difficultés rencontrées, les modalités de réalisation et les limites de cette approche.

Références

- Barðdal, J. & Eypórsson, Þ. (2009). The origin of the oblique-subject construction. An Indo-European comparison. In: Bubenik, V., Hewson, J. & Rose, S. (eds.), *Grammatical Change in Indo-European Languages: Papers presented at the workshop on Indo-European Linguistics at the XVIIIth International Conference on Historical Linguistics*, Montreal, 2007. Amsterdam & Philadelphia: John Benjamins, pp. 179–193.
- Barðdal, J. & Eypórsson, Þ. (2012). 'Hungering and Lusting for Women and Fleshly Delicacies': Reconstructing Grammatical Relations for Proto-Germanic. *Transactions of the Philological Society*, 110(3): 363–393.
- Dressler, W. (1971). Über die Rekonstruktion der indogermanischen Syntax. *Zeitschrift für vergleichende Sprachforschung*, 85: 5–22.
- Krisch, T. (1984). *Konstruktionsmuster und Bedeutungswandel indogermanischer Verben: Anwendungsversuche von Valenztheorie und Kasusgrammatik auf Diachronie und Rekonstruktion*. Frankfurt am Main, Bern, New York & Nancy: Lang.
- Panagl, O. (1973). Mykenologische Marginalien zur 'Rekonstruktion der indogermanischen Syntax'. *Zeitschrift für vergleichende Sprachforschung*, 87(2): 199–206.
- Panagl, O. (1975). Methoden der modernen Linguistik und ihre Anwendung in der Mykenologie. *Živa antika*, 25: 361–375.
- Panagl, O. (1977). Etymologie und Syntaxtheorie. Die Bedeutung der Valenzgrammatik für die Rekonstruktion. *Klagenfurter Beiträge zur Sprachwissenschaft*, 3: 39–58.
- Tesnière, L. (1959). *Éléments de syntaxe structurale*. Paris: Klincksieck.
- Zanchi, Ch. & Tarsi, M. (2021). Valency Patterns and Alternations in Gothic. In: Luraghi, S. & Roma, E. (eds.), *Valency over Time: Diachronic Perspectives on Valency Patterns and Valency Orientation*. Berlin & Boston: de Gruyter Mouton (Trends in Linguistics: Studies and Monographs 368), pp. 31–87.
- Zifonun, G. (2000). Grundlagen der Valenz. In: Ägel, V., Eichinger, L. M., Eroms, H.-W., Hellwig, P., Heringer, H. J. & Lobin H. (eds.), *Dependenz und Valenz: Ein internationales Handbuch der zeitgenössischen Forschung/Dependency and Valency: An International Handbook of Contemporary Research*. Berlin & New York: Walter de Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft 25(1)), pp. 352–377.

Nigar Ali Hussein (Université Paul Valéry, Montpellier)

Langue kurde et identité dans le Grand Kurdistan : entre diversité linguistique et revendications identitaires

Mots-clés : Langue kurde, identité, sociolinguistique, politiques linguistiques, diversité culturelle

Résumé

La langue kurde constitue un vecteur central de l'identité du peuple kurde, réparti entre quatre États — Turquie, Iran, Irak et Syrie — depuis le démantèlement de l'Empire ottoman. Cette répartition géopolitique a entraîné une fragmentation linguistique (kurmancî, soranî, zazakî, etc.) et des politiques linguistiques disparates qui influencent directement les pratiques langagières, la transmission intergénérationnelle et les représentations identitaires des locuteurs. Cette communication s'inscrit dans une approche sociolinguistique comparée, à l'intersection des dynamiques linguistiques, des politiques publiques et des enjeux identitaires. Elle s'appuie sur des données empiriques recueillies dans le cadre d'une thèse de doctorat en Sciences du langage – Sociolinguistique. L'enquête de terrain a été conduite auprès de 400 participants dans les quatre régions du « Grand Kurdistan », avec une méthodologie qualitative et comparative visant à analyser les usages, les attitudes et les représentations autour de la langue kurde.

Trois axes d'analyse :

1. Politiques linguistiques et pratiques langagières : L'analyse met en lumière l'impact différencié des politiques linguistiques sur la vitalité des variétés kurdes : co-officialité du kurde en Irak (depuis 2005), introduction expérimentale dans l'éducation en Turquie (depuis 2012), généralisation dans les écoles au Rojava (depuis 2014), et forte répression linguistique en Iran. Ces politiques ont des effets tangibles sur la transmission linguistique, la scolarisation en langue kurde et les usages dans la sphère publique.
2. Langue kurde comme marqueur identitaire : En examinant les discours des participants, la recherche met en évidence une forte corrélation entre sentiment d'appartenance kurde et valorisation de la langue kurde, perçue à la fois comme héritage culturel, outil de résistance et emblème d'une identité transnationale.
3. Diaspora et réappropriation linguistique : La dernière partie explore le rôle des communautés kurdes en diaspora dans les efforts de revitalisation, d'enseignement et de diffusion de la langue, révélant des stratégies de réappropriation souvent plus libres des contraintes politiques nationales.

L'étude mobilise une triangulation méthodologique combinant entretiens semi-directifs, questionnaires et observations ethnographiques. Elle s'appuie sur des cadres théoriques issus de la sociolinguistique critique, de la sociologie du langage et des études sur le nationalisme linguistique. Les résultats mettent en lumière les tensions entre fragmentation linguistique et unité identitaire, entre assimilation et résistance, ainsi que les dynamiques complexes de réappropriation linguistique face aux politiques d'effacement ou de reconnaissance.

Cette communication contribue à nourrir la réflexion sur les rapports entre langue, pouvoir et identité dans des contextes de minoration, et dialogue avec la thématique générale du colloque, notamment dans la perspective des processus d'appropriation et de réappropriation des langues.

Keywords: Kurdish language, identity, sociolinguistics, language policies, cultural diversity

Abstract

The Kurdish language constitutes a central vehicle of the identity of the Kurdish people, divided among four states – Turkey, Iran, Iraq, and Syria – since the dismantling of the Ottoman Empire. This geopolitical partition has led to linguistic fragmentation (Kurmanji, Sorani, Zazaki, etc.) and disparate language policies that directly influence linguistic practices, intergenerational transmission, and the identity representations of speakers. This paper is situated within a comparative sociolinguistic approach, at the intersection of linguistic dynamics, public policies, and identity issues. It draws on empirical data collected as part of a doctoral dissertation in Language Sciences – Sociolinguistics. Fieldwork was conducted with 400 participants across the four regions of “Greater Kurdistan”, using a qualitative and comparative methodology aimed at analyzing uses, attitudes, and representations surrounding the Kurdish language.

Three axes of analysis:

1. Language policies and linguistic practices: The analysis highlights the differentiated impact of language policies on the vitality of Kurdish varieties: co-official status of Kurdish in Iraq (since 2005), experimental introduction in education in Turkey (since 2012), generalization in schools in Rojava (since 2014), and severe linguistic repression in Iran. These policies have tangible effects on language transmission, schooling in Kurdish, and usage in the public sphere.
2. Kurdish language as an identity marker: By examining the participants’ discourse, the research underscores a strong correlation between the sense of Kurdish belonging and the valorization of the Kurdish language, perceived simultaneously as cultural heritage, a tool of resistance, and the emblem of a transnational identity.
3. Diaspora and linguistic reappropriation: The final part explores the role of Kurdish diaspora communities in the revitalization, teaching, and dissemination of the language, revealing reappropriation strategies that are often freer from national political constraints.

The study employs a methodological triangulation combining semi-structured interviews, questionnaires, and ethnographic observations. It is grounded in theoretical frameworks from critical sociolinguistics, the sociology of language, and studies on linguistic nationalism. The results shed light on the tensions between linguistic fragmentation and identity unity, between assimilation and resistance, as well as on the complex dynamics of linguistic reappropriation in the face of policies of erasure or recognition.

This paper contributes to the broader reflection on the relationships between language, power, and identity in contexts of minoritization, and engages with the general theme of the conference, particularly in relation to processes of appropriation and reappropriation of languages.

Références

- Boyer, H. (2008). *Langue et identité: Sur le nationalisme linguistique*. Limoges: Lambert-Lucas.
- Carceles, R. (2021). « La kurdicité, tentative de définition d'une identité ethnique sans frontières ». Doctoriales Mesopolhis, Aix-en-Provence.
- Lagarde, C. (2008). *Identité, langue et nation. Qu'est-ce qui se joue avec les langues ?* Perpignan: Trabucaire.
- Piguemal, M., (2016). « Philippe Blanchet : Rejeter un accent, c'est toucher à l'identité de l'être — 24 avril 2016 », en ligne : Libération, consulté le 27/01/2025.

Oksana Ivchenko & Natalia Grabar (University of Lille)

Eye-Tracking and Surprisal: Predicting Reading Behavior Across Text Complexities

Keywords: Eye-tracking; Surprisal; Text simplification; Medical discourse; Reading behavior

Abstract

This study investigates the relationship between text complexity, reading behavior, and cognitive processing using a novel approach combining eye-tracking methodology and surprisal analysis. Eye-tracking technology objectively captures real-time reading behavior, making it a reliable method to assess text complexity. Indeed, several studies confirm correlations between key eye-tracking metrics and reading ease or text complexity (Duchowski, 2007; Cromley et al., 2010; Radach & Kennedy, 2013). Surprisal quantifies how unexpected a word is based on its preceding context: words with high surprisal values indicate greater cognitive load because they deviate significantly from reader expectations (Hale, 2001; Levy, 2008). Our research addresses a fundamental question: To what extent can surprisal, as a measure of linguistic complexity, account for or predict these differences in reading behavior?

We constructed a corpus of French texts across three categories - general and medical Wikipedia articles, and clinical case reports - each presented in original and simplified versions. The simplification process involved syntactic, lexical, and semantic modifications according to established guidelines (OCDE, 2015). Twenty-three native French speakers without medical training participated in an eye-tracking experiment using a Tobii Pro Spectrum camera, reading the texts paragraph by paragraph while their eye movements were recorded. Surprisal values were calculated using a pre-trained French GPT-2 model. We analyzed six eye-tracking indicators that capture distinct aspects of reading behavior: first-pass first fixation duration, first-pass duration, regression-path duration, re-reading duration, total fixation duration, and number of fixations (Hyönä & Kaakinen, 2019; Rayner, 1998). Linear Mixed Effects Models were used to examine the relationships among surprisal, text type, simplification, and eye-tracking measures.

Our findings (Fig.1) reveal systematic relationships between word surprisal and eye-tracking measures across all text types, confirming that less predictable words require greater cognitive effort during reading. The strength of these correlations varied significantly by text type: clinical texts showed the strongest correlations with surprisal ($r = 0.36-0.47$), followed by general texts ($r = 0.30-0.42$), and medical texts exhibiting the weakest correlations ($r = 0.25-0.38$). The benefits of simplification were notably text-type dependent: medical texts showed substantial reductions in reading times after simplification, whereas clinical texts demonstrated minimal changes.

The most important theoretical finding emerged from the significant three-way interaction among surprisal, text type, and simplification across multiple eye-tracking measures. This interaction indicated that simplification specifically reduced the effect of surprisal in medical texts but not in clinical texts.

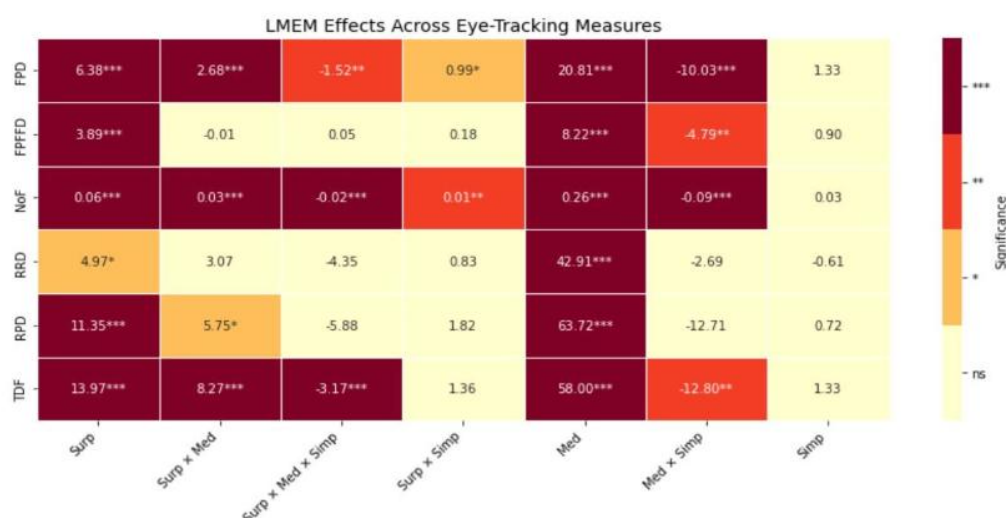


Fig. 1: Coefficient values and significance levels (* $p < .05$, ** $p < .01$, *** $p < .001$) for Linear Mixed Effects Models across six eye-tracking measures.

This research contributes to psycholinguistic theory and practical text adaptation by demonstrating that simplification benefits vary significantly by text type and that surprisal effects differ across text genres. Our findings suggest that text adaptation strategies should be tailored to specific domains, with particular attention to how simplification interacts with word predictability in context.

Future work will expand the corpus, include domain experts as participants, employ multiple language models for surprisal calculation, and explore additional psycholinguistic phenomena to uncover subtler aspects of how predictability influences reading behavior across text types and simplification versions.

Mots-clés : Oculométrie ; Surprise ; Simplification textuelle ; Discours médical ; Comportement de lecture

Résumé

Cette étude examine la relation entre la complexité textuelle, le comportement de lecture et le traitement cognitif en utilisant une approche novatrice combinant la méthodologie d'oculométrie et l'analyse de l'« effet de surprise » (*surprisal*). La technologie d'oculométrie capture objectivement le comportement de lecture en temps réel, ce qui en fait une méthode fiable pour évaluer la complexité textuelle. En effet, plusieurs études confirment les corrélations entre les métriques clés d'oculométrie et la facilité de lecture ou la complexité textuelle (Duchowski, 2007 ; Cromley et al., 2010 ; Radach & Kennedy, 2013). L'« effet de surprise » quantifie à quel point un mot est inattendu en fonction de son contexte précédent : les mots avec des valeurs de surprise élevées indiquent une charge cognitive plus importante car ils dévient significativement des attentes du lecteur (Hale, 2001 ; Levy, 2008). Notre recherche aborde une question

fondamentale : Dans quelle mesure l'effet de surprise, en tant que mesure de complexité linguistique, peut-elle expliquer ou prédire ces différences dans le comportement de lecture ?

Nous avons construit un corpus de textes français dans trois catégories - articles Wikipédia généraux et médicaux, et rapports de cas cliniques - chacun présenté dans des versions originales et simplifiées. Le processus de simplification a impliqué des modifications syntaxiques, lexicales et sémantiques selon des directives établies (OCDE, 2015). Vingt-trois locuteurs natifs français sans formation médicale ont participé à une expérience d'oculométrie utilisant une caméra Tobii Pro Spectrum, lisant les textes paragraphe par paragraphe tandis que leurs mouvements oculaires étaient enregistrés. Les valeurs de surprise ont été calculées en utilisant un modèle GPT-2 français pré-entraîné. Nous avons analysé six indicateurs d'oculométrie qui capturent des aspects distincts du comportement de lecture : durée de première fixation au premier passage, durée du premier passage, durée du chemin de régression, durée de relecture, durée totale de fixation, et nombre de fixations (Hyönä & Kaakinen, 2019 ; Rayner, 1998). Des Modèles Linéaires à Effets Mixtes ont été utilisés pour examiner les relations entre effet de surprise, type de texte, simplification, et mesures d'oculométrie.

Nos résultats (Fig.1) révèlent des relations systématiques entre l'effet de surprise des mots et les mesures d'oculométrie à travers tous les types de textes, confirmant que les mots moins prévisibles nécessitent un effort cognitif plus important pendant la lecture. La force de ces corrélations variait significativement selon le type de texte : les textes cliniques ont montré les corrélations les plus fortes avec l'effet de surprise ($r = 0,36-0,47$), suivis par les textes généraux ($r = 0,30-0,42$), et les textes médicaux présentant les corrélations les plus faibles ($r = 0,25-0,38$). Les bénéfices de la simplification étaient notamment dépendants du type de texte : les textes médicaux ont montré des réductions substantielles des temps de lecture après simplification, tandis que les textes cliniques ont démontré des changements minimaux.

Le résultat théorique le plus important a émergé de l'interaction triple significative entre effet de surprise, type de texte, et simplification à travers plusieurs mesures d'oculométrie. Cette interaction a indiqué que la simplification a spécifiquement réduit l'effet de surprise dans les textes médicaux mais pas dans les textes cliniques.

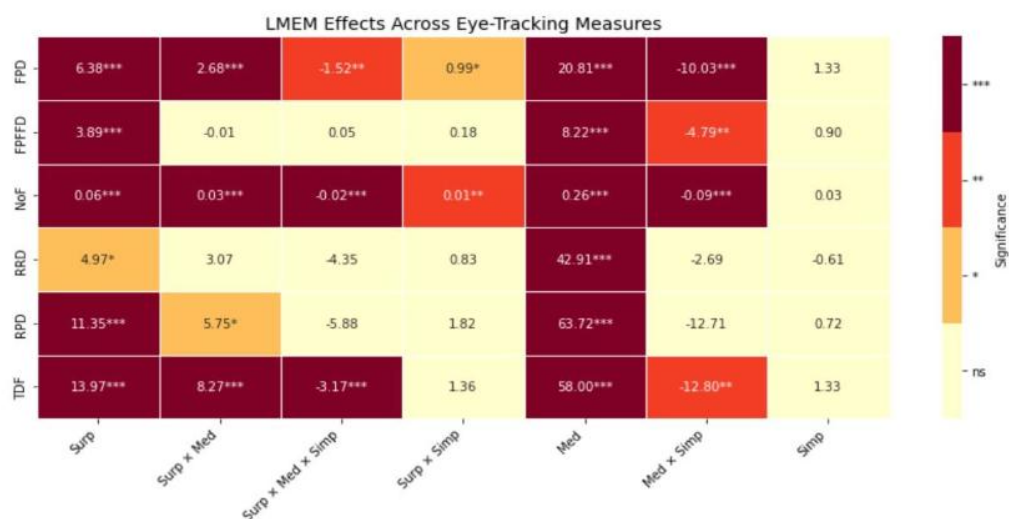


Fig. 1 : Valeurs de coefficients et niveaux de significativité (* $p < .05$, ** $p < .01$, *** $p < .001$) pour les Modèles Linéaires à Effets Mixtes à travers six mesures d'oculométrie.

Cette recherche contribue à la théorie psycholinguistique et à l'adaptation textuelle pratique en démontrant que les bénéfices de la simplification varient significativement selon le type de texte et que les effets de surprise diffèrent à travers les genres textuels. Nos résultats suggèrent que les stratégies d'adaptation textuelle devraient être adaptées à des domaines spécifiques, avec une attention particulière à la façon dont la simplification interagit avec la prévisibilité des mots en contexte.

Les travaux futurs élargiront le corpus, incluront des experts du domaine comme participants, emploieront plusieurs modèles de langue pour le calcul de l'effet de surprise, et exploreront des phénomènes psycholinguistiques supplémentaires pour découvrir des aspects plus subtils de la façon dont la prévisibilité influence le comportement de lecture à travers les types de textes et les versions de simplification.

Références

- CROMLEY J. G., SNYDER-HOGAN L. E. & LUCIW-DUBAS U. A. (2010). Reading comprehension of scientific text : A domain-specific test of the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, 102(3), 687–700. <https://doi.org/10.1037/a0019452>
- DUCHOWSKI A. (2007). Eye Tracking Methodology : Theory and Practice. <https://doi.org/10.1007/978-1-84628-609-4>
- HALE J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*.
- HYÖNÄ J. & KAAKINEN J. (2019). Eye Movements During Reading, p. 239–274. https://doi.org/10.1007/978-3-030-20085-5_7

LEVY R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177. <https://doi.org/10.1016/j.cognition.2007.05.006>

OCDE (2015). Guide de style de l'OCDE Troisième édition : Troisième édition. OECD Publishing.

RADACH R. & KENNEDY A. (2013). Eye movements in reading : Some theoretical context. *Quarterly journal of experimental psychology* (2006), 66. <https://doi.org/10.1080/17470218.2012.750676>

RAYNER K. (1998). Eye movements in reading and information processing : 20 years of research. *Psychological Bulletin*, 124(3), 372–422. <https://doi.org/10.1037/0033-2909.124.3.372>

Maria Karali (Aristotle University of Thessaloniki)

To rob the year of its spring

Keywords: lament; figuration; experience; performativity; healing of trauma; artistic diction

Abstract

'To rob the year of its spring' is Aristotle's metaphor for *mors immatura*. This paper is a socio-cognitive study of the *metaphorical* expression of grief in Hellenistic (ca 300-end of 1st c. B.C.) Greek grave epigrams for children. Lamenting, a way to bear the unbearable excruciation of the untimely death of a young child is characterized by various thematic and linguistic topoi. Diverse cognitive, linguistic strategies are employed to convey emotion: evaluative adjectives in the superlative degree to praise the dead, semantically heavy verbs (weep), a strong sense of memory that leads to a fictional dialogue with the dead person or personified Death, and many novel metaphors, to which I concentrate, in order to study, within the frame of Cognitive Metaphor Theory, the relationship of figuration and emotions, even though it may not be possible to cover exhaustively all of them in these epigrams (approximately 100).

Grief for *mors immatura* involves ontological, strongly imagistic, naturalistic or cultural, metaphors and spins around three main axes: the unjust death (OPPOSITION BETWEEN LIFE AS LIGHT, DEATH AS DARKNESS, DEATH IS ALL-COVERER), the young age and beauty of the deceased child (DECEASED CHILDREN ARE YOUNG BLOSSOMY PLANTS/UPROOTED TREES BY THE WIND/BEAUTIFUL BIRDS) and the bereaved parents' sense of utter despair (e.g. *shall I ever see light without you? Which daemon has snatched away my sweet nightingale?*).

Lament metaphors serve as consolidation, and not only reveal experientially motivated aspects of meaning, but emphatically point to the CAUSE of the emotion (the dead child). Their source domains contain prototypical experiential forms (cf. Foolen, 2012) of BEAUTY (to refer to the child) and of LIFE and DEATH (LIGHT/DARKNESS).

Metaphors as creation of new meanings reveal a new way of interpreting the world (Ricoeur, 1975: 369; cf. Guité-Verret, 2021) and of insightfully coping with traumatic loss (Dickie, 2019). Their emotional force enables bereaved parents to express their disturbed sense of parental identity and their eternal attachment to the deceased (cf. Neimeier et al., 2002: 237, 241).

The lamenting metaphors elucidate the functionality and communicativity of the trope: its performativity, since all such metaphors are deliberate, expressive speech acts of pain (cf. Silvonon et al., 2019), they evoke pathos (Gitay, 2007), they enable the socialization of inner anguish, and represent a poetic healing to the trauma, as they constitute instances of artistic elegiac diction.

Emotional meaning, as a least tangible notion, can be explicated in terms psychologically real to people, and verifiable by them through figuration, which is more intense or efficient than its literal counterpart. There seems to be a correlation between the intensity of an emotion and the expressive power of the metaphorical rendering.

This study will also contribute to the comparison of lament themes in the frame of ethnographic, cultural anthropology -for instance many Hellenistic themes are not the same as the ones in Hungarian laments (Sztó, 2024).

Mots-clés : lamentation ; figuration ; expérience ; performativité ; guérison du traumatisme ; fiction artistique

Résumé

« Priver l'année de son printemps » est la métaphore d'Aristote pour *mors immatura*. Cet article est une étude socio-cognitive de l'expression métaphorique du chagrin dans les épigrammes funéraires grecques de la fin de l'époque hellénistique (ca 300-fin du 1er s. av. J.-C.) pour enfants. La lamentation, une façon de supporter l'expression insoutenable de la mort prématurée d'un jeune enfant, est caractérisée par divers topoï thématiques et linguistiques. Diverses stratégies cognitives et linguistiques sont employées pour transmettre l'émotion : des adjectifs évaluatifs dans le degré superlatif pour louer les morts, des verbes sémantiquement lourds (pleurer), un fort sentiment de mémoire qui mène à un dialogue fictif avec la personne décédée ou la Mort personnifiée, et de nombreuses métaphores nouvelles, sur lesquelles je me concentre, dans le but d'étudier, dans le cadre de la Théorie Métaphorique Cognitive, la relation de figuration et d'émotions, même s'il n'est peut-être pas possible de couvrir exhaustivement toutes ces métaphores dans ces épigrammes (environ 100).

Le chagrin pour *mors immatura* implique des métaphores ontologiques, fortement imagées, naturalistes ou culturelles, et s'articule autour de trois axes majeurs : la mort injuste (OPPOSITION ENTRE LA VIE COMME LUMIÈRE, LA MORT COMME OBSCURITÉ, LA MORT TOUT-DÉVORANTE), la jeunesse et la beauté du défunt (L'ENFANT DÉCÉDÉ COMME JEUNES FLEURS ÉPANOUIES/ARBRES DÉRACINÉS PAR LE VENT/BEAUX OISEAUX) et le sentiment d'abandon total des parents endeuillés (par ex. ne verrai-je jamais la lumière sans toi ? Lequel des démons a arraché mon doux rossignol ?).

Les métaphores de lamentation servent de consolation, et ne révèlent pas seulement des aspects expérientiellement motivés du sens, mais pointent aussi vers l'ORIGINE de l'émotion (l'enfant mort). Leurs domaines sources contiennent des formes expérientielles prototypiques (cf. Foolen, 2012) de BEAUTÉ (pour faire référence à l'enfant) et de VIE et MORT (LUMIÈRE/OBSCURITÉ).

Les métaphores comme création de nouvelles significations révèlent une nouvelle façon d'interpréter le monde (Ricoeur, 1975 : 369 ; Gibbs-Verret, 2021) et des parents en deuil imaginatif face à la perte traumatique (Dickie, 2019). Leur force émotionnelle permet aux parents endeuillés d'exprimer leur sens perturbé de l'identité parentale et leur attachement éternel au défunt (cf. Neimeier et al., 2002 : 237, 241).

Les métaphores de lamentation élucident la fonctionnalité et la communicativité du trope : sa performativité, puisque toutes ces métaphores sont des actes de parole expressifs et délibératifs de douleur (cf. Slivonen et al., 2019), elles évoquent le pathos (Gitay, 2007), elles permettent la socialisation de l'angoisse intérieure, et représentent une guérison poétique du traumatisme, puisqu'elles constituent des instances de fiction élégiaque artistique.

La signification émotionnelle, comme notion du moins tangible, peut être explicitée en termes psychologiquement réels pour les gens, et vérifiable par leur figuration intense, qui est plus intense ou efficace que son équivalent littéral. Il semble y avoir une corrélation entre l'intensité d'une émotion et le pouvoir expressif du rendu métaphorique.

Cette étude contribuera également à la comparaison des thèmes de lamentation dans le cadre de l'anthropologie ethnographique culturelle - par exemple, de nombreux thèmes hellénistiques ne sont pas les mêmes que ceux des lamentations hongroises (Szitó, 2024).

Références

Dickie, J. F. (2019). Lament as a contributor to the healing of trauma: An application of poetry in the form of biblical lament. *Pastoral Psychology*, 68(2), 145-156. <https://doi.org/10.1007/s11089-018-0851-z>

Foolen, A. (2012). The relevance of emotion for language and linguistics. In: Foolen, A., Zlatev, J., Lüdtke, U. & Racine, T. (eds.), *Moving Ourselves, Moving Others: Motion and Emotion in Consciousness, Intersubjectivity and Language*. John Benjamins, Amsterdam/Philadelphia, pp. 347-368.

Gitay, Y. (2007). The Poetics of Exile and Suffering: Memory and Perceptions A Cognitive-Linguistics Study of Lamentations. In: Becking, B. & Human, D. (eds.), *Exile and Suffering: A Selection of Papers Read at the 50th Anniversary Meeting of the Old Testament Society of South Africa*, pp. 203-212. <https://doi.org/10.1163/ej.9789004171046.i-286.74>

Guité-Verret, A., Vachon, M., Ummel, D., Lessard, E., & Francoeur-Carron, C. (2021). Expressing grief through metaphors: Family caregivers' experience of care and grief during the Covid-19 pandemic. *International journal of qualitative studies on health and well-being*, 16(1), 1996872. <https://doi.org/10.1080/17482631.2021.1996872>

Neimeyer, R. A., Prigerson, H. G. & Davies, B. (2002). Mourning and meaning. *American Behavioral Scientist*, 46(2), 235-251.

Ricoeur, P. (1975). *La métaphore vive*. Éditions du Seuil.

Silvonen, V. & Stepanov, E. (2019). Language, Music and Emotion in Lament Poetry: The Embodiment and Performativity of Emotions in Karelian Laments. In: Pritzker, S. E. J. Fenigsen, J. & Wilce, J. M. (eds.), *The Routledge Handbook of Language and Emotion*. Routledge Handbooks in Linguistics, Routledge, London, pp. 203-222. <https://doi.org/10.4324/9780367855093-12> <http://hdl.handle.net/10138/337521>

Szitó, J. (2024). Metaphors and Metonymies for death and grief in Hungarian Laments from the Upper-Tisza Region. In: Kóczy, J. B. & Szelid, V. (eds.) *Cultural Linguistics and (Re)conceptualized Tradition. Cultural Linguistics*. Springer, Singapore. https://doi.org/10.1007/978-981-97-6325-2_10

Wilce, J. M., & Fenigsen, J. (2015). Mourning and Honor: Register in Karelian Lament. Registers of Communication, *Studia Fennica Linguistica* 18. Helsinki: Finnish Literature Society: 187-209.

Hadil Karawani (University of Konstanz)

Past and Perfect in Palestinian Arabic

Abstract

In this talk, I explore the relation between form and meaning in the temporal domain of Arabic, focusing on the interaction between the past, the perfective, and the perfect. Following Frege's principle of compositionality, we examine how sentence meaning is built from lexical semantics and morphosyntactic structure. Yet, as is well known, and as the Arabic data illustrate, this alignment is far from straightforward. In Palestinian Arabic, the perfect can yield what resembles a post-state or extended eventuality reading, a reading that cannot be reduced to the inherent aspectual properties of the root, but rather emerges from the interaction of aspectual templates with higher functional material and temporal modifiers. The interplay between root properties and clausal aspect reveals a layer of semantic flexibility not readily captured by canonical analyses of the perfect (cf. Iatridou et al. 2001). I further examine additional cases of variation shaping the temporal and modal system of Arabic varieties, including form-meaning non-alignments such as the use of perfect morphology in evidential contexts, the use of the simple past for counterfactuality, and the use of the perfective in actuality entailments. These patterns provide a rich empirical ground for thinking about compositional semantics, raising important questions about the role of grammaticalization patterns and language change.

Résumé

Dans cette présentation, j'explore la relation entre forme et sens dans le domaine temporel de l'arabe, en me concentrant sur l'interaction entre le passé, le perfectif et le parfait. En suivant le principe de compositionnalité de Frege, nous examinons comment le sens des phrases se construit à partir de la sémantique lexicale et de la structure morphosyntaxique. Pourtant, comme cela est bien connu – et comme l'illustrent les données de l'arabe –, cet alignement est loin d'être direct. En arabe palestinien, le parfait peut donner lieu à une lecture qui s'apparente à un état subséquent ou à un événement prolongé, une lecture qui ne peut être réduite aux propriétés aspectuelles inhérentes de la racine, mais qui émerge plutôt de l'interaction entre les patrons aspectuels, le matériel fonctionnel d'ordre supérieur et les modificateurs temporels. L'articulation entre les propriétés de la racine et l'aspect propositionnel révèle un degré de flexibilité sémantique qui échappe aux analyses canoniques du parfait (cf. Iatridou et al. 2001). J'examine également d'autres cas de variation qui façonnent le système temporel et modal des variétés de l'arabe, notamment des désalignements entre forme et sens, tels que l'usage de la morphologie du parfait dans des contextes évidentiels, l'emploi du passé simple pour exprimer la contrefactualité, ou encore celui du perfectif dans des inférences d'actualité. Ces phénomènes constituent un terrain empirique riche pour la réflexion sur la sémantique compositionnelle, en soulevant des questions importantes quant au rôle des processus de grammaticalisation et du changement linguistique.

Locality determines pattern selection

Keywords: Root-and-pattern morphology; Distributed Morphology; gerundial patterns; pattern variation; locality restrictions

Abstract

Introduction: Arabic languages are characterized by a vast number of morphological patterns, with many of them showing idiosyncratic behavior. In this paper, I show that by looking at the syntactic composition of words a very clear system emerges. More precisely, the choice of morphological pattern – parallel to the transparency of lexical roots – is restricted by locality. Patterns attaching low in the derivational structure show great variability in shape, whereas patterns attaching high are fully predictable. I adopt the Root-and-Pattern approach of Semitic languages (McCarthy 1981, Prunet et al. 2000, Idrissi 2001, Faust and Hever 2010) and the theoretical framework of Distributed Morphology where roots and patterns are stored as abstract entities in the lexicon and pronounceable words are the result of syntactic derivation (Doron 2003, Al Kaabi 2015).

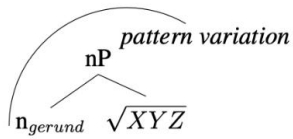
Gerundial patterns: In Syrian Arabic, gerunds of Measure I verbs (XvYvZ) are realized by about ten different and arbitrarily chosen patterns. On the other hand, the gerunds tied to augmented verbs (Measures II-X) show fully predictable shapes (Cowell 1964, 289). ((1)-(2) illustrates this fragmentary.)

- | | | | |
|-----|------------------------------|-------------------------------|---------|
| (1) | Gerunds of Measure I verbs: | | |
| | a. kasar ‘to break’ | → kasər ‘breaking’ | XaYəZ |
| | b. nažah ‘to succeed’ | → nažaah ‘succeeding/success’ | XaYaaZ |
| | c. nəzel ‘to descend’ | → nzuul ‘descending’ | XYuuZ |
| (2) | Gerunds of Measure II verbs: | | |
| | a. ʕallam ‘to teach’ | → taʕliim ‘teaching’ | taXYiiZ |
| | b. sakkar ‘to close’ | → taskiir ‘closing’ | taXYiiZ |
| | c. sabbab ‘to cause’ | → tasbiib ‘causing’ | taXYiiZ |

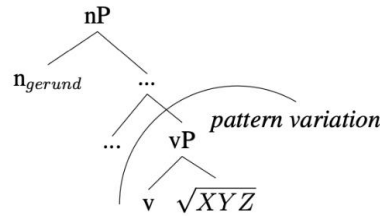
Locality determines form and meaning: Following a vast body of literature (Arad 2003, 2005, Marantz 2000), word formation from roots differs systematically from word formation from words. A root can get multiple and idiosyncratic interpretations until the first categorizing head combines with it, whereas derivations that take words as their input are semantically transparent. I argue that the very same locality constraints apply to patterns too. The patterns combining with roots exhibit large variation, whereas patterns that attach to words have a regular shape. More precisely, I propose that patterns implement categorizing heads and the categorizing heads that combine with roots can be realized by various patterns, whereas categorizing heads that attach to words are manifested by a consistent pattern. I argue that the asymmetry in (1)-(2) results precisely from these two different word formation processes. Gerunds of Measure I verbs are formed with nominalizers that combine directly with the root (see (3a)), whereas gerunds of augmented verbs are formed with nominalizers attaching to verbs thus categorized words (see (3b)).

(3)

a. Gerunds of Measure I verbs



b. Gerunds of Measure II-X verbs



The generalization and its predictions: The predictions of the generalization that pattern variation is only possible as long as the root isn't categorized yet are borne out. I assume that broken plural patterns instantiate (plural) nominalizers and may combine with roots or words. The choice of broken plural pattern is arbitrary when it attaches to the root, but predictable when it combines with an (augmented) word. E.g. instrumental words of the shape məXYaaZ are always pluralized by the same pattern, namely maXaaYiiZ.

Mots-clés : Morphologie racine-et-schème ; Morphologie Distribuée ; schèmes gérondifs ; variation de schèmes ; restrictions de localité

Résumé

Introduction : Les langues arabes se caractérisent par un vaste nombre de schèmes morphologiques, dont beaucoup présentent un comportement idiosyncrasique. Dans cet article, je montre qu'en examinant la composition syntaxique des mots, un système très clair émerge. Plus précisément, le choix du schème morphologique – parallèlement à la transparence des racines lexicales – est restreint par la localité. Les schèmes qui s'attachent bas dans la structure dérivationnelle montrent une grande variabilité de forme, tandis que les schèmes qui s'attachent haut sont entièrement prévisibles. J'adopte l'approche Racine-et-Schème des langues sémitiques (McCarthy 1981, Prunet et al. 2000, Idrissi 2001, Faust et Hever 2010) et le cadre théorique de la Morphologie Distribuée où les racines et schèmes sont stockés comme entités abstraites dans le lexique et les mots prononçables sont le résultat d'une dérivation syntaxique (Doron 2003, Al Kaabi 2015).

Schèmes gérondifs : En arabe syrien, les gérondifs des verbes de Mesure I (XvYvZ) sont réalisés par environ dix schèmes différents et arbitrairement choisis. D'autre part, les gérondifs liés aux verbes augmentés (Mesures II-X) montrent des formes entièrement prévisibles (Cowell 1964, 289). Les exemples (1)-(2) illustrent cette fragmentation.

(1) Gérondifs des verbes de Forme I:

a. kasar 'casser' → kasər 'cassant'

XaYəZ

b. nažah 'réussir' → nažaah 'réussissant/réussite'

XaYaaZ

c. nəzel 'descendre' → nzuul 'descendant'

XYuuZ

(2) Gérondifs des verbes de Forme II:

a. ʕallam 'enseigner' → taʕliim 'enseignant'

taXYiiZ

b. sakkar 'fermer' → taskiir 'fermant'

taXYiiZ

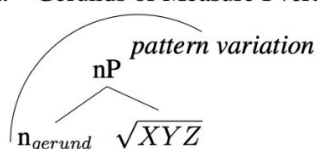
c. sabbab 'causer' → tasbiib 'causant'

taXYiiZ

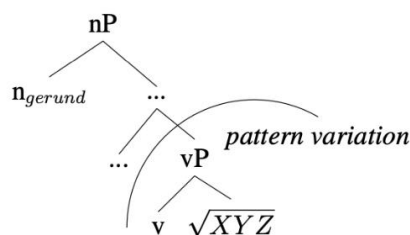
La localité détermine la forme et le sens : Suivant un vaste corpus de littérature (Arad 2003, 2005, Marantz 2000), la formation de mots à partir de racines diffère systématiquement de la formation de mots à partir de mots. Une racine peut recevoir des interprétations multiples et idiosyncrasiques jusqu'à ce que la première tête catégorisante se combine avec elle, tandis que les dérivations qui prennent des mots comme input sont sémantiquement transparentes. Je soutiens que les mêmes contraintes de localité s'appliquent aussi aux schèmes. Les schèmes qui se combinent avec des racines présentent une grande variation, tandis que les schèmes qui s'attachent à des mots ont une forme régulière. Plus précisément, je propose que les schèmes implémentent des têtes catégorisantes et que les têtes catégorisantes qui se combinent avec des racines peuvent être réalisées par divers schèmes, tandis que les têtes catégorisantes qui s'attachent à des mots se manifestent par un schème cohérent. Je soutiens que l'asymétrie dans (1)-(2) résulte précisément de ces deux processus différents de formation de mots. Les gérondifs des verbes de Mesure I sont formés avec des nominalisateurs qui se combinent directement avec la racine (voir (3a)), tandis que les gérondifs des verbes augmentés sont formés avec des nominalisateurs s'attachant à des verbes donc à des mots catégorisés (voir (3b)).

(3)

a. Gerunds of Measure I verbs



b. Gerunds of Measure II-X verbs



La généralisation et ses prédictions : Les prédictions de la généralisation selon laquelle la variation de schèmes n'est possible que tant que la racine n'est pas encore catégorisée se confirment. Je suppose que les schèmes de pluriel brisé instancient des nominalisateurs (pluriel) et peuvent se combiner avec des racines ou des mots. Le choix du schème de pluriel brisé est arbitraire quand il s'attache à la racine, mais prévisible quand il se combine avec un mot (augmenté). Par exemple, les mots instrumentaux de la forme maXYaaZ sont toujours pluralisés par le même schème, à savoir maXaaYiiZ.

Références

Al Kaabi, Meera. 2015. The Anatomy of Arabic words: The Role of the Root in Representation and Processing. Doctoral dissertation, New York University.

Arad, Maya. 2003. Locality Constraints on the Interpretation of Roots: The Case of Hebrew Denominal Verbs. *Natural Language and Linguistic Theory* 21:737–778.

Arad, Maya. 2005. Roots and patterns: Hebrew morpho-syntax. Number 63 in *Studies in Natural Language & Linguistic Theory*. Dordrecht: Springer Science & Business Media.

Cowell, Mark. 1964. A Reference Grammar of Syrian Arabic. Georgetown University Press.

Doron, Edith. 2003. The Semantics of the Semitic Templates. *Natural Language Semantics* 11:1–67.

Faust, Noam, and Ya'ar Hever. 2010. Empirical and Theoretical Arguments in Favor of the Discontinuous Root in Semitic Languages. *Brill's Journal of Afroasiatic Languages and Linguistics* 80–118.

Idrissi, Ali. 2001. Towards a Root-and-Template Approach to Shape-Invariant Morphology. Doctoral dissertation, Montreal.

Marantz, Alec. 2000. Roots: The Universality of Root and Pattern Morphology. In *Conference on Afro Asiatic languages*.

McCarthy, John. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic inquiry* 12:373–418.

Prunet, Jean-François, Renée Béland, and Ali Idrissi. 2000. The Mental Representation of Semitic Words. *Linguistic Inquiry* 31:609–648.

Madeline Ladore (Universität Göttingen)

Understanding the mass-count distinction in Asante-Twi

Keywords: Semantics; morphology; Niger-Congo; mass-count distinction

Abstract

Though the noun class system in Asante-Twi is no longer productive, noun class prefixes are still used to indicate the singular-plural distinction in count nouns (1a) (Osam 1993). Mass nouns, however, lack a morphological number distinction (MND), though they retain their noun class prefix⁷ (1b). Intermediate nouns also lack an MND and pattern in some respects with count nouns and in others with mass nouns (1c).

- | | | |
|-------------------------|-------------------|------------------|
| 1. a. [Count] a/n-konya | b. [Mass] m-frama | c. [Int.] n-homa |
| SG/PL-chair | NC-air | NC-book |

Intermediate nouns can combine with what Quine (1960) calls ‘count adjectives’ (cf. Schwarzschild 2011). These adjectives (big, small, and round) only select nouns with minimal parts. Only certain mass nouns can combine with count adjectives, and get an idiomatic interpretation, as in (2b).

- | | | |
|------------------------------|-------------------------|-------------------------|
| 2. a. [Count] n-koterε kesee | b. [Mass] m-frama kesee | c. [Int.] a-kutuo kesee |
| PL-lizard big | NC-air big | NC-orange big |
| ‘big lizards’ | ‘strong wind’ | ‘big orange(s)’ |

Conversely, mass quantifiers such as *dodo* ‘much’ and *kakra* ‘a bit of’ can also combine with intermediate nouns (3b, c), but not count nouns (3a).

- | | | |
|--|----------------------|-------------------------|
| 3. a. [Count] | b. [Mass] | c. [Int.] |
| a-/n-konya * dodo /* kakra | m-frama dodo/kakra | a-kutuo dodo/kakra |
| SG-/PL-chair much/a.bit.of | NC-air much/a.bit.of | NC-orange much/a.bit.of |

Mass and count nouns cannot be compared quantitatively, as count nouns are compared on a cardinal scale, and mass nouns are compared on a volumetric scale. As seen in (4), a comparison between ε-wɔɔ ‘honey’ and *nkonya* ‘chairs’ is infelicitous.

- | | | | | | | |
|---|-----|----------|-----------|------------|---------|----------|
| 4. #Kofi | wɔ | ε-wɔɔ | bebre sen | Akosua a | ɔ-wɔ | n-konya. |
| Kofi | has | NC-honey | more COMP | Akosua REL | 3SG-has | PL-chair |
| ‘Kofi has more honey than Akosua has chairs.’ | | | | | | |

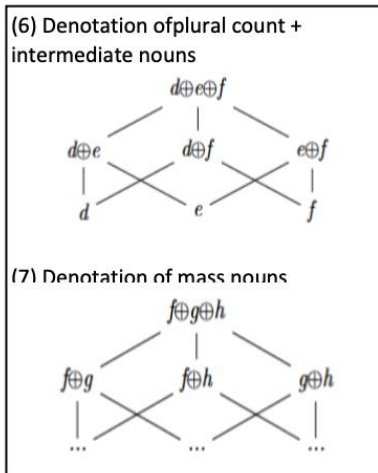
⁷ Glossed as NC.

Interestingly, intermediate nouns can be compared on both cardinal and volumetric scales. *n-yadoa* ‘garden eggs’, an intermediate noun is felicitous in a comparison with either *nkonya* ‘chairs’ or *ε-wɔɔ* ‘honey’.

5. Kofi *ɔ* **n-yadoa** bebree *sen* Akosua *a* *ɔ*-*ɔ* **n-konya/ε-wɔɔ**.
 Kofi has NC-garden.eggs more COMP Akosua REL she-has PL-chair/NC-honey
 ‘Kofi has more garden eggs than Akosua has chairs/honey.’

Intermediate nouns once belonged to Proto Niger-Congo’s class 5/6, characterized by masses and liquids, though it also contains some count nouns (Maho 1999, Good 2020). The domains of count nouns in PNC5/6 and those of Intermediate nouns are closely aligned (agricultural produce, body parts, paired things, things that tend to come in multiples). This common beginning lends itself to intermediate nouns fitting felicitously in contexts typically reserved for mass nouns. Given that “the mass-count distinction seems to have a role similar to that of nominal features such as gender ($\pm$ feminine, $\pm$ masculine) or animacy ($\pm$ animate),” a feature such as noun class determining the syntactic distribution of nouns is not such a far-fetched idea (Bale & Gillon 2020, p.2).

I propose that due to their inherent properties, intermediate nouns have the same semantic denotation as plural count nouns: A full join semi-lattice with atoms (Ladore 2024) (6). Following Link (1983), mass nouns are cumulative, but lack minimal parts. This difference in denotations explains why count adjectives, which select nouns with minimal parts, can combine with count and intermediate nouns, but not mass nouns (unless an idiomatic reading via coercion is available).



This research presents a novel description of the mass-count distinction in Asante-Twi, which has, to this point, gone uninvestigated at the semantic level. The attrition of the noun class system has introduced intermediate nouns into the mass-count divide. This third category presents an opportunity to elaborate on current theories of mass-count, take a closer look at the impact of the loss of Asante-Twi’s noun class system, as well as to incorporate data from an under-researched language into our understanding of the ways a noun’s denotation and distribution can vary.

Mots-clés : Sémantique ; morphologie ; Niger-Congo ; distinction massif-comptable

Résumé

Bien que le système de classes nominales en asante-twi ne soit plus productif, les préfixes de classes nominales sont encore utilisés pour indiquer la distinction singulier-pluriel dans les noms comptables (1a) (Osam 1993). Les noms massifs, cependant, manquent d'une distinction morphologique nombre (MND), bien qu'ils conservent leur préfixe de classe nominale⁸ (1b). Les noms intermédiaires manquent également d'une MND et s'alignent à certains égards avec les noms comptables et à d'autres avec les noms massifs (1c).

- | | | |
|-------------------------|-------------------|------------------|
| 1. a. [Count] a/n-konya | b. [Mass] m-frama | c. [Int.] n-homa |
| SG/PL-chaise | NC-air | NC-livre |

Les noms intermédiaires peuvent se combiner avec ce que Quine (1960) appelle les 'adjectifs comptables' (cf. Schwarzschild 2011). Ces adjectifs (grand, petit, et rond) ne sélectionnent que des noms avec des parties minimales. Seuls certains noms massifs peuvent se combiner avec des adjectifs comptables, et obtiennent une interprétation idiomatique, comme en (2b).

- | | | |
|------------------------------|-------------------------|-------------------------|
| 2. a. [Count] n-koterε kesεε | b. [Mass] m-frama kesεε | c. [Int.] a-kutuo kesεε |
| PL-lézard gros | NC-air gros | NC-orange gros |
| 'gros lézards' | 'strong wind' | 'grosse(s) orange(s)' |

Inversement, les quantificateurs massifs tels que *dodo* 'beaucoup' et *kakra* 'un peu' peuvent aussi se combiner avec les noms intermédiaires (3b, c), mais pas avec les noms comptables (3a).

- | | | |
|--|---------------------------|------------------------------|
| 3. a. [Count] | b. [Mass] | c. [Int.] |
| a-/n-konya * dodo /* kakra | m-frama dodo/kakra | -kutuo dodo/kakra |
| SG-/PL-chaise beaucoup/un.peu.de | NC-air beaucoup/un.peu.de | NC-orange beaucoup/un.peu.de |

Les noms massifs et comptables ne peuvent pas être comparés quantitativement, car les noms comptables sont comparés sur une échelle cardinale, et les noms massifs sont comparés sur une échelle volumétrique. Comme on le voit en (4), une comparaison entre *e-wɔ* 'miel' et *nkonya* 'chaises' est inappropriée.

- | | | | | | | |
|-------------|--------------|------------|-------------|-----|---------|-----------------|
| 4. #Kofi wɔ | ε-wuɔ | bebreε sen | Akosua | a | ɔ-wɔ | n-konya. |
| Kofi has | NC-honey | more | COMP Akosua | REL | 3SG-has | PL-chair |
- 'Kofi a plus de miel qu'Akosua n'a de chaises.'

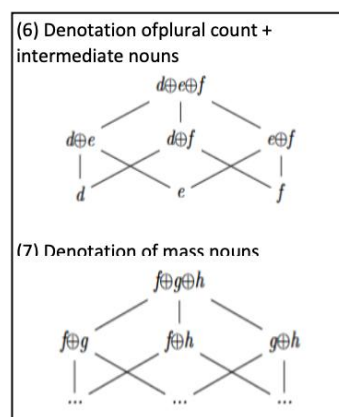
De manière intéressante, les noms intermédiaires peuvent être comparés sur les échelles cardinale et volumétrique. *n-yadoa* 'œufs de jardin', un nom intermédiaire, est approprié dans une comparaison avec soit *nkonya* 'chaises' soit *e-wɔ* 'miel'.

- | | | | | | |
|------------|----------------|------------|-----------------|---------|-----------------------|
| 5. Kofi wɔ | n-yadoa | bebreε sen | Akosua a | ɔ-wɔ | n-konya/ε-wuɔ. |
| Kofi has | NC-garden.eggs | more | COMP Akosua REL | she-has | PL-chair/NC-honey |
- « Kofi a plus d'aubergines qu'Akosua n'a de chaises / de miel. »

⁸ Glosés en tant que NC.

Les noms intermédiaires appartenait autrefois à la classe 5/6 du proto niger-congo, caractérisée par les masses et les liquides, bien qu'elle contienne aussi quelques noms comptables (Maho 1999, Good 2020). Les domaines des noms comptables dans PNC5/6 et ceux des noms intermédiaires sont étroitement alignés (produits agricoles, parties du corps, choses appariées, choses qui tendent à venir en multiples). Ce début commun se prête à ce que les noms intermédiaires s'intègrent de manière appropriée dans des contextes typiquement réservés aux noms massifs. Étant donné que « la distinction massif-comptable semble avoir un rôle similaire à celui des traits nominaux tels que le genre ($\pm$ féminin, $\pm$ masculin) ou l'animéité ($\pm$ animé) », un trait tel que la classe nominale déterminant la distribution syntaxique des noms n'est pas une idée si farfelue (Bale & Gillon 2020, p.2).

Je propose qu'en raison de leurs propriétés inhérentes, les noms intermédiaires ont la même dénotation sémantique que les noms comptables pluriels : un semi-treillis de jonction complet avec des atomes (Ladore 2024) (6). Suivant Link (1983), les noms massifs sont cumulatifs, mais manquent de parties minimales. Cette différence de dénnotations explique pourquoi les adjectifs comptables, qui sélectionnent des noms avec des parties minimales, peuvent se combiner avec des noms comptables et intermédiaires, mais pas avec des noms massifs (à moins qu'une lecture idiomatique via la coercition ne soit disponible).



Cette recherche présente une description novatrice de la distinction massif-comptable en asante-twi, qui a, jusqu'à présent, été non-investigée au niveau sémantique. L'attrition du système de classes nominales a introduit les noms intermédiaires dans la division massif-comptable. Cette troisième catégorie présente une opportunité d'élaborer sur les théories actuelles du massif-comptable, d'examiner de plus près l'impact de la perte du système de classes nominales de l'asante-twi, ainsi que d'incorporer des données d'une langue sous-étudiée dans notre compréhension des façons dont la dénotation et la distribution d'un nom peuvent varier.

Références

- Bunt, Harry C. 1985. *Mass terms and model-theoretic semantics*. Cambridge University Press.
- Deal, Amy Rose. 'Countability Distinctions and Semantic Variation'. *Natural Language Semantics* 25, no. 2 (June 2017): 125–71. <https://doi.org/10.1007/s11050-017-9132-0>.
- Krifka, Manfred. 1989. Nominal reference, temporal constitution and quantification in event semantics. In *Semantics and Contextual Expression*, eds. R. Bartsch, J.F.A.K. van Benthem, and P. von Emde Boas, 75–115. Dordrecht: Foris Publications.
- Lima, S. (2014, August). All notional mass nouns are count nouns in Yudja. In *Semantics and Linguistic Theory* (pp. 534-554).
- Osam, Emmanuel Kweku. 'The Loss of the Noun Glass System in Akan'. *Acta Linguistica Hafniensia* 26, no. 1 (1993): 81–106.

Pierre Larrivée (Université de Caen)

A tree-bank approach to Null subjects in Medieval Gascon

Abstract

The purpose of this paper is to report on subject expression in Gascon. Gascon is a distinct variety in the Occitan continuum which has largely been understudied, especially where syntax is concerned (an opinion shared by Sibille and Bach 2024). The received wisdom is that Gascon is a V2 language with null subjects of the consistent type, and that it was always thus (Ledgeway 2020, Vance et al 2010). It is not clear from the literature however whether the null subject is of the V2 type, with a preponderant distribution in main clauses as in Medieval French (Larrivée et al 2024), or of the consistent type, as in Iberian languages. An empirical issue arises as to the data to be used, in its representativeness, both in terms of the texts fully representing the variety (see Glessgen 2021) and in terms of ordinary language (see, for French, Pinzin and Goux 2024). In this exploratory study, we refer to the 15th century *Fors du Béarn*: on the one hand, the customary text type has been shown to be less conservative for Medieval French, and it might be so for Gascon as well; on the other, the later date allows any development from expectations to be pinned down. We use a treebank approach to a fully UD-annotated version of the text to be part of the SOLIDGRAM corpus under development. The dimension that is explored in this quantitative approach is specifically the distribution of null subjects in root as compared to embedded environments. The preliminary results appear to confirm the consistent nature of null subject language status, and interestingly not a V2 status, given the absence of a quantitative asymmetry in the distribution of null subjects between root and embedded. The results we compare to available quantitative studies for historical Spanish (Dufter 2011) and Portuguese (Galves in preparation), and find a closely comparable lack of asymmetry. This absence is all the more significant that the rate of expressed subject in post-verbal position is indicative of a V2 distribution for Gascon. We hope that these results provide a baseline for comparisons to other stages of Gascon, surrounding Occitan varieties and other Romance varieties.

Résumé

L'objectif de cet article est de rendre compte de l'expression du sujet en gascon. Le gascon est une variété distincte dans le continuum occitan qui a été largement sous-étudiée, en particulier du point de vue syntaxique (une opinion partagée par Sibille et Bach 2024). La thèse communément admise est que le gascon est une langue V2 avec des sujets nuls de type constant, et que cela a toujours été le cas (Ledgeway 2020, Vance et al. 2010). Toutefois, la littérature ne permet pas de déterminer si le sujet nul est du type V2, avec une distribution prépondérante dans les propositions principales comme en français médiéval (Larrivée et al. 2024), ou du type constant, comme dans les langues ibériques. Une question empirique se pose quant aux données à utiliser, tant en termes de représentativité des textes (voir Glessgen 2021) qu'en termes de langue ordinaire (voir, pour le français, Pinzin et Goux 2024). Dans cette étude exploratoire, nous nous référons au *Fors du Béarn* du XVe siècle : d'une part, ce type de texte coutumier s'est révélé moins conservateur pour le français médiéval, et il pourrait en être de même pour le gascon ; d'autre part, la date plus tardive permet de dégager tout développement par rapport aux attentes. Nous adoptons une approche par *treebank* sur une version entièrement annotée UD du texte, destinée à faire partie du corpus SOLIDGRAM en cours de développement.

La dimension explorée dans cette approche quantitative est spécifiquement la distribution des sujets nuls en position de racine par rapport aux environnements subordonnés. Les résultats préliminaires semblent confirmer la nature constante du statut des sujets nuls, et de manière intéressante, non un statut V2, étant donné l'absence d'une asymétrie quantitative dans la distribution des sujets nuls entre les propositions racines et subordonnées. Nous comparons ces résultats aux études quantitatives disponibles pour l'espagnol historique (Dufter 2011) et le portugais (Galves en préparation), et nous constatons une absence d'asymétrie comparable. Cette absence est d'autant plus significative que le taux de sujets exprimés en position postverbale est indicatif d'une distribution V2 pour le gascon. Nous espérons que ces résultats fourniront une base de comparaison pour d'autres stades du gascon, les variétés occitanes environnantes et d'autres variétés romanes.

Références

- Dufter, A. 2011. Sujetos pronominales y estructura informativa en el español tardomedieval. Andreas Dufter and Daniel Jacob (Eds), *Syntaxe, structure informationnelle et organisation du discours dans les langues romanes*. Berlin : Peter Lang. 49-78.
- Frascarelli, M. 2018. The interpretation of pro in consistent and partial null-subject languages. Federica Cognola and Jan Casalicchio (Eds), *Null subjects in Generative Grammar: A synchronic and diachronic perspective*. Oxford: Oxford University Press. 211-239.
- Glessgen, M. 2021. Pour une histoire textuelle du gascon médiéval. *Revue de linguistique romane*, 85(339-340), 325-384.
- Larrivée, P., C.Poletto, F. Pinzin and M. Goux Asymmetry as a general cue for V2 (loss). *Isogloss* 10,7.
- Ledgeway, A. 2020. Variation in the Gallo-Romance left-periphery: V2, complementizers, and the Gascon enunciative system. Sam Wolfe and Martin Maiden (Eds), *Variation and Change in Gallo-Romance Grammar*. Oxford: Oxford University Press.
- Pinzin, F. and M. Goux. 2024. How genre affects word order: a diachronic analysis of French. MS, Frankfurt and Caen.
- Sibille, J. and X. Bach. 2024. Syntax. L. Esher and J. Sibille (Eds). *Manuel de linguistique occitane*. Berlin: De Gruyter. 351-398.
- Vance, B., B. Donaldson and B. D. Steiner. 2010. V2 loss in Old French and Old Occitan. In *Romance Linguistics 2009: Selected Papers from the 39th Linguistic Symposium on Romance Languages (LSRL)*, Tucson, Arizona, March 2009 (Vol. 315, p. 301). Philadelphia : Benjamins.

**Alexander Letuchiy (Université HSE, Moscou / Institut de la langue russe
Vinogradov de l'Académie des sciences de Russie, Moscou)**

Russian transitive verbs with an omitted object: A-labile or strictly transitive?

Keywords: transitivity, A-lability, object omission, depictives, quantification

Abstract

In my talk I study the status of the omitted argument of transitive verbs in Russian. The question is whether the use of these verbs with an omitted direct object is really intransitive (as with A-labile verbs) or the omitted object is present in syntax and found by syntactic processes (antipassives, typologically described by Cooreman (1994), posit similar questions). The first test I use builds on depictives in the instrumental case (like *pjanym* '(being) drunk' or *gorjačim* '(being) hot') that represent the property as a secondary predication (see Bailyn 2001 for details). It turns out that this criterion works differently for two verb classes. Verbs with animate objects allow depictives referring to a zero / omitted object ((1)), while verbs with inanimate objects do not allow it ((3)), though depictives are possible with the direct object expressed, as in (2).

- (1) *Sjuda pjan-ymii Ø_i ne puskaj-ut.*
here drunk-PL.INS NEG admit-PRS.3PL
'Nobody is allowed to enter here drunk.'
- (2) *On je-l-Ø mjas-o_i gorjač-im_i*
he.NOM eat-PST-SG.M meat-SG.ACC hot-N.SG.INS
'He ate meat hot.'
- (3) **On je-l-Ø Ø_i gorjač-im_i*
he.NOM eat-PST-SG.M hot-N.SG.INS
'He ate (everything) hot.'

Another, less reliable test, checks the possibility of numerals and quantifiers marked for instrumental case and denoting the quantity of the object. It turns out that this construction is possible both with inanimate (4) and animate objects (5).

- (4) *I es-t Ø_i cel-ymi misk-am_i.*
and eat-PRS.3SG whole-INS.PL plate-INS.PL
'And (the cat) eats by whole plates.'
- (5) *Uvol'nja-l-Ø Ø_i sotnj-am_i i tysjač-am_i.*
fire-PST-PL hundred-INS.PL and thousand-INS.PL
'(They) fired by hundreds and thousands.'

'(They) fired by hundreds and thousands.' The difference in the results of two tests can be explained by the fact that only the first of them is syntactic in the proper sense (depictives, as Bailyn (2001) shows, require a syntactic licenser). The second test is syntactic, rather than semantic. It shows only the presence of the second participant in semantics, but not its presence in syntax. I conclude that the status and the syntactic visibility of the zero object is different for animate and inanimate objects. Only verbs with inanimate objects behave as A-labile, while with

verbs having animate objects, the object is present in syntax, even if omitted, thus, they should be regarded as transitive with object omission. The reason can be that animate participants are more salient in discourse and are easier to refer to even when expressed by zero. Another explanation comes from the fact that animate (especially human) object class is more homogenous and is easier to be recovered by the speakers if omitted (e.g., in (1), 'they let [a person] drunk inside'). The class of inanimate objects includes heterogeneous objects that are more difficult to be recovered.

Mots-clés : transitivité, A-labilité, omission d'objet, prédicatifs dépicatifs, quantification

Résumé

Dans ma présentation, j'étudie le statut de l'argument omis des verbes transitifs en russe. La question est de savoir si l'usage de ces verbes avec un objet direct omis est réellement intransitif (comme avec les verbes A-labiles) ou si l'objet omis est présent dans la syntaxe et identifié par des processus syntaxiques (les antipassifs, décrits typologiquement par Cooreman (1994), soulèvent des questions similaires).

Le premier test que j'utilise s'appuie sur les prédicatifs dépicatifs à l'instrumental (comme *ljanym* « (étant) ivre » ou *gorjačim* « (étant) chaud ») qui représentent la propriété comme une prédication secondaire (voir Bailyn 2001 pour plus de détails).

Il s'avère que ce critère fonctionne différemment pour deux classes de verbes. Les verbes avec objets animés permettent des prédicatifs dépicatifs se référant à un objet zéro/omis ((1)), tandis que les verbes avec objets inanimés ne le permettent pas ((3)), bien que les prédicatifs dépicatifs soient possibles avec l'objet direct exprimé, comme en (2).

- (1) *Sjuda ljan-ymii Ø_i ne puskaj-ut.*
 ici ivre-PL.INS NEG autoriser-PRS.3PL
 « Personne n'est autorisé à entrer ici en état d'ivresse. »
- (2) *On je-l-Ø mjas-o_i gorjač-imi_i*
 il.NOM manger-PST-SG.M viande-SG.ACC chaud-N.SG.INS
 « Il a mangé la viande chaude. »
- (3) **On je-l-Ø Ø_i gorjač-imi_i*
 il.NOM manger-PST-SG.M chaud-N.SG.INS
 « Il a mangé (tout) chaud. »

Un autre test, moins fiable, vérifie la possibilité d'utiliser des numéraux et quantificateurs marqués à l'instrumental et dénotant la quantité de l'objet. Il s'avère que cette construction est possible aussi bien avec des objets inanimés (4) qu'avec des objets animés (5).

- (4) *I es-t Ø_i cel-ymi misk-ami_i.*
 et manger-PRS.3SG entier-INS.PL assiette-INS.PL
 « Et (le chat) mange par assiettes entières. »

- (5) Uvol'nja-I-I Ø_i sotnj-ami_i i tysjač-ami.
 licencier-PST-PL centaine-INS.PL et millier-INS.PL
 (Ils) ont licencié par centaines et par milliers. »

La différence dans les résultats des deux tests peut s'expliquer par le fait que seul le premier d'entre eux est syntaxique au sens propre (les prédicatifs déictifs, comme le montre Bailyn (2001), requièrent un licencié syntaxique). Le second test est syntaxique plutôt que sémantique. Il ne montre que la présence du second participant en sémantique, mais pas sa présence en syntaxe.

Je conclus que le statut et la visibilité syntaxique de l'objet zéro sont différents pour les objets animés et inanimés. Seuls les verbes avec objets inanimés se comportent comme des A-labiles, tandis qu'avec les verbes ayant des objets animés, l'objet est présent en syntaxe, même s'il est omis ; ainsi, ils devraient être considérés comme transitifs avec omission d'objet. La raison peut être que les participants animés sont plus saillants dans le discours et sont plus faciles à référencer même lorsqu'ils sont exprimés par zéro. Une autre explication vient du fait que la classe d'objets animés (en particulier humains) est plus homogène et est plus facile à récupérer par les locuteurs si elle est omise (par exemple, en (1), « ils laissent entrer [une personne] ivre »). La classe des objets inanimés inclut des objets hétérogènes qui sont plus difficiles à récupérer.

References

- Bailyn, John. 2001. The syntax of Slavic Predicate case. *ZAS Papers in Linguistics*. 2001. 22 January. P. 1–23.
- Cooreman, Ann. 1994. A functional typology of antipassives. In *Voice: Form and Function*, Barbara Fox & Paul J. Hopper (eds), 49-82. Amsterdam, Philadelphia: John Benjamins.

Svetlana Leuranguer (Université Toulouse Jean Jaurès), Juliette Thuilier
(Université Toulouse Jean Jaurès, CLLE) & Cyrille Granget (Université Toulouse
Jean Jaurès, LNPL)

L'alternance actif/passif des verbes psychologiques en français et en russe :

Une étude sur corpus

Mots-clés : Verbes psychologiques ; passif ; animéité ; français ; russe

Abstract

L'objectif de ce travail est de comparer l'usage de l'actif et du passif avec les verbes psychologiques transitifs à *Expérienceur* objet en français et en russe. D'après l'étude de Ferreira (1994) sur ce type de verbes en anglais, l'*Expérienceur*, typiquement animé, attire la fonction sujet, ce qui favorise l'utilisation de la voix passive, notamment quand le *Stimulus* est inanimé (*The manager was worried by the layoffs*). Des résultats similaires ont été obtenus pour le français (Thuilier et al., 2024). Nous cherchons à comparer ces effets en français et en russe, une langue à cas avec un ordre des arguments flexible.

Nous avons retenu cinq couples de verbes appariés en sens et ayant des fréquences relatives comparables dans les deux langues : *surprendre/udivit'*, *décevoir/razočarovat'*, *énervier/razdražit'*, *impressionner/vpečatlit'*, *fasciner/očarovat'*. Nos données, issues de corpus écrit et oraux, correspondent aux occurrences de ces verbes à l'actif et au passif, dans les corpus CEFC (Debaisieux & Benzitoun, 2020) *frTenTen23* (Kilgarrieff et al., 2014) pour le français ; et le corpus *Russian National Corpus* (Savčuk et al., 2024) pour le russe. Nous n'avons conservé que les phrases avec des sujets et des objets nominaux ou pronominaux, et des compléments d'agent introduits par *de* ou *par*, en français, et au cas instrumental, en russe. Afin de privilégier une interprétation éventive du passif, nous avons sélectionné les phrases au passé composé et au plus-que-parfait pour le français, et au passé pour le russe. Les 2686 occurrences obtenues (français : 57 oral/1972 écrit, russe : 90 oral/571 écrit) ont été annotées selon la voix, l'ordre des arguments et l'animéité du *Stimulus*.

En nous appuyant sur la modélisation statistique du passif français proposée dans Da Cunha & Abeillé (2020), nous avons construit un modèle de régression logistique à effets mixtes, avec 1 effet aléatoire (Sens verbal) et 3 effets fixes (animéité du *Stimulus*, Langue et Modalité). Les résultats montrent l'effet de l'asymétrie d'animéité entre l'*Expérienceur* et le *Stimulus* sur le choix du passif. Le *Stimulus* inanimé favorise le passif (1.2756, $p\text{-value} < 0.001$) dans les deux langues. Nous avons également observé les effets de la modalité et de la langue. Selon le modèle, l'oral favorise l'actif (-0.8800, $p\text{-value} < 0.001$). Cet effet doit néanmoins être étudié plus en détail car nos données orales sont sous-représentées par rapport aux données écrites. Quant à la langue, le français favorise le passif (1.5229, $p\text{-value} < 0.001$). Ces résultats sont cohérents avec les données descriptives (Figure 1) et le fait que le russe permet d'avoir l'ordre *Expérienceur-Stimulus* sans avoir recours au passif. D'ailleurs, cet ordre est attesté dans 37% des phrases actives russes et l'animéité du *Stimulus* participe aussi à déterminer l'ordre des arguments : un *Stimulus* inanimé favorise l'ordre *Expérienceur-Stimulus* ($\chi^2 = 10.684$, $df = 1$, $p\text{-value} = 0.00108$) (Tableau 1).

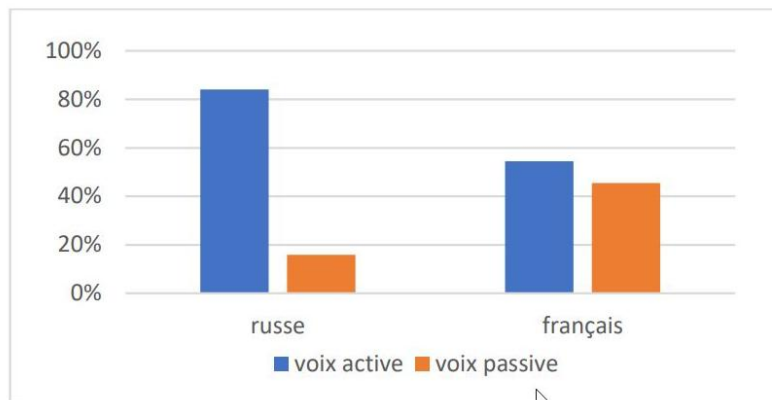


Fig. 1 : Distribution actif/passif en français et en russe

	Expérienceur-Stimulus	Stimulus-Expérienceur
Stimulus animé	21 (21,9%)	75 (78,1%)
Stimulus inanimé	185 (40,2%)	275 (59,8%)

Tableau 1 : Ordre entre *Expérienceur* et *Stimulus* dans les phrases actives du russe

Keywords: Psychological verbs; passive voice; animacy; French; Russian

Abstract

The aim of this work is to compare the use of the active and passive voices with transitive psychological verbs with an *Experiencer* object in French and Russian. According to Ferreira's (1994) study on this type of verbs in English, the *Experiencer*, typically animate, tends to attract the subject function, which favors the use of the passive voice, especially when the *Stimulus* is inanimate (*The manager was worried by the layoffs*). Similar results have been obtained for French (Thuilier et al., 2024). Our goal is to compare these effects in French and in Russian, a case-marking language with flexible argument order.

We selected five pairs of verbs matched in meaning and with comparable relative frequencies in the two languages: *surprendre/udivit'*, *décevoir/razočarovat'*, *énervier/razdražit'*, *impressionner/vpečatlit'*, *fasciner/očarovat'*. Our data, drawn from written and spoken corpora, include occurrences of these verbs in active and passive constructions, from CEFC (Debaisieux & Benzitoun, 2020) and *frTenTen23* (Kilgarriff et al., 2014) for French, and from the Russian National Corpus (Savčuk et al., 2024) for Russian. We retained only sentences with nominal or pronominal subjects and objects, and agentive complements introduced by *de* or *par* in French, and in the instrumental case in Russian. In order to prioritize an eventive interpretation of the passive, we selected sentences in the passé composé and plus-que-parfait for French, and in the past tense for Russian. The 2,686 occurrences obtained (French: 57 spoken / 1,972 written; Russian: 90 spoken / 571 written) were annotated for voice, argument order, and animacy of the *Stimulus*.

Building on the statistical modeling of French passives proposed in Da Cunha & Abeillé (2020), we constructed a mixed-effects logistic regression model, with one random effect (Verb meaning) and three fixed effects (Stimulus animacy, Language, and Modality). The results show the effect of animacy asymmetry between *Experiencer* and *Stimulus* on the choice of the passive. An inanimate *Stimulus* favors the passive (1.2756, p -value < 0.001) in both languages. We also observed effects of modality and language. According to the model, spoken data favor the active (-0.8800, p -value < 0.001). However, this effect requires further investigation since spoken data are under-represented compared to written data. As for language, French favors the passive (1.5229, p -value < 0.001). These results are consistent with the descriptive data (Figure 1) and with the fact that Russian allows for the *Experiencer*–*Stimulus* order without resorting to the passive. Indeed, this order is attested in 37% of Russian active sentences, and the animacy of the *Stimulus* also helps determine argument order: an inanimate *Stimulus* favors the *Experiencer*–*Stimulus* order ($\chi^2 = 10.684$, $df = 1$, p -value = 0.00108) (Table 1).

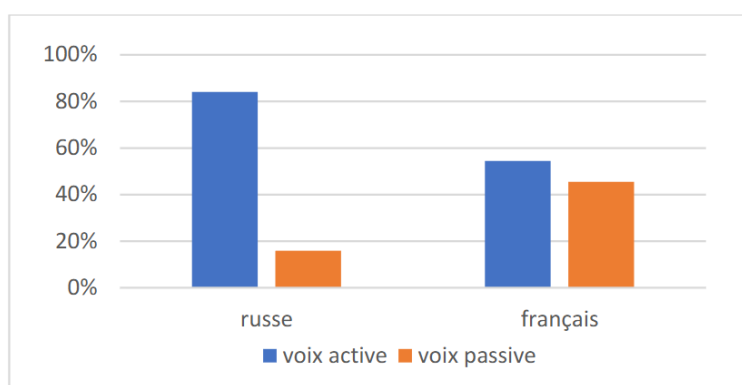


Fig. 1: Distribution of active/passive in French and Russian

	Expérienceur-Stimulus	Stimulus-Expérienceur
Stimulus animé	21 (21,9%)	75 (78,1%)
Stimulus inanimé	185 (40,2%)	275 (59,8%)

Table 1: Experiencer–Stimulus word order in Russian active sentences

Références

- 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. A journal of linguistics, psycholinguistics and computational linguistics*, 27, Article 27. <https://doi.org/10.4000/discours.10956>
- Debaisieux J.-M. & Benzitoun C. (Éds.). (2020). *Langages no 219 : Orféo : un corpus et une plateforme pour l'étude du français contemporain*, Malakoff, Armand Colin/Dunod.
- Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of memory and language*, 33, 715-763.
- Kilgarrieff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1: 7-36.

Savčuk S. O., Arhangelsky T. A., Bonč-Osmolovskaya A. A., Donina O. V., Kuznetsova Y. N., Lyaševskaya O. N., Orehov B. V., Podryadčikova M. V. (2024). Национальный корпус русского языка 2.0: новые возможности и перспективы развития. *Voprosy Yazykoznaniiya*, 2 : 7–34. <https://doi.org/10.31857/0373-658X.2024.2.7-34>

Thuilier, J., Faghiri, P., da Cunha, Y., & Pineau, L. (2024, septembre). Alternance de voix des verbes psychologiques : Une étude expérimentale. *Langues et Langage à la croisée des disciplines*. <https://hal.science/hal-04717256> (date de consultation : 12 décembre 2024).

Do geminate consonants exist in contemporary French? A corpus-based analysis

Keywords: Geminate, large corpus, morphology, orthography, French

Abstract

Geminates are fortis consonants, generally longer than their singleton counterparts (Davis 2011). Although geminates inherited from Latin were simplified around the 7th century (Zink 1986), descriptive works on French have noted the emergence of long consonants in the 20th century in certain morphological or orthographic contexts, such as *irréel*, *grammaire*, or *allégation* (Grammont 1914, Fagyal et al. 2006). In some cases, the phenomenon creates potential minimal pairs, such as *courais* (imperfect) vs. *courrais* (conditional, with -r- suffix). Does French exhibit a (marginal) singleton-geminate contrast? Surprisingly, the phenomenon has been largely unstudied. Meisenburg (2006) finds that contrasts such as <*courais-courrais*>, <*frappa-frappe pas*> are produced and perceived by some speakers, but with high variation.

As the first step in a broader project on the existence and distribution of geminates in French, we analyzed the effects of morphology and orthography on the duration of four sonorant consonants (/r/, /l/, /m/, /n/) in large, automatically aligned corpora of spontaneous speech. These corpora include broadcast news, radio and TV shows and spontaneous conversations. We examined consonant duration across six groups, each defined by a specific prefix (*ir-*, *sur-*, *il-*, *al-*, *im-* and *in-*). Consonant durations were normalized relative to phoneme rate to account for speech tempo variability. Within each group, we conducted comparisons of normalized durations across categories defined by their respective morphological and/or orthographic exponents (see Table 1)⁹. Statistical analysis was performed using Kruskal–Wallis rank sum tests to assess overall differences among categories. Where significant effects were found, post hoc pairwise comparisons were conducted using Dunn’s test.

Category N°	Category	Number of occurrence	Word example	Prefix meaning
<i>ir-</i>	prefix<ir-r>	113	irrespect	negation
	onset<ir>	798	ironie	/
<i>sur-</i>	prefix<sur-r>	25	surréagi	above, in excess
	prefix<sur>	22	surenchère	above, in excess
<i>il-</i>	prefix<il-l>	121	illogique	negation
	onset<ill>	139	illusion	/
<i>al-</i>	prefix<al-l>	415	alléger	towards
	onset<all>	2715	allergique	/
<i>im-</i>	prefix<im-m>	475	immorale	negation
	onset<im>	923	image	/
<i>in-</i>	prefix<in-n>	115	innombrables	negation
	prefix<in>	527	inutile	negation
	onset<in>	209	initier	/

⁹ Categories containing fewer than 20 instances were excluded. Thus, the final categories analyzed in this study reflect only those with sufficient occurrences in the examined corpora.

The results are presented in Figure 1, with groups distinguished by color. Overall, the cross-category comparisons suggest three potential effects that may contribute to increased consonant duration:

- 1) The **orthography** effect, revealed by comparing *pr.*<sur-r> with *pr.*<sur> and *pr.*<in-n> with *pr.*<in>;
- 2) The **morphology** effect, revealed by comparing *pr.*<al-l> with *on.*<all> and *pr.*<in> with *on.*<in>;
- 3) The **combined** orthography-morphology effect, revealed by comparing *pr.*<im-m> with *on.*<im> and *pr.*<in-n> with *on.*<in>.

It is worth noting that words in the category *pr.*<al-l>, although historically prefixed by the Latin prefix *ad-* (with *al-* as one of its assimilated forms), no longer exhibit a transparent relationship between base and derivative in contemporary French. Nevertheless, the /l/ remains phonetically longer in *pr.*<al-l> compared to *on.*<all>. This observation motivates a broader examination of other assimilated variants of *ad-*, such as *af-*, *ag-*, *ad-*, *ac-*, and *ap-*. Accordingly, the scope of our investigation will be extended to include other types of consonants (e.g., fricatives and plosives). Additionally, we aim to explore consonant duration in other morphosyntactic contexts, as well as the potential influence of communicative context, as reflected in various speech corpora.

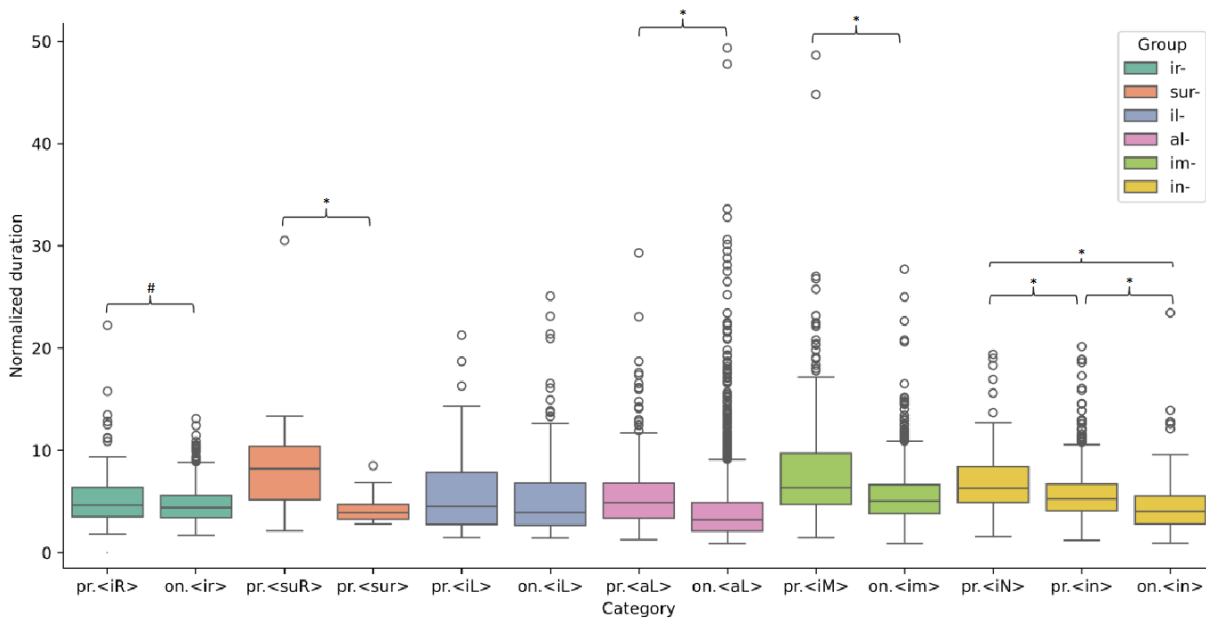


Fig. 1: Normalized duration of the sonorant consonants /r/, /m/, /n/, and /l/ across different word categories. For “Category,” the uppercase letter within angle brackets indicates a graphic double consonant (e.g., *pr.*<iR> as in *irr  el*). An asterisk (*) indicates a significant duration difference between two categories; a hash (#) indicates a marginally significant difference.

Mots-clés : Gémignée ; grand corpus ; morphologie ; orthographe ; français

Résumé

Les gémignées sont des consonnes *fortis*, généralement plus longues que leurs homologues simples (Davis 2011). Bien que les gémignées héritées du latin aient été simplifiées vers le 7^e siècle (Zink 1986), les travaux descriptifs sur le français ont noté l'émergence de consonnes longues au 20^e siècle dans certains contextes morphologiques ou orthographiques, comme dans irréel, grammaire, ou allégation (Grammont 1914, Fagyal et al. 2006). Dans certains cas, ce phénomène crée des paires minimales potentielles, comme *courais* (imparfait) vs. *courrais* (conditionnel, avec le suffixe -r-). Le français présente-t-il un contraste (marginal) entre consonnes simples et gémignées ? De manière surprenante, ce phénomène a été largement peu étudié. Meisenburg (2006) constate que des contrastes tels que ⟨*courais-courrais*⟩, ⟨*frappa-frappe pas*⟩ sont produits et perçus par certains locuteurs, mais avec une grande variation.

Comme première étape d'un projet plus large sur l'existence et la distribution des gémignées en français, nous avons analysé les effets de la morphologie et de l'orthographe sur la durée de quatre consonnes sonantes (/r/, /l/, /m/, /n/) dans de grands corpus de parole spontanée alignés automatiquement. Ces corpus incluent des journaux télévisés, des émissions de radio et de télévision, et des conversations spontanées. Nous avons examiné la durée des consonnes dans six groupes, chacun défini par un préfixe spécifique (*ir-*, *sur-*, *il-*, *al-*, *im-* et *in-*). Les durées consonantiques ont été normalisées par rapport au débit phonémique pour tenir compte de la variabilité du tempo de parole. Au sein de chaque groupe, nous avons effectué des comparaisons des durées normalisées entre les catégories définies par leurs composants morphologiques et/ou orthographiques respectifs (voir Tableau 1).¹⁰ L'analyse statistique a été réalisée à l'aide de tests de rang de Kruskal-Wallis pour évaluer les différences globales entre les catégories. Lorsque des effets significatifs ont été trouvés, des comparaisons post hoc par paires ont été effectuées en utilisant le test de Dunn.

Category N°	Category	Number of occurrence	Word example	Prefix meaning
<i>ir-</i>	prefix<ir-r>	113	irrespect	negation
	onset<ir>	798	ironie	/
<i>sur-</i>	prefix<sur-r>	25	surréagi	above, in excess
	prefix<sur>	22	surenchère	above, in excess
<i>il-</i>	prefix<il-l>	121	illogique	negation
	onset<ill>	139	illusion	/
<i>al-</i>	prefix<al-l>	415	alléger	towards
	onset<all>	2715	allergique	/
<i>im-</i>	prefix<im-m>	475	immorale	negation
	onset<im>	923	image	/
<i>in-</i>	prefix<in-n>	115	innombrables	negation
	prefix<in>	527	inutile	negation
	onset<in>	209	initier	/

¹⁰ Les catégories contenant moins de 20 occurrences ont été exclues. Ainsi, les catégories finales analysées dans cette étude ne reflètent que celles présentant un nombre suffisant d'occurrences dans les corpus examinés.

Les résultats sont présentés dans la Figure 1, avec les groupes distingués par couleur. Dans l'ensemble, les comparaisons inter-catégories suggèrent trois effets potentiels qui peuvent contribuer à l'augmentation de la durée consonantique :

- 1) **L'effet orthographique**, révélé en comparant *pr.*⟨*sur-r*⟩ avec *pr.*⟨*sur*⟩ et *pr.*⟨*in-n*⟩ avec *pr.*⟨*in*⟩ ;
- 2) **L'effet morphologique**, révélé en comparant *pr.*⟨*al-l*⟩ avec *on.*⟨*all*⟩ et *pr.*⟨*in*⟩ avec *on.*⟨*in*⟩ ;
- 3) **L'effet combiné orthographe-morphologie**, révélé en comparant *pr.*⟨*im-m*⟩ avec *on.*⟨*im*⟩ et *pr.*⟨*in-n*⟩ avec *on.*⟨*in*⟩.

Il convient de noter que les mots de la catégorie *pr.*⟨*al-l*⟩, bien qu'historiquement préfixés par le préfixe latin *ad-* (avec *al-* comme l'une de ses formes assimilées), n'exhibent plus de relation transparente entre base et dérivé en français contemporain. Néanmoins, le // reste phonétiquement plus long dans *pr.*⟨*al-l*⟩ comparé à *on.*⟨*all*⟩. Cette observation motive un examen plus approfondi d'autres variantes assimilées d'*ad-*, telles que *af-*, *ag-*, *ad-*, *ac-*, et *ap-*. En conséquence, la portée de notre investigation sera étendue pour inclure d'autres types de consonnes (par exemple, fricatives et occlusives). De plus, nous visons à explorer la durée consonantique dans d'autres contextes morphosyntaxiques, ainsi que l'influence potentielle du contexte communicatif, comme reflété dans divers corpus de parole.

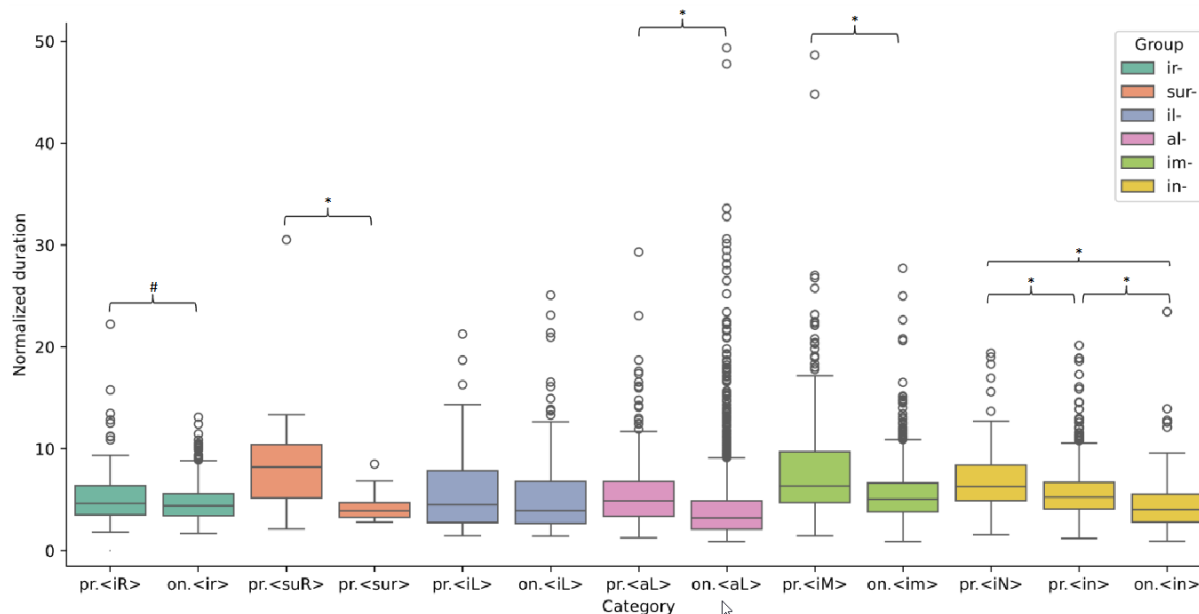


Fig. 1 : Durée normalisée des consonnes sonantes /r/, /m/, /n/ et /l/ selon différentes catégories de mots. Pour la rubrique « Catégorie », la lettre majuscule entre chevrons indique une consonne double graphique (par ex. *pr.*⟨*iR*⟩ comme dans *irréel*). Un astérisque (*) indique une différence de durée significative entre deux catégories ; un dièse (#) indique une différence marginalement significative.

Références

- Davis, S. (2011). Geminate. In: M. Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.). *The Blackwell Companion to Phonology*. Wiley, pp.1-25.
- Fagyal, Z., Kibbee, D., & Jenkins, F. (2006). *French: A linguistic introduction*. Cambridge University Press.
- Grammont, M. (1914). *Traité pratique de prononciation française*. Delagrave.
- Meisenburg, T. (2006). Fake geminates in French: A production and perception study. *Speech Prosody Conference* (Dresden). Citeseer.
- Zink, G. (1986). *Phonétique historique du français*. Paris : Presses Universitaires de France.

Elsa Liste Lamas (ZHAW, Switzerland)

Motion events in L2 German: exploring task variation, group and individual differences

Keywords: lexicalization patterns; motion events; German L2; path encoding; learner varieties; inter- and intra-individual variation

Abstract

While previous research on the encoding of motion events in an L2 has primarily focused on the role of L1 typological classification and L2 proficiency (e.g., Cadierno, 2004 & 2010; Filipović & Vidaković, 2010; Liste Lamas, 2024; Treffers-Daller & Tidball, 2016), less attention has been paid to task variation as well as to inter- and intra-individual variation.

Against this background, the present contribution aims to investigate group and individual differences and their relation to L1 typological classification and L2 proficiency, thus contributing to a better understanding of the encoding of motion events in an L2.

To this end, oral retellings of eight scenes from the cartoon *Canary Row* (Freleng, 1950) are compared with written descriptions of 42 video clips depicting 14 different types of paths (Liste Lamas, 2024). Data were collected from 50 L1 Spanish and 43 L1 Danish users of L2 German (ranging from level A2 to C2 of the CEFR) as well as from 33 native speakers of German, Spanish, and Danish, respectively.

As a first step, I will examine differences in the frequency of the encoding of motion events — more specifically of the path component— across the two tasks (oral retelling and written description). I will show that, in both L2 groups, path is encoded more frequently in the description task than in the retelling task. Furthermore, I will show that while the L2 oral retellings appear to be shaped by the typological classification of the L1 and its associated narrative routines, in the written description task, L1 typology alone does not fully account for the frequency and characteristics of motion event verbalization in L2 German. Instead, proficiency emerges as a more crucial factor.

In a next step, I will focus on inter-individual variation and show that, in both tasks and regardless of proficiency level, L1 Spanish users of L2 German exhibit greater variability in the linguistic means used (locus, constructions and path elements) compared to L1 Danish users of L2 German. In this context, I will discuss whether this variability can be attributed to the typological distance between the L1 and the L2.

Finally, I will address intra-individual variation as a feature of learners' varieties and examine its relationship with L2 proficiency. Preliminary results suggest that L2 intra-individual variation tends to decrease with increasing proficiency. However, the extent of intra-individual variation also appears to differ among L2 speakers and to depend on the type of motion event encoded.

Mots-clés : schémas de lexicalisation ; événements de mouvement ; allemand L2 ; encodage du parcours ; variétés d'apprenants ; variation inter- et intra-individuelle

Résumé

Alors que les recherches antérieures sur l'encodage des événements de mouvement en L2 se sont principalement concentrées sur le rôle de la classification typologique de la L1 et de la compétence en L2 (par ex., Cadierno, 2004 & 2010 ; Filipović & Vidaković, 2010 ; Liste Lamas, 2024 ; Treffers-Daller & Tidball, 2016), moins d'attention a été accordée à la variation des tâches ainsi qu'à la variation inter- et intra-individuelle.

Dans ce contexte, la présente contribution vise à étudier les différences de groupe et individuelles et leur relation avec la classification typologique de la L1 et la compétence en L2, contribuant ainsi à une meilleure compréhension de l'encodage des événements de mouvement en L2.

À cette fin, des récits oraux de huit scènes du dessin animé *Canary Row* (Freleng, 1950) sont comparés avec des descriptions écrites de 42 clips vidéo représentant 14 types différents de parcours (Liste Lamas, 2024). Les données ont été collectées auprès de 50 utilisateurs d'allemand L2 de L1 espagnole et 43 de L1 danoise (allant du niveau A2 à C2 du CECR) ainsi qu'auprès de 33 locuteurs natifs d'allemand, d'espagnol et de danois, respectivement.

Dans un premier temps, j'examinerai les différences dans la fréquence de l'encodage des événements de mouvement — plus spécifiquement de la composante parcours — entre les deux tâches (récit oral et description écrite). Je montrerai que, dans les deux groupes L2, la Trajectoire est encodée plus fréquemment dans la tâche de description que dans la tâche de récit. De plus, je montrerai que tandis que les récits oraux L2 semblent être façonnés par la classification typologique de la L1 et ses routines narratives associées, dans la tâche de description écrite, la typologie L1 seule ne rend pas pleinement compte de la fréquence et des caractéristiques de la verbalisation des événements de mouvement en allemand L2. Au lieu de cela, la compétence émerge comme un facteur plus crucial.

Dans une étape suivante, je me concentrerai sur la variation inter-individuelle et montrerai que, dans les deux tâches et quel que soit le niveau de compétence, les utilisateurs d'allemand L2 de L1 espagnole présentent une plus grande variabilité dans les moyens linguistiques utilisés (locus, constructions et éléments de Trajectoire) comparés aux utilisateurs d'allemand L2 de L1 danoise. Dans ce contexte, je discuterai si cette variabilité peut être attribuée à la distance typologique entre la L1 et la L2.

Enfin, j'aborderai la variation intra-individuelle comme une caractéristique des variétés d'apprenants et examinerai sa relation avec la compétence en L2. Les résultats préliminaires suggèrent que la variation intra-individuelle en L2 tend à diminuer avec l'augmentation de la compétence. Cependant, l'ampleur de la variation intra-individuelle semble également différer entre les locuteurs L2 et dépendre du type d'événement de mouvement encodé.

Références

- Cadierno, T. (2004). Expressing Motion Events in a Second Language: A Cognitive Typological Perspective. In: Achard, M. & Niemeier, S. (eds.), *Cognitive Linguistics, Second Language Acquisition, and Foreign Language Teaching*. Berlin/New York: De Gruyter Mouton, pp. 13–50.
- Cadierno, T. (2010). Motion in Danish as a Second Language: Does the Learner's L1 Make a Difference?. In: Han, Z. & Cadierno, T. (eds.), *Linguistic Relativity in SLA: Thinking for Speaking*. Bristol: Multilingual Matters, pp. 1-33.
- Filipovic, L. & Vidakovic, I. (2010). Typology in the L2 classroom: Second language acquisition from a typological perspective. In Putz, M. & Sicola, L. (eds.), *Inside the Learner's Mind: Cognitive Processing in Second Language Acquisition*. Amsterdam: John Benjamins, pp. 269–291.
- Freleng, F. (1950). Canary Row. New York: Time Warner.
- Liste Lamas, E. (2024). *Bewegungseignisse im Deutschen als Fremdsprache: die Kodierung der Wegkomponente bei Lernenden mit Spanisch und Dänisch als Muttersprache*. Santiago de Compostela: Universidade de Santiago de Compostela.
- Treffers-Daller, J. & Tidball, F. (2015). Can L2 learners learn new ways to conceptualise events? Evidence from motion event construal among English-speaking learners of French. In Guijarro-Fuentes, P., Schmitz, K. & Müller, N. (eds.), *The acquisition of French in multi-lingual contexts*. Bristol: Multilingual Matters, pp. 145–184.

Sylvain Loiseau (Université Sorbonne Paris Nord) & Maximilien Guérin (Université de Lille)

Description of dialects in the Croissant (north Occitan) area:

A differential approach

Keywords : Croissant; marchois; dialect; doculect; differential method; variation

Abstract

This communication focuses on the description of the dialects (doculects) within the Croissant, the northernmost occitan dialect continuum (Guerin 2020; Esher *et al.* 2021; Quint 2023). This area is marked by significant internal variation as well as notable influence from Oïl dialects and/or French language. To characterize the specific linguistic variation of this region, we propose a methodology for describing and comparing dialects based on an adaptation of the differential method (Poirier 1995, Thibault 2020). Originally developed to describe secondary dialects, the differential method involves documenting differences between a regional variety and the central standardized variety.

We have adapted this approach to describe primary dialects in a small area, applying it to pairs of adjacent dialects. Specifically, we examine three doculects within the Croissant (Dompierré-les-Églises (Guérin 2019), Lourdoueix-Saint-Michel (Guylaine Brun-Trigault, field data), Saint-Hilaire-la-Treille (Dupueux 2017), by contrast with a fourth one, Saint-Agnant-de-Versillat (Sylvain Loiseau, field data). Of course, that dialect is not a standard, it is only a *tertium comparationis*. Three of these varieties (Dompierré, Saint-Hilaire, Saint-Agnant) belong to a so called “Bas-Marchois” dialect group, according to a set of grammatical features (Guérin *et al.* 2023). We extracted all the differences in each of the three pairs through onomasiological comparison. These differences were categorized into types and subtypes (phonetic, lexical, semantic, phraseological and morphological).

This approach first makes it possible to better characterize dialectal variation within a small area: the differences between two doculects can be summarized by the number of lexical, phonetic, morphological, semantic, differences observed across the entire lexicon, which, from which one can derive 'variation profiles'. Furthermore, this approach helps reveal several recurring patterns of contrast. For instance, in Saint-Hilaire, a set of forms has acquired an initial /a-/, likely due to the generalization of a pattern arising from either local innovation or borrowing.

Saint-Agnant	Saint-Hilaire	
lyzere	alyzer	'lizard'
wɛj	awɛj	'ewe', 'ewe flock'
arkyla	rkyla / arkyla	'to go backwards'
rmyda	armyda	'to stir'
rikakwɛn	arikan	'rainbow'
vødla	avødla	'to calve'
kā kə	akā kə	'when...'

A systematic contrast between Saint-Agnant and Dompierre is the final /-e/ vs. /-aj/, as seen in Dompierre /ʃapaj/ 'hat' vs. Saint-Agnant /ʃape/, and Dompierre /rataj/ 'rake' vs. Saint-Agnant /rate/, etc. These regular patterns are not easily identifiable in linguistic atlases, which tend to depict the overall structure of a large area rather than the contrasts between two adjacent doculects. However, as with atlases, geography here provides historical insights: these contrasts may originate from local reanalysis, serve as indicators of ancient phonetic changes, and help distinguish between features that arise from diffusion and those that are inherited. They can also offer relative chronology insights.

More generally, this approach provides a measure of the intensity of the variation across a geographic area, and allows for a comprehensive yet non-redundant description of close doculects, that highlights their differences.

Mots-clés : Croissant ; marchois ; dialecte ; doclecte ; méthode différentielle ; variation

Résumé

Cette communication se concentre sur la description des dialectes (doclectes) au sein du Croissant, le continuum dialectal occitan le plus septentrional (Guérin 2020 ; Esher et al. 2021 ; Quint 2023). Cette zone est marquée par une variation interne significative ainsi qu'une influence notable des dialectes d'oïl et/ou de la langue française. Pour caractériser la variation linguistique spécifique de cette région, nous proposons une méthodologie de description et de comparaison des dialectes basée sur une adaptation de la méthode différentielle (Poirier 1995, Thibault 2020). Originellement développée pour décrire les dialectes secondaires, la méthode différentielle implique de documenter les différences entre une variété régionale et la variété centrale standardisée.

Nous avons adapté cette approche pour décrire les dialectes primaires dans une petite zone, en l'appliquant à des paires de dialectes adjacents. Spécifiquement, nous examinons trois doclectes au sein du Croissant (Dompierre-les-Églises (Guérin 2019), Lourdoueix-Saint-Michel (Guylaine Brun-Trigault, données de terrain), Saint-Hilaire-la-Treille (Dupueux 2017)), par contraste avec un quatrième, Saint-Agnant-de-Versillat (Sylvain Loiseau, données de terrain). Bien sûr, ce dialecte n'est pas un standard, c'est seulement un tertium comparationis. Trois de ces variétés (Dompierre, Saint-Hilaire, Saint-Agnant) appartiennent à un groupe dialectal dit « Bas-

Marchois », selon un ensemble de traits grammaticaux (Guérin et al. 2023). Nous avons extrait toutes les différences de chacune des trois paires lors de la comparaison onomasiologique. Ces différences ont été catégorisées en types et sous-types (phonétiques, lexicaux, sémantiques, phraséologiques et morphologiques).

Cette approche permet d'abord de mieux caractériser la variation dialectale au sein d'une petite zone : les différences entre deux doctes peuvent être résumées par le nombre de différences lexicales, phonétiques, morphologiques, sémantiques observées dans l'ensemble du lexique, à partir desquelles on peut dériver des « profils de variation ». De plus, cette approche aide à révéler plusieurs schémas récurrents de contraste. Par exemple, à Saint-Hilaire, un ensemble de formes a acquis un /a-/ initial, probablement dû à la généralisation d'un schéma résultant soit d'une innovation locale soit d'un emprunt.

Saint-Agnant	Saint-Hilaire	
lyzere	alyzer	'lézard'
wɛj	awɛj	'brebis', 'troupeau de brebis'
arkyla	rkyla / arkyla	'reculer'
rmyda	armyda	'remuer'
rikakwɛn	arikan	'arc-en-ciel'
vødla	avødla	'vêler'
kã kə	akã kə	'quand...'

Un contraste systématique entre Saint-Agnant et Dompierre est le /-e/ final vs. /-aj/, comme on le voit dans Dompierre /ʃapaj/ 'chapeau' vs. Saint-Agnant /ʃape/, et Dompierre /rataj/ 'râteau' vs. Saint-Agnant /rate/, etc. Ces schémas réguliers ne sont pas facilement identifiables dans les atlas linguistiques, qui tendent à dépeindre la structure globale d'une large zone plutôt que les contrastes entre deux doctes adjacents. Cependant, comme avec les atlas, la géographie fournit ici des aperçus historiques : ces contrastes peuvent provenir de réanalyses locales, servir d'indicateurs de changements phonétiques anciens, et aider à distinguer les traits qui résultent de la diffusion de ceux qui sont hérités. Ils peuvent aussi offrir des aperçus de chronologie relative.

Plus généralement, cette approche fournit une mesure de l'intensité de la variation à travers une zone géographique, et permet une description exhaustive mais non redondante de doctes proches, qui met en évidence leurs différences.

Références

Dupeux Michel. 2017. *Lexique (Saint-Hillaire-La-Treille)*, 4e édition, pas d'éditeur.

Esher, Louise, Maximilien Guérin, Nicolas Quint et Michela Russo. 2021. *Le Croissant linguistique : entre oc, oil et francoprovençal*. Paris: L'Harmattan.

Guérin, Maximilien. 2019. *Grammaire du parler marchois de Dompierre-les-Églises (Haute-Vienne)*. Paris: L'Harmattan.

Guérin, Maximilien. 2020. Les parlers du Croissant : des parlers de transition au cœur de l'aire gallo-romane. *Feuille de Philologie Comparée Lituanienne et Française* XI. 13-33.

Guérin, Maximilien, Nicolas Quint, Guylaine Brun-Trigaud. 2023. Les parlers du Croissant en Haute-Vienne. *Archives en Limousin* 56, 76-84.

Poirier, Claude. 1995. Les variantes topolectales du lexique français: Propositions de classement à partir d'exemples québécois. In Michel C. Francard & Danièle Latin (eds.), *Le Régionalisme lexical*, 13-56. Louvain-La-Neuve: Duculot.

Quint, Nicolas. 2023. Les parlers du Croissant : un aperçu des actions actuelles de documentation et de promotion d'un patrimoine linguistique menacé. In Annie Rialland & Michela Russo (eds.), *Les langues régionales de France Nouvelles approches, nouvelles méthodologies, revitalisation*, 213-245. Paris : Société De Linguistique De Paris.

Thibault André 2020. « Lexicographie historique et différentielle », in Gaétane Dostie, *Combats pour la linguistique au Québec (1960-2000)*. *Courants, théories, domaines*, Peter Lang, pp. 57-79.

Usage patterns and challenges in motion event encoding in German as a second language

Keywords: motion events, second language acquisition, learned attention, information packaging, information density

Abstract

Languages display a certain range of options for encoding motion events; however, preferred configurations in terms of *information packaging* result in language-specific *lexicalization patterns* (Talmy 2000), such that languages differ with respect to the habitual main *information focus* and/or *locus* (cf. Slobin 1996; Harr 2012). Differences in information packaging may also result in differences in *information density* (constructional complexity), i.e., the number of semantic components typically encoded in one clause (cf. Madlener-Charpentier 2022; Madlener-Charpentier, Skoruppa & Behrens 2017), and *tightness of packaging* (cf. Allen et al. 2007; Hendriks et al. 2022).

Typological differences may lead to challenges in second language (L2) use, due to effects of so-called *learned attention* (Ellis 2006): After years of entrenchment of first language (L1) patterns, restructuring may be challenging even for advanced L2 users of typologically different languages, e.g., German and French (Berthele & Stocker 2016; Madlener-Charpentier 2025), which differ with respect to information focus, locus, and density: German is a satellite-framed language (path of motion being typically encoded in a satellite, e.g., *raus* 'out', manner in the main verb, e.g., *rennen* 'run') with high levels of manner salience and constructional complexity (Madlener-Charpentier 2022, 2025) and a preference for tight-fit constructions. French is a verb-framed language (path being typically encoded in the main verb, e.g., *sortir* 'exit', manner being only verbalized if salient) with lower levels of manner salience and information density and a preference for semi-tight/loose-fit constructions.

This presentation examines motion event encoding in oral cartoon and picture book retellings in L2 German (L1 French; 6 participants, 571 clauses; B2-C2) as compared to L1 German (6 participants, 775 clauses), L1 French (6 participants, 676 clauses), and L2 French (L1 German; 6 participants, 610 clauses; B2-C1). Across two data sessions, each participant was presented with each stimulus twice, once in German (L1 or L2), once in French (L1 or L2). We investigate L2 users' constructional repertoires and potential learned attention effects with respect to *information packaging* (RQ1) and *information density* (RQ2):

1. RQ1: How do elicited retellings by L2 users compare to those of L1 users with respect to *information focus* and *information locus*? Do (advanced) L2 users attune to the target language's lexicalization patterns and preferred degree of manner salience?
2. RQ2: How do L2 users compare to L1 users with respect to *constructional complexity*? Do (advanced) L2 users differ from L1 users with respect to the combinatorial potential of linguistic means of expression (information density) and/or tightness of packaging?

Our analyses show that advanced L2 users of German (1) display good restructuring, overall, in terms of *information packaging* (increasing manner salience à RQ 1), but (2) still struggle, to some extent, with specific aspects of *information density* (i.e., constructional complexity and the combinatorial potential of the linguistic means available in the target language à RQ 2); (3) seem to implicitly adhere to the principle of *Maximising Common Ground* (Filipović 2021), i.e., to gravitate towards constructions that are shared/convergent between the L1 and L2 (cf. Schroeder 2009; Yilmaz 2018).

Mots-clés : événements de mouvement, acquisition d'une langue seconde, attention apprise, empaquetage informationnel, densité informationnelle

Résumé

Les langues offrent une certaine gamme d'options pour encoder les événements de mouvement ; cependant, les configurations préférées en termes d'*empaquetage informationnel* résultent en des *patterns de lexicalisation* spécifiques aux langues (Talmy 2000), de sorte que les langues diffèrent par rapport au *focus informationnel* habituel et/ou au *locus* (cf. Slobin 1996 ; Harr 2012). Les différences dans l'empaquetage informationnel peuvent également entraîner des différences de *densité informationnelle* (complexité constructionnelle), c'est-à-dire le nombre de composants sémantiques typiquement encodés dans une clause (cf. Madlener-Charpentier 2022 ; Madlener-Charpentier, Skoruppa & Behrens 2017), et de *densité d'empaquetage* (cf. Allen et al. 2007 ; Hendriks et al. 2022).

Les différences typologiques peuvent conduire à des défis en langue seconde (L2) en raison des effets de ce qu'on appelle l'*attention apprise* (Ellis 2006) : après des années d'ancrage des patterns de langue première (L1), la restructuration peut être difficile même pour les utilisateurs avancés de L2 de langues typologiquement différentes, par exemple l'allemand et le français (Berthele & Stocker 2016 ; Madlener-Charpentier 2025), qui diffèrent par rapport au focus informationnel, au locus et à la densité : l'allemand est une langue à cadrage satellitaire (la trajectoire du mouvement étant typiquement encodée dans un satellite, par ex., *raus* 'dehors', la manière dans le verbe principal, par ex., *rennen* 'courir') avec des niveaux élevés de saillance de la manière et de complexité constructionnelle (Madlener-Charpentier 2022, 2025) et une préférence pour les constructions à ajustement serré. Le français est une langue à cadrage verbal (la trajectoire étant typiquement encodée dans le verbe principal, par ex., *sortir*, la manière n'étant verbalisée que si elle est saillante) avec des niveaux plus faibles de saillance de la manière et de densité informationnelle et une préférence pour les constructions semi-serrées/lâches.

Cette présentation examine l'encodage d'événements de mouvement dans des restitutions orales de bande dessinée et de livre d'images en allemand L2 (L1 français ; 6 participants, 571 clauses ; B2-C2) comparé à l'allemand L1 (6 participants, 775 clauses), au français L1 (6 participants, 676 clauses) et au français L2 (L1 allemand ; 6 participants, 610 clauses ; B2-C1). Au cours de deux sessions de données, chaque participant s'est vu présenter chaque stimulus deux fois, une fois en allemand (L1 ou L2), une fois en français (L1 ou L2). Nous étudions les répertoires constructionnels des utilisateurs L2 et les effets potentiels d'attention apprise par rapport à l'*empaquetage informationnel* (RQ1) et à la *densité informationnelle* (RQ2) :

1. RQ1 : Comment les restitutions suscitées par les utilisateurs L2 se comparent-elles à celles des utilisateurs L1 par rapport au *focus informationnel* et au *locus informationnel* ? Les utilisateurs L2 s'accordent-ils au degré cible de saillance de la manière ?
2. RQ2 : Comment les utilisateurs L2 se comparent-ils aux utilisateurs L1 par rapport à la *complexité constructionnelle* ? Les utilisateurs L2 (avancés) diffèrent-ils des utilisateurs L1 par rapport au potentiel combinatoire des moyens linguistiques d'expression (densité informationnelle) et/ou à la densité d'empaquetage ?

Nos analyses montrent que les utilisateurs L2 avancés d'allemand (1) présentent une bonne restructuration, globalement, en termes d'*empaquetage informationnel* (saillance croissante de la manière → RQ1), mais (2) luttent encore, dans une certaine mesure, avec des aspects spécifiques de la *densité informationnelle* (c'est-à-dire la complexité constructionnelle et le potentiel combinatoire des moyens linguistiques disponibles dans la langue cible → RQ2) ; (3) semblent adhérer implicitement au principe de *Maximisation du Terrain Commun* (Filipović 2021), c'est-à-dire graviter vers des constructions qui sont partagées/convergentes entre la L1 et la L2 (cf. Schroeder 2009 ; Yilmaz 2018).

Références

Allen, S.; Özyürek, A.; Kita, S.; Brown, A.; Furman, R.; Ishizuka, T. & Fujii, M. (2007). Language-specific and universal influences in children's syntactic packaging of Manner and Path: A comparison of English, Japanese, and Turkish. *Cognition* 102, 16-48.

Madlener-Charpenler, Karin (2022). Learned attention beyond typological bootstrapping: Constructional repertoires and constructional complexity in the spatial language domain. In H. C. Boas (Hrsg.), *Directions for Pedagogical Construction Grammar: Learning and Teaching (with)Constructions*. Berlin, Boston: De Gruyter, 217-261.

Madlener-Charpenler, Karin (2025). Information packaging and constructional complexity: Challenges in motion event encoding in L2 German and L2 French. *Constructions* 17, <https://doi.org/10.24338/cons-706>

Madlener, Karin; Skoruppa, Katrin & Behrens, Heike (2017). Gradual development of constructional complexity in German spatial language. *Cognitive Linguistics* 28 (4), 757-798, <https://doi.org/10.1515/cog-2016-0089>

Berthele, R. & Stocker, L. (2016). The effect of language mode on motion event descriptions in German-French bilinguals. *Language and Cognition* 9, 1–29.

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics* 27(2), 164-194.

Filipović, L. (2021). First Language versus Second Language Effect on Memory for Motion Events: The Role of Language Type and Proficiency. *International Journal of Bilingualism*, <https://doi.org/10.1177/13670069211022863>

Harr, A.-K. (2012). Language-specific factors in first language acquisition: The expression of motion events in French and German. De Gruyter. Hendriks, H.; Hickmann, M. & Pastorino-

- Campos, C. (2022). Running or crossing? Children's expression of voluntary motion in English, German, and French. *Journal of Child Language* 49, 578-601.
- Schroeder, C. (2009). Gehen, laufen, torkeln: Eine typologisch gegründete Hypothese für den Schriftspracherwerb in der Zweitsprache Deutsch mit Erstsprache Türkisch. In K. Schramm & C. Schroeder (Eds.), *Empirische Zugänge zu Sprachförderung und Spracherwerb in Deutsch als Zweitsprache*, Waxmann, 185–202.
- Slobin, D. I. (1996). From “thought and language” to “thinking for speaking”. In J. J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity*. Cambridge University Press, 70–96.
- Talmy, L. (2000). *Toward a Cognitive Semantics*. MIT Press.
- Yilmaz, S. D. (2018). The crosslinguistic effect in motion event and judgement of typologically different conceptualization patterns by Turkish prospective teachers and native speakers of English. *European Journal of Education Studies* 4 (8), 208-243.

Beth Malory (University College London)

Translating metalinguistic attitudes into lexical reform in healthcare and policy?

Pregnancy Loss terminology in British English

Keywords: bereavement; pregnancy loss; baby loss; language policy; healthcare terminology; medical communication; mixed-methods research

Abstract

It has long been recognised that reproductive health terminology in British English impacts patients' experiences of events such as pregnancy loss (Smith et al., 2020) as well as social attitudes to such events (Austin et al., 2021; [author], 2022). However, until recently any efforts to intervene to put linguistic resources in place to alleviate these difficulties have been minimal. This is due in part to a lack of empirical data and tendency for only clinicians to highlight the negative impacts of such terminology, for example via published 'consensus statements' on diagnostic nomenclature. Such 'consensus statements' (e.g. Johnson et al., 2020) and language guidance (RCOG, 2022), have tended to draw only on anecdotal and impressionistic evidence, and have not always been focused on improving patient experience but rather on research accessibility and comparability (e.g. Kolte et al., 2015). Meanwhile, people affected by the language used in healthcare settings around experiences such as pregnancy loss in Britain have lacked a formal outlet for expressing dissatisfaction with linguistic resources available, which has instead arisen on social media. Here, many have contended that the lexicon of adverse pregnancy outcomes, such as loss, perpetuates outdated attitudes and problematic ideologies, making them unacceptable for use in clinical settings or elsewhere.

This paper draws on methodologically innovative empirical research on language attitudes and the experiential function of diagnostic lexis in contexts of pregnancy loss ([author], 2025; [author & co-authors], 2024, 2025; [author and co-authors], 2025). Combining insights from qualitative and quantitative participatory research, the paper shows that British English diagnostic terminology such as *miscarriage*, *cervical incompetence*, and *termination* are experienced as implying blame, and thus increasing self-reproach, as well as being associated with stigma. It also reports that clinical jargon, such as *empty sac*, *products of conception*, and *fetal demise* can impede comprehensibility and communicative clarity, acting as a barrier to health literacy and inclusion ([author], 2025), and can be regarded as cold and lacking in empathy. Moreover, language perceived to hierarchize experiences of loss, such as *full-term stillbirth* and *early stillbirth* prove divisive, and are embraced by some but rejected by others ([author], 2025). These findings indicate that terminology used by healthcare professionals can have a significant impact on the process of bereavement.

These findings have also facilitated evidence-based recommendations for individualized communication in clinical settings ([author], 2024), and concrete, consensus-based recommendations for mass communication ([author & co-authors], 2024), as well as the development of a health communication training course for clinicians ([author & co-authors, 2025). This paper thus argues for the crucial role of empirical interdisciplinary research and evidence-based lexical reform in intervening to improve experiences of reproductive healthcare.

Mots-clés : deuil ; perte périnatale ; perte de bébé ; politique linguistique ; terminologie des soins de santé ; communication médicale ; recherche à méthodes mixtes

Résumé

Il est reconnu depuis longtemps que la terminologie de la santé reproductive en anglais britannique impacte les expériences des patients lors d'événements tels que la perte de grossesse (Smith et al., 2020) ainsi que les attitudes sociales envers de tels événements (Austin et al., 2021 ; [auteur], 2022). Cependant, jusqu'à récemment, les efforts d'intervention pour mettre en place des ressources linguistiques visant à atténuer ces difficultés ont été minimales. Cela est dû en partie à un manque de données empiriques et une tendance des cliniciens à souligner uniquement l'impact négatif d'une telle terminologie, par exemple via des « déclarations de consensus » publiées sur la nomenclature diagnostique. De telles « déclarations de consensus » (par ex. Johnson et al., 2020) et orientations linguistiques (RCOG, 2022), ont eu tendance à ne s'appuyer que sur des preuves anecdotiques et impressionnistes, et n'ont pas toujours été centrées sur l'amélioration de l'expérience patient mais plutôt sur l'accessibilité et la comparabilité de la recherche (par ex. Kolte et al., 2015). Pendant ce temps, les personnes affectées par le langage utilisé dans les environnements de soins de santé autour d'expériences telles que la perte de grossesse en Grande-Bretagne ont manqué d'un exutoire formel pour exprimer leur insatisfaction avec les ressources linguistiques disponibles, ce qui a plutôt émergé sur les réseaux sociaux. Ici, beaucoup ont soutenu que le lexique des issues défavorables de grossesse, telles que la perte, perpétue des attitudes dépassées et des idéologies problématiques, les rendant inacceptables pour une utilisation en milieu clinique ou ailleurs.

Cet article s'appuie sur une recherche empirique méthodologiquement innovante sur les attitudes linguistiques et la fonction expérientielle du lexique diagnostique dans les contextes de perte de grossesse ([auteur], 2025 ; [auteur & co-auteurs], 2024, 2025 ; [auteur et co-auteurs], 2025). Combinant des perspectives de recherche qualitative et quantitative participative, l'article montre que la terminologie diagnostique anglaise britannique telle que *miscarriage*, *cervical incompetence*, et *termination* est vécue comme impliquant le blâme, et donc augmentant l'auto-reproche, ainsi qu'étant associée à la stigmatisation. Il rapporte également que le jargon clinique, tel que *empty sac*, *products of conception*, et *fetal demise* peut entraver la compréhensibilité et la clarté communicative, agissant comme une barrière à la littératie en santé et à l'inclusion ([auteur], 2025), et peut être perçu comme froid et manquant d'empathie. De plus, le langage perçu comme hiérarchisant les expériences de perte, tel que *full-term stillbirth* et *early stillbirth* s'avère diviseur, et est embrassé par certains mais rejeté par d'autres ([auteur], 2025). Ces résultats indiquent que la terminologie utilisée par les professionnels de santé peut avoir un impact significatif sur le processus de deuil.

Ces résultats ont également facilité des recommandations fondées sur des preuves pour une communication individualisée en milieu clinique ([auteur], 2024), et des recommandations concrètes, basées sur le consensus pour la communication de masse ([auteur & co-auteurs], 2024), ainsi que le développement d'un cours de formation en communication de santé pour les cliniciens ([auteur & co-auteurs], 2025). Cet article soutient ainsi le rôle crucial de la recherche interdisciplinaire empirique et de la réforme lexicale fondée sur des preuves pour intervenir afin d'améliorer les expériences des soins de santé reproductive.

Références

- Austin, L., Littlemore, J., McGuinness, S., Turner, S., Fuller, D., & Kuberska, K. (2021). Effective Communication Following Pregnancy Loss: A Study in England. *Cambridge Quarterly of Healthcare Ethics*, 30(1), 175–187. <https://doi.org/10.1017/S0963180120000651>
- Beard, R. W., Mowbray, J. F., & Pinker, G. D. (1985). Miscarriage or abortion. *Lancet (London, England)*, 2(8464), 1122–1123. [https://doi.org/10.1016/s0140-6736\(85\)90709-3](https://doi.org/10.1016/s0140-6736(85)90709-3)
- Gorfinkel, I. (2015, October 15). It's time to stop calling pregnancy loss 'miscarriage'. *The Globe and Mail*. <https://www.theglobeandmail.com/life/health-and-fitness/health/its-time-to-stop-calling-pregnancy-loss-miscarriage/article26823539/> (last access: 28 March 2025)
- Johnson, J., Arezina, J., Tomlin, L., Alt, S., Arnold, J., Bailey, S., Beety, H., Bender-Atik, R., Bryant, L., Coates, J., Collinge, S., Fishburn, J., Fisher, J., Fowler, J., Glanville, T., Hallett, J., Harley-Roberts, A., Harrison, G., Horwood, K., ... Hardicre, N. K. (2020). UK consensus guidelines for the delivery of unexpected news in obstetric ultrasound: The ASCKS framework. *Ultrasound*, 28(4), 235–245. <https://doi.org/10.1177/1742271X20935911>
- Kolte, A. M., Bernardi, L. A., Christiansen, O. B., Quenby, S., Farquharson, R. G., Goddijn, M., Stephenson, M. D., & on behalf of the ESHRE Special Interest Group, Early Pregnancy. (2015). Terminology for pregnancy loss prior to viability: A consensus statement from the ESHRE early pregnancy special interest group. *Human Reproduction*, 30(3), 495–498. <https://doi.org/10.1093/humrep/deu299>
- Oré, M. (2020, January 10). Women Are Calling for the Word Miscarriage to Be Banished For Good. *Glamour*. <https://www.glamour.com/story/women-are-calling-for-the-word-miscarriage-to-be-banished-for-good> (last access: 28 March 2025)
- Royal College of Obstetricians and Gynaecologists. (2022, November). *RCOG Language Guide*. Royal College of Obstetricians and Gynaecologists.
- Smith, L., Dickens, J., Bender Atik, R., Bevan, C., Fisher, J., & Hinton, L. (2020). Parents' experiences of care following the loss of a baby at the margins between miscarriage, stillbirth and neonatal death: A UK qualitative study. *BJOG: An International Journal of Obstetrics & Gynaecology*, 127(7), 868–874. <https://doi.org/10.1111/1471-0528.16113>

Michail Marinis (Harvard University & The Ohio State University)

Special Toponymic Morphology: Some lessons from the diachrony of Greek pluralia tantum toponyms

Keywords: toponyms; Special Toponymic Morphology; Special Toponymic Grammar; pluralia tantum; anti-defectiveness

Abstract

This study focuses on the diachronic examination of Greek toponyms that are defective in the singular subparadigm, building on a framework I recently proposed in Marinis (2025), according to which toponyms are governed by a *Special Toponymic Morphology*, a distinct component of Special Toponymic Grammar (Marinis 2025; Nintemann to appear; Stolz & Nintemann 2024), which in turn forms part of the broader Special Onymic Grammar (Nübling et al. 2015). Special Toponymic Morphology sets toponyms apart both from homonymous common nouns and from other types of proper names, such as anthroponyms, in that their paradigm in Greek (a) is consistently defective in one of the two numbers and (b) may exhibit *anti-defectiveness*, a notion I introduced (Marinis 2025: 87) defining it as “*The property of a subset of elements to display the inverse characteristics of defectiveness with respect to the broader set to which they belong*”.

In the present paper, I examine Modern Greek toponyms that were defective in the singular subparadigm in Classical Greek and are realized today as defective either in the singular (1a, 1b) or in the plural subparadigm (1c). Although most of these toponyms are not realized today with the same lexical form as in antiquity—in fact, many are the result of renaming—the manner in which they have re-emerged offers valuable insight into certain aspects of Special Toponymic Morphology.

(1) Classical Greek

- (a) del'foi Δελφοί 'Delphi'
- (b) my'ce:nai Μυκῆναι 'Mycenae'
- (c) a'the:nai Ἀθῆναι 'Athens'

Modern Greek

- ðel'fi Δελφοί 'Delphi'
- mi'cines Μυκῆνες 'Mycenae'
- a'θina Αθήνα 'Athens'

I examine the factors that determine why some toponyms that were defective in the singular subparadigm in Classical Greek are realized as pluralia tantum in Modern Greek, while others are not. I show that a crucial role in the selection of number in Modern Greek is played by an interaction between morphology and phonology: final segments that resemble inflectional class markers of Modern Greek contribute to the preservation of the toponym's Classical defective-in-the-singular character. For instance, the final segment *-i* in (1a, Classical Greek) coincides with the Modern Greek inflectional suffix *-i*¹¹, which marks the nominative plural cell of a particular class of masculine nouns (1a, Modern Greek). In contrast, other Classical Greek toponyms with final segments such as *-e* (1b, 1c, Classical Greek), though inflectional in the ancient system, do not correspond to any recognizable Modern Greek plural suffix. In such cases, toponyms are (re)integrated into the Modern Greek inflectional system either by shifting to the *-es* class (1b, Modern Greek), thereby maintaining defectiveness in the singular, or by altering the type of

¹¹ As part of a regular phonological change in the subsequent years, all Classical Greek /ai/ became /e/ and all /oi/ became /i/.

defectiveness and becoming defective in the plural subparadigm (1c, Modern Greek). Although their obligatory defectiveness appears to destabilize during certain periods (e.g., the era of diglossia in the 20th century), toponyms consistently end up being defective in one of the two numbers. Overall, the findings offer new insights into the morphological behaviour of proper names across time and support the theoretical validity of a distinct morphological domain for toponyms.

Mots-clés: toponymes ; morphologie toponymique spéciale ; grammaire toponymique spéciale ; pluralia tantum ; anti-défectivité

Résumé

Cette étude est consacrée à l'examen diachronique des toponymes grecs qui présentent des anomalies dans le sous-paradigme singulier, en s'appuyant sur un cadre que j'ai récemment proposé dans Marinis (2025), selon lequel les toponymes sont régis par une morphologie toponymique spéciale, une composante distincte de la grammaire toponymique spéciale (Marinis 2025 ; Nintemann à paraître ; Stolz & Nintemann 2024), qui à son tour fait partie de la grammaire onymique spéciale plus large (Nübling et al. 2015). La morphologie toponymique spéciale distingue les toponymes à la fois des noms communs homonymes et des autres types de noms propres, tels que les anthroponymes, en ce que leur paradigme en grec (a) est systématiquement défective dans l'un des deux nombres et (b) peut présenter une *anti-défectivité*, une notion que j'ai introduite (Marinis 2025 : 87) en la définissant comme « *la propriété d'un sous-ensemble d'éléments d'afficher les caractéristiques inverses de la défectivité par rapport à l'ensemble plus large auquel ils appartiennent* ».

Dans le présent article, j'examine les toponymes du grec moderne qui étaient défectifs dans le sous-paradigme du singulier du grec classique et qui sont aujourd'hui réalisés comme défectifs soit au singulier (1a, 1b), soit au pluriel (1c). Bien que la plupart de ces toponymes ne soient plus utilisés aujourd'hui sous la même forme lexicale qu'à l'Antiquité – en fait, beaucoup sont le résultat d'un changement de nom –, la manière dont ils ont réapparu offre un aperçu précieux de certains aspects de la morphologie toponymique spéciale.

(1) Grec classique

- (a) del'foi Δελφοί 'Delphi'
- (b) my'ce:nai Μυκῆναι 'Mycenae'
- (c) a'the:nai Ἀθῆναι 'Athens'

Grec moderne

- ðel'fi Δελφοί 'Delphi'
- mi'cines Μυκῆνες 'Mycenae'
- a'θina Αθήνα 'Athens'

J'examine les facteurs qui déterminent pourquoi certains toponymes qui étaient défectifs au singulier dans le sous-paradigme du grec classique sont réalisés comme pluralia tantum en grec moderne, tandis que d'autres ne le sont pas. Je montre qu'un rôle crucial dans la sélection du nombre en grec moderne est joué par une interaction entre la morphologie et la phonologie : les segments finaux qui ressemblent aux marqueurs de classe flexionnelle du grec moderne contribuent à la préservation du caractère défectif au singulier du toponyme classique. Par exemple, le segment final *-i* dans (1a, grec classique) coïncide avec le suffixe flexionnel grec moderne *-i*¹², qui marque la cellule nominative plurielle d'une classe particulière de noms

¹² Dans le cadre d'un changement phonologique régulier au cours des années suivantes, tous les /ai/ du grec classique sont devenus des /e/ et tous les /oi/ sont devenus des /i/.

masculins (1a, grec moderne). En revanche, d'autres toponymes du grec classique avec des segments finaux tels que -e (1b, 1c, grec classique), bien que flexionnels dans le système ancien, ne correspondent à aucun suffixe pluriel reconnaissable en grec moderne. Dans de tels cas, les toponymes sont (ré)intégrés dans le système flexionnel du grec moderne soit en passant à la classe -es (1b, grec moderne), conservant ainsi leur défektivité au singulier, en modifiant le type de défektivité et en devenant défectifs dans le sous-paradigme pluriel (1c, grec moderne). Bien que leur défektivité obligatoire semble se déstabiliser à certaines périodes (par exemple, l'ère de la diglossie au XXe siècle), les toponymes finissent systématiquement par être défectifs dans l'un des deux nombres. Dans l'ensemble, ces résultats offrent un nouvel éclairage sur le comportement morphologique des noms propres à travers le temps et soutiennent la validité théorique d'un domaine morphologique distinct pour les toponymes.

Références

- Marinis, M. I. 2025. On the morphology of toponyms: What (Greek) inflectional paradigms can teach us. *Transactions of the Philological Society*, 123(1), 77-96. <https://doi.org/10.1111/1467-968X.12313>
- Nintemann, J. to press. *Toponymische Sondergrammatik: Was Ortsnamen so besonders macht*.
- Nübling, D., F. Fahlbusch & R. Heuser. 2015. *Namen: eine Einführung in die Onomastik*. Tübingen: Narr Francke Attempto Verlag.
- Stolz, Th. & Nintemann, J. 2024. Special Onymic Grammar in Typological Perspective. *Cross-Linguist Data, Recurrent Patterns, Functional Explanations*. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783111331874>

Clayton Marr (Ohio State University)

Introducing the electronic Neogrammarian

Keywords: Neogrammarianism, regularity, sound-change, comparative method, forward reconstruction

Abstract

Since the comparative method reconstructs proto-forms on the basis of regularly development their cognate reflexes via exceptionless sound change, it is assumed that realizing each language's ordered chronology of sound changes on proto-forms should produce their respective observed cognate reflexes without unwanted side effects. Although Neogrammarianism is central to the comparative method, mapping (often interfering) effects of ordered sound changes in parallel over thousands of etyma is a painstakingly long error-prone task for humans. Thankfully, the regularity of sound change makes it a classical algorithm, so computers can, in a matter of seconds, perform Computerized Forward Reconstruction (or 'CFR'; Sims-Williams (2018)) to empirically evaluate whether a hypothesized relative chronology of sound changes predicts the right outcomes. The CFR-diagnostic system DiaSim (Marr and Mortensen, 2020) implements CFR and provides a suite of diagnostics for 'debugging' a hypothesized relative chronology (cascade), such as correlations of 'error' to segments and their contexts, which mismatches of aligned segments are most frequent between predicted and observed outcomes, etc. These diagnostics can help identify errors in rule ordering, rule formulation, and even hint at missed regular sound changes. A pilot study debugged a French baseline drawn from the seminal work of Pope (1934) and both independently reproduced insights from subsequent work on French (Marr and Mortensen, 2023) and detected at least sound change apparently missed by past literature (Marr, 2024). Given that CFR has demonstrated both the ability to arrive more quickly at the same conclusions as the traditional manual method and facilitate novel contributions, it is now time to bring CFR from the 'lab' into the 'field'. CFR is now being applied in other projects, including reconstructing Albanian historical phonology (Marr, 2023), informing statistical examination of analogical patterns in Romance (Herce and Marr, 2024), and validating LLM approaches to reflex prediction (Naik et al., 2024). I argue that, by empowering analysis and simulation of phonological diachrony, CFR also has much to offer for validating reconstructions, and can inform phylogeny by helping chronologize splits within the tree. Thus, while much current diachronic modeling has hitherto favored lexicostatistics over Neogrammarian models (Piwowarczyk, 2022; Ringe, 2022), I suggest that the comparative method has much to gain from a computerized refraction of Neogrammarianism.

Mots-clés: néogrammarisme, régularité, changement de son, méthode comparative, reconstruction directe

Résumé

Puisque la méthode comparative reconstruit les protoformes sur la base du développement régulier de leurs réflexes apparentés via des changements phonétiques sans exception, on suppose que la prise en compte de la chronologie ordonnée des changements phonétiques de chaque langue sur les protoformes devrait produire leurs réflexes apparentés observés sans effets secondaires indésirables. Bien que le néogrammarisme soit au cœur de la méthode comparative, cartographier les effets (souvent interférents) des changements phonétiques ordonnés en parallèle sur des milliers de racines est une tâche laborieuse et sujette aux erreurs pour les humains. Heureusement, la régularité des changements phonétiques en fait un algorithme classique, permettant aux ordinateurs d'effectuer, en quelques secondes, une reconstruction avant informatisée (ou « REI » ; Sims-Williams (2018)) afin d'évaluer empiriquement si une chronologie relative hypothétique des changements phonétiques prédit les résultats attendus. Le système de diagnostic CFR DiaSim (Marr et Mortensen, 2020) implémente la CFR et fournit une suite de diagnostics permettant de « déboguer » une chronologie relative hypothétique (cascade), comme les corrélations d'« erreur » avec les segments et leurs contextes, les incohérences de segments alignés les plus fréquentes entre les résultats prédits et observés, etc. Ces diagnostics peuvent aider à identifier les erreurs d'ordonnancement et de formulation des règles, et même à identifier des changements phonétiques réguliers manqués.

Une étude pilote a débogué une base de référence française tirée des travaux fondateurs de Pope (1934) et a reproduit indépendamment les enseignements de travaux ultérieurs sur le français (Marr et Mortensen, 2023) et a détecté au moins des changements phonétiques apparemment ignorés par la littérature antérieure (Marr, 2024). Étant donné que la CFR a démontré sa capacité à parvenir plus rapidement aux mêmes conclusions que la méthode manuelle traditionnelle et à faciliter de nouvelles contributions, il est temps de faire passer la CFR du « laboratoire » au « terrain ». La CFR est actuellement appliquée à d'autres projets, notamment la reconstruction de la phonologie historique albanaise (Marr, 2023), l'analyse statistique des schémas analogiques en langue romane (Herce et Marr, 2024) et la validation des approches LLM de prédiction réflexe (Naik et al., 2024).

Je soutiens qu'en facilitant l'analyse et la simulation de la diachronie phonologique, la CFR a également beaucoup à offrir pour la validation des reconstructions et peut éclairer la phylogénie en aidant à chronologiser les divisions au sein de l'arbre. Ainsi, alors qu'une grande partie de la modélisation diachronique actuelle a jusqu'à présent privilégié la lexicostatistique aux modèles néogrammairiens (Piwowarczyk, 2022 ; Ringe, 2022), je suggère que la méthode comparative a beaucoup à gagner d'une réflexion informatisée sur le néogrammarisme.

Références

- Fox, A. (2015). Phonological Reconstruction. In Honeybone, P. and Salmons, J., editors, *The Oxford Handbook of Historical Phonology*. Oxford University Press.
- Herce, B. and Marr, C. G. S. (2024). Diachronizing WP morphology: quantifying the complexity effects of sound change vs analogy. Presented at LLCD Workshop WS8: Future directions in Word and Paradigm morphology, 1st meeting of Langues et langage à la croisée des disciplines (LLCD) in Paris, 2024.
- Marr, C. and Mortensen, D. (2023). Large-Scale Computerized Forward Reconstruction Yields New Perspectives in French Diachronic Phonology. *Diachronica*, 40(2):238–285.
- Marr, C. and Mortensen, D. R. (2020). Computerized Forward Reconstruction for Analysis in Diachronic Phonology, and Latin to French Reflex Prediction. In *Proceedings of LT4HALA 2020-1st Workshop on Language Technologies for Historical and Ancient Languages*, pages 28–36.
- Marr, C. G. S. (2023). A computerized investigation of Albanian diachronic phonology. Presented at ICHL 26, Workshop 3: Computational models of diachronic language change, 26th meeting of International Conference of Historical Linguistics (ICHL) in Heidelberg, 2023.
- Marr, C. G. S. (2024). A missed regular sound change between Latin and French: velar onset voicing. *Indogermanische Forschungen*, 129:7–43.
- Naik, A., Zhang, K., Robinson, N., Mysore, A., Marr, C., Byrnes, H. S. R., Cai, A., Chang, K., and Mortensen, D. (2024). Can Large Language Models Code Like a Linguist?: A Case Study in Low Resource Sound Law Induction. Osthoff, H. and Brugmann, K. (1878). *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*, vol. 1. Hirzel.
- Piowarczyk, D. (2022). Computational Approaches to Linguistic Chronology and Subgrouping. In Olander, T., editor, *The Indo-European Family: A Phylogenetic Perspective*, pages 33–51. Cambridge University Press.
- Pope, M. K. (1934). *From Latin to Modern French with especial consideration of Anglo-Norman: Phonology and morphology*. Manchester University Press.
- Ringe, D. (2022). What We Can (and Can't) Learn from Computational Cladistics. In Olander, T., editor, *The Indo-European Family: A Phylogenetic Perspective*, pages 52–62. Cambridge University Press.
- Rubehn, A., Montemagni, S., and Nerbonne, J. (2024). Extracting Tuscan phonetic correspondences from dialect pronunciations automatically. *Language Dynamics and Change*, 14(1):1–33.
- Sims-Williams, P. (2018). Mechanising Historical Phonology. *Transactions of the Philological Society*, 116(3):555–573. 2/2.

Andrei Munteanu (McGill University)

Why proto-world is (likely) not reconstructible

Keywords: historical linguistics, comparative reconstruction, computational

Abstract

When using the comparative method, a researcher posits the properties of an unattested ancestor language by performing a feature-by-feature comparison of its presumed descendants and inferring the changes in between. The process is inherently temporally bound, as many diachronic processes (e.g. mergers) are not information-preserving. This notion is seconded in the literature, where it is widely assumed that the comparative method is ineffectual beyond a certain point, from 5,000 to 10,000 years approximately (Kroeber, 1960; Nichols, 1992; Sicoli & Holton, 2014; Macklin-Cordes et al., 2021). However, these limits seem to mirror the estimated age of the oldest reconstructed proto-language, typically proto-Afro-Asiatic, rather than any principled reasoning.

Herein, I approach the question of time depth and the comparative method theoretically. To this end, I employ Wordlist Distortion Theory (WDT) (Munteanu, 2024), a framework for evaluating comparative reconstructions based on the likelihood of observing the inferred changes in random data. In effect, WDT allows us to speculate about the number of diachronic transformations required to erode evidence of descent.

Diachronic transformation, for the purposes of this project comprise sound change and lexical replacement, with the rate of change for the two types approximated based on existing literature (Pagel et al., 2019; Jackson et al., 2023; Munteanu, 2024). Certain combinations of sound change and lexical replacement result in a WDT estimate above the chosen cut-off of reliability, here .05. The rate of incidence for both types of transformations is assumed to be constant though this is a simplification (Hruschka et al., 2015). However, because multiple instances of replacements of the same entry can be reanalyzed as a single instance, the number of lexical replacements required by a reconstruction grows only logarithmically. For both sound change and lexical replacement, the observed rates vary. As such, there were two estimates of the cut-off, a liberal estimate – with low rates of change for both and a deep temporal cut-off – and a conservative estimate – with high rates of change and a shallow temporal cut-off.

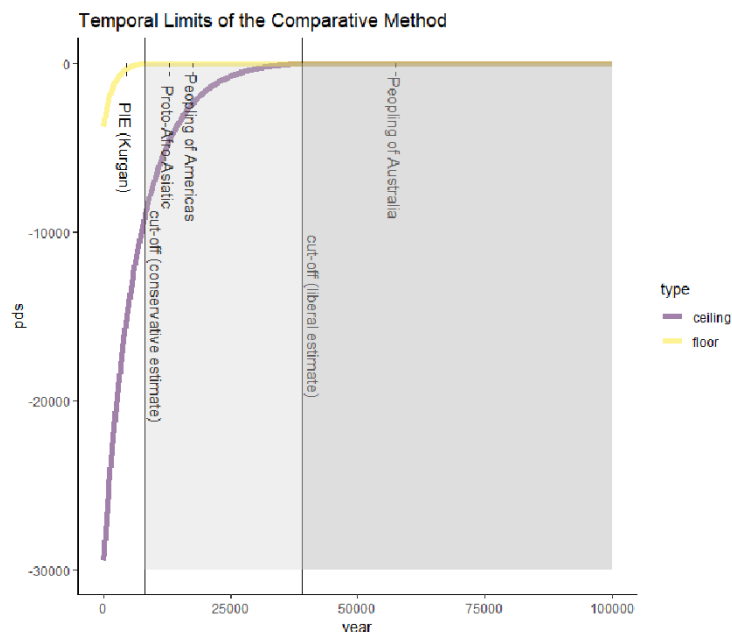


Figure 1: probability (logarithmic, y-axis) of a reconstruction from a proto-language being evidenced by chance after a give number of years (x-axis) for a conservative estimate of synchronic and diachronic factors (light yellow) and a liberal estimate of the same (dark purple).

Figure 1 presents the likelihood of observing evidence for a reconstruction in random chance for the two estimates. The results imply that the temporal limit of comparative reconstruction is between 8,300 and 39,000 years. Proto-languages that split after the conservative estimate are expected to be reliably evidenced in their descendants; proto-languages that split before the liberal estimates are likely theoretically unreconstructible; proto-languages that split in between the two may be reconstructible depending on the rates of change exhibited in their descendants. Interestingly, despite the fact that previous informal estimates about the temporal limits of the comparative method do not figure in the analysis, they are largely in line with results. As such, this method seems to indirectly confirm previous informal estimates of the comparative reconstruction limits.

Mots-clés: linguistique historique, reconstruction comparative, computationnel

Résumé

Lorsqu'il utilise la méthode comparative, le chercheur postule les propriétés d'une langue ancestrale non attestée en comparant trait par trait ses descendants présumés et en déduisant les changements intervenus entre les deux. Ce processus est intrinsèquement limité dans le temps, car de nombreux processus diachroniques (par exemple, les fusions) ne préservent pas l'information. Cette idée est corroborée par la littérature, où il est largement admis que la méthode comparative est inefficace au-delà d'un certain point, entre 5 000 et 10 000 ans environ (Kroeber, 1960 ; Nichols, 1992 ; Sicoli et Holton, 2014 ; Macklin-Cordes et al., 2021). Cependant, ces limites semblent refléter l'âge estimé de la plus ancienne proto-langue reconstruite, généralement proto-afro-asiatique, plutôt qu'un raisonnement fondé sur des principes.

J'aborde ici la question de la profondeur temporelle et de la méthode comparative de manière théorique. À cette fin, j'utilise la théorie de la distorsion des listes de mots (WDT) (Munteanu, 2024), un cadre d'évaluation des reconstructions comparatives basé sur la probabilité d'observer les changements inférés dans des données aléatoires. En effet, la WDT nous permet de spéculer sur le nombre de transformations diachroniques nécessaires pour éroder les preuves de filiation.

Dans le cadre de ce projet, la transformation diachronique comprend le changement phonétique et le remplacement lexical. Le taux de changement pour les deux types est estimé à partir de la littérature existante (Pagel et al., 2019 ; Jackson et al., 2023 ; Munteanu, 2024). Certaines combinaisons de changement phonétique et de remplacement lexical donnent une estimation du WDT supérieure au seuil de fiabilité choisi, ici 0,05. Le taux d'incidence pour les deux types de transformations est supposé constant, bien qu'il s'agisse d'une simplification (Hruschka et al., 2015). Cependant, comme plusieurs instances de remplacement d'une même entrée peuvent être réanalysées comme une seule instance, le nombre de remplacements lexicaux requis par une reconstruction n'augmente que de manière logarithmique. Pour le changement phonétique, comme pour le remplacement lexical, les taux observés varient. Il existe donc deux estimations du seuil : une estimation libérale (avec de faibles taux de variation pour les deux et un seuil temporel profond) et une estimation prudente (avec des taux de variation élevés et un seuil temporel peu profond).

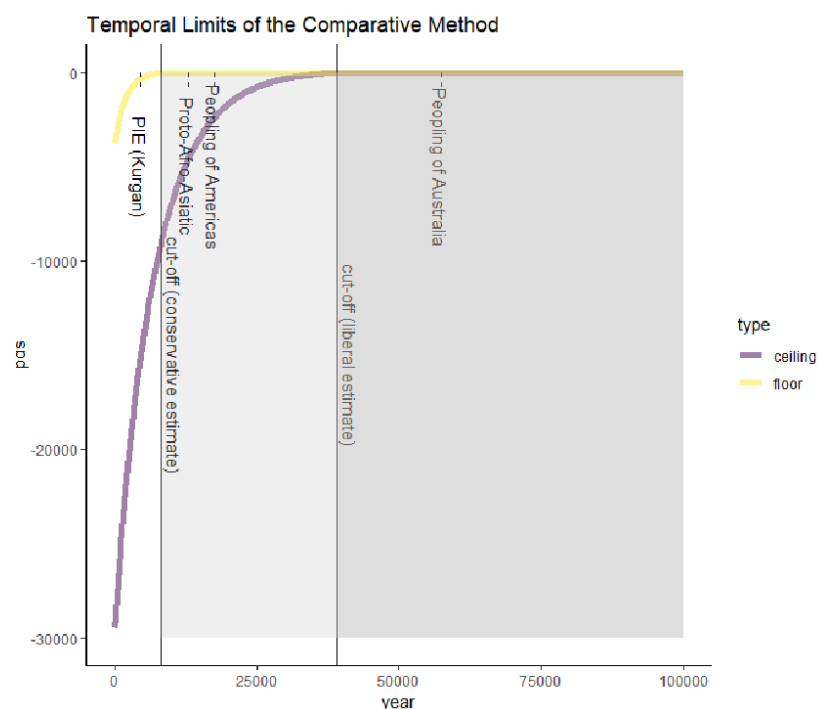


Figure 1 : Probabilité (logarithmique, axe des ordonnées) qu'une reconstruction d'une proto-langue soit mise en évidence par hasard après un nombre d'années donné (axe des abscisses) pour une estimation prudente des facteurs synchroniques et diachroniques (jaune clair) et une estimation libérale de ces mêmes facteurs (violet foncé)

La figure 1 présente la probabilité d'observer des preuves d'une reconstruction aléatoire pour les deux estimations. Les résultats impliquent que la limite temporelle de la reconstruction comparative se situe entre 8 300 et 39 000 ans. Les proto-langues qui se sont divisées après l'estimation prudente devraient être confirmées de manière fiable chez leurs descendantes ; les proto-langues qui se sont divisées avant les estimations libérales sont probablement théoriquement non restructurables ; les proto-langues qui se sont divisées entre les deux pourraient être restructurables selon les taux de changement observés chez leurs descendantes. Il est intéressant de noter que, bien que les estimations informelles antérieures concernant les limites temporelles de la méthode comparative ne figurent pas dans l'analyse, elles concordent largement avec les résultats. Ainsi, cette méthode semble confirmer indirectement les estimations informelles antérieures des limites de la reconstruction comparative.

Références

- Hruschka, D., Brantford, S., Smith, E., Wilkins, J., Meade, A., Pagel, M. & T. Bhattacharya (2015). Detecting Regular Sound Changes in Linguistics as Events of Concerted Evolution. *Current Biology*, 25(1), 1-9. <https://doi.org/10.1016/j.cub.2014.10.064>
- Jackson, J.; Lindquist, K.; Drabble, R.; Atkinson, Q. & Watts, J. (2022). Valence-dependent mutation in lexical evolution. *Nature Human Behaviour*, 7(2), 190-199. <https://doi.org/10.1038/s41562-022-01483-8>
- Kroeber, A. (1960). Statistics, Indo-European, and Taxonomy. *Language*, 36 (1), 1-21. <https://doi.org/10.2307/410621>
- Machlin-Cordes, J.; Bowern, C. & Round, E. (2021). Phylogenetic signal in phonotactics. *Diachronica*, 38(2), 210-258. <https://doi.org/10.1075/dia.20004.mac>
- Munteanu (2024). Probabilistic Evaluation of Comparative Reconstruction [doctoral dissertation]. University of Toronto.
- Nichols, J. (1992). *Linguistic Diversity in Space and Time*. University of Chicago.
- Pagel, M.; Beaumont, M.; Meade, A.; Verkerk, A. & Calude, A. (2019). Dominant words rise to the top by positive frequency-dependent selection. *PNAS*, 116(15), 7397-7402. <https://doi.org/10.1073/pnas.1816994116>
- Sicoli, M. & Holton, G. (2014). Linguistic phylogenies support back-migration from Beringia to Asia. *PLoS One*, 9(3), e91722.

Tautological constructions of the type 'What + Verb1 + Verb1' in Modern Greek and English

Keywords: constructions; tautological constructions; correlative constructions; idiomaticity; phraseology; Modern Greek; English

Abstract

1. Introduction

Tautological constructions, existing in different languages, constitute an interesting object for study. Such constructions, “totally noninformative at the level of what is said, while informative at the level of what is implicated” [Grice 1991: 33], attract the attention of both linguists and philosophers [Grice 1991, Wierzbicka 1987, Fraser 1988]. While the researchers analyse most actively “classical” nominal tautologies, such as *X copula X* (Engl. *Business is business*, Fr. *La loi, c'est la loi*), some other types of tautologies have been studied to a much lesser extent.

2. Object and methodology

This study provides a comparative analysis of correlative tautological constructions with repeating verbs of the type *What + Verb1 + Verb1* in Modern Greek (1) and English (2):

- (1) *O,ti xathike xathike.*
What be.lost.PST.3SG be.lost.PST.3SG
'What is lost, is lost.'
- (2) *What happened, happened.*

These constructions are of interest not only in the synchronic aspect, due to their productivity in Modern Greek and English respectively, as corpus data show, but also in the diachronic aspect, due to the existence of some famous textual precedents. One of the instances of the Greek construction *O,ti X X* in its archaic form *O X X* can already be found in the text of the Gospel of John (19: 21-22), widely translated into other languages, including English:

- (3) *“Μη γράφεις ο βασιλεύς των Ιουδαίων, ἀλλ' ὅτι ἐκεῖνος εἶπεν, βασιλεύς εἶμαι των Ιουδαίων”.*
ἀπεκρίθη ὁ Πιλᾶτος· ὁ γέγραφα, γέγραφα.

«“Do not write, ‘The King of the Jews,’ but, ‘He said, “I am the King of the Jews.”” Pilate answered, “**What I have written, I have written.**”» (NKJV)

The objectives of this study are: to examine tautological constructions of this type existing in Greek and English and to identify any differences between them; to analyse the internal syntax of each construction and the types of actants filling its slots; to compare the semantics of the constructions in question and to determine whether they are polysemic or have only one conventional meaning.

The study is carried out in the theoretical framework of the Construction Grammar with the use of the modern web-corpora *eTenTen19* and *enTenTen21* (www.sketchengine.eu), as well as historical corpora of Greek and English. Parallel fragments of literary texts and their translations in Greek and English are also used.

3. Analysis

The analysis of corpus data shows that the archaic form of the Greek construction, *O X X*, as in (3), survives in Modern Greek only in the form of several fixed idiomatic expressions: *O jeyrafa jeyrafa* 'What I have written, I have written', *O jeyone jeyone* 'What happened, happened'. However, the underlying syntactic pattern in its modern form, *O,ti X X*, demonstrates productivity and high compatibility with semantically and syntactically different types of actants:

- (4) *O,ti kaname kaname.*
What do.PST.1PL do.PST.1PL
'We did what we did, we can change nothing about it.'
- (5) *O,ti katalava katalava.*
What understand.PST.1SG understand.PST.1SG
'I understood what I could understand, that's it.'

While the English construction *What X X* seems also productive in corpus (2, 6–7), it no longer allows its slots to be filled with verbs having a formal subject different from what, as in (3). In the latter case, a change of construction to *X what X* is demanded, as in (8):

- (6) *What is done is done.*
(7) *What is said is said.*
(8) *You wrote what you wrote.*

We can observe this change of construction or an overall substitution of the Greek construction *O,ti X X* by another construction in English in some translations of Greek literary texts to English:

- (9) "— Έτσι; Άκου, **ό,τι είδα, είδα**. Φτάνουν και περισσεύουν." (I. Καρυσιάνη, "Σουέλ")
"— That's so? Listen, **I saw what was mine to see**. It's been plenty." (I. Karystiani, "Swell")

4. Conclusions

The functions of the Greek tautological construction *O,ti X X* are distributed in English between two different constructions: *What X X* and *X what X*. The difference of syntax of these English constructions possibly entails some semantic differences between them. Tautological constructions in different languages may have common features, but only the analysis of as many of these constructions as possible will help us to better understand their nature and see whether they constitute linguistic universals or are language-specific.

Mots-clés: constructions ; constructions tautologiques ; constructions corrélatives ; idiomaticité ; phraséologie ; grec moderne ; anglais

Résumé

1. Introduction

Les constructions tautologiques, présentes dans différentes langues, constituent un sujet d'étude intéressant. Ces constructions, « totalement non informatives au niveau de ce qui est dit, mais informatives au niveau de ce qui est sous-entendu » [Grice 1991 : 33], attirent l'attention des linguistes et des philosophes [Grice 1991, Wierzbicka 1987, Fraser 1988]. Alors que les

chercheurs analysent plus activement les tautologies nominales « classiques », telles que X copule X (en anglais **Business is business**, en français *La loi, c'est la loi*), d'autres types de tautologies ont été étudiés dans une moindre mesure.

2. Objet et méthodologie

Cette étude propose une analyse comparative des constructions tautologiques corrélatives avec répétition de verbes du type *What + Verbe1 + Verbe1* en grec moderne (1) et en anglais (2) :

(1) *O,ti xathike, xathike.*
ce qui est.perdu.PST.3SG est.perdu.PST.3SG
‘Ce qui est perdu, est perdu.’

(2) *What happened, happened.*

Ces constructions présentent un intérêt non seulement sur le plan synchronique, en raison de leur productivité en grec moderne et en anglais respectivement, comme le montrent les données du corpus, mais aussi sur le plan diachronique, en raison de l'existence de quelques précédents textuels célèbres. L'un des exemples de la construction grecque *O,ti X X* dans sa forme archaïque *O X X* se trouve déjà dans le texte de l'Évangile de Jean (19: 21-22), largement traduit dans d'autres langues, y compris l'anglais :

(3) « *Μη γράφεις ο βασιλεύς των Ιουδαίων, ἀλλ' ὅτι ἐκεῖνος εἶπεν, βασιλεὺς εἶμαι τῶν Ιουδαίων* ». ἀπεκρίθη ὁ Πιλάτος· ὁ γέγραφα, γέγραφα.

« N'écris pas : « Le roi des Juifs », mais : « Il a dit : "Je suis le roi des Juifs" ». Pilate répondit : « **Ce que j'ai écrit, je l'ai écrit** ». (NKJV)

Les objectifs de cette étude sont les suivants : examiner les constructions tautologiques de ce type existant en grec et en anglais et identifier les différences entre elles ; analyser la syntaxe interne de chaque construction et les types d'actants qui occupent ses positions ; comparer la sémantique des constructions en question et déterminer si elles sont polysémiques ou si elles n'ont qu'une seule signification conventionnelle.

L'étude est menée dans le cadre théorique de la grammaire de construction à l'aide des corpus web modernes *eITenTen19* et *enTenTen21* (www.sketchengine.eu), ainsi que de corpus historiques de grec et d'anglais. Des fragments parallèles de textes littéraires et leurs traductions en grec et en anglais sont également utilisés.

3. Analyse

L'analyse des données du corpus montre que la forme archaïque de la construction grecque, *O X X*, comme dans (3), ne survit dans le grec moderne que sous la forme de plusieurs expressions idiomatiques fixes : *O jeyrafa jeyrafa* ‘Ce que j'ai écrit, je l'ai écrit’, *O jeyone jeyone* ‘Ce qui s'est passé, s'est passé’. Cependant, le modèle syntaxique sous-jacent dans sa forme moderne, *O,ti X X*, démontre une productivité et une grande compatibilité avec des types d'actants différents sur le plan sémantique et syntaxique :

(4) *O,ti kaname kaname.*
ce que faire.PST.1PL faire.PST.1PL
‘Nous avons fait ce que nous avons fait, nous ne pouvons rien y changer.’

- (5) *O,ti katalava katalava.*
 ce que comprendre.PST.1SG comprendre.PST.1SG
 'J'ai compris ce que je pouvais comprendre, c'est tout.'

Bien que la construction anglaise *What X X* semble également productive dans le corpus (2, 6–7), elle ne permet plus que ses positions soient occupées par des verbes ayant un sujet formel différent de *what*, comme dans (3). Dans ce dernier cas, un changement de construction en *X what X* est nécessaire, comme dans (8) :

(6) *What is done is done.*

(7) *What is said is said.*

(8) *You wrote what you wrote.*

Nous pouvons observer ce changement de construction ou une substitution globale de la construction grecque *O,ti X X* par une autre construction en anglais dans certaines traductions de textes littéraires grecs :

- (9) « — *Έτσι ; Άκου, ό,τι είδα, είδα. Φτάνουν και περισσεύουν.* » (I. Καρυσιάνη, « Σουέλ »)
 « — *That's so? Listen, I saw what was mine to see. It's been plenty.* » (I. Karystiani, « Swell »)

4. Conclusions

Les fonctions de la construction tautologique grecque *O,ti X X* sont réparties en anglais entre deux constructions différentes : *What X X* et *X what X*. La différence syntaxique entre ces constructions anglaises peut entraîner certaines différences sémantiques entre elles. Les constructions tautologiques dans différentes langues peuvent avoir des caractéristiques communes, mais seule l'analyse d'un maximum de ces constructions nous aidera à mieux comprendre leur nature et à déterminer si elles constituent des universaux linguistiques ou si elles sont spécifiques à chaque langue.

Références

- Fraser, B. (1988). Motor oil is motor oil. An account of English nominal tautologies. *Journal of Pragmatics*, 12: 215-220.
- Grice, H. P. (1991). *Studies in the way of words*. Cambridge: Harvard University Press.
- Wierzbicka, A. (1987). Boys will be boys: 'Radical Semantics' vs 'Radical Pragmatics', *Language*, 63:1, 95-114.

Elena Nikishina (HSE University, Vinogradov Russian Language Institute of the Russian Academy of Sciences)

Expérience de rédaction du Dictionnaire du langage familier russe : comment décrire l'oral avec des instruments de lexicographie

Mots-clés : langage familier ; dictionnaire ; discours oral ; syntaxe ; pragmatique

Résumé

Dans mon exposé, je parlerai du *Dictionnaire du langage familier russe* (TSRRR 2014–2022), qui est créé par l'équipe d'auteurs (dont je fais partie) à l'Institut de langue russe Vinogradov de l'ASR. Ces chercheurs représentent l'École de sociolinguistique fonctionnelle de Moscou (RRR 1973 ; RRR 1978 ; RRR 1981 ; RRR 1983).

Ce dictionnaire reflète les diverses propriétés des moyens lexicaux et phraséologiques utilisés dans le discours quotidien par un citoyen d'aujourd'hui, ainsi que les particularités de leur utilisation dans différentes situations de communication spontanée informelle et non préparée. L'article de ce dictionnaire est une séquence de zones, chacune d'entre elles communiquant un certain type d'informations sur le mot (DEF, MORPH, SYNT, STYL, SYN, ANT, ANALOG, CONV, PHRAS, PRAGM). Je me focaliserai surtout sur les zones SYNT et PRAGM, qui présentent un intérêt particulier et prouvent l'originalité du dictionnaire.

La zone SYNT fournit des informations sur la structure d'arguments des verbes et d'autres prédicats. En outre, cette zone contient des informations selon lesquelles le mot peut être utilisé comme un énoncé indépendant, former des unités discursives minimales, se combiner avec certains mots. La zone SYNT peut contenir des commentaires sur le comportement syntaxique d'un mot en tant que partie d'un énoncé. À cet égard, il est particulièrement intéressant de noter les cas où la structure d'arguments du mot familier diffère de la structure d'arguments du même mot utilisé dans le discours standard littéraire :

On im malo otstegnul/ vot oni ego i sdali! 'Il leur a peu **donné** [d'argent], alors ils l'ont **dénoncé**'

Dans le langage standard littéraire, le verbe *otstegnul'* signifie 'détacher, décrocher' : *On otstegnul remen'* 'Il a **détaché** sa ceinture, *On otstegnul parašjut* 'Il a **décroché** son parachute'. Le sens 'donner' n'existe que dans le langage familier, cela veut dire que le verbe *otstegnul'* dans le langage littéraire ne peut pas avoir le COI (complément d'objet indirect) au datif (*im* 'leur'), ni être accompagné du mot ayant la sémantique quantitative en position du COD (complément d'objet direct) (*malo* 'peu'). Et pourtant, tout cela est absolument normal pour le langage familier. En ce qui concerne le verbe *sdat'*, dans le sens 'dénoncer' qui est aussi familier, il a le COD animé, tandis que dans le langage standard littéraire il ne peut avoir que le COD inanimé aux sens 'rendre qqch.' (*Esli vy uže zakončili, vy možete sdāt' vaši raboty* 'Si vous avez déjà fini, vous pouvez **rendre** vos copies') ou 'passer [un examen]'. Cette information est présentée dans la zone SYNT.

La zone PRAGM contient des informations les plus diverses (y compris extralinguistiques) sur les conditions déterminant l'utilisation du mot dans le discours oral. Entre autres, nous pouvons y trouver des cas où l'apparence phonétique des mots change où bien ils passent de la catégorie des mots dénominatifs à la catégorie des marqueurs discursifs, des particules ou des connecteurs logiques.

Voici quelques exemples des zones discutées :

KAKOJ.

dans le 1^e sens) DEF : aucun, pas du tout. *Da kakie tut mogut byt' razgovory!* 'Quel genre de conversation peut-on y avoir !'

SYNT : le mot est généralement utilisé dans une réponse sous une forme de question rhétorique ; souvent avec la particule *da* 'et' en préposition et les adverbes inaccentués *tam* 'là' ou *tut* 'ici' après *kakoj*.

PRAGM : le mot est intonné d'une manière spécifique avec une montée de ton sur la voyelle accentuée.

TAK.

(dans le 6^e sens). DEF : pas tout à fait bien, comme ci comme ça. — *Kak tam mama/ Lenočka? Mama kak? — Tak, po-vsâkomu// — Po-vsâkomu? — Konečno/ kogda boleet/ kogda...* '— Comment va maman, Lenočka ? Comment va maman ? — Comme ci comme ça, tant bien que mal. — Tant bien que mal ? — Bien sûr, soit elle est malade, soit...'

PRAGM : réponse évasive du locuteur à propos de quelque chose qui ne correspond pas à ses exigences, ses désirs, ses attentes (pas assez bien, pas assez beau, pas assez confortable, etc.) ; le mot a généralement un schéma d'intonation particulier : la voyelle est prononcée avec une plus grande intensité et un coup de glotte et est accompagnée d'une mimique appropriée.

La variété des informations présentées dans ce dictionnaire le rend intéressant et utile aux utilisateurs ordinaires aussi bien qu'aux linguistes qui étudient l'oral.

Keywords: colloquial language; dictionary; spoken discourse; syntax; pragmatics

Abstract

In my talk, I will discuss the *Dictionary of Russian Colloquial Language* (TSRRR 2014–2022), created by a team of authors (of which I am a member) at the Vinogradov Institute of the Russian Language of the Russian Academy of Sciences. These researchers represent the Moscow School of Functional Sociolinguistics (RRR 1973; RRR 1978; RRR 1981; RRR 1983).

This dictionary reflects the various properties of lexical and phraseological means used in everyday discourse by today's urban speakers, as well as the particularities of their use in different situations of informal and unprepared spontaneous communication. Each dictionary entry consists of a sequence of zones, each providing a specific type of information about the word (DEF, MORPH, SYNT, STYL, SYN, ANT, ANALOG, CONV, PHRAS, PRAGM). I will focus in particular on the SYNT and PRAGM zones, which are of special interest and demonstrate the originality of the dictionary.

The SYNT zone provides information about the argument structure of verbs and other predicates. In addition, this zone contains information on whether the word can be used as an independent utterance, form minimal discourse units, or combine with specific words. The SYNT zone may

also include comments on the syntactic behavior of a word within an utterance. In this respect, it is particularly noteworthy to observe cases where the argument structure of a colloquial word differs from the argument structure of the same word when used in standard literary discourse.

On im malo otstegnul/ vot oni ego i sdali! "He **gave** them little [money], so they **denounced** him."

In standard literary language, the verb *otstegnul'* means 'to unfasten, to detach': *On otstegnul remen* 'He **unfastened** his belt', *On otstegnul parašjut* 'He **detached** his parachute'. The meaning 'to give' exists only in colloquial language; this means that the verb *otstegnul'* in literary usage cannot take an indirect object in the dative (e.g., *im* 'to them') nor be accompanied by a quantitatively semantic word in the direct object position (e.g., *malo* 'a little'). And yet, all this is absolutely normal in colloquial usage.

As for the verb *sdal'*, in the sense 'to denounce', which is also colloquial, it takes an animate direct object, whereas in standard literary language it can only take an inanimate direct object with the meanings 'to hand something over' (e.g., *Esli vy uže zakončili, vy možete sdal' vaši raboty* 'If you have already finished, you can **hand in** your assignments') or 'to pass [an exam]'. This information is presented in the SYNT zone.

The PRAGM zone contains a wide range of information (including extralinguistic) about the conditions determining the use of a word in oral discourse. Among other things, it can indicate cases where the phonetic form of words changes, or where they shift from the category of denominative words to that of discourse markers, particles, or logical connectors.

Here are some examples of the discussed zones:

KAKOJ.

(in the 1st sense) DEF: none, not at all. *Da kakie tut mogut byt' razgovory!* 'What kind of conversation could there possibly be here!'

SYNT: the word is generally used in an answer in the form of a rhetorical question; often with the particle *da* 'and' in preposition and the unstressed adverbs *tam* 'there' or *tut* 'here' after *kakoj*.

PRAGM: the word is pronounced in a specific way with a pitch rise on the stressed vowel.

TAK.

(in the 6th sense) DEF: not quite well, so-so. — *Kak tam mama/Lenočka? Mama kak? — Tak, po-vsâkomu! — Po-vsâkomu? — Konečno/ kogda boleet/ kogda...* '— How is mother, Lenočka? How is mother? — **So-so**, in different ways. — In different ways? — Of course, sometimes she is ill, sometimes...'

PRAGM: evasive response of the speaker about something that does not meet their requirements, desires, expectations (not good enough, not pretty enough, not comfortable enough, etc.); the word generally has a particular intonation pattern: the vowel is pronounced with greater intensity and a glottal stop, and is accompanied by an appropriate facial expression.

La variété des informations présentées dans ce dictionnaire le rend intéressant et utile aux utilisateurs ordinaires aussi bien qu'aux linguistes qui étudient l'oral.

Références

TSRRR 2014–2022 : Glovinskaâ, Marina Â., Golanova, Elena I., Ermakova, Olga P., Židkova, Elena G., Zanadvorova, Anna V., Kakorina, Elena V., Kitajgorodskaâ, Margarita V., Krysin, Leonid P., Kuz'mina, Svetlana M., Nečaeva, Iâ V., Nikišina, Elena A., Pestova, Anna R., Rozanova, Nina N., Rozina, Raisa I., Šarykina, Olga A. Tolkovyj slovar' russkoj razgovornoj reči [Le dictionnaire du langage familier russe]. Volumes 1–5. Sous la direction de Leonid P. Krysin. Moscou : Maison d'édition ÂSK, 2014, 2017, 2019, 2021. Moscou, Saint-Pétérbourg : Nestor-historia Publisher, 2022.

RRR 1973 – Russkaâ razgovornaâ reč' [Le langage familier russe]. Sous la direction de E. A. Zemskaâ. Moscou : Science, 1973.

RRR 1978 – Russkaâ razgovornaâ reč' : teksty [Le langage familier russe : textes]. Sous la direction de E. A. Zemskaâ et de L. A. Kapanadze. Moscou : Science, 1978.

RRR 1981 – Zemskaâ E. A., Kitaygorodskaâ M. V., Širyaev E. N. Russkaâ razgovornaâ reč'. Obšie voprosy. Slovoobrazovanie. Sintaksis [Le langage familier russe. Questions générales. Formation des mots. Syntaxe]. Moscou : Science, 1981.

RRR 1983 – Russkaâ razgovornaâ reč' : Fonetika. Morfologiâ. Leksika. Žest [Le langage familier russe : Phonétique. Morphologie. Vocabulaire. Geste]. Sous la direction de E. A. Zemskaâ. Moscou : Science, 1983.

Yuanlong Peng (Sorbonne Nouvelle) & Xin Li (Sorbonne Université, STIH)

***Les marqueurs discursifs d'alignement en FLE : usages d'apprenants
sinophones et perspectives didactiques fondées sur corpus***

Mots-clés : marqueurs discursifs d'alignement, enseignement du FLE, apprenants sinophones, interaction.

Résumé

Les marqueurs discursifs d'alignements (désormais MDA) constituent un sous-type des marqueurs discursifs, définis comme des éléments linguistiques « qui sont syntaxiquement optionnels, hautement grammaticalisés, n'ont pas de contenu propositionnel et portent sur le niveau discursif ». Ils « expriment une fonction liée à l'accord (partage ou non d'opinions) et à l'intercompréhension (partage ou non de représentations mentales », tels que *voilà, oui, c'est clair*. (Crible, 2023 ; Crible et al., 2024).

Notre étude, fondée sur une approche expérimentale, s'intéresse à la façon dont les apprenants sinophones utilisent les MDA dans la gestion de l'interaction, au sein d'une tâche orientée. Elle vise à comparer ces usages à ceux de locuteurs natifs francophones (voir Crible, 2024), afin de dégager des pistes de didactisation pour développer chez les apprenants une compétence interactionnelle leur permettant de s'adapter à des situations réelles de communication en français. Inspirée des travaux de Bangerter et Clark (2003) et de Branigan et al. (2011), cette étude repose sur la constitution d'un corpus de productions orales recueillies auprès de 40 apprenants sinophones inscrits en troisième année de licence en langue et littérature française, dans deux universités chinoises. Les données ont été collectées en présentiel, à travers un jeu de description d'image de type *tangram* réalisée en binôme. Dans chaque interaction, un participant joue le rôle de *directeur*, chargé de décrire dix figures géométriques, tandis que l'autre, le *matcher*, doit les retrouver parmi seize images. Ce type de tâche permet d'observer en contexte réel les mécanismes d'alignement et la négociation de terrain commun (Danino et al., 2020). Les échanges, soit 20 dialogues, ont été intégralement filmés, transcrits à l'aide du logiciel ELAN, puis annotés manuellement dans TXM. L'annotation porte spécifiquement sur les MDA produits par les apprenants, en se fondant sur la typologie fonctionnelle proposée par Crible (2024), qui distingue quatre sous-fonctions interactionnelles : l'alignement temporaire, l'alignement final, la fonction négative et l'affiliation.

Nos données montrent que, comparés aux francophones natifs, les apprenants sinophones produisent moins de MDA, tant en quantité qu'en variété. L'alignement est souvent assuré par des reprises lexicales ou syntaxiques des propos de l'interlocuteur, ce qui révèle une compétence interactionnelle encore en cours de développement. Par ailleurs, les fonctions les plus fréquemment mobilisées sont l'alignement final et la non-compréhension, tandis que l'alignement temporaire et l'affiliation restent marginaux. Enfin, l'analyse met en évidence des divergences fonctionnelles dans l'usage de certains marqueurs, dont la répartition varie selon le profil des locuteurs (natifs vs apprenants sinophones). Ces différences ont été recensées sous forme d'un tableau comparatif, servant de base à notre réflexion pédagogique. Sur cette base, nous proposons d'élaborer des outils pédagogiques structurés, tels que des schémas fonctionnels et

formels (Delahaie, 2009), afin d'accompagner ces apprenants dans l'appropriation progressive de ces marqueurs, essentiels à la maîtrise de la compétence interactionnelle en français.

Keywords: alignment discourse markers, teaching French as a foreign language, Sinophone learners, interaction

Abstract

Alignment discourse markers (hereafter ADM) constitute a sub-type of discourse markers, defined as linguistic elements “which are syntactically optional, highly grammaticalized, lack propositional content, and operate at the discourse level.” They “express a function related to agreement (sharing or not of opinions) and intercomprehension (sharing or not of mental representations),” such as *voilà, oui, c'est clair* (Crible, 2023; Crible et al., 2024).

Our study, based on an experimental approach, investigates how Sinophone learners use ADM in the management of interaction within a task-based setting. It aims to compare these uses with those of native French speakers (see Crible, 2024), in order to identify pedagogical pathways for developing learners' interactional competence, enabling them to adapt to real-life communication situations in French. Inspired by the work of Bangerter & Clark (2003) and Branigan et al. (2011), this study is based on the compilation of an oral production corpus collected from 40 Sinophone students enrolled in their third year of a bachelor's degree in French language and literature at two Chinese universities. Data were collected face-to-face, through a tangram-style image description game carried out in pairs. In each interaction, one participant acted as the director, describing ten geometric figures, while the other, the matcher, had to identify them among sixteen images. This type of task makes it possible to observe, in a real context, mechanisms of alignment and the negotiation of common ground (Danino et al., 2020).

The exchanges – 20 dialogues in total – were fully video-recorded, transcribed using ELAN, and then manually annotated in TXM. The annotation focused specifically on the ADM produced by learners, drawing on the functional typology proposed by Crible (2024), which distinguishes four interactional sub-functions: temporary alignment, final alignment, negative function, and affiliation.

Our data show that, compared with native French speakers, Sinophone learners produce fewer ADM, both in quantity and in variety. Alignment is often achieved through lexical or syntactic repetition of the interlocutor's utterances, revealing an interactional competence still under development. Furthermore, the most frequently mobilized functions are final alignment and non-understanding, whereas temporary alignment and affiliation remain marginal. Finally, the analysis highlights functional divergences in the use of certain markers, whose distribution varies according to speaker profile (natives vs. Sinophone learners). These differences were systematized in a comparative table, serving as the basis for our pedagogical reflection. On this basis, we propose to develop structured teaching tools, such as functional and formal schemes (Delahaie, 2009), in order to support these learners in the progressive appropriation of these markers, which are essential for mastering interactional competence in French.

Références

- Bangerter, A. & Herbert, C. H. (2003). Navigating joint projects with dialogue. *Cognitive Science*, 27, 195– 225.
- Branigan, H. P., Catchpole, C. M. & Pickering, M. J. (2011). What makes dialogues easy to understand? *Language and Cognitive Processes*, 26(10), 1667–1686.
- Crible, L. (2023). “New data, new markers? Opening up discourse markers to expressions of alignment.” Paper presented at the International Conference on “Discourse Markers: Theory and Method”, Paris, France, May 24-26th 2024.
- Crible, L. (2024). *Ah oui d'accord okay* : marqueurs discursifs et expression de l'alignement conversationnel. In *SHS Web of Conferences* (Vol. 191, p. 01002). EDP Sciences.
- Crible, L., Gandolfi, G. & Pickering, M. J. (2024). Feedback quality and divided attention: exploring commentaries on alignment in task-oriented dialogue. *Language & Cognition* (online): 1-29.
- Danino, C., Knutsen, D. & Col, G. (2020). Naviguer dans le dialogue et faire voir ce que l'on dit : Approches linguistique et psycholinguistique de voilà. Dans M. Saiz-Sanchez, A. Rodriguez Somolinos & S. Gomez-Jordana Ferary (eds), *Marques d'oralité et représentation de l'oral en français*. Presses universitaires Savoie Mont Blanc.
- Delahaie, J. (2009). *Oui, voilà ou d'accord?* Enseigner les marqueurs d'accord en classe de FLE. *Synergies pays scandinaves*, 4, 17-34.

Quentin Peyras (Université Paul Valéry Montpellier 3 - Laboratoire ReSO)

Éléments de description d'un autre "Croissant" gallo-roman : La bande hybride de Romans-sur-Isère entre occitan et francoprovençal

Mots-clés : Francoprovençal ; morphologie ; lexique ; phonétique ; occitan

Résumé

Grâce aux travaux récents consacrés au Croissant linguistique entre les diasystèmes occitan et français (Esher & al. 2021), nous commençons à mieux percevoir l'originalité de cette frontière, qualifiée par Jakob Wüest (1979 : 354) comme « la plus importante de toute la Romania ». Divers indices laissent penser qu'un phénomène comparable de chevauchement diasystémique s'observe entre l'occitan et le francoprovençal (désormais FP), notamment dans la vallée du Rhône.

Dans la *Grammaire istorique des parlers provençaux modernes* (1930, § 12), Jules Ronjat considère la région de Romans-sur-Isère comme extérieure à l'aire occitane. Il note qu'une zone comprise entre le Rhône et Saint-Marcellin, incluant Romans, ne manifeste pas la palatalisation FP de -a après palatale (ibid. : 20). Gaston Tuaillon (2007 : 80–85) souligne également la singularité du secteur situé immédiatement au nord-ouest de Romans, où le seul critère de palatalisation FP permettant une identification claire est la double conjugaison des verbes de la première classe latine. Cette observation conduit Ronjat et Tuaillon à tracer des limites divergentes entre occitan et FP dans cette région (Fig.1).

Jean-Claude Bouvier (1971), en s'appuyant sur le critère des clitiques sujets, a classé aussi les parlers autour de Romans comme non-occitans. Ce critère se révèle toutefois peu pertinent, dans la mesure où de larges zones occitanes présentent de tels clitiques. On observe ainsi un désaccord latent dans la classification de cette aire, souvent rejetée par les spécialistes des deux diasystèmes, comme appartenant à l'autre. Jean-Michel Effantin, spécialiste des zones de contact occitan–FP, évoque pour sa part un possible brouillage du parler romanais par l'influence du français (Garnier 2020 : 16).

Pour aborder la spécificité de cette bande de transition, nous proposons une analyse contrastive de ses traits phonétiques et morphologiques, en comparaison avec l'occitan vivaro-alpin et le FP limitrophe. Nous chercherons à comprendre pourquoi ce parler demeure difficile à catégoriser, notamment à travers l'étude de sa morphologie verbale. Le dialecte romanais ne présente pas certains marqueurs morphologiques caractéristiques de l'occitan, tels que les infixations verbales typiques, tout en partageant de nombreux traits avec le FP.

Sur le plan lexical, nous proposons une cartographie de quelques adverbes caractéristiques de l'occitan, permettant d'évaluer les continuités (ou discontinuités) lexicales dans le parler de Romans.

Notre analyse s'appuie à la fois sur les enquêtes orales de Jean-Claude Bouvier pour l'*Atlas linguistique de la Provence* et pour son étude de géographie phonétique de la Drôme (1976), déposées à la Maison Méditerranéenne des Sciences de l'Homme, et sur l'abondante documentation écrite en dialecte romanais issue de la presse des XIX^e et début XX^e siècles (Effantin 1978, 2002).

Cette dernière source permet d'observer plusieurs formes des paradigmes verbaux. L'objectif de cette communication est de mieux situer ce parler hybride dans l'ensemble gallo-roman, en le comparant à d'autres zones de transition entre occitan et FP, telles que le plateau de Saint-Bonnet (Loire), le Valjouffrey ou encore l'Oisans (Isère) (Girard 2010, François & al. 2021).

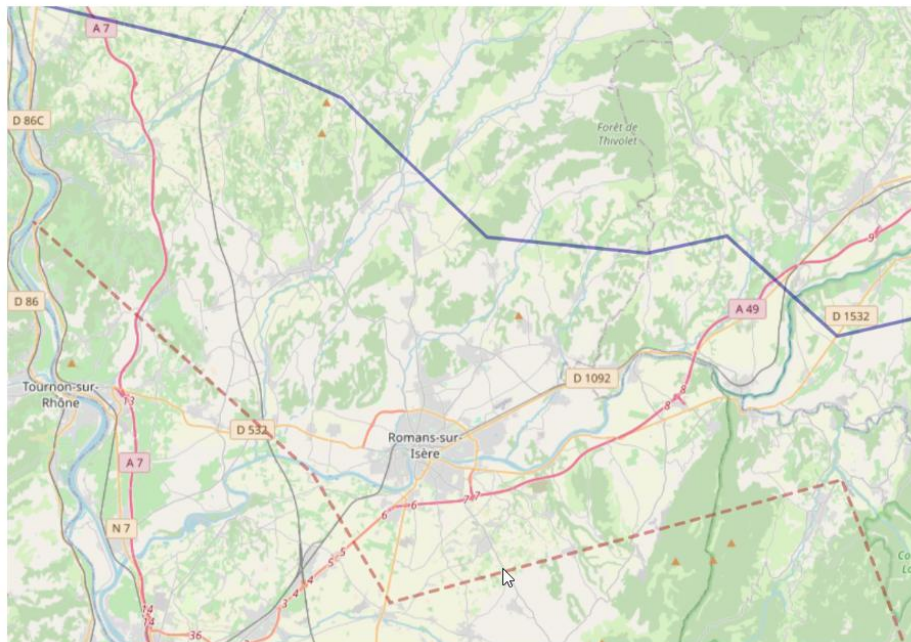


Fig. 1: Deux limites entre OC et FP. Ligne bleue continue établie par Tuaillon, rouge discontinue établie par Ronjat

Keywords: Francoprovençal; morphology; lexicon; phonetics; Occitan

Abstract

Thanks to recent studies devoted to the *Croissant linguistique* between the *Occitan* and French diasystems (Esher et al. 2021), we are beginning to better perceive the originality of this border, described by Jakob Wüest (1979: 354) as “the most important in the whole of *Romania*.” Several indications suggest that a comparable phenomenon of diasystemic overlap can be observed between *Occitan* and *Francoprovençal* (hereafter FP), particularly in the Rhône valley.

In his *Grammaire istorique des parlers provençaux modernes* (1930, §12), Jules Ronjat considered the region of Romans-sur-Isère as lying outside the *Occitan* area. He noted that the zone between the Rhône and Saint-Marcellin, including Romans, does not exhibit the FP palatalization of -a after a palatal (ibid.: 20). Gaston Tuaillon (2007: 80–85) likewise underlined the singularity of the sector located immediately northwest of Romans, where the only FP palatalization criterion allowing a clear identification is the double conjugation of verbs of the first Latin class. This observation led Ronjat and Tuaillon to draw divergent boundaries between *Occitan* and FP in this region (Fig. 1).

Jean-Claude Bouvier (1971), relying on the criterion of subject clitics, also classified the dialects around Romans as non-*Occitan*. This criterion, however, proves of limited relevance, insofar as wide *Occitan* areas also display such clitics. A latent disagreement can thus be observed in the classification of this area, often rejected by specialists of both diasystems as belonging to the other. Jean-Michel Effantin, a specialist of *Occitan*–FP contact zones, has meanwhile suggested a possible blurring of the Romans dialect under the influence of French (Garnier 2020: 16).

To address the specificity of this transition zone, we propose a contrastive analysis of its phonetic and morphological features, in comparison with *Vivaro-Alpine Occitan* and the neighboring FP. We aim to understand why this dialect remains difficult to categorize, particularly through the study of its verbal morphology. The Romans dialect lacks certain morphological markers characteristic of *Occitan*, such as typical verbal infixations, while at the same time sharing many features with FP.

On the lexical level, we propose a mapping of several adverbs characteristic of *Occitan*, in order to assess lexical continuities (or discontinuities) in the Romans dialect.

Our analysis is based both on Jean-Claude Bouvier's oral surveys for the *Atlas linguistique de la Provence* and for his study of the phonetic geography of the Drôme (1976), deposited at the *Maison Méditerranéenne des Sciences de l'Homme*, and on the abundant written documentation in the Romans dialect drawn from the press of the 19th and early 20th centuries (Effantin 1978, 2002).

This latter source makes it possible to observe several forms of verbal paradigms. The aim of this paper is to better situate this hybrid dialect within the Gallo-Roman domain, by comparing it to other transition zones between *Occitan* and FP, such as the Saint-Bonnet plateau (Loire), the Valjouffrey, or the Oisans (Isère) (Girard 2010, François et al. 2021).

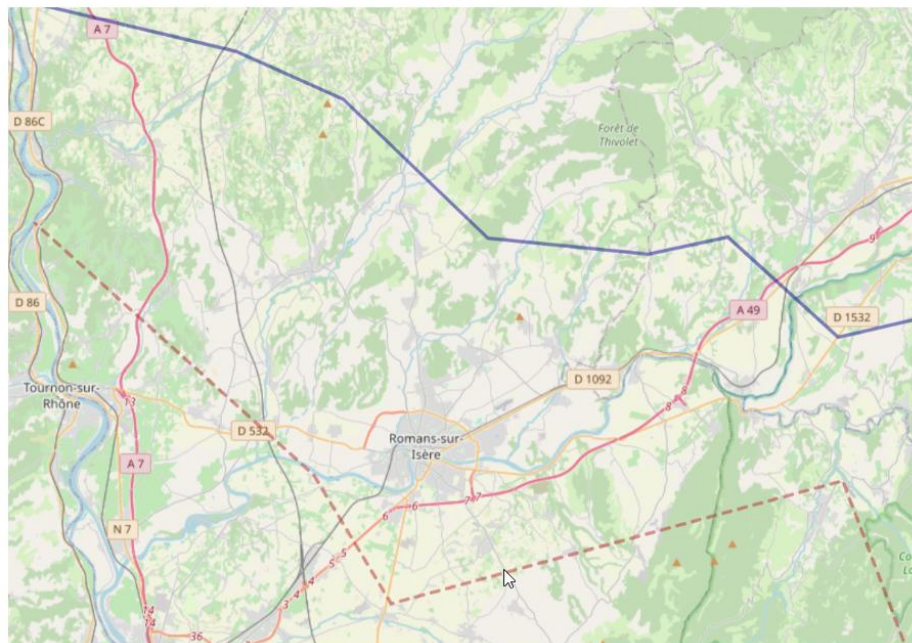


Fig. 1: Two boundaries between OC and FP. Solid blue line established by Tuailon, dashed red line established by Ronjat

Références

- Bouvier, J-C. (1971). « Le pronom personnel sujet et la frontière linguistique entre provençal et francoprovençal ». *Revue de Linguistique Romane*. 35. 1971. 1-16.
- Bouvier, J-C. & Martel, C. (1975-1986). *Atlas linguistique et ethnographique de la Provence*. CNRS.
- Bouvier, J-C. (1976). *Les parlers provençaux de la Drôme*. C. Klincksieck. Paris.
- Effantin, J-M. 1978, « Jacquemart ». *Novel Temp – Quaderno di cultura e studi occitani alpini*. 6. janvier 1978. 35-39.
- Effantin, J-M. 2003, « L'écriture de la langue locale: l'occitan a Romans, du Moyen Age a Calixte Lafosse », in: Bouvier J.C, Rixte J.C & Coste-Rixte M.C. (eds.). (2003). *Calixte Lafosse, journaliste et poète-artisan romansais*, *Revue drômoise*. T. XCV. n°507-508 (mars-juin 2003).
- Esher, L., Guérin, M., Quint, N., Russo, M. (2021). *Le Croissant linguistique : entre oc, oïl et franco-provençal - Des mots à la grammaire, des parlers aux aires*. Collection "Les Parlers du Croissant". L'Harmattan. Paris.
- François B., Orcel G. & Albertino L. (eds.). (2021). *À la recherche des patois de l'Oisans*. Association Coutumes & Traditions de l'Oisans – Groupe Patois d'Huez et Alpe d'Huez. Grenoble.
- Garnier, Q. (2020). « Le vivaro-alpin : progrès d'une définition », *Géolinguistique* [En ligne], 20 | 2020, <http://journals.openedition.org/geolinguistique/1992> (date de consultation : 1 avril 2025)
- Girard, C. (2010). *Patois du Valjouffrey* [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage) - www.ortolang.fr, v3, <https://hdl.handle.net/11403/sldr000006/v3>.
- Ronjat, J. (1930). *Grammaire istorique des parlers provençaux modernes*. T1. Vol.1. Société pour l'étude des langues romanes. Montpellier.
- Tuaillon, G. (2007). *Le Francoprovençal*. Musumeci. Quart.
- Wuest, J. (1979). *La dialectalisation de la Gallo-Romania. Problèmes phonologiques*. Romancia Helvética Vol. 91. Francke. Berne.

Morphosyntactic variation in multiverbal constructions in colloquial Arabic

Abstract

Varieties of Arabic exhibit a vast range of strategies to connect two verbs into one syntagm, some of which are demonstrated in (1a-d).

(1) Syrian Arabic

- a. *ʔləʕ-ət rakəḍ / ərkāḍ mən əl-bēt*
leave.PF-1SG run.MSD / run.MSD from DEF-house
- b. *ʔləʕ-ət rākiḍ mən əl-bēt*
leave.PF-1SG run.PTCP.M.SG from DEF-house
- c. *ʔləʕ-ət ʔə-rkuḍ mən əl-bēt*
leave.PF-1SG 1SG-run.IPF from DEF-house
- d. *ʔləʕ-ət (w-)ʕam-ʔə-rkuḍ mən əl-bēt*
leave.PF-1SG and-PROG-1SG-run.IPF from DEF-house
- 'I left home running (hastily/to go for a run...)

While not all of the patterns are available in all Arabic varieties (some of 1a-d would be claimed agrammatical by speakers of certain Syrian varieties), most of the varieties offer at least two or three of them, with functional differences of various kind (as suggested by the English translation). The goal of this paper is to present an analysis of the existing variability across a sample of Arabic varieties (several Levantine varieties, Egyptian, Moroccan and Algerian Arabic) and examine the pragmatic, semantic and syntactic factors that determine the distribution of these patterns.

While it turns out that to some extent the availability of these patterns in individual varieties is given by morphosyntactic restrictions (like that of excluding the participial form from such a structure), in many cases it is conditioned lexically, which in turn reflects pragmatic and contextual factors – for example in Moroccan Arabic a pattern analogical to (1b) is unavailable for the verb 'run', but works well for 'fly' (2):

(2) Moroccan Arabic

- ħrəž-t ʔāyir l-d-dār*
leave.PF-1SG fly.PTCP.SG.M to-DEF-house
- 'I flew out to [get] home.

Importantly, the distribution of the patterns relates strongly to particular conceptual domains – while they are most prolific in the domain of motion verbs, as seen in the examples above, the paper considers their availability in multiverbal structures in general. Among other functions, they can serve deverbal adverbial modification, as in (3):

(3) Egyptian Arabic

Baʕat-li *kitāba* *wi* *ʔal-li* *ʔe* *raʔy-ik?*
 send.PF.3SG.M-IO.1SG write.MSD and say-IO.1SG what opinion-POSS.2SG.F
 'He texted me saying "What do you think?"' ["he sent to me (in) writing..."]

The paper shows that, given the formal diversity of the structures under study, their analysis and comparison allowed by the variability across Arabic varieties can contribute importantly to general considerations about crucial syntactic notions such as dependency (Nichols 1986), coordination, subordination, verb seriation (Haspelmath 2016) and even compounding. It also argues that to adequately account for many of these structures one needs to take advantage of concepts which attempt to link notions which are a priori conceived of as opposed to each other, such as pseudocoordination (Ross 2022) or co-subordination (Foley & Van Valin 1984).

The analysis is based on natural data from the varieties in question and consultations with native Speakers.

Résumé

Les variétés de l'arabe présentent une grande diversité de stratégies pour relier deux verbes dans un même syntagme, comme illustré dans les exemples (1a–d).

(1) Arabe syrien

a. ʔləʕ-ət *rakəʔ / ərkāʔ* *mən* *əl-bēt*
 partir.PF-1SG courir.MSD / courir.MSD de DEF-maison
b. ʔləʕ-ət *rākiʔ* *mən* *əl-bēt*
 partir.PF-1SG courir.PTCP.M.SG de DEF-maison
c. ʔləʕ-ət *ʔə-rkuʔ* *mən* *əl-bēt*
 partir.PF-1SG 1SG-courir.IPF de DEF-maison
d. ʔləʕ-ət *(w-)ʕam-ʔə-rkuʔ* *mən* *əl-bēt*
 partir.PF-1SG et-PROG-1SG-courir.IPF de DEF-maison

« Je suis parti de chez moi en courant (à la hâte / pour aller courir...) »

Alors que tous les schémas ne sont pas disponibles dans toutes les variétés de l'arabe (certaines constructions comme 1a–d seraient considérées comme grammaticales uniquement en arabe syrien), la plupart des variétés en offrent au moins deux ou trois, avec des différences fonctionnelles de diverses natures (comme le suggère la traduction anglaise). L'objectif de cet article est d'analyser la variabilité existante à travers un échantillon de variétés arabes (plusieurs variétés du levant, l'arabe égyptien, marocain et algérien), et d'examiner les facteurs pragmatiques, sémantiques et syntaxiques qui déterminent la distribution de ces constructions.

Il s'avère que, dans une certaine mesure, la disponibilité de ces constructions dans chaque variété est régie par des restrictions morphosyntaxiques (par exemple, l'exclusion de la forme participiale dans une telle structure), mais dans de nombreux cas, elle est conditionnée lexicalement, ce qui reflète des facteurs pragmatiques et contextuels. Par exemple, en arabe marocain, un schéma analogue à (1b) est impossible avec le verbe "courir", mais fonctionne bien avec "voler", comme dans l'exemple (2) :

(2) Arabe marocain

ḥraḏ-t ṭāyir l-d-dār
partir.PF-1SG voler.PTCP.SG.M vers-DEF-maison

« Je me suis envolé jusqu'à chez moi. »

Fait important : la distribution des schémas est étroitement liée à certains domaines conceptuels – les constructions multiverbales sont particulièrement fréquentes dans le domaine des verbes de mouvement, comme le montrent les exemples. L'article examine aussi leur présence dans des contextes plus généraux, comme la modification adverbiale, illustrée en (3) :

(3) Arabe égyptien

Baṣat-li kitāba wi ʔal-li ʔe raʔy-ik?
envoyer.PF.3SG.M-IO.1SG écrire.MSD et dire-IO.1SG quoi opinion-POSS.2SG.F

« Il m'a envoyé un message en me disant : "Qu'en penses-tu ?" » ["Il m'a envoyé (par écrit)"]

L'article montre que, compte tenu de la diversité formelle des structures étudiées, leur analyse et la comparaison rendue possible par la variabilité entre les variétés arabes peuvent contribuer de manière significative à des réflexions plus générales sur des notions syntaxiques cruciales, telles que la dépendance (Nichols 1986), la coordination, la subordination, la sérialisation verbale (Haspelmath 2016) et même la composition. Il plaide également pour une prise en compte de concepts qui permettent de relier des unités traditionnellement conçues comme séparées, tels que la pseudo-coordination (Ross 2022) ou la co-subordination (Foley & Van Valin 1984).

L'analyse repose sur des données naturelles issues des variétés étudiées et sur des entretiens avec des locuteurs natifs.

Références

- Foley, William A. & Van Valin, Robert D., Jr. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.
- Haspelmath, Martin. (2016). The serial verb construction: Comparative concept and cross-linguistic generalizations. *Language and Linguistics* 17(3). 291–319.
- Nichols, Johanna. 1986. Head-marking and dependent-marking grammar. *Language* 62. 56–119.
- Ross, Daniel. (2022). Chapter 14. Pseudocoordination and Serial Verb Constructions as Multi-Verb Predicates. In G. Giusti, V. Di Caro & D. Ross (Ed.), *Pseudo-Coordination and Multiple Agreement Constructions* (pp. 315-336). John Benjamins Publishing Company.

**Elena Riekhakaynen (Saint-Petersburg State University) and Lada Shumakova
(Saint-Petersburg State University; HSE University)**

Towards Automatic Phonetic Transcription of Natural Speech: Can Kaldi Help?

Keywords: automatic speech recognition; natural speech; phonetic transcription; Kaldi; Russian

Abstract

We have been developing the corpora of spoken Russian to get the data for studying spoken word recognition. Their annotation includes the level of acoustic-phonetic transcription, where we differentiate the allophones that are reported to be different perceptual variants for native Russian speakers, i.e. all consonant phonemes, vowels that are between hard consonants, after or between soft consonants and before soft consonants, and [ə] for strongly reduced vowels. Such phonetic annotation, if performed manually, requires a large amount of human and time resources. For this reason, our spoken corpora are quite small (less than 5 hours of duration each). As far as we know, there are no other open spoken corpora for the Russian language provided with an acoustic-phonetic transcription, so from this point of view, Russian turns out to be a low-resource language and we are dealing with a low-resource scenario. It would be useful to have a tool for an automatic (or at least semi-automatic) phonetic transcription to enlarge the amount of phonetically annotated Russian data. To develop an algorithm of automatic speech recognition, we chose the Kaldi tool (Povey et al., 2011), which showed good results in different languages (Bu et al., 2017; Kew et al., 2020; Linke et al., 2023). We used Kaldi instead of end-to-end models because control over acoustic features and flexibility in pipeline customization are critical to our task, not just the final transcription accuracy.

As the material for testing the algorithm, we used the recordings of radio interviews and talk shows from one of our corpora. The training set included 2466 interpausal intervals, and the test set – 617 ones (we used the time codes for the pauses longer than 100 ms as the borders of interpausal intervals).

Mel-frequency cepstrum coefficients were used to extract the features, and Gaussian mixture models were used for acoustic modeling. The dictionary contained 95 different allophones, which appeared in both the training and test sets, since our task was to train the model to recognize allophones, not words. 15 models for monophone training and 15 models for triphone training were tested. The models differed from each other in the length of N-grams, which could be specified as one of the model parameters.

Triphone models coped better with the task than the monophone ones. Increasing the length of N-grams generally results in a reduction of PER (Phone Error Rate): the results of monophone and triphone models improved by about 4%. The best results were obtained with N-gram lengths of 6 or more (PER around 36%), presumably because it was the most common length of interpausal intervals in our data.

We also analyzed specific errors of the models. The frequency of allophones in the data did not affect the accuracy of their recognition. Vowels are recognized worse than consonants, which is also the case when trained phoneticians transcribe the data manually.

In our further studies, we plan to set other parameters for extracting acoustic features and to check how the speaker factor affects the quality of speech recognition.

Mots-clés: reconnaissance vocale automatique ; langage naturel ; transcription phonétique ; Kaldi ; russe

Résumé

Nous avons développé des corpus de russe parlé afin d'obtenir des données pour étudier la reconnaissance des mots parlés. Leur annotation comprend le niveau de transcription acoustico-phonétique, où nous différencions les allophones qui sont considérés comme des variantes perceptuelles différentes pour les locuteurs natifs russes, c'est-à-dire tous les phonèmes consonantiques, les voyelles situées entre des consonnes dures, après ou entre des consonnes douces et avant des consonnes douces, et [ə] pour les voyelles fortement réduites. Une telle annotation phonétique, si elle est effectuée manuellement, nécessite d'importantes ressources humaines et temporelles. C'est pourquoi nos corpus parlés sont assez petits (moins de 5 heures chacun). À notre connaissance, il n'existe aucun autre corpus oral ouvert pour la langue russe fourni avec une transcription acoustique-phonétique. De ce point de vue, le russe s'avère donc être une langue à faibles ressources et nous sommes confrontés à un scénario à faibles ressources. Il serait utile de disposer d'un outil de transcription phonétique automatique (ou au moins semi-automatique) afin d'augmenter la quantité de données russes annotées phonétiquement. Pour développer un algorithme de reconnaissance vocale automatique, nous avons choisi l'outil Kaldi (Povey et al., 2011), qui a donné de bons résultats dans différentes langues (Bu et al., 2017 ; Kew et al., 2020 ; Linke et al., 2023). Nous avons utilisé Kaldi plutôt que des modèles de bout en bout, car le contrôle des caractéristiques acoustiques et la flexibilité dans la personnalisation du pipeline sont essentiels à notre tâche, et pas seulement la précision finale de la transcription.

Pour tester l'algorithme, nous avons utilisé les enregistrements d'interviews radio et d'émissions-débats provenant de l'un de nos corpus. L'ensemble d'apprentissage comprenait 2 466 intervalles interpausaux, et l'ensemble du test 617 (nous avons utilisé les codes temporels des pauses de plus de 100 ms comme limites des intervalles interpausaux).

Les coefficients du cepstre de fréquence Mel ont été utilisés pour extraire les caractéristiques, et des modèles de mélange gaussien ont été utilisés pour la modélisation acoustique. Le dictionnaire contenait 95 allophones différents, qui apparaissaient à la fois dans les ensembles d'entraînement et de test, car notre tâche consistait à entraîner le modèle à reconnaître des allophones, et non des mots. 15 modèles pour l'entraînement monophonique et 15 modèles pour l'entraînement triphonique ont été testés. Les modèles différaient les uns des autres par la longueur des N-grammes, qui pouvait être spécifiée comme l'un des paramètres du modèle.

Les modèles triphoniques ont mieux réussi cette tâche que les modèles monophoniques. L'augmentation de la longueur des N-grammes entraîne généralement une réduction du PER (taux d'erreur phonétique) : les résultats des modèles monophoniques et triphoniques se sont améliorés d'environ 4 %. Les meilleurs résultats ont été obtenus avec des longueurs de N-grammes de 6 ou plus (PER d'environ 36 %), probablement parce qu'il s'agissait de la longueur la plus courante des intervalles interpausaux dans nos données.

Nous avons également analysé les erreurs spécifiques des modèles. La fréquence des allophones dans les données n'a pas affecté la précision de leur reconnaissance. Les voyelles sont moins bien reconnues que les consonnes, ce qui est également le cas lorsque des phonéticiens expérimentés transcrivent les données manuellement.

Dans le cadre de nos études futures, nous prévoyons de définir d'autres paramètres pour l'extraction des caractéristiques acoustiques et de vérifier comment le facteur locuteur affecte la qualité de la reconnaissance vocale.

Références

- Bu, H. Du, J., Na, X., Wu, B., & Zheng, H. (2017). AISHELL-1: An Open-Source Mandarin Speech Corpus and A Speech Recognition Baseline. In: *Proceedings of Conference of The Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Technique*. Bali: IEEE, pp. 58–62. <https://doi.org/10.1109/ICSDA.2017.8384449>
- Kew, T., Nigmatulina, I., Nagele, L., & Samardzic, T. (2020). UZH TILT: A Kaldi Recipe for Swiss German Speech to Standard German Text. In: *Proceedings of the 7th Workshop on NLP for Similar Languages, Varieties and Dialects*. Barcelona: ICCL, pp. 15–24.
- Linke, J., Wepner, S., Kubin, G., & Schupple, B. (2023). Using Kaldi for Automatic Speech Recognition of Conversational Austrian German. <https://arxiv.org/abs/2301.06475> (last access: 6 April 2025).
- Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hanneman, M., Motlíček, P., Qian, Y., Schwarz, P., Silovský, J., Stemmer, G. & Veselý, K. (2011). The Kaldi Speech Recognition Toolkit. In: *Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding, Waikoloa, 11-15 December 2011*. Waikoloa Village: IEEE Signal Processing Society.
- https://www.fit.vut.cz/research/group/speech/public/publi/2011/povey_asru2011_Kaldi%20toolkit.pdf (last access: 6 April 2025).

Jean Rohleder (LACITO-CNRS)

Grammatical change in Northern New Caledonia

Keywords: South Oceanic, syntax, New Caledonia, contact, historical linguistics, pronouns, articles

Abstract

This talk will give an overview of pronominal paradigms and articles in the available descriptions of North New Caledonian languages and compare them to map language contact. It showcases the potential of areal typology and historical syntax to map grammatical change, with examples of cognacy and analogy where published.

Northern New Caledonian languages form a language family within South Oceanic that comprises around 19 varieties, spoken in a space smaller than the Gironde département. Their speakers were historically multilingual and in regular contact with each other. This situation is typical of Melanesia, as is the diversity within the language family.

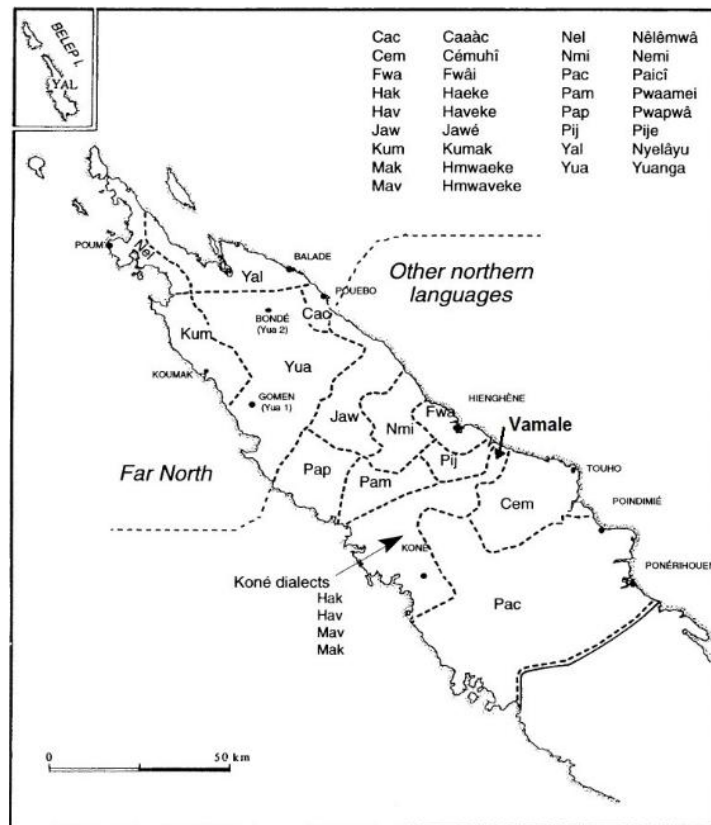


Figure 1: Northern New Caledonian languages after (Ozanne-Rivierre, 1992)

A family tree based on phonological innovations exists (Ozanne-Rivierre, 1992), but contact has a significant impact on the lexicon and grammar of these languages. For example, Pwapwā's article and pronominal paradigms are strongly influenced by its closest neighbour Jao (see Fig. 1), to which it is not closely related, rather than by its neighbour and sister Pwaamei (Ozanne-Rivierre, 2007). Thus lexical and grammatical change should be considered to better understand

the history of the languages and the contact situations in the North. This remains to be done, though efforts towards broad lexical comparison are underway.

Pronouns not only map verbal alignment but also contact situations. For example, Hienghène languages (Pije, Fwâi, Nemi) all show tripartite alignment (Haudricourt & Ozanne-Rivierre, 1982). Neighbouring, not closely related Cèmuhî and Vamale also show this alignment, but use different forms, which they share (Rivierre, 1980; Rohleder, 2024). This suggests that Vamale borrowed the tripartite alignment through Cèmuhî, which would have formed it in analogy to Hienghène systems, prior to Vamale incursion into the area between Pije and Cèmuhî.

Article systems similarly react to neighboring systems. Jao and Pwapwâ share functions and forms, as do eastern Nemi and Fwâi, and Vamale and Pije, regardless of genealogical ties. An example is the visible specific definite article that marks the highest salience in the sentence. Since the eastern Nemi form is a Fwâi loan, it seems likely that the function originated in Fwâi and then spread. Given contact geography, it would have spread in pre-colonial times to Jao via eastern Nemi, to Usa via Pije and Nemi, and to Pije via Fwâi and Nemi. As Cèmuhî has unique relative articles and lacks the visible article, the function is likely relatively recent.

Furthermore, Cèmuhî and western Voh-Koné articles are similar in form (Rivierre, 1980; Rivierre & Ehrhardt, 2006). This suggests a zone of contact from which Vamale departed and adapted to the languages further north.

Retained material across the area, especially the definite article **vi* and the indefinite **ca~ra*, hints at periods of regionally established patterns, which languages like Paicî and Jao diverged from later on, likely consciously so (Rivierre, 1983).

Other domains such as TAM markers, prepositions, and classifiers follow these analogical tendencies as well; but since the documentation is in many cases sparse, we propose two well-known paradigms as an example to be built on in the future.

Mots-clés: Océanie du Sud, syntaxe, Nouvelle-Calédonie, contact, linguistique historique, pronoms, articles

Résumé

Cette communication présentera un aperçu des paradigmes pronominaux et des articles figurant dans les descriptions disponibles des langues du nord de la Nouvelle-Calédonie, et les comparer aux contacts linguistiques. Il mettra en évidence le potentiel de la typologie spatiale et de la syntaxe historique pour cartographier les changements grammaticaux, avec des exemples de cognats et d'analogie, le cas échéant.

Les langues du nord de la Nouvelle-Calédonie forment une famille linguistique au sein de l'océanique sud, comprenant environ 19 variétés, parlées dans un espace plus restreint que le département de la Gironde. Leurs locuteurs étaient historiquement multilingues et en contact régulier. Cette situation est typique de la Mélanésie, tout comme la diversité au sein de cette famille linguistique.

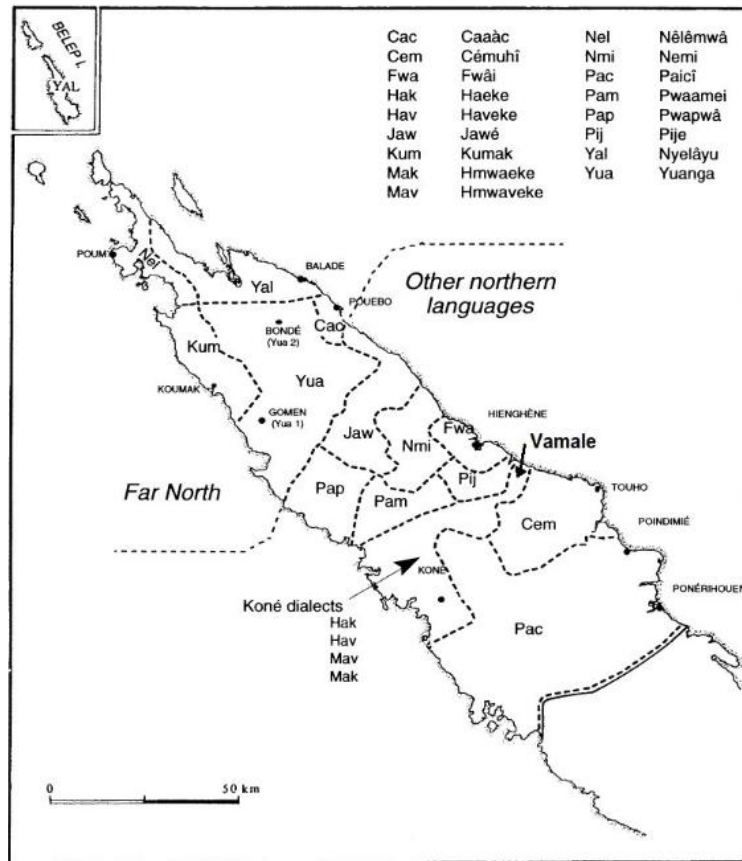


Figure 1 : Langues du nord de la Nouvelle-Calédonie d'après (Ozanne-Rivierre, 1992)

Il existe un arbre généalogique basé sur les innovations phonologiques (Ozanne-Rivierre, 1992), mais le contact a un impact significatif sur le lexique et la grammaire de ces langues. Par exemple, l'article et les paradigmes pronominaux du pwapwâ sont fortement influencés par son plus proche voisin, le jao (voir figure 1), auquel il n'est pas étroitement apparenté, plutôt que par son voisin et frère, le pwaamei (Ozanne-Rivierre, 2007).

Il convient donc de prendre en compte les changements lexicaux et grammaticaux pour mieux comprendre l'histoire des langues et les situations de contact au nord. Cela reste à faire, bien que des efforts de comparaison lexicale élargie soient en cours.

Les pronoms ne représentent pas seulement l'alignement verbal, mais aussi les situations de contact. Par exemple, les langues ienghène (pije, fwâi, nemi) présentent toutes un alignement tripartite (Haudricourt et Ozanne-Rivierre, 1982).

Le cêmuhi et le vamale, qui sont des langues voisines mais sans lien de parentalité étroit, présentent également cet alignement, mais utilisent des formes différentes, qu'ils partagent (Rivierre, 1980 ; Rohleder, 2024). Cela suggère que le vamale a emprunté l'alignement tripartite au cêmuhi, qui l'aurait formé par analogie avec les systèmes du hienghène, avant l'incursion du vamale dans la zone située entre le pije et le cêmuhi.

Les systèmes d'articles réagissent de manière similaire aux systèmes voisins. Le jao et le pwapwâ partagent des fonctions et des formes, tout comme le nemi et le fwâi orientaux, ainsi que le vamale et le pije, indépendamment des liens généalogiques. Un exemple est l'article défini spécifique visible qui marque la plus grande saillance dans la phrase. Puisque la forme du nemi orientale est un emprunt au fwâi, il semble probable que la fonction soit originaire de fwâi et se soit ensuite propagée. Compte tenu du contact géographique, elle se serait propagée à l'époque précoloniale au jao via le nemi oriental, au usa via le pije et le nemi, et au pije via le fwâi et le nemi. Comme le cèmuhi possède des articles relatifs uniques et est dépourvu d'article visible, la fonction est probablement relativement récente.

De plus, les articles en cèmuhi et en voh-koné occidental présentent une forme similaire (Rivierre, 1980 ; Rivierre et Ehrhardt, 2006). Cela suggère une zone de contact d'où le vamale s'est éloigné pour s'adapter aux langues plus au nord.

Les éléments conservés dans la région, notamment l'article défini *vi et l'article indéfini *ca~ra, suggèrent des périodes de modèles régionaux établis, dont des langues comme le paicî et le jao se sont éloignées plus tard, probablement consciemment (Rivierre, 1983).

D'autres domaines, tels que les marqueurs TAM, les prépositions et les classificateurs, suivent également ces tendances analogiques. Cependant, la documentation étant souvent rare, nous proposons deux paradigmes bien connus comme exemples à suivre à l'avenir.

Références

- Ozanne-Rivierre, F. (1992). The Proto-Oceanic Consonantal System and the Languages of New Caledonia. *Oceanic Linguistics*, 31(2), 191–207. <https://doi.org/10.2307/3623014>
- Ozanne-Rivierre, F. (1995). Structural Changes in the Languages of Northern New Caledonia. *Oceanic Linguistics*, 34(1), 45–72.
- Ozanne-Rivierre, F. (2007). *Dictionnaire thématique des langues pwaamei hnaakâ (1), pwaamei yaak (2) et pwapwâ (Nouvelle-Calédonie)* [Book].
- Haudricourt, A.-G., & Ozanne-Rivierre, F. (1982). *Dictionnaire thématique des langues de la région de Hienghène (Nouvelle-Calédonie)—Pije—Fwâi—Nemi—Jawe*. Peeters Publishers.
- Rivierre, J.-C. (1980). *La langue de Touho: Phonologie et grammaire du cèmuhi*. SELAF.
- Rivierre, J.-C. (1983). *Dictionnaire paicî-français, suivi d'un lexique français-paicî*. Peeters Publishers.
- Rivierre, J.-C., & Ehrhardt, S. (2006). *Le Bwato et les dialectes de la région de Koné (Nouvelle-Calédonie)*. Peeters.
- Rohleder, J. (2024). *A grammar of Vamale* (9). Language Science Press. <https://doi.org/10.5281/zenodo.126060>

Catherine Ruchon (Université de Lille) & Nada Jonchere (Université de Montpellier Paul-Valéry)

Questionner la nomination dans le discours institutionnel sur le palliatif. “les gens sont vivants jusqu’au bout”

Mots-clés : analyse du discours, éthique, nomination, palliatif, représentations

Résumé

Cette étude porte sur l’analyse du discours institutionnel dans le domaine palliatif. Elle consiste en une étape préliminaire d’une plus vaste étude. Il s’agit ici de circonscrire les désignations posant problème. Le corpus initial sera travaillé dans un second temps avec le concours des personnes concernées, des institutions en lien et des associations accompagnantes (telles que *Le mouvement du Nid* ou JAMALV).

Dans un domaine aussi sensible que celui des soins palliatifs, comment nommer sans heurter ? La nomination en règle générale pose des problèmes éthiques, sociaux, politiques, psychologiques. Elle a des résonances à des niveaux intimes comme à l’échelle sociétale. Une « mal-nomination » peut stigmatiser de façon insidieuse et le stigmate (Goffman, 1963) peut rester longtemps invisible. Par exemple, des mots « ordinaires » tels que *gériatre* ou *gériatrie* créent une catégorie, faisant des personnes âgées une classe à part. C’est aussi le cas d’expressions comme *mourant*, *palliatif*, ou *fin de vie*. Ainsi, peut-on désigner des individus encore en vie comme des personnes « en fin de vie » ? Ces désignations assignent une place, contribuent à la position d’acteur faible (Payet *et al.*, 2008) et à la construction de la vulnérabilité (Ghliiss *et al.*, 2019). On oublie qu’il fut un temps sans « palliatif » et que ce terme s’est répandu tardivement, à partir des années 1960 et des travaux d’Elisabeth Kübler-Ross (Vovelle, 2015 : p. 17). Ont suivi des expressions comme « en phase terminale » que l’on peut trouver en 1986 dans la première circulaire ministérielle « relative à l’organisation des soins et à l’accompagnement des malades en phase terminale » (Laroque 1986). Le terme *palliatif* est entouré de représentations morbides et ne traduit pas ce qui est à l’œuvre dans les unités de soins. La médecine palliative est une « médecine du vivant et non pas du “mourir” » (Viallard & Landrin, 2015 : p. 266). Comme le dit une infirmière d’une unité de soins palliatifs « Je bosse avec les vivants. Le mourant ne devrait pas exister. Ils sont vivants jusqu’au bout. » L’emploi du paradigme lexical *mort/mourant/mourir* traduit ainsi le même embarras lexicologique.

Les institutions sont ainsi amenées à utiliser des désignations consensuelles qui cherchent à gommer la polémique sur l’euthanasie et qui véhiculent l’idée de mort, contraire à la philosophie des unités de soin palliatif et du droit à la vie. Il incombe aux chercheur-es de visibiliser et d’analyser ces désaccords linguistiques, comme celui autour de l’« aide à mourir »¹³. Cette désignation est le fruit d’une réflexion lexicale dirigée par l’académicien Éric Orsenna à la demande de l’exécutif afin d’établir un glossaire des termes sur la fin de vie¹⁴. Ces termes posent problème. Une expression comme *aide à mourir* génère incompréhension et contestation,

¹³ <https://www.vie-publique.fr/loi/293752-fin-de-vie-projet-de-loi-soins-palliatifs-aide-mourir-pour-malades> (date de consultation : 28 mars 2025)

¹⁴ https://www.lemonde.fr/societe/article/2022/12/23/fin-de-vie-un-groupe-d-experts-mandate-pour-concevoir-un-glossaire-sur-le-sujet_6155502_3224.html (date de consultation : 28 mars 2025).

apparaissant pour certains comme l'euphémisme d'une euthanasie consentie, pour d'autres comme la tentative d'en faire un soin parmi d'autres. Institutions, grand public, communautés religieuses... tous emploient le mot *mourir* mais sans parler de la même chose. Dans une approche fonctionnaliste, la mort survient lorsque le cœur cesse de battre. Mais la mort est évidemment bien plus qu'un arrêt des fonctions organiques, elle est « irréductible à la fin de vie » (Baudry, 2015 : p. 262).

L'enjeu linguistique dans le palliatif est de trouver les mots qui fassent consensus sans en diminuer la portée symbolique et philosophique... Les tentatives inabouties de glossaires montrent cette difficulté à nommer.¹⁵ Souhaitant contribuer à une éthique de la nomination (Ruchon, 2018), cette recherche qui porte sur le lexique s'inscrit plus largement dans une analyse de discours. Le corpus regroupe des discours institutionnels de différentes sources : textes de lois en lien avec le domaine palliatif, discours politiques, ouvrages écrits par des soignants et des représentants d'associations, le Rapport de la Convention citoyenne 2023, le glossaire du Centre National des Soins Palliatifs et de la Fin de Vie (NCSPFV), notes d'entretiens. Dans le cadre d'une analyse de discours qualitative, l'objectif est de repérer les désignations dont l'usage ne semble pas aller de soi. L'attention sera notamment portée aux marqueurs de dialogisme, qui permettent de repérer du discours « autre » et le refus de son appropriation (Bres et Novakowska, 2008). Une approche croisée entre les différentes sciences humaines et sociales permettra de répondre aux attentes multiples, sociétales et éthiques.

Keywords: discourse analysis, ethics, naming, palliative care, representations

Abstract

This study focuses on the analysis of institutional discourse in the field of palliative care. It represents a preliminary stage of a broader study. The aim here is to delimit problematic designations. The initial corpus will later be examined in collaboration with the people concerned, related institutions, and supporting associations (such as *Le mouvement du Nid* or JAMALV).

In a field as sensitive as palliative care, how can one name without offending? Naming in general raises ethical, social, political, and psychological issues. It resonates on intimate as well as societal levels. A “mis-naming” can insidiously stigmatize, and stigma (Goffman, 1963) can remain invisible for a long time. For example, seemingly “ordinary” words such as *geriatrician* or *geriatrics* create a category, setting elderly people apart as a separate class. The same is true of expressions such as *dying*, *palliative*, or *end of life*. Can we designate individuals who are still alive as persons “at the end of life”? Such designations assign a place, contributing to the position of a weak actor (Payet et al., 2008) and to the construction of vulnerability (Ghliiss et al., 2019). It is often forgotten that there was a time without “palliative,” and that the term only spread belatedly, starting in the 1960s with the work of Elisabeth Kübler-Ross (Vovelle, 2015: 17). Expressions such as *in the terminal phase* soon followed, as seen in the first ministerial circular of 1986 “concerning the organization of care and support for patients in the terminal phase” (Laroque 1986). The term *palliative* is surrounded by morbid representations and does not convey what is actually at work in care units. Palliative medicine is “a medicine of the living and not of ‘dying’”

¹⁵ Notamment les projets de glossaire envisagés par le Conseil de l'Europe, la convention OVIEDO, Palliative care Europe et la SFAP.

(Viallard & Landrin, 2015: 266). As a nurse in a palliative care unit put it: “I work with the living. The dying should not exist. They are alive until the end.” The use of the lexical paradigm *dead/dying/die* thus reflects the same lexicological discomfort.

Institutions are therefore led to use consensual designations intended to erase the polemic surrounding euthanasia and to carry the idea of death – contrary to the philosophy of palliative care units and the right to life. It is up to researchers to render these linguistic disagreements visible and to analyze them, as in the debate around the *aide à mourir*¹⁶ (“*assisted dying*”). This designation resulted from a lexical reflection led by academician Éric Orsenna at the request of the executive, with the aim of establishing a glossary of terms relating to the end of life¹⁷. These terms are problematic. An expression such as *assisted dying* generates misunderstanding and contestation: for some it appears as an euphemism for consented euthanasia, for others as an attempt to make it one care practice among others. Institutions, the general public, religious communities – all use the word *die*, but without referring to the same thing. From a functionalist perspective, death occurs when the heart stops beating. But death is obviously far more than the cessation of organic functions; it is “irreducible to the end of life” (Baudry, 2015: 262).

The linguistic challenge in palliative care is to find words that foster consensus without diminishing their symbolic and philosophical weight. The unsuccessful attempts at glossaries highlight the difficulty of naming.¹⁸ Aiming to contribute to an ethics of naming (Ruchon, 2018), this research – focused on lexicon – fits more broadly within discourse analysis. The corpus brings together institutional discourse from various sources: legal texts related to palliative care, political discourse, writings by healthcare professionals and association representatives, the 2023 Citizens’ Convention Report, the glossary of the *Centre National des Soins Palliatifs et de la Fin de Vie* (NCSPFV), and interview notes. Within the framework of qualitative discourse analysis, the goal is to identify designations whose use does not seem self-evident. Special attention will be paid to dialogism markers, which signal “other” discourse and the refusal of its appropriation (Bres & Novakowska, 2008). A cross-disciplinary approach, drawing on the various human and social sciences, will help address multiple societal and ethical expectations.

¹⁶ <https://www.vie-publique.fr/loi/293752-fin-de-vie-projet-de-loi-soins-palliatifs-aide-mourir-pour-malades> (last consulted: 28 March 2025)

¹⁷ https://www.lemonde.fr/societe/article/2022/12/23/fin-de-vie-un-groupe-d-experts-mandate-pour-concevoir-un-glossaire-sur-le-sujet_6155502_3224.html (last consulted: 28 March 2025).

¹⁸ In particular, the glossary projects envisaged by the Council of Europe, the Oviedo Convention, Palliative Care Europe, and the SFAP (French Society for Palliative Care).

Références

- Baszanger, I., Salamagne, M.-H. et Lecourt, D. (dir.). (2004) *Dictionnaire de la pensée médicale*. PUF, pp. 1058-1063.
- Baudry, P. (2015). Soins palliatifs et thanatologie. Vers une mort pacifiée. In : Salamagne, M.-H. et Thominet, P. (dir.). *Accompagner. Trente ans de soins palliatifs en France*. Paris : Demopolis, pp. 257-274.
- Bres J. et Novakowska A. (2008). « J'exagère ? ... Du dialogisme interlocutif ». In : Birkelund M., Mosegaard Hansen M. B. et Norén C. (éd), *L'énonciation dans tous ses états*, Berne, Peter Lang, pp. 1-27.
- Butler Judith, 2004 [1997], *Le Pouvoir des mots. Politique du performatif*, Paris, Éditions Amsterdam.
- Fourcade, C. (2025). *Journal de la fin de vie*. Paris. Fayard.
- Fourcade, C. (2019). *Les patients au cœur, la vie dans un service de soins palliatifs*. Montrouge. Bayard.
- Fourcade, C. (2012). *1001 vies en soins palliatifs*. Montrouge. Bayard.
- Ghliss, Y., Paveau, M.-A. et Ruchon, C., (2019). Dynamiques discursives de la vulnérabilité. *Signes, discours, sociétés*, [en ligne], <<http://revue-signes.gsu.edu.tr/>>.
- Goffman, E. (1975 [1963]), *Stigmate. Les usages sociaux des handicaps*, trad. par Alain Kihm. Paris : Éditions de minuit.
- Higgins, R-W. (2006). Le statut du mourant. In P. Baudry, R-W. Higgins et J. Ricot, *Le Mourant*, M-Éditeur, Nantes.
- Laroque, G. (1986). Circulaire du 26 août 1986 ou circulaire Laroque, relative « à l'organisation des soins et à l'accompagnement des malades en phase terminale ». En ligne. <https://sfap.org/system/files/circulaire-laroque.pdf> (date de consultation : 28 mars 2025).
- Lavoie, M., Koninck, T. D. et Blondeau, D. (2009). Frontière entre la mort et le mourir. In *Laval théologique et philosophique*, 65(1), pp. 67–81. <https://doi.org/10.7202/037941ar> (date de consultation : 28 mars 2025).
- Mino, J. et Frattini, M. (2007). Les soins palliatifs en France : « mettre en pratiques » une politique de santé. *Revue française des affaires sociales*, pp. 137-156. <https://doi.org/10.3917/rfas.072.0137> (date de consultation : 28 mars 2025).
- Payet J.-P., Giuliani F. et Laforgue D., (2008). *La voix des acteurs faibles. De l'indignité à la reconnaissance*, Rennes, PUR.
- Ricœur, P. (2007 [1996]). *Vivant jusqu'à la mort* suivi de *Fragments*. Editions du Seuil.
- Ruchon, C. (2018). L'être et le nom : éthique de la nomination dans le cadre du deuil périnatal. *Langage & Société* 163, pp. 101–119.
- Salamagne, M.-H. et Hirsch, É. (1992). *Accompagner jusqu'au bout de la vie – Manifeste pour les soins palliatifs*, Cerf, Paris.

Salamagne, Michèle-H. et Thominet, P. (dir.) (2015). *Accompagner. Trente ans de soins palliatifs en France*.

Vovelle, M. (2015). Regards sur la fin de vie, histoire des soins palliatifs et des mentalités. In : Salamagne, M.-H. et Thominet, P. (dir.). *Accompagner. Trente ans de soins palliatifs en France*. Paris : Demopolis, pp. 12-20.

Viallard, M.-L. et Landrin, T. (2015). Des soins palliatifs à la médecine palliative. Un changement de paradigme. In : Salamagne, M.-H. et Thominet, P. (dir.), *Accompagner. Trente ans de soins palliatifs en France*. Paris : Demopolis, pp. 263-274.

Worms, F. (2010). *Le moment du soin : À quoi tenons-nous ?* Presses Universitaires de France.
<https://doi.org/10.3917/puf.worms.2010.01> (date de consultation : 28 mars 2025).

Anindita Sahoo (Indian Institute of Technology Madras), Celia Shiyu He (University of Bristol), Krishna Raj S. R. (Indian Institute of Technology Madras) & Dagmar Divjak (University of Birmingham)

Investigating Cognitive Processing Using Eye-Tracking: A Study on Punjabi Grammatical Voice Constructions

Keywords: Voice constructions, Visual world paradigm, Eye-movements, Generalized Additive Mixed models

Abstract

Grammatical voice, as a morphosyntactic phenomenon, signifies a specific relationship between the meaning and form of clauses. Its values correspond to particular diatheses marked in the form of predicates (Siewierska 1984; Shibatani 1985; Zuñiga & Kittilä 2019). This current study takes a construction grammar based psycholinguistic approach (Divjak et al. 2022) and examines the passive voice constructions in Punjabi along with the associated cognitive load that they carry. It also explores the cognitive correlates of alternating constructions involving active and passive voice.

To assess cognitive processing, we followed Divjak, Milin, & Medimorec (2020) and used a visual world paradigm, focusing on visual processing differences in comprehension among active voice (AV), passive voice (PV), and active voice with light verb (ALV) constructions in Punjabi (1-3).

Forty-two L1 speakers (16 male and 26 female) of Punjabi, with ages ranging from 19 to 79 years (mean age: 41.57, SD = 18.69), residing in Birmingham, UK, were recruited and tested in the community using a Gazepoint GP3 eye-tracker with a 60Hz sampling rate. The experiment consists of a free viewing phase followed by the display of scenes in three different voice constructions before revisiting the images.

Eye-tracking data, including gaze duration and saccade patterns, alongside pupil dilation metrics, were recorded to assess participants' attention and cognitive load during image viewing. The analysis determined the influences of syntactic structure on visual attention and cognitive effort, as reflected by gaze behavior and pupil dynamics.

Following Divjak, Milin, & Medimorec (2020) the *order of access*, *time to first view*, and *total dwell time* in the *area of interest* (AOI) were analyzed using Generalized Additive Models (GAM). The preliminary modeling of *total dwell time* revealed several significant predictors among the fixed effects, including the interaction of *voice type* and *argument relation*. Additionally, while *gender* emerged as a significant factor, *age* did not demonstrate much significance. The influence of *gender* in this study builds on prior research indicating differences in gaze patterns between men and women when viewing images (Alexander & Son 2009). The analysis of the *time to first view* revealed that among the fixed effects, the patient triggers a longer *time to first view*. Lastly, the *order of access* shows that the *argument relation* (*agent/patient dichotomy*) is influential. *Agents* are associated with a *lower order of access* than patients, holding all else constant.

This study reveals that Punjabi grammatical voice constructions impose varying cognitive demands, with syntactic structure and argument relations significantly shaping visual attention and cognitive load. These findings contribute to understanding the cognitive processing of

morphosyntactic phenomena and emphasize the utility of psycholinguistic methods in exploring lesser-studied languages.

Examples

(1) *loka-ne munde-nu buri tarah kuTe-y-a* (Active voice)

people-ERG man-ACC bad like hit-PFV-3PL

'People hit the man badly.'

(2) *munde-nu buri tarah kuT-ya-a gayaa* (Passive voice)

man-ACC bad like hit-PST go-PASS

'The man was hit badly.'

(3) *loka-ne munde-nu buri tarah kuT dita* (Active voice with light verb)

people-ERG man-ACC bad like hit give-PST-3PL

'People hit the man badly.' (Mild assertive reading)

Mots-clés : Constructions de voix, Paradigme de monde visuel, Mouvements oculaires, Modèles additifs mixtes généralisés

Résumé

La voix grammaticale, en tant que phénomène morphosyntaxique, signifie une relation spécifique entre le sens et la forme des propositions. Ses valeurs correspondent à des diathèses particulières marquées dans la forme des prédicats (Siewierska 1984 ; Shibatani 1985 ; Zuñiga & Kittilä 2019). Cette étude adopte une approche psycholinguistique basée sur la grammaire constructionnelle (Divjak et al. 2022) et examine les constructions de voix passive en punjabi ainsi que la charge cognitive associée qu'elles portent. Elle explore également les corrélats cognitifs des constructions alternées impliquant la voix active et passive.

Pour évaluer le traitement cognitif, nous avons suivi Divjak, Milin, & Medimorec (2020) et utilisé un paradigme de monde visuel, en nous concentrant sur les différences de traitement visuel dans la compréhension parmi les constructions de voix active (VA), voix passive (VP), et voix active avec verbe léger (VAL) en punjabi (1-3).

Quarante-deux locuteurs natifs de punjabi L1 (16 hommes et 26 femmes), âgés de 19 à 79 ans (âge moyen : 41,57, ET = 18,69), résidant à Birmingham, Royaume-Uni, ont été recrutés et testés dans la communauté en utilisant un oculomètre Gazepoint GP3 avec un taux d'échantillonnage de 60Hz. L'expérience consiste en une phase de visualisation libre suivie de l'affichage de scènes dans trois constructions de voix différentes avant de revisiter les images.

Les données d'oculométrie, incluant la durée de fixation et les patterns de saccades, ainsi que les métriques de dilatation pupillaire, ont été enregistrées pour évaluer l'attention et la charge cognitive des participants pendant la visualisation d'images. L'analyse a déterminé les influences de la structure syntaxique sur l'attention visuelle et l'effort cognitif, tel que reflété par le comportement de fixation et la dynamique pupillaire.

En suivant Divjak, Milin, & Medimorec (2020), l'*ordre d'accès*, le *temps jusqu'à la première visualisation*, et le *temps total de fixation* dans la *zone d'intérêt (ZI)* ont été analysés en utilisant des Modèles Additifs Généralisés (MAG). La modélisation préliminaire du *temps total de fixation*

a révélé plusieurs prédicteurs significatifs parmi les effets fixes, incluant l'interaction du *type de voix* et de la *relation d'argument*. De plus, alors que le *genre* est apparu comme un facteur significatif, l'*âge* n'a pas démontré beaucoup de significativité. L'influence du *genre* dans cette étude s'appuie sur des recherches antérieures indiquant des différences dans les *patterns* de fixation entre hommes et femmes lors de la visualisation d'images (Alexander & Son 2009). L'analyse du *temps jusqu'à la première visualisation* a révélé que parmi les effets fixes, le *patient* déclenche un *temps jusqu'à la première visualisation* plus long. Enfin, l'*ordre d'accès* montre que la *relation d'argument* (dichotomie agent/patient) est influente. Les agents sont associés à un *ordre d'accès* plus faible que les patients, toutes choses égales par ailleurs.

Cette étude révèle que les constructions de voix grammaticales en punjabi imposent des demandes cognitives variables, avec la structure syntaxique et les relations d'arguments façonnant significativement l'attention visuelle et la charge cognitive. Ces résultats contribuent à la compréhension du traitement cognitif des phénomènes morphosyntaxiques et soulignent l'utilité des méthodes psycholinguistiques dans l'exploration des langues moins étudiées.

Exemples

(1) *loka-ne munde-nu buri tarah kuTe-y-a* (Active voice)

people-ERG man-ACC bad like hit-PFV-3PL

« Les gens ont frappé l'homme violemment. »

(2) *munde-nu buri tarah kuT-ya-a gayaa* (Passive voice)

man-ACC bad like hit-PST go-PASS

« L'homme a été frappé violemment. »

(3) *loka-ne munde-nu buri tarah kuT dita* (Active voice with light verb)

people-ERG man-ACC bad like hit give-PST-3PL

« Les gens ont frappé l'homme violemment. » (Lecture assertive atténuée)

Références

Alexander, G. M., and Son, T. (2007). Androgens and eye movements in women and men during a test of mental rotation ability. *Hormones and Behavior*, 52(2), 197-204.

Divjak, D., Milin, P., and Medimorec, S. (2020). Construal in language: A visual-world approach to the effects of linguistic alternations on event perception and conception. *Cognitive Linguistics*, 31(1), 37–72. <https://doi.org/10.1515/cog-2018-0103>

Divjak, D., Milin, P., Medimorec, S., & Borowski, M. (2022). Behavioral signatures of memory resources for language: Looking beyond the lexicon/grammar divide. *Cognitive Science*, 46(11): e13206.

Shibatani, M. (1985). Passive and related constructions: a prototype analysis. *Language* 61: 821-48.

Siewierska, A. (1984). *The passive: a contrastive linguistic analysis*. London: Croom Helm.

Zúñiga, F. and Seppo K. (2019). *Grammatical voice*. Cambridge: Cambridge University Press.

Maguelone Sauzet (Université de Neuchâtel)

Le marquage du genre grammatical dans les parlers francoprovençaux

Mots-clés : dialectologie ; francoprovençal ; morphologie ; genre grammatical ; atlas linguistique

Résumé

Parlé à l'intérieur d'un domaine à cheval sur trois pays (France, Italie et Suisse), le francoprovençal est encore souvent méconnu. Cette langue a été mise en évidence par un linguiste, G. I. Ascoli (1874), à partir de ses évolutions phonétiques propres. Ayant échappé à toute tentative de standardisation, elle connaît une forte variation interne. Bien qu'ils aient aujourd'hui massivement régressé en France et en Suisse romande, les parlers francoprovençaux continuent d'être une réalité, toutefois fragile (Kristol 2016).

Cette communication, issue d'une thèse de doctorat en cours, s'intéresse à un aspect spécifique de la morphologie nominale du francoprovençal : l'expression du genre grammatical. En effet, le francoprovençal se distingue parmi les langues romanes en possédant deux marques de féminin au singulier et au masculin, une marque réalisée ou l'absence de marque. Dans les textes anciens du Dauphiné (13^e siècle), on observe ainsi au masc. sodo 'soudain' adj. en face de novel 'nouveau', et au fém. novella 'nouvelle' en face de freschi 'fraîche' (Devaux 1892 : 363). Ces particularités résultent d'évolutions du vocalisme posttonique qui font partie des traits qui définissent la langue (Tuailon 2007). Les -U et -E du latin se sont conservés après un groupe de consonnes lourd, comme en français et en occitan, mais en restant différenciés : -o et -e. Par ailleurs, -A latin final atone s'est conservé sous forme de -a, sauf en contexte palatal, où il a abouti à -i, et devant -S/T finaux où il a abouti à -e. De ce profil phonologique unique, qui distingue le francoprovençal des langues galloromanes voisines, résultent cinq morphèmes possibles pour marquer le genre grammatical et le nombre, c'est-à-dire pour réaliser l'accord. Ces marques permettent de répartir les formes adjectivales en quatre classes de flexion (Tableau 1). Un adjectif peut appartenir indépendamment à une des deux classes de masculin (1 ou 2) et à une des deux classes de féminin (A ou B). Soit par combinaison 4 classes : 1A, 1B, 21, 2B.

Genre	Classe	Singulier		Pluriel
Masculin	1	-o	→	-o
	2	∅	→	∅
Féminin	A	-a	→	-e
	B	-i	→	

Tableau 1 : Marques d'accord de l'adjectif francoprovençal

Il n'y a bien que deux genres : masculin et féminin, mais deux façons de marquer chacun de ces genres pour les mots variables que sont les adjectifs. L'identification de désinences de genre dans les adjectifs permet en retour de repérer dans les noms des finales qui les spécifient ou les prédisposent pour le genre. Mais c'est l'accord et la variabilité qui font le genre, d'où la priorité que nous accordons aux adjectifs dans notre approche.

On observe encore le système à quatre classes (« canonique », défini d'après formes médiévales) à l'époque moderne. Ainsi à Saint-Jean-de-Bournay (Isère, point 921 de l'ALF), on a :

- (1) Classe 1A : [vɛ], ['vɛrta] 'vert, verte'
- (2) Classe 1B : [blā], ['blā θi] 'blanc, blanche'
- (3) Classe 2A : ['bravo], ['brava] 'joli, jolie'
- (4) Classe 2B : ['ru:ðo], ['ru:ði] 'rouge' (m., f.)

Ailleurs, le modèle canonique connaît des évolutions, comme à Bernex (ALF 936) où on a les formes: [vɛ:r], ['vɛrɔa] 'vert, verte' (classe 1A) mais [blā], ['blā θ] 'blanc, blanche' (classe 1B) et une tendance à voir disparaître les masculins en -o, [ðo:n], [ðo:na] 'jaune (m., f.)'. Nous explorerons ces évolutions dans les matériaux de l'ALF, qui constitue notre premier corpus, en les complétant si possible par d'autres relevés, comme les *Tableaux phonétiques des patois suisses romands* (1925), ou des atlas régionaux plus récents, comme l'ALJA (1971-1978). Une asymétrie nous apparaît d'ores et déjà claire parmi les marques de genre du francoprovençal moderne: en face de deux marques fortes, -a et -o, très largement conservées et productives, les autres marques évoluent fréquemment vers d'autres timbres, [e] ou [ə], ou un statut de voyelle abstraite comme le schwa ou « e muet » français engendrant ainsi un tableau complexe que nous travaillons à décrire.

Keywords: dialectology; Francoprovençal; morphology; grammatical gender; linguistic atlas

Abstract

Spoken within a territory spanning three countries (France, Italy, and Switzerland), Francoprovençal remains largely unfamiliar to many. This language was first identified by the linguist G. I. Ascoli (1874) on the basis of its distinctive phonological developments. Having escaped all attempts at standardization, it exhibits considerable internal variation. Although Francoprovençal dialects have massively receded today in France and in French-speaking Switzerland, they nonetheless continue to exist, albeit in a fragile state (Kristol 2016).

This paper, drawn from ongoing doctoral research, focuses on a specific aspect of nominal morphology in Francoprovençal: the expression of grammatical gender. Francoprovençal stands out among the Romance languages in possessing two possible feminine markers in the singular and, in the masculine, either a realized marker or the absence of a marker. In medieval texts from Dauphiné (13th century), for instance, one finds masc. *sodo* 'sudden' (adj.) alongside *novel* 'new', and fem. *novella* 'new' alongside *freschi* 'fresh' (Devaux 1892: 363). These features result from the evolution of post-tonic vowels, which form part of the defining traits of the language (Tuailon 2007). Latin -U and -E were preserved after heavy consonant clusters, as in French and Occitan, though they remained distinct: -o and -e. Likewise, final unstressed Latin -A was retained as -a, except in palatal context, where it developed into -i, and before final -S/T, where it developed into -e. From this unique phonological profile – distinguishing Francoprovençal from its neighboring Gallo-Romance languages – emerge five possible morphemes to mark grammatical gender and number, i.e. to realize agreement. These markers distribute adjectival forms into four inflectional classes (Table 1). An adjective may independently belong to one of the two masculine classes (1

or 2) and to one of the two feminine classes (A or B). By combination, this yields four classes: 1A, 1B, 2A, 2B.

Gender	Class	Singular		Plural
Masculine	1	-o	→	-o
Masculine	2	∅	→	∅
Feminine	A	-a	→	-e
Feminine	B	-i	→	

Table 1: Agreement markers of the Franco-Provençal adjective

There are indeed only two genders, masculine and feminine, but two ways of marking each of these genders for variable words such as adjectives. The identification of gender endings in adjectives in turn makes it possible to detect in nouns certain final vowels that specify or predispose them for gender. Yet it is agreement and variability that constitute gender, which is why we place priority on adjectives in our approach.

The four-class system (“canonical,” defined on the basis of medieval forms) can still be observed in the modern period. Thus, in Saint-Jean-de-Bournay (Isère, point 921 of the *ALF*), we find:

- (1) Class 1A : [vɛ], [ˈvɛrta] ‘green’ (m., f.)
- (2) Class 1B : [blã], [ˈblã θi] ‘white’ (m., f.)
- (3) Class 2A : [ˈbravo], [ˈbrava] ‘pretty’ (m., f.)
- (4) Class 2B : [ˈru:ðo], [ˈru:ði] ‘red’ (m., f.)

Elsewhere, the canonical model undergoes changes, as in Bernex (*ALF* 936), where we find the forms [vɛ:r], [ˈvɛrɔ] ‘green (m., f.)’ (Class 1A) but [bλã], [ˈbλãθ] ‘white (m., f.)’ (Class 1B), along with a tendency for masculine forms in -o to disappear, e.g. [ðo:n], [ˈðo:na] ‘yellow (m., f.)’. We will explore these developments in the materials of the *ALF*, which constitutes our primary corpus, complementing them where possible with other sources, such as the *Tableaux phonétiques des patois suisses romands* (1925), or more recent regional atlases such as the *ALJA* (1971–1978). An asymmetry already emerges clearly among the gender markers of modern Francoprovençal: alongside the two strong markers, -a and -o, which are widely preserved and productive, the other markers frequently evolve toward different qualities, [e] or [ə], or toward the status of an abstract vowel such as the French schwa or *e muet*, thereby generating a complex system that we are working to describe.

Références

- ALF = Gillieron, J. & Edmont, E. (1902-1910). *Atlas Linguistique de la France*. Paris: Champion.
- ALJA = Martin, J.-B. & Tuaillon, G. (1971-1978). *Atlaslinguistique et ethnolinguistique du Jura et des Alpes du Nord*, éd. du CNRS.
- Ascoli, G. I. (1874). Schizzi franco-provenzali. *Archivio Glottologico Italiano*, 3, 61-120.
- Corbett, G. G. (1991). *Gender*. Cambridge University Press.
- Devaux, A. (1892). *Essai sur la langue vulgaire du Dauphiné septentrional au Moyen Âge*, Lyon/Paris: A. Cote/H. Welter.
- Hasselrot, B. (1974). Adieu au francoprovençal. *Revue de linguistique romane*, 38, 265-275.
- Kristol, A. (2013). Le francoprovençal, laboratoire des virtualités linguistiques de la Romania occidentale: le système bicasuel des parlers valaisan. In: Casanova, E. & Calvo, C. (eds), *Actes del 26é Congrès internacional de lingüística i filologia romàniques*, València, 6-11 septembre 2010, vol. 1, Berlin, etc. : De Gruyter.
- Kristol, A. (2014). Les grammaires du francoprovençal : l'expression de la partitivité. Quelques leçons du projet ALAVAL. In : *La Géolinguistique dans les Alpes au XXe siècle : méthodes, défis et perspectives. Actes de la Conférence annuelle sur l'activité scientifique du Centre d'études francoprovençales « René Willien »*. Saint Nicolas: BREL.
- Kristol A. (2016) .Francoprovençal. In Ledgeway, A. & Maiden, M. (eds), *The Oxford Guide to the Romance Languages*. Oxford/New York: Oxford University Press, 350-362.
- Loporcaro, M. (2018). *Gender from Latin to Romance*. Oxford: Oxford University Press.
- Gauchat, L., Jeanjaquet, J. & Tappolet, E. (1925). *Tableaux phonétiques des patois suisses romands: relevés comparatifs d'environ 500 mots dans 62 patois-types*. Neuchâtel: P. Attinger.
- Rivoira, M. & Régis, R. (2023). *Dialetti d'Italia : Piemonte e Valle d'Aosta*. Roma : Carocci editore.
- Tuaillon, G. (2007). *Le francoprovençal*. Quart (Vallée d'Aoste): Musumeci.

Haykanush Sazhumyan & Alice Blumenthal-Dramé (Albert Ludwig University of Freiburg)

The effects of aspect on meaning interpretation in the online and offline processing of modal constructions

Keywords: modality; aspect; construction grammar; self-paced reading; online and offline experimentation

Abstract

This psycholinguistic study investigates the effects of aspect (i.e., the grammatical distinction between progressive and simple) on the interpretation of English modals (epistemic vs. deontic) to gain insights into the mental representation of modal constructions, specifically their schematicity degree and node structure. Research (e.g., Wright 1995; Frajzyngier et al. 2008; Anthonissen et al. 2016) suggests that changing the grammatical structure of a verb phrase from present simple to progressive (*The choir must sing/be singing live*) makes an epistemic interpretation of the modal more likely than the deontic one. While a great amount of research has been published on modality (Depraetere et al. 2023: 1), it remains understudied within CxG (Hilpert and Flach 2023: 254).

First, the study aims to provide confirmatory evidence that, contrary to the simple aspect, the progressive adds a subjective (i.e., epistemic) meaning to modals. Second, it investigates whether the strength of this effect varies across different modals. This provides insights into whether the epistemic meaning of modals is processed exclusively via a highly abstract constructional schema subsuming all modals (i.e., the higher-level construction [SBJ V_{mod} *be* V_{ing}]_{epistemic}) or whether language users rely on additional modal-specific subschemas (e.g., the mid-level construction [SBJ *must be* V_{ing}]_{epistemic}).

To measure the entrenchment of certain necessity modals (*must*, *should*, and *have to*) with certain grammatical constructions, our study measured word-by-word reaction times in self-paced reading (online processing) combined with decision times and forced choices for follow-up judgment questions (offline processing). Online tasks can measure how the mental representation for a stimulus unfolds incrementally, tapping into more spontaneous and less controlled cognitive processes, whereas offline methods reveal the final interpretation, also taking into account metalinguistic, more strategic judgements. The experiment was presented via PClbex farm (Zehr and Schwarz 2018) with 87 participants recruited via Prolific (www.prolific.com).

The results show that, while the progressive is generally more inclined towards epistemic modality than the simple aspect in the offline task (see Figure 1), the strength of this effect varies across different modals in online processing (see Figure 2). From a constructionist perspective, this suggests that some modals have developed mid-level constructional schemas for certain grammatical forms (e.g., [SBJ *must be* V_{ing}]_{epistemic}), while others are processed via higher-level schemas (e.g., [SBJ V_{mod} *be* V_{ing}]_{epistemic}). This demonstrates that aspectual context is a crucial meaning-defining component within constructional networks that contributes to the formation of mid-level constructions for certain modals.

Fig. 1: Marginal effects from the mixed-effects model predicting the probability of epistemic interpretation based on structure in the judgment questions.

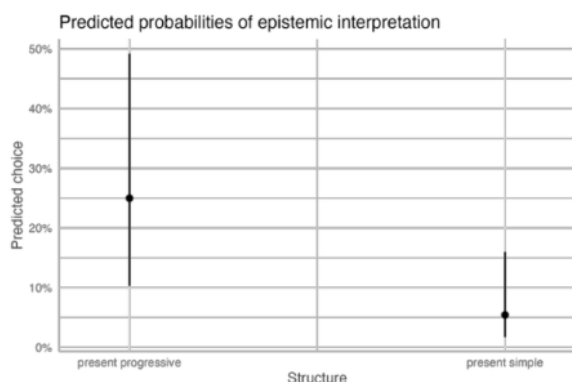
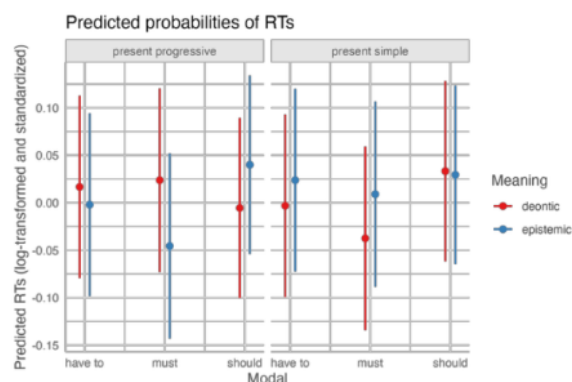


Fig. 2: Marginal effects from the mixed-effects model for the log-transformed and standardized RTs to the critical regions across all conditions.



Mots-clés : modalité ; aspect ; grammaire constructionnelle ; lecture auto-rythmée ; expérimentation online et offline

Résumé

Cette étude psycholinguistique examine les effets de l'aspect (c'est-à-dire la distinction grammaticale entre progressif et simple) sur l'interprétation des modaux anglais (épistémique vs. déontique) afin d'obtenir des informations sur la représentation mentale des constructions modales, spécifiquement leur degré de schématicité et leur structure nodale. Les recherches (par exemple, Wright 1995 ; Frajzyngier et al. 2008 ; Anthonissen et al. 2016) suggèrent que modifier la structure grammaticale d'une expression verbale du présent simple au progressif (*The choir must sing/be singing live*) rend une interprétation épistémique du modal plus probable que l'interprétation déontique. Bien qu'une quantité importante de recherches ait été publiée sur la modalité (Depraetere et al. 2023 : 1), elle reste sous-étudiée dans la CxG (Hilpert and Flach 2023 : 254).

Premièrement, l'étude vise à fournir des preuves confirmatoires que, contrairement à l'aspect simple, le progressif ajoute un sens subjectif (c'est-à-dire épistémique) aux modaux. Deuxièmement, elle examine si la force de cet effet varie selon différents modaux. Cela permet d'obtenir des informations sur la question de savoir si le sens épistémique des modaux est traité exclusivement via un schéma constructionnel hautement abstrait subsumant tous les modaux (c'est-à-dire la construction de niveau supérieur [SBJ V_{mod} be V_{ing}]_{épistémique}) ou si les utilisateurs de la langue s'appuient sur des schémas spécifiques aux modaux supplémentaires (par exemple, la construction de niveau intermédiaire [SBJ must be V_{ing}]_{épistémique}).

Pour mesurer l'enracinement de certains modaux de nécessité (must, should, et have to) avec certaines constructions grammaticales, notre étude a mesuré les temps de réaction mot par mot en lecture auto-rythmée (traitement en ligne) combinés aux temps de décision et aux choix forcés pour les questions de jugement de suivi (traitement offline). Les tâches online peuvent mesurer comment la représentation mentale d'un stimulus se déploie de manière incrémentale, exploitant des processus cognitifs plus spontanés et moins contrôlés, tandis que les méthodes offline

révèlent l'interprétation finale, tenant également compte des jugements métalinguistiques plus stratégiques. L'expérience a été présentée via PCIBex farm (Zehr and Schwarz 2018) avec 87 participants recrutés via Prolific (www.prolific.com).

Les résultats montrent que, bien que le progressif soit généralement plus enclin vers la modalité épistémique que l'aspect simple dans la tâche offline (voir Figure 1), la force de cet effet varie selon différents modaux dans le traitement online (voir Figure 2). D'un point de vue constructionniste, cela suggère que certains modaux ont développé des schémas constructionnels de niveau intermédiaire pour certaines formes grammaticales (par exemple, [SBJ must be V_{ing}]_{épistémique}), tandis que d'autres sont traités via des schémas de niveau supérieur (par exemple, [SBJ V_{mod} be V_{ing}]_{épistémique}). Cela démontre que le contexte aspectuel est un composant crucial de définition du sens au sein des réseaux constructionnels qui contribue à la formation de constructions de niveau intermédiaire pour certains modaux.

Fig. 1: Marginal effects from the mixed-effects model predicting the probability of epistemic interpretation based on structure in the judgment questions.

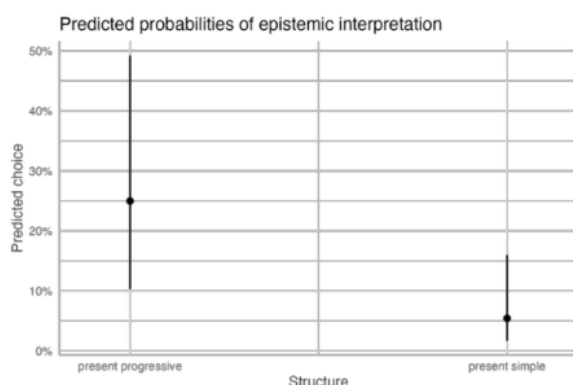
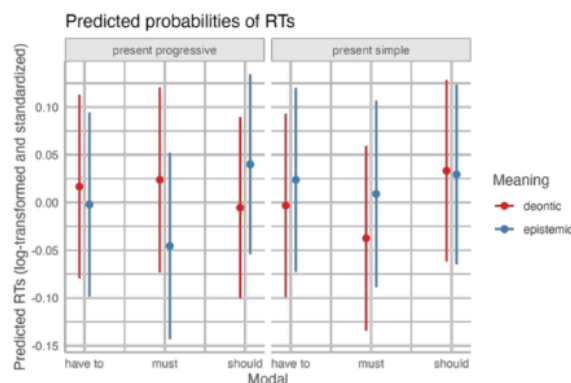


Fig. 2: Marginal effects from the mixed-effects model for the log-transformed and standardized RTs to the critical regions across all conditions.



Références

Anthonissen, L., de Wit A. & Mortelmans, T. (2016). Aspect meets modality: A semantic analysis of the German *am*-progressive. *Journal of Germanic Linguistics*, 28(1), 1-30. <https://doi.org/10.1017/S1470542715000185>

Depraetere, I., Cappelle B. & Hilpert, M. (2023). Introduction. In: Depraetere, I., Cappelle, B., Hilpert, M., De Cuypere, L., Dehouck, M., Denis, P., Flach, S., Grabar, N., Grandin, C., Hamon, T., Hufeld, C., Leclercq B. & Schmid, H. (eds.), *Models of Modals: From Pragmatics and Corpus Linguistics to Machine Learning*. Berlin/Boston: De Gruyter Mouton (Topics in English Linguistics), pp. 1–13.

Frajzyngier, Z., Bond, M., Heintzelman, L., Keller, D., Ogihara, S. & Shay, E. (2008). Towards an understanding of the progressive form in English: The imperative as a heuristic tool. In: Abraham W. & Leiss, E. (eds.), *Modality-aspect interfaces: Implications and typological solutions*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 81-96.

Grabar, N., Grandin, C., Hamon, T., Hufeld, C., Leclercq B. & Schmid, H. (eds.), *Models of Modals: From Pragmatics and Corpus Linguistics to Machine Learning*. Berlin/Boston: De Gruyter Mouton (Topics in English Linguistics), pp. 254-269.

Hilpert, M. & Flach, S. (2023). Modals in the network model of Construction Grammar. In: Depraetere, I., Cappelle, B., Hilpert, M., De Cuypere, L., Dehouck, M., Denis, P., Flach, S.,

Wright, S. (1995). Subjectivity and experiential syntax. In: Stein D. & Wright S. (eds.), *Subjectivity and subjectification: Linguistic perspectives*. Cambridge: Cambridge University Press, pp. 151-172.

Zehr, J. & Schwarz, F. (2018). *PennController for Internet Based Experiment (IBEX)*. <https://doi.org/10.17605/OSF.IO/MD832> (last access: 10 March 2025).

Lien vers l'étude : <https://farm.pcibex.net/p/purfna/>

Nadia Scanavino (Università di Torino) & Guglielmo Inglese (Università di Torino)

Ditransitive verbs in the history of Greek. A typological and diachronic approach

Keywords: Ancient Greek; Modern Greek; ditransitive construction; logistic regression

Abstract

Ditransitives prototypically express transfer events involving three participants: an Agent (A), a Theme (T), and a Recipient (R). Ditransitives display a cross-linguistic tendency to alternate between more than one argument structure construction, specifically concerning the coding of T and R. Typologists have identified different alignment patterns based on how T and R roles are encoded with respect to the Patient (P) of transitive verbs (Haspelmath 2005; Malchukov et al. 2010), whereas language-specific studies, have tried to pinpoint possible (extra-)linguistic factors responsible for the variable coding of T and R (e.g., Zehentner 2019; De Vaere 2023). Ditransitives have also increasingly been studied from a diachronic perspective (Napoli & Fedriani 2020; Zehentner et al. 2023).

The growing interest in ditransitives has also concerned (ancient) Indo-European languages (e.g. Napoli 2018, Fedriani 2020 on Latin and Dahl 2020 on Vedic), with the notable and unfortunate exception of Greek. Studies on Ancient Greek have focused either on specific constructions (Jaquinod 1989, Luraghi & Zanchi 2018) or individual verbs (e.g. Benedetti & Bruno 2022), whereas the literature on Modern Greek mainly addresses the topic from a formal perspective (see Bouba 1998; Kordoni 2004; Georgala et al. 2012; Anagnostopoulou & Sevdali 2016). This means that we still lack a comprehensive account of the behavior of ditransitive constructions in Greek. To fill this gap, we undertake a systematic exploration of ditransitive constructions in Homeric (HG), Classical (CG) and Modern Greek (MG).

We focus on 27 ditransitive verbs belonging to three semantic subclasses: verbs of physical transfer, verbs of mental transfer, and verbs of dispossession. For each verb, we extracted a random sample of up to 100 occurrences from three corpora: HG data was retrieved through *HoDeL*, whereas for CG we explored a sub-corpus of prose texts using the TLG. MG data was extracted from the *Greek Web Corpus*. Each token was annotated based on several factors, including verbal voice, animacy and coding of T and R, their weight, position, and their realization (full NPs, pronouns, zero).

Based on this data, our goal is two-fold. First, we offer a comprehensive description of the attested ditransitive alignment patterns at different stages of Greek, highlighting the main patterns of variation and continuity and the language-specific historical accident (e.g. case syncretism) that shape the diachrony of ditransitive alignment in Greek. Our results shows that CG features a richer inventory of alignment patterns as compared to MG, which is largely due to the spread of indirective alignment to verbs that originally featured a different pattern, e.g. *dōréō* ‘donate’ featuring secundative alignment in CG.

Second, we focus on the variation in the coding of R. Already in HG, R can be expressed either by the bare dative case or with a prepositional phrase. The prepositional strategy gains ground over time, becoming widespread in MG. Preliminary results from regression models show that in HG/CG animacy plays a key role, with animate R favoring the dative and animate T the

prepositional pattern, respectively. We plan to extend the analysis to MG, so as to gain a better understanding of the factors that shape the diachrony of the coding of R over time.

Mots-clés: Grec ancien ; grec moderne ; construction ditransitive ; régression logistique

Résumé

Les verbes ditransitifs expriment généralement des événements de transfert impliquant trois participants : un agent (A), un thème (T) et un récepteur (R). Les verbes ditransitifs présentent une tendance interlinguistique à alterner entre plusieurs structures argumentatives, notamment en ce qui concerne le codage de T et R. Les typologues ont identifié différents modèles d'alignement en fonction de la manière dont les rôles T et R sont codés par rapport au patient (P) des verbes transitifs (Haspelmath 2005 ; Malchukov et al. 2010), tandis que des études spécifiques à certaines langues ont tenté d'identifier les facteurs (extra-)linguistiques possibles responsables du codage variable de T et R (par exemple, Zehentner 2019 ; De Vaere 2023). Les verbes ditransitifs ont également fait l'objet d'études de plus en plus nombreuses dans une perspective diachronique (Napoli & Fedriani 2020 ; Zehentner et al. 2023).

L'intérêt croissant pour les verbes ditransitifs a également concerné les langues indo-européennes (anciennes) (par exemple Napoli 2018 et Fedriani 2020 sur le latin, Dahl 2020 sur le védique), à l'exception notable et regrettable du grec. Les études sur le grec ancien se sont concentrées soit sur des constructions spécifiques (Jaquinod 1989, Luragu & Zanchi 2018) ou sur des verbes individuels (Benedetti & Buno 2022), tandis que la littérature sur le grec moderne aborde principalement le sujet d'un point de vue formel (voir Bouba 1998 ; Kordoni 2004 ; Georgala et al. 2012 ; Anagnostopolou & Sevdali 2016). Cela signifie que nous ne disposons toujours pas d'une description complète du comportement des constructions ditransitives en grec. Pour combler cette lacune, nous entreprenons une exploration systématique des constructions ditransitives en grec homérique (HG), classique (CG) et moderne (MG).

Nous nous concentrons sur 27 verbes ditransitifs appartenant à trois sous-classes sémantiques : les verbes de transfert physique, les verbes de transfert mental et les verbes de dépossession. Pour chaque verbe, nous avons extrait un échantillon aléatoire de 100 occurrences maximum à partir de trois corpus : les données HG ont été récupérées via HoDeL, tandis que pour CG, nous avons exploré un sous-corpus de textes en prose à l'aide du TLG. Les données MG ont été extraites du *Greek Web Corpus*. Chaque token a été annoté en fonction de plusieurs facteurs, notamment la voix verbale, l'animéité et le codage de T et R, leur poids, leur position et leur réalisation (NP complets, pronoms, zéro).

Sur la base de ces données, notre objectif est double. Premièrement, nous proposons une description complète des modèles d'alignement ditransitifs attestés à différentes étapes du grec, en mettant en évidence les principaux modèles de variation et de continuité ainsi que les accidents historiques spécifiques à la langue (par exemple, le syncrétisme des cas) qui façonnent la diachronie de l'alignement ditransitif en grec. Nos résultats montrent que le grec classique présente un inventaire plus riche de modèles d'alignement que le grec moderne, ce qui s'explique en grande partie par la propagation de l'alignement indirectif aux verbes qui présentaient à l'origine un modèle différent, par exemple *dōréō* « donner » qui présente un alignement secondatif en grec classique.

Deuxièmement, nous nous concentrons sur la variation dans le codage de R. Déjà dans le HG, R peut être exprimé soit par le cas datif seul, soit par une locution prépositionnelle. La stratégie prépositionnelle gagne du terrain au fil du temps et se généralise dans le MG. Les résultats préliminaires des modèles de régression montrent que dans le HG/CG, l'animé joue un rôle clé, R animé favorisant le datif et T animé le modèle prépositionnel, respectivement. Nous prévoyons d'étendre l'analyse au MG, afin de mieux comprendre les facteurs qui façonnent la diachronie du codage de R au fil du temps.

Ressources

HoDeL: <https://su-lab.unipv.it/tasf/index.php/hodel/>

TLG: <https://stephanus.tlg.uci.edu/>

EITenTen19: <https://www.sketchengine.eu/eltenten-greek-corpus/>

Références

Anagnostopoulou, Elena & Sevali, Christina. 2016. Two modes of genitive and dative case assignment: Evidences from two stages of Greek. *Natural Language Linguist Theory* 38. 987-1051.

Benedetti Marina & Bruno Carla. 2022. Costrutti risultativi in greco antico? Spunti su διδάσκω. *Studi e saggi linguistici* 60: 9-27.

Bouba, Maria. 1998. Ditransitive Verbs in M. Greek. *Selected papers on theoretical and applied linguistics* 12 (1): 56-70.

Dahl, Eystein. 2020. The diachrony of ditransitives in Vedic Sanskrit. In *The Diachrony of Ditransitives*, Maria Napoli & Chiara Fedriani (eds.), 61-96. Berlin & New York: de Gruyter.

De Vaere, Hilde. 2023. *The Ditransitive Alternation in Present-Day German: A corpus-based analysis*. Amsterdam & Philadelphia: John Benjamins.

Fedriani, Chiara. 2020. The spread of ad construction in Merovingian Latin: identifying semantic paths in the domain of ditransitives. In *The Diachrony of Ditransitives*, Maria Napoli & Chiara Fedriani (eds.), 61-96. Berlin & New York: de Gruyter

Georgala Effi, Özsoy Sumru & Gürel Ayşe. 2012. Short object shift and ditransitive structure in Greek. *Current Issues in Mediterranean Syntax*. Amsterdam & Philadelphia: John Benjamins.

Haspelmath, Martin. 2005. Argument marking in ditransitive alignment types. *Linguistic Discovery* 3: 1–21.

Jacquiod, Bernard. 1989. *Le double accusatif en grec d'Homère à la fin du V siècle avant J.-C.* Louvain-la-Neuve: Peeters.

Kordoni, Valia. 2004. *Modern Greek ditransitives in LMT. Proceedings of the International Conference on Head-Driven Phrase Structure Grammar*. 151-167.

Luraghi, Silvia & Zanchi, Chiara. 2018. Double accusative constructions and ditransitives in Ancient Greek. In *Ditransitive constructions in a cross-linguistic perspective*, Agnes Korn & Andrej Malchukov (eds.), 13–35. Wiesbaden: Reichert.

Malchukov Andrej, Haspelmath Martin & Comrie Bernard (eds.). 2010. *Studies in Ditransitive Constructions: A Comparative Handbook*. Berlin & New York: de Gruyter.

Napoli, Maria. 2018. Ditransitive verbs in *Latin: A typological approach*. *Journal of Latin Linguistics* 17(1), 51-91.

Napoli, Maria. 2020. Old italian ditransitive constructions: between alternations and change. In *The Diachrony of Ditransitives*, Maria Napoli & Chiara Fedriani (eds.), 97–136. Berlin & New York: de Gruyter.

Napoli Maria & Fedriani Chiara. 2020. *The Diachrony of Ditransitives*. Berlin & New York: de Gruyter.

Zehentner, Eva. 2019. *Competition in Language Change: The Rise of the English Dative Alternation*. Berlin & New York: de Gruyter.

Zehentner Eva, Röthlisberger Melanie & Coleman Timothy (eds.). 2023. *Ditransitive constructions in Germanic languages*. Amsterdam & Philadelphia: John Benjamins.

Tobias Scheer (Université Côte d'Azur)

Rule crossing in the lexical variation period

Abstract

The purpose of this talk is to draw attention to a pattern of diachronic change, rule crossing in the lexical variation period, which on the face of it is a clear case of multiple outputs, i.e. where a given input produces more than one output through diachronic evolution. Under neogrammarian exceptionlessness, this should not occur. It turns out, however, that all sound laws involved apply without exception, the paradoxical result being multiple outputs obeying neogrammarian exceptionlessness.

We know from the Labovian study of ongoing change that a sound law is implemented over several generations of speakers: it needs time before the new pronunciation affects all words concerned in all instances of their production by all speakers. Consider the case where during this variation period two rules apply successively in bleeding order: rule 1 has only yet transformed a subset of words that it is competent for when rule 2, freshly innovated, destroys its context of application. Thus in the evolution from Latin to Old French, intervocalic voicing applies to the *t* of *cōg(i)tāre* > *cuidier* (mod. *cuidier*), but not to the *t* of *com(i)te* > *comte*. The multiple outputs at hand *t/d* are only found when the vowel preceding the voiceless intervocalic obstruent is subject to syncope (indicated by brackets). Application of syncope bleeds voicing since it removes the obstruent from its intervocalic position. What happened is thus that at a stage where syncope had already affected some words (*com(i)te* > °*comte*), but not yet others (*cōg(i)tāre*), intervocalic voicing was innovated and applied to the latter, but not to the former.

Note that both sound laws at hand were working without any exceptions: syncope was implemented by lexical diffusion, and ended up applying to all words. Looked at from hindsight, it is a clean neogrammarian change. Its diffusing character is only indirectly visible because it was crossed by intervocalic voicing, which applied to all items that were in its purview at the diachronic stage at hand.

Thus rule crossing during the lexical variation period makes evaporate a number of multiple output patterns that on the face of it challenge exceptionlessness. Exactly how much of the multiple outputs pool is concerned remains to be seen. The empirical material presented is drawn from French diachrony (Grande Grammaire Historique du Français GGHF), where rule crossing is massive. The analytic absurdities that the neogrammarians were driven into (and which fueled their critics) when trying to account for the multiple outputs of rule crossing patterns are also discussed. Finally, the talk reviews the question whether all change involves lexical diffusion (e.g. Phillips 2006, contra Labov 1994 and Bernúdez-Otero 2007), that is, whether all sound laws are eligible for the rule crossing pattern.

Résumé

L'objectif de cet exposé est d'attirer l'attention sur un modèle de changement diachronique, le franchissement de règles au cours de la période de variation lexicale, qui constitue a priori un cas évident de sorties multiples, c'est-à-dire lorsqu'une entrée donnée produit plusieurs sorties par évolution diachronique. En vertu de l'absence d'exception néogrammairienne, cela ne devrait pas se produire. Or, il s'avère que toutes les lois sonores concernées s'appliquent sans exception, ce qui, paradoxalement, conduit à des sorties multiples obéissant à l'absence d'exception néogrammairienne.

L'étude labovienne des changements en cours nous apprend qu'une loi phonétique s'applique sur plusieurs générations de locuteurs : il faut du temps avant que la nouvelle prononciation affecte tous les mots concernés, quelle que soit leur production par tous les locuteurs. Prenons le cas où, pendant cette période de variation, deux règles s'appliquent successivement, l'une après l'autre : la règle 1 n'a encore transformé qu'un sous-ensemble de mots pour lesquels elle est compétente, lorsque la règle 2, nouvellement innovée, détruit son contexte d'application. Ainsi, dans l'évolution du latin vers l'ancien français, le voisement intervocalique s'applique au *t* de *cōg(i)tāre* > *cuidier* (mod. *cuidier*), mais pas au *t* de *cōm(i)te* > *comte*. Les multiples sorties *t/d* en question ne se rencontrent que lorsque la voyelle précédant l'obstruante intervocalique sourde est sujette à une syncope (indiquée par des crochets). L'application d'une syncope altère le voisement puisqu'elle retire l'obstruante de sa position intervocalique. Ce qui s'est passé, c'est qu'à un stade où la syncope avait déjà affecté certains mots (*cōm(i)te* > °*cōmte*), mais pas encore d'autres (*cōg(i)tāre*), le voisement intervocalique a été innové et appliqué à ces derniers, mais pas aux premiers.

Il convient de noter que les deux lois phonétiques en question fonctionnaient sans exception : la syncope a été mise en œuvre par diffusion lexicale et s'est finalement appliquée à tous les mots. Rétrospectivement, il s'agit d'un changement néogrammairien évident. Son caractère diffusant n'est visible qu'indirectement, car elle a été traversée par un voisement intervocalique, qui s'est appliqué à tous les éléments relevant de son champ d'application au stade diachronique considéré.

Ainsi, le franchissement de règles pendant la période de variation lexicale fait disparaître un certain nombre de modèles de sorties multiples qui, à première vue, remettent en cause l'absence d'exception. Il reste à déterminer dans quelle mesure ce réservoir de sorties multiples est concerné. Le matériel empirique présenté est tiré de la diachronie française (Grande Grammaire Historique du Français GGHF), où le franchissement de règles est massif. Les absurdités analytiques auxquelles les néogrammairiens ont été contraints (et qui ont alimenté leurs critiques) en tentant de rendre compte des sorties multiples des modèles de franchissement de règles sont également abordées. Enfin, l'exposé examine la question de savoir si tout changement implique une diffusion lexicale (par exemple, Phillips 2006, contre Labov 1994 et Bernúdez-Otero 2007), c'est-à-dire si toutes les lois sonores sont éligibles au modèle de franchissement de règles.

Jian Shen Shih (National Chengchi University Graduate Institute of Linguistics)

Semantic Decoding Function of the Matrix in Closed Texts: Validating Riffaterre's Semiotic Decoding Framework for Poetic Language

Keywords: Matrix Interpretation, Poetic Language, Semiotic Decoding, Semantic Deviation

Abstract

This research, centered on Michael Riffaterre's poetics, investigates the semantic integrative and semiotic generative roles of the matrix within closed poetic texts. Riffaterre's theory asserts that poetic language consistently creates interpretative obstacles through semantic deviation and syntactic tension, forcing readers to retrospectively search for recurring, semantically cohesive words within the text to reconstruct deeper poetic structures. Consequently, poetry emerges as not just an artistic linguistic form but also as a closed, autonomous, circularly inter-referential semiotic system. Within this system, meaning is generated via processes such as repetition of codes, symbolic substitution, and structural counterpoints. Incorporating Tsai Hsiu-Chih's (2006) interpretation of hypograms and textual inter-referentiality, this study posits that the matrix functions not merely as a semantic core but also as a pivotal node for cultural semiotic memory and symbolic cognition. Thus, poetic language should not be simplified to sentence-level comprehension. Rather, it involves an ongoing, dynamic process of constructing and integrating symbolic relationships between readers and language.

To empirically verify the applicability and relevance of Riffaterre's theory in contemporary poetry reading, this study conducted a semantic cognition experiment involving participants from three groups with different educational backgrounds. It aimed to observe students' reading processes under the guidance of Riffaterre's framework, as well as their ability to identify matrices within poems. Additionally, the study examined how each group processed instances of semantic deviation in poetry. The findings revealed that, with appropriate instructional support, readers were able to effectively identify recurring elements and implicit structural frameworks within the text, thereby actively engaging in semantic integration and deepening their understanding of poetic meaning.

This study underscores the robust theoretical potential of Riffaterre's poetic model in linguistic analyses, particularly by providing clear analytical tools for understanding linguistic closure, structural coherence, and the intricate logic of meaning generation in poetry. Furthermore, this research emphasizes the significance of semiotic decoding within semantic studies, establishing that decoding poetic language is an intricate, complex yet highly structured semantic interpretive process. Ultimately, the insights derived from this investigation not only confirm Riffaterre's theoretical propositions but also illuminate broader implications for semiotic and poetic studies.

Mots-clés: interprétation matricielle, langage poétique, décodage sémiotique, déviation sémantique

Résumé

Cette recherche, centrée sur la poétique de Michael Riffaterre, examine les rôles sémantiques intégratifs et sémiotiques génératifs de la matrice dans les textes poétiques fermés. La théorie de Riffaterre affirme que le langage poétique crée systématiquement des obstacles interprétatifs par le biais de déviations sémantiques et de tensions syntaxiques, obligeant les lecteurs à rechercher rétrospectivement des mots récurrents et sémantiquement cohérents dans le texte afin de reconstruire des structures poétiques plus profondes. Par conséquent, la poésie apparaît non seulement comme une forme linguistique artistique, mais aussi comme un système sémiotique fermé, autonome et circulairement interréférentiel. Au sein de ce système, le sens est généré par des processus tels que la répétition de codes, la substitution symbolique et les contrepoints structurels.

En intégrant l'interprétation de Tsai Hsiu-Chih (2006) des hypogrammes et de l'interréférentialité textuelle, cette étude postule que la matrice fonctionne non seulement comme un noyau sémantique, mais aussi comme un nœud central pour la mémoire sémiotique culturelle et la cognition symbolique. Ainsi, le langage poétique ne doit pas être simplifié à une compréhension au niveau de la phrase. Il implique plutôt un processus continu et dynamique de construction et d'intégration de relations symboliques entre les lecteurs et le langage.

Afin de vérifier empiriquement l'applicabilité et la pertinence de la théorie de Riffaterre dans la lecture de la poésie contemporaine, cette étude a mené une expérience de cognition sémantique impliquant des participants issus de trois groupes ayant des parcours éducatifs différents. Elle visait à observer les processus de lecture des étudiants en mobilisant le cadre théorique de Riffaterre, ainsi que leur capacité à identifier les matrices dans les poèmes. En outre, l'étude a examiné la manière dont chaque groupe traitait les cas de déviation sémantique dans la poésie. Les résultats ont révélé qu'avec un soutien pédagogique approprié, les lecteurs étaient capables d'identifier efficacement les éléments récurrents et les cadres structurels implicites dans le texte, s'engageant ainsi activement dans l'intégration sémantique et approfondissant leur compréhension du sens poétique.

Cette étude souligne le solide potentiel théorique du modèle poétique de Riffaterre dans les analyses linguistiques, notamment en fournissant des outils analytiques clairs pour comprendre la fermeture linguistique, la cohérence structurelle et la logique complexe de la génération de sens dans la poésie. En outre, cette recherche souligne l'importance du décodage sémiotique dans les études sémantiques, établissant que le décodage du langage poétique est un processus d'interprétation sémantique complexe, mais hautement structuré. En fin de compte, les conclusions tirées de cette enquête confirment non seulement les propositions théoriques de Riffaterre, mais mettent également en lumière des implications plus larges pour les études sémiotiques et poétiques.

Références

- Saussure, F. de. (1916). *Course in general linguistics* (C. Bally & A. Sechehaye, Eds.; W. Baskin, Trans.). McGraw-Hill.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643–662.
- Barthes, R. (1972). *Mythologies* (A. Lavers, Trans.). Hill and Wang. (Original work published 1957)
- Riffaterre, M. (1978). *Semiotics of poetry*. Indiana University Press.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Derrida, J. (1978). *Writing and difference* (A. Bass, Trans.). University of Chicago Press. (Original work published 1967)
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge.
- Tsai, S.-C. (2006). 利法代《詩的符號學》中的潛藏符譜與文本的相互指涉性. 文山評論: 文學與文化, 2(1), 85–117.
- Fauconnier, G., & Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities*. Basic Books.
- Lai, K.-J. (2005). 符號傳達的基礎之探討：巴特《神話學》與符號學之研究 [Unpublished master's thesis]. Tunghai University.
- Chen, Y.-J. (2015). 運用認知語言學為高中學生搭建閱讀詩的認知階梯 [Unpublished master's thesis]. National Taiwan Normal University.
- Patterson, K. (2015). *Lexical priming and metaphor* [Doctoral dissertation, Liverpool John Moores University].
- Kharkhurin, A. V. (2016). Cognitive poetry: Theoretical framework for the application of cognitive psychology techniques to poetic text. *Culture and Text*, 6(1), 38–55.
- Chiang, C.-C. (2020). 兒童故事中的隱喻框架和概念整合：哲學與認知的關係 [Unpublished master's thesis]. National Chung Cheng University.

Abstract

Narrative discourse can be viewed as organized in episodes, i.e. semantic units, larger than a clause, exhibiting thematic concord in the set of identical participants, time, location, or prevailing event or action (Longacre 1979, Chafe 1980, van Dijk 1981). Segmentation of the narrative into episodes is often marked by linguistic means such as discourse markers, spatial and temporal adverbs, or explicit meta-comments, which signal episode boundaries corresponding to a change in time, location, participants, action sequence, or perspective (Brinton 1993: 74–75).

This study aims to identify grammatical categories and phenomena that are sensitive to episodic structure in narratives. We use annotated corpora of three unrelated languages with different typological profiles: Chuvash (Turkic, Central Russia), Selkup (Uralic, Western Siberia), and Wan (Mande, Côte d'Ivoire). Our choice of languages is determined by the availability of narrative corpora of sufficient size as well as by considerations of structural diversity. Chuvash and Selkup are both agglutinative languages with flexible word order and a tendency towards SOV, while Wan is a highly isolating language with a rigid SOVX word order. Thus, we include two languages with more structural similarity (Chuvash and Selkup) as well as a structurally very different language (Wan) to achieve a broader comparison.

Our annotation is based on a combination of criteria for narrative analysis from Labov (1972) and criteria for thematic unity from Dijk (1981). It is aimed at identifying narrative event sequences, and at distinguishing sequence-initial clauses from clauses that appear sequence-internally. The annotation is based on the interpretation of clauses in discourse and is independent of their particular form or the choice of linking devices.

To establish correlations between the position of the clause in the narrative sequence and its structural features, we examine the frequency of use of the major grammatical markers and discourse particles in sequence-initial and sequence-internal clauses, looking specifically for asymmetries in the distributions that would suggest that a particular marker is sensitive to episodic structure. This method allows us to identify markers that signal segment boundaries as well as markers that signal continuity within segments (if any such markers exist in the language).

We discuss the results of this systematic comparison for the phenomena commonly associated with narrative-structuring functions in previous studies. The common suspects include temporal conjunctions and discourse particles (Bakker 1993), lexical expressions such as 'it happened that' (Brinton 1993), "backgrounding" devices such as nominalization (Mithun 2002), the choice of referential expressions (pronouns vs. nouns, Fox 1987), and verbal categories such as tense and aspect (Schiffrin 1981).

Our preliminary findings suggest that the phenomena associated with the marking of episodic boundaries vary widely across the three languages, ranging from the use of special linking devices such as conjunctions to shifts in the use of tense-aspect forms and specific patterns of pronominal use. We propose possible sources of the cross-linguistic differences in the marking of episodic boundaries and discuss general implications of our findings for narrative typology and the study of discourse.

Résumé

Le discours narratif peut être considéré comme organisé en épisodes, c'est-à-dire en unités sémantiques plus grandes qu'une clause, présentant une concordance thématique dans l'ensemble des participants, du temps, du lieu ou l'événement ou de l'action prédominant (Longacre 1979, Chafe 1980, van Dijk 1981). La segmentation du récit en épisodes est souvent marquée par des moyens linguistiques tels que des marqueurs discursifs, des adverbes spatiaux et temporels, ou des méta-commentaires explicites, qui signalent les limites des épisodes correspondant à un changement de temps, de lieu, de participants, de séquence d'actions ou de perspective (Brinton 1993 : 74-75).

Cette étude vise à identifier les catégories grammaticales et les phénomènes sensibles à la structure épisodique dans les récits. Nous utilisons des corpus annotés de trois langues non apparentées présentant des profils typologiques différents : le tchouvache (turc, Russie centrale), le selkoupe (ouralique, Sibérie occidentale) et le wan (mande, Côte d'Ivoire). Notre choix de langues est déterminé par la disponibilité de corpus narratifs de taille suffisante ainsi que par des considérations de diversité structurelle. Le tchouvache et le selkoupe sont des langues agglutinantes avec un ordre des mots flexible et une tendance vers le SOV, tandis que le wan est une langue très isolante avec un ordre des mots SOVX rigide. Ainsi, nous incluons deux langues présentant une plus grande similitude structurelle (le tchouvache et le selkoupe) et une langue très différente sur le plan structurel (le wan) afin d'obtenir une comparaison plus large.

Notre annotation repose sur une combinaison des critères d'analyse narrative de Labov (1972) et des critères d'unité thématique de Dijk (1981). Elle vise à identifier les séquences d'événements narratifs et à distinguer les clauses initiales de la séquence des clauses qui apparaissent à l'intérieur de la séquence. L'annotation est basée sur l'interprétation des clauses dans le discours et est indépendante de leur forme particulière ou du choix des dispositifs de liaison.

Pour établir des corrélations entre la position de la clause dans la séquence narrative et ses caractéristiques structurelles, nous examinons la fréquence d'utilisation des principaux marqueurs grammaticaux et particules discursives dans les clauses initiales et internes à la séquence, en recherchant spécifiquement les asymétries dans les distributions qui suggéreraient qu'un marqueur particulier est sensible à la structure épisodique. Cette méthode nous permet d'identifier les marqueurs qui signalent les limites des segments ainsi que ceux qui signalent la continuité au sein des segments (si de tels marqueurs existent dans la langue).

Nous discutons les résultats de cette comparaison systématique pour les phénomènes généralement associés aux fonctions de structuration narrative dans les études précédentes. Les suspects habituels comprennent les conjonctions temporelles et les particules discursives (Bakker 1993), les expressions lexicales telles que « il arriva que » (Brinton 1993), les dispositifs de « mise en arrière-plan » tels que la nominalisation (Mithun 2002), le choix des expressions référentielles (pronoms vs noms, Fox 1987) et les catégories verbales telles que le temps et l'aspect (Schiffrin 1981).

Nos conclusions préliminaires suggèrent que les phénomènes associés au marquage des limites épisodiques varient considérablement entre les trois langues, allant de l'utilisation de dispositifs de liaison spéciaux, tels que les conjonctions, à des changements dans l'utilisation des formes de temps et d'aspect et à des schémas spécifiques d'utilisation des pronoms. Nous proposons

des sources possibles des différences interlinguistiques dans le marquage des limites épisodiques et discutons des implications générales de nos conclusions pour la typologie narrative et l'étude du discours.

Références

- Bakker, Egbert J. 1993. Boundaries, topics, and the structure of discourse: An investigation of the Ancient Greek particle *dé*. *Studies in Language* 17(2): 275-311.
- Brinton, Laurel J. 1993. Episode boundary markers in Old English discourse. R. J. Jeffers & H. Aertsen (eds.) *Historical Linguistics 1989: Papers from the 9th International Conference on Historical Linguistics, Rutgers University, 14-18 August 1989*. Amsterdam: John Benjamins, pp. 73-89.
- Chafe, Wallace. 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.
- van Dijk, Teun A. 1981. Episodes as units of Discourse Analysis. Deborah Tannen (ed.) *Analyzing Discourse: Text and Talk*. Georgetown: Georgetown University Press, pp. 177–195.
- Fox, Barbara. 1987. Anaphora in popular written English narratives. Russell S. Tomlin (ed.) *Coherence and Grounding in Discourse: Outcome of a Symposium, Eugene, Oregon, June 1984*. Amsterdam: John Benjamins, pp. 157-174.
- Labov, William. 1972. *Language in the Inner City*. Philadelphia: University of Pennsylvania Press.
- Longacre, Robert E. 1979. The paragraph as a grammatical unit. Talmy Givón (ed.) *Syntax and semantics* 12. *Discourse and syntax*. New York: Academic Press, 115–134.
- Mithun, Marianne. 2002. Rhetorical nominalization in Barbareño Chumash. *Survey Reports, Survey of California and Other Indian Languages* 12: 55–63.
- Schiffrin, Deborah. 1981. Tense variation in narrative. *Language* 57(1): 45–62.

Alina Tsikulina & Eva Soroli (University of Lille, STL)

L2 Reading comprehension and Eye Movements:

Tough constructions' decoding by bilingual and trilingual speakers

Keywords: second language acquisition, L3 influence, eye-tracking, reading strategies, fixation measures, evaluative constructions, cross-linguistic transfers, French, English, Russian

Abstract

Reading in a first language (L1) is typically an automatised and efficient process. In contrast, second language (L2) reading, especially when involving syntactically complex or cross-linguistically asymmetrical constructions, tends to be slower and cognitively more demanding (Parshina, et al., 2021; Whitford & Joanisse, 2018; Cheng et al., 2021; Patterson et al., 2014; Hawkins, 2004). One such structure is the *Tough Construction* (TC), as in “*This pet is easy to wash*”, which requires readers to interpret the grammatical subject (“*this pet*”) as the inferred object of an embedded non-finite verb (“*to wash*”) (Chung, 2001; Guérin, 2006). Although TCs are available in English and French, they are absent in languages such as Russian, which offer functional analogues instead (Comrie & Matthews, 1990; Paykin & Van Peteghem, 2020; Tsikulina et al., 2025). Previous research suggests that L1 syntactic structures significantly influence L2 processing, particularly when constructions diverge across languages (Booth et al., 2020; De Angelis et al., 2015). Similarly, from a usage-based perspective (Höder, 2019), experience with a second or third language (L3) can lead to competing, hybridized, or overlapping constructions, which may in some cases provide compensatory processing support – particularly when the languages involved are typologically close (Hui, 2010; Cheung et al., 2011).

This study explores whether bilingual and trilingual language knowledge facilitates the processing of TCs. Specifically, it investigates: (a) how advanced French-English (FR-EN) and Russian-English (RU-EN) late bilinguals process English TCs; (b) whether Russian-English-French (RU-EN-FR) trilinguals benefit from their L3-knowledge in decoding L2 constructions; and (c) how these multilingual groups compare to English monolinguals (EN MNs). The study focuses on the role of language experience and typological proximity in syntactic and semantic processing, as well as in potential transfer effects.

Thirty-six late bilinguals (FR-EN, RU-EN), fourteen RU-EN-FR trilinguals, and seven EN MNs completed a Sentence-Picture Matching Task with parallel eye-tracking recording. Participants read English TC sentences (e.g., “*This cat is easy to clean*”) involving animate/inanimate NPs, double/single-scope adjectives, and verbs of variable transitivity, and selected one of two images: one reflecting an object, TC-congruent, interpretation (e.g., a woman cleaning a cat), and the other reflecting a subject interpretation (the cat cleaning itself).

All groups consistently preferred the object interpretations. However, eye-tracking data revealed distinct processing strategies. Fixation duration analyses showed that EN MNs and FR-EN bilinguals fixated significantly more on the infinitive (INF) area, suggesting efficient object gap-filling. RU-EN bilinguals, however, distributed attention between the noun phrase and the INF areas, suggesting increased construction processing effort and need for syntactic reassignment. RU-EN-FR trilinguals displayed an intermediate, hybridized, pattern, indicating partial facilitation from L3/French. Fixation count analysis revealed overall significantly more fixations in the RU-EN

and RU-EN-FR groups than in the EN MNs and FR-EN groups, and while all L2 participants fixated most on the tough-adjective, with RU-EN bilinguals showing the strongest effect, FR-EN bilinguals again closely resembled the gaze patterns of EN MNs. Exploratory gaze-path analysis further revealed that RU-EN bilinguals engaged in more regressions, indicative of greater syntactic reanalysis needs during TC processing, as opposed to more linear readings by FR-EN and EN MNs.

These findings suggest that L2 processing difficulty is amplified by structural asymmetry and typological distance from L1. However, L3 knowledge, when typologically aligned with L2, can mitigate some of these challenges, supporting more efficient construction integration during L2 reading.

Mots-clés : acquisition d'une langue seconde, influence de la L3, oculométrie, stratégies de lecture, mesures de fixation, constructions évaluatives, transferts inter-linguistiques, français, anglais, russe

Résumé

La lecture dans une langue première (L1) est généralement un processus automatisé et efficace. En revanche, la lecture en langue seconde (L2), notamment lorsqu'elle implique des constructions syntaxiquement complexes ou asymétriques d'un point de vue inter-linguistique, tend à être plus lente et cognitivement plus exigeante (Parshina, et al, 2021 ; Whitford & Joanisse, 2018 ; Cheng et al., 2021 ; Patterson et al., 2014 ; Hawkins, 2004). L'une de ces structures est la *Tough Construction* (TC), comme dans « *This pet is easy to wash* », qui exige des lecteurs qu'ils interprètent le sujet grammatical (« this pet ») comme l'objet inféré d'un verbe non-fini enchâssé (« *to wash* ») (Chung, 2001 ; Guérin, 2006). Bien que les TC soient disponibles en anglais et en français, elles sont absentes dans des langues comme le russe, qui offrent des analogues fonctionnels à la place (Comrie & Matthews, 1990 ; Paykin & Van Peteghem, 2020 ; Tsikulina et al., 2025). Des recherches antérieures suggèrent que les structures syntaxiques de la L1 influencent significativement le traitement en L2, particulièrement lorsque les constructions divergent entre les langues (Booth et al., 2020 ; De Angelis et al., 2015). De même, d'un point de vue basé sur l'usage (Höder, 2019), l'expérience avec une langue seconde ou troisième (L3) peut conduire à des constructions concurrentes, hybrides ou qui se chevauchent, ce qui peut dans certains cas fournir un soutien compensatoire au traitement – particulièrement lorsque les langues impliquées sont typologiquement proches (Hui, 2010 ; Cheung et al., 2011).

Cette étude explore si la connaissance linguistique bilingue et trilingue facilite le traitement des TC. Spécifiquement, elle examine : (a) comment des bilingues tardifs avancés français-anglais (FR-EN) et russe-anglais (RU-EN) traitent les TC anglaises ; (b) si les trilingues russe-anglais-français (RU-EN-FR) bénéficient de leur connaissance de la L3 dans le décodage des constructions en L2 ; et (c) comment ces groupes multilingues peuvent être comparés aux monolingues anglais (EN MNs). L'étude se concentre sur le rôle de l'expérience linguistique et de la proximité typologique dans le traitement syntaxique et sémantique, ainsi que sur les effets de transfert potentiels.

Trente-six bilingues tardifs (FR-EN, RU-EN), quatorze trilingues RU-EN-FR, et sept monolingues anglais ont complété une tâche d'appariement phrase-image avec enregistrement oculométrique en parallèle. Les participants ont lu des phrases TC anglaises (par ex., « *This cat is easy to*

clean ») impliquant des syntagmes nominaux animés/inanimés, des adjectifs à double/simple portée, et des verbes de transitivité variable, et ont sélectionné l'une de deux images : l'une reflétant une interprétation objet, congruente avec la TC (par ex., une femme nettoyant un chat), et l'autre reflétant une interprétation sujet (le chat se nettoyant lui-même).

Tous les groupes ont systématiquement préféré les interprétations objet. Cependant, les données oculométriques ont révélé des stratégies de traitement distinctes. Les analyses de durée de fixation ont montré que les monolingues anglais et les bilingues FR-EN ont fixé significativement plus sur la zone infinitive (INF), suggérant un comblement efficace de lacune objet. Les bilingues RU-EN, cependant, ont distribué l'attention entre le syntagme nominal et les zones INF, suggérant un effort de traitement de construction accru et un besoin de réassignement syntaxique. Les trilingues RU-EN-FR ont affiché un modèle intermédiaire, hybride, indiquant une facilitation partielle de la L3/français. L'analyse du nombre de fixations a révélé globalement significativement plus de fixations dans les groupes RU-EN et RU-EN-FR que dans les groupes monolingues anglais et FR-EN, et bien que tous les participants L2 aient fixé le plus sur l'adjectif "tough", les bilingues RU-EN montrant l'effet le plus fort, les bilingues FR-EN ont à nouveau étroitement employé les mêmes stratégies oculaires que les monolingues anglais. L'analyse exploratoire du parcours du regard a en outre révélé que les bilingues RU-EN se sont engagés dans plus de régressions, indicatives de besoins de réanalyse syntaxique plus importants pendant le traitement des TC, par opposition à des lectures plus linéaires par les FR-EN et les monolingues anglais.

Ces résultats suggèrent que la difficulté de traitement en L2 est amplifiée par l'asymétrie structurelle et la distance typologique par rapport à la L1. Cependant, la connaissance de la L3, lorsqu'elle est typologiquement alignée avec la L2, peut atténuer certains de ces défis, soutenant une intégration de construction plus efficace pendant la lecture en L2.

Références

- Booth, P., Clenton, J., & Van Herwegen, J. (2020). L1–L2 semantic and syntactic processing: The influence of language proximity. In *First Language Influences on Multilingual Lexicons* (pp. 35–52). Routledge.
- Cheng, Y. S., Rothman, J., & Cunnings, I. (2021). Parsing preferences and individual differences in nonnative sentence processing: Evidence from eye movements. *Applied Psycholinguistics*, 42(1), 129–151.
- Cheung S. C., Matthews S. & Tsang W. L. (2011). Transfer from L3 German to L2 English in the domain of tense/aspect. In Dewaele J.-M., De Angelis G. (Eds.), *New trends in crosslinguistic influence and multilingualism research* (pp. 53–73). *Multilingual Matters*.
- Chung, Y.-S. (2001). *Tough construction in English: A construction grammar approach* (Doctoral dissertation). University of California.
- Comrie, B., & Matthews, S. (1990). Prolegomena to a typology of tough movement. In *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75th birthday* (pp. 43–58). Amsterdam: John Benjamins.

- De Angelis, G, Jessner, U., & Kresic, M (2015). Cross-linguistic influence and cross-linguistic interaction in multilingual language learning. London, U.K.: Bloomsbury Academic.
- Guérin, V. (2006). On tough constructions in French. *Manoa Working Papers in Linguistics*, 37(1), 1–21.
- Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. OUP Oxford.
- Höder, S., Prentice, J. & Tingsell, S. (2021). Additional language acquisition as emerging multilingualism: a Construction Grammar approach. In: Hans C. Boas & Steffen Höder. (Eds.), *Constructions in Contact 2: Language Change, Multilingual Practices, and Additional Language Acquisition*. John Benjamins Publishing Company, Amsterdam, Philadelphia, 309–338.
- Hui, B. (2010). Backward transfer from L3 French to L2 English production of relative clauses by L1 Cantonese speakers in Hong Kong. *Hong Kong Journal of Applied Linguistics*, 12(2), 45–60.
- Parshina, O., Laurinavichyute, A. K., & Sekerina, I. A. (2021). Eye-movement benchmarks in heritage language reading. *Bilingualism: Language and Cognition*, 24(1), 69–82.
- Patterson, C., Trompelt, H., & Felser, C. (2014). The online application of binding condition B in native and non-native pronoun resolution. *Frontiers in Psychology*, 5, 147.
- Paykin, K., & Van Peteghem, M. (2020). Des adjectifs tough dans des langues sans construction Tough? Le cas du russe. *Langages*, 218(2), 75–88.
- Tsikulina, A., Tayalati, F. & Soroli, E. (2025). English tough-constructions and their analogues in French and Russian: A parallel corpus investigation. *Languages in Contrast*, 25(1), 83–121.
- Whitford, V. & Joanisse, M. F. (2018). Do eye movements reveal differences between monolingual and bilingual children's first-language and second-language reading? A focus on word frequency effects. *Journal of Experimental Child Psychology*, 173, 318–337.

Sarah Turner (Coventry University) & Jeannette Littlemore (University of Birmingham)

What can metaphor teach us about sensitive and effective communication with bereaved parents and carers?

Keywords: metaphor; bereavement; healthcare communication; professional communication

Abstract

This talk introduces findings from two related research projects in which we interviewed parents and carers who considered themselves bereaved following a) pregnancy loss (Littlemore & Turner, 2020) or b) the death of their child of any age (Turner et al., 2022). The aim of both projects was to reach a better understanding of the nature of these experiences, and to consider ways in which the various professionals with whom bereaved parents and carers come into contact may best care for and communicate with the bereaved. In both projects, participants were asked to describe their experiences and to reflect on the care and communication they received from the range of professionals who supported them following the loss of their pregnancy or the death of their child. Our interview transcripts formed the datasets for each project.

We drew on metaphor analysis in analysing our interview transcripts. Previous research has shown that people frequently draw on metaphor when the subject under consideration is particularly personal or emotional, and has highlighted the utility of the methodology in reaching a deeper understanding of the nature of such experiences (e.g. Semino et al., 2017). In identifying metaphor, we employed an adapted version of the PRAGGLEJAZ Group's (2007) Metaphor Identification Procedure (MIP), which we combined with Cameron's (2003) vehicle identification procedure to identify metaphors at the level of the phrase. We then classified the metaphors into broad categories that reflected the semantic fields to which the vehicle terms belonged. For example, one participant produced a 'movement' related metaphor when she explained:

(1) "It feels impossible to shift and move forward."

From this analysis, we were able to consider the implications of the metaphors used, and the insights they might provide into the ways in which our participants were experiencing their losses. In Example (1) above, for example, the inability to 'move forward' indicates a sense of powerlessness or difficulties coming to terms with the experience.

In this talk, I share examples of the metaphorical language used by participants relating to three key themes: a) physical and emotional responses; b) altered relationships with time and space, and; b) continuing parental identity. I discuss cases where participants evaluated the care and communication they received from healthcare practitioners in light of these three themes, proposing that effective communication can be considered that which coheres with the experiences of the bereaved, as conveyed through metaphor. I conclude by demonstrating how we used our findings to inform suggestions for improved practice.

Mots-clés : métaphore ; deuil ; communication en santé ; communication professionnelle

Résumé

Cette présentation introduit les résultats de deux projets de recherche connexes dans lesquels nous avons interrogé des parents et aidants qui se considéraient en deuil suite à a) une perte de grossesse (Littlemore & Turner, 2020) ou b) au décès de leur enfant quel que soit son âge (Turner et al., 2022). L'objectif des deux projets était de parvenir à une meilleure compréhension de la nature de ces expériences, et d'examiner les façons dont les divers professionnels avec lesquels les parents et aidants en deuil entrent en contact peuvent mieux prendre soin des personnes endeuillées et communiquer avec elles. Dans les deux projets, les participants ont été invités à décrire leurs expériences et à réfléchir sur les soins et la communication qu'ils ont reçus de la part de l'éventail de professionnels qui les ont soutenus suite à la perte de leur grossesse ou au décès de leur enfant. Nos transcriptions d'entretiens ont constitué les jeux de données pour chaque projet.

Nous avons eu recours à l'analyse métaphorique dans l'analyse de nos transcriptions d'entretiens. Des recherches antérieures ont montré que les gens utilisent fréquemment la métaphore lorsque le sujet considéré est particulièrement personnel ou émotionnel, et ont souligné l'utilité de cette méthodologie pour parvenir à une compréhension plus approfondie de la nature de telles expériences (par ex. Semino et al., 2017). Pour identifier les métaphores, nous avons employé une version adaptée de la Procédure d'Identification des Métaphores (MIP) du Groupe PRAGGLEJAZ (2007), que nous avons combinée avec la procédure d'identification des véhicules de Cameron (2003) pour identifier les métaphores au niveau de la phrase. Nous avons ensuite classé les métaphores en catégories générales qui reflétaient les champs sémantiques auxquels appartenaient les termes véhicules. Par exemple, une participante a produit une métaphore liée au « mouvement » lorsqu'elle a expliqué :

(1) « Il semble impossible d'évoluer et d'aller de l'avant. »

À partir de cette analyse, nous avons pu considérer les implications des métaphores utilisées, et les éclairages qu'elles pourraient apporter sur les façons dont nos participants vivaient leurs pertes. Dans l'Exemple (1) ci-dessus, par exemple, l'incapacité à « aller de l'avant » indique un sentiment d'impuissance ou des difficultés à composer avec l'expérience.

Dans cette présentation, je partage des exemples du langage métaphorique utilisé par les participants en rapport avec trois thèmes clés : a) les réponses physiques et émotionnelles ; b) les relations altérées avec le temps et l'espace ; et c) l'identité parentale continue. Je discute des cas où les participants ont évalué les soins et la communication qu'ils ont reçus de la part des praticiens de santé à la lumière de ces trois thèmes, proposant qu'une communication efficace peut être considérée comme celle qui s'accorde avec les expériences des personnes endeuillées, telles qu'elles sont transmises par la métaphore. Je conclus en démontrant comment nous avons utilisé nos résultats pour éclairer des suggestions d'amélioration des pratiques.

Références

Cameron, L. (2003). *Metaphor in Educational Discourse*. Continuum.

Littlemore, J., & Turner, S. (2020). Metaphors in communication about pregnancy loss. *Metaphor and the Social World*, 10(1), 45–75.

Pragglejaz. (2007). MIP: A Method for Identifying Metaphorically Used Words in Discourse. *Metaphor and Symbol*, 22(1), 1–40.

Semino, E., Demjén, Z., Hardie, A., Payne, S., & Rayson, P. (2017). *Metaphor, cancer and the end of life: A corpus-based study*. Routledge.

Turner, S., Littlemore, J., Taylor, J., Parr, E., & Topping, A. (2022). Metaphors that shape parents' perceptions of effective communication with healthcare practitioners following child death: A qualitative UK study. *BMJ Open*, 12(1), 1–8.

**Tobias Ungerer (University of Toronto), Theresa Matzinger (University of Vienna),
Michael Pleyer (Nicolaus Copernicus University) & Stefan Hartmann (Heinrich
Heine University Düsseldorf)**

***Testing extravagance with experiments. An artificial language learning approach
to constructional creativity***

Keywords: Extravagance, creativity, artificial language learning, communication game experiments

Abstract

Languages are socially constructed entities, and as such, the ways in which they evolve over time are subject to various social pressures. The role of one such factor – extravagance, defined as speakers’ desire to talk in such a way that they are noticed (Haspelmath 1999) – has seen a surge of research interest in recent years, especially so in recent Construction Grammar approaches to linguistic creativity (e.g., Petré 2017; De Wit et al. 2020; Ungerer & Hartmann 2020; Eitelmann & Haumann [eds.] 2022; Neels et al. 2023; Koliopoulou & Walker 2024; Hoffmann 2024). According to the extravagance view, speakers aim to stand out from their peers by creating novel linguistic forms and meanings, which may then be adopted by others due to the competing social pressure towards conformity. However, as a concept that refers to language users’ intentions, extravagance has proved difficult to operationalize, especially since most studies to date have been corpus-based (but see Baumann & Mühlenbernd 2022 for a computational modelling approach).

We present the first experimental study of the role of extravagance as a driver of linguistic innovation, using an artificial language-learning approach (e.g., Kirby et al. 2008, Fedzechkina et al. 2023). In our pre-registered experiment (https://osf.io/8hea6/?view_only=db75f10e9aa64b8b81ce04ad72ceb52f), 100 participants (as determined by a power analysis with G*Power; Faul et al. 2007) will be instructed, in a game-like setting, to learn words of an alien language, while pursuing either of two social goals: (i) in the “extravagance bias” condition, participants’ goal is to be recognized by the aliens as special; (ii) in the “conformity bias” condition, their goal is to be recognized by the aliens as one of their own.

Participants are trained on an artificial miniature language consisting of four adjectives, each of which will be presented in a “non-extravagant” and an “extravagant” form. “Extravagant” forms contain diacritics that are absent from English orthography (e.g., *ranu* vs. *rānũ*). During training, participants will see the eight written forms paired with pictures, followed by a comprehension (i.e. picture- selection) task. During testing, participants will be presented with pictures and asked to type the alien words (using a virtual keyboard that includes diacritics). The test phase will include both the familiar (trained) pictures and four novel (untrained) pictures, for which the participants are asked to freely invent words.

We predict that participants in the “extravagance bias” condition will produce more extravagant forms (as measured by the number of diacritics) and match the training items less faithfully (as measured by Levenshtein distance; Levenshtein 1966) than participants in the “conformity bias” condition. This would suggest that the social pressure to stand out from one’s peers gives rise to increased linguistic innovation. We will also investigate differences between familiar and novel

items, for example whether participants produce more extravagant forms for novel items than for trained ones. Our findings will provide the first experimental evidence about how social pressures towards extravagance and conformity shape linguistic innovations at the lexical level.

Mots-clés: extravagance, créativité, apprentissage des langues artificielles, expériences de jeux de communication

Résumé

Les langues sont des entités construites socialement et, à ce titre, leur évolution au fil du temps est soumise à diverses pressions sociales. Le rôle d'un de ces facteurs – l'extravagance, définie comme le désir des locuteurs de parler de manière à se faire remarquer (Haspelmath 1999) – a suscité un regain d'intérêt de la part des chercheurs ces dernières années, en particulier dans les approches récentes de la grammaire de construction appliquées à la créativité linguistique (par exemple, Petr   2017 ; De Wit et al. 2020 ; Ungerer & Hartmann 2020 ; Eitelmann & Haumann [  d.] 2022 ; Neels et al. 2023 ; Koliopoulou & Walker 2024 ; Hoffmann 2024). Selon la th  orie de l'extravagance, les locuteurs cherchent    se d  marquer de leurs pairs en cr  ant des formes et des significations linguistiques nouvelles, qui peuvent ensuite   tre adopt  es par d'autres en raison de la pression sociale concurrentielle vers la conformit  . Cependant, en tant que concept qui fait r  f  rence aux intentions des utilisateurs de la langue, l'extravagance s'est av  r  e difficile    op  rationnaliser, d'autant plus que la plupart des   tudes men  es    ce jour ont   t   bas  es sur des corpus (mais voir Baumann & M  hlenbernd 2022 pour une approche de mod  lisation computationnelle).

Nous pr  sentons la premi  re   tude exp  rimentale sur le r  le de l'extravagance en tant que moteur de l'innovation linguistique, en utilisant une approche d'apprentissage artificiel des langues (par exemple, Kirby et al. 2008, Fedzechkina et al. 2023). Dans notre exp  rience pr  enregistr  e (https://osf.io/8hea6/?view_only=db75f10e9aa64b8b81ce04ad72ceb52f), 100 participants (d  termin  s par une analyse de puissance avec G*Power ; Faul et al. 2007) seront invit  s, dans un cadre ludique,    apprendre des mots d'une langue extraterrestre, tout en poursuivant l'un des deux objectifs sociaux suivants : (i) dans la condition « biais d'extravagance », l'objectif des participants est d'  tre reconnus comme sp  ciaux par les extraterrestres ; (ii) dans la condition « biais de conformit   », leur objectif est d'  tre reconnus par les extraterrestres comme l'un des leurs. Les participants sont form  s    une langue miniature artificielle compos  e de quatre adjectifs, chacun d'entre eux   tant pr  sent   sous une forme « non extravagante » et une forme « extravagante ». Les formes « extravagantes » contiennent des signes diacritiques qui n'existent pas dans l'orthographe anglaise (par exemple, ranu vs ra  n  ). Pendant la formation, les participants verront les huit formes   crites associ  es    des images, suivies d'une t  che de compr  hension (c'est-  -dire une s  lection d'images). Pendant le test, les participants se verront pr  senter des images et devront taper les mots extraterrestres (   l'aide d'un clavier virtuel comprenant des signes diacritiques). La phase de test comprendra    la fois les images famili  res (appries) et quatre images nouvelles (non appries), pour lesquelles les participants devront inventer librement des mots.

Nous pr  voyons que les participants dans la condition « biais d'extravagance » produiront des formes plus extravagantes (mesur  es par le nombre de signes diacritiques) et correspondront moins fid  lement aux   l  ments d'entra  nement (mesur  s par la distance de Levenshtein ;

Levenshtein 1966) que les participants dans la condition « biais de conformité ». Cela suggérerait que la pression sociale pour se démarquer de ses pairs donne lieu à une innovation linguistique accrue. Nous étudierons également les différences entre les éléments familiers et les éléments nouveaux, par exemple si les participants produisent des formes plus extravagantes pour les éléments nouveaux que pour les éléments appris. Nos résultats fourniront les premières preuves expérimentales sur la manière dont les pressions sociales en faveur de l'extravagance et de la conformité façonnent les innovations linguistiques au niveau lexical.

Références

- Baumann, Andreas & Roland Mühlenbernd. 2022. Less of the same: Modeling horror aequi and extravagance as mechanisms of negative frequency dependence in linguistic diversification. In Andrea Ravignani, Rie Asano, Daria Valente, Francesco Ferretti, Stefan Hartmann, Misato Hayashi, Yannick Jadoul, et al. (eds.), *The Evolution of Language: Proceedings of the Joint Conference on Language Evolution (JCoLE)*, 50–57. Nijmegen: MPI.
- De Wit, Astrid, Peter Petré & Frank Brisard. 2020. Standing out with the progressive. *Journal of Linguistics* 1–36. <https://doi.org/10.1017/S0022226719000501>
- Eitelmann, Matthias & Dagmar Haumann (eds.). 2022. *Extravagant morphology: Studies in rule-bending, pattern-extending and theory-challenging morphology*. Amsterdam & Philadelphia: John Benjamins.
- Faul, Franz, Edgar Erdfelder, Albert-Georg Lang & Axel Buchner. 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39(2). 175–191. <https://doi.org/10.3758/BF03193146>
- Fedzechkina, Masha, Lucy Hall Hartley & Gareth Roberts. 2023. Social biases can lead to less communicatively efficient languages. *Language Acquisition* 30(3–4). 230–255. <https://doi.org/10.1080/10489223.2022.2057229>
- Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37(6). 1043–1068. <https://doi.org/10.1515/ling.37.6.1043>
- Hoffmann, Thomas. 2024. Cognitive approaches to linguistic creativity. *Cambridge Encyclopedia of Cognitive Linguistics*. <https://doi.org/10.33774/coe-2024-8tnps>
- Kirby, Simon, Hannah Cornish & Kenny Smith. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *PNAS* 105(31). 10681–10686. <https://doi.org/10.1073/pnas.0707835105>
- Koliopoulou, Maria & Jim Walker. 2024. Exploring creativity and extravagance: The case of double suffixation in English. *Languages* 9(3). 88. <https://doi.org/10.3390/languages9030088>
- Levenshtein, V. I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* 10. 707–710.
- Neels, Jakob, Stefan Hartmann & Tobias Ungerer. 2023. A quantum of salience: Reconsidering the role of extravagance in grammaticalization. In Hendrik De Smet, Peter Petré & Benedikt Szendrői (eds.), *Context, intent and variation in grammaticalization*, 47–78. Berlin, Boston: De Gruyter. <https://doi.org/10.1515/9783110753059-003>

Petré, Peter. 2017. The extravagant progressive: An experimental corpus study on the history of emphatic [be V-ing]. *English Language and Linguistics* 21(2). 227–250.
<https://doi.org/10.1017/S1360674317000107>

Traugott, Elizabeth C. 2017. Low salience as an enabling factor in morphosyntactic change. In Marianne Hundt, Sandra Mollin & Simone E. Pfenninger (eds.), *The changing English language: Psycholinguistic perspectives*, 93–110. Cambridge: Cambridge University Press.

Ungerer, Tobias & Stefan Hartmann. 2020. Delineating extravagance: Assessing speakers' perceptions of imaginative constructional patterns. *Belgian Journal of Linguistics* 34. 345–356.
<https://doi.org/10.1075/bjl.00058.ung>

Agathe Wallet, Johanna Cordova, Samuel Chakmakjian, Damien Nouvel, Ilaine Wang & Victoria Khurshudyan (Inalco)

Transcription automatique de langues peu dotées : Optimisation de l'apprentissage de modèles de reconnaissance automatique de la parole dans un contexte de ressources limitées

Mots-clés : ASR ; LID ; écriture phonétique ; langue pluricentrique ; détection de variant

Résumé

Pour améliorer les systèmes de reconnaissance automatique de la parole, des données abondantes et diversifiées sont cruciales. Les langues moins dotées, comme l'arménien occidental, manquent de données, ce qui pose un défi. Parlé uniquement par les communautés diasporiques, l'arménien occidental (SWA pour Standard Western Armenian) est une variante minoritaire par rapport à l'arménien oriental (SEA pour Standard Eastern Armenian), bénéficiant d'un ancrage territorial. L'absence de code ISO distinct avant 2018 complique également la collecte des données dans une variante spécifique, notamment en arménien occidental dont les données sont d'autant plus rares et éparses.

Les modèles de reconnaissance vocale existants (MMS-1B-all, wav2vec2-large-xls-r-300m-armenian et AssemblyAI) montrent des performances limitées sur cette langue, avec des taux d'erreur par caractère élevés.

Le contact de l'arménien occidental avec d'autres langues introduit des variations de prononciation, rendant la tâche encore plus complexe. Pour surmonter ces obstacles, une approche basée sur la transcription phonémique a été adoptée, permettant de mieux gérer les variations linguistiques malgré le peu de données disponibles. Un outil de détection des variantes a également été développé pour créer des jeux de données plus homogènes.

Identification des variantes

Un modèle d'identification de langue (LID) basé sur le modèle statistique de mélange gaussien (GMM) (Sarmah, Bhattacharjee 2014) a été entraîné sur diverses données audio en arménien oriental et occidental (podcasts, émissions télévisées, livres audio, etc), qui ne font pas partie de nos données d'entraînement ou de test pour l'entraînement des modèles de reconnaissance.

Ce modèle calcule la log-vraisemblance associée à la prédiction pour chaque extrait audio. On mesure l'écart entre les log-vraisemblances associées aux deux variantes standards (SWA et SEA), qu'on interprète comme un "degré d'appartenance" à une des deux variantes linguistiques.

Transcription phonétique

Pour obtenir des données en arménien occidental, nous avons utilisé des témoignages de Syriens d'origine arménienne¹⁹, totalisant 8 heures d'enregistrement annotées. Le détecteur de variants a permis de créer cinq jeux de données d'entraînement et de validation de 2h47 chacun, ainsi qu'un jeu de test de 33 minutes. Le score médian des écarts de log-vraisemblances (ici, 4,4) a été utilisé pour déterminer la confiance du détecteur dans la détection des variants (un score

¹⁹ Données issues du projet Rerooted, utilisables dans le cadre de travaux académiques : <https://www.rerooted.org/>

supérieur à 4,4 indiquant que le détecteur est confiant dans sa détection) et ainsi créer des jeux de données plus ou moins homogènes mais quantitativement équivalents (voir tableau 1).

Trois modèles ont été entraînés sur ces données : Wav2Vec2-XLS-R (300M et 1B de paramètres) (Babu et al., 2021) non affinés et Multipa²⁰ (Taguchi et al. 2023), optimisé pour la reconnaissance de phones. Les résultats obtenus sur le jeu test sont présentés dans le tableau 2.

Jeux de données	Score du détecteur	Nombre de fichiers
1	100% <i>score</i> $\geq 4,4$	4151
2	100% <i>score</i> $< 4,4$	4310
3	50% <i>score</i> $\geq 4,4$ et 50% <i>score</i> $< 4,4$	4214
4	75% <i>score</i> $\geq 4,4$ et 25% $2,6 < \textit{score} < 4,4$	4164
5	75% <i>score</i> $\geq 4,4$ et 25% <i>score</i> $< 2,6$	4206
Test	Choisies aléatoirement	1168
SWA détecté comme étant SEA		310

Tableau 1 : Présentation des jeux de données, les jeux 1 à 5 étant les différents jeux d'entraînement sur lesquels les modèles de reconnaissance ont été entraînés

Modèle	Utilisation GPU	Jeu d'entraînement	PER (en %)
Multipa	20 à 22 Go	1	23,13
		2	23,59
		3	<u>22,25</u>
		4	22,68
		5	22,60
XLS-R-300M	7 à 9 Go	1	21,07
		2	21,33
		3	<u>20,19</u>
		4	20,58
		5	20,35
XLS-R-1B	24 à 26 Go	1	24,14
		2	24,44
		3	<u>23,00</u>
		4	23,78
		5	22,97

Tableau 2 : Taux d'erreur par phone des différents modèles sur le jeu de test, en fonction des données d'entraînement

Malgré le peu de données, nous obtenons des résultats encourageants avec des ressources limitées : un GPU de 16Go et 2h45 d'audio suffisent pour approcher les performances de l'état de l'art pour les langues peu dotées (Babu et al. obtiennent un PER de 17% pour le cantonais de Hong-Kong (zh-HK) avec 51h d'audio pour le pré-entraînement et 1h de données annotées). Les données et scripts sont disponibles sur le répertoire du projet²¹.

²⁰ <https://github.com/ctaguchi/multipa/tree/main?tab=readme-ov-file>

²¹ https://gitlab.huma-num.fr/ertim/dalih/asr_dalih/-/tree/main?ref_type=heads

Keywords: ASR; LID; phonetic transcription; pluricentric language; variant detection

Abstract

To improve automatic speech recognition (ASR) systems, abundant and diverse data are crucial. Lesser-resourced languages, such as Western Armenian, lack such data, which poses a challenge. Spoken only by diasporic communities, Western Armenian (SWA, Standard Western Armenian) is a minority variety compared to Eastern Armenian (SEA, Standard Eastern Armenian), which benefits from a territorial base. The absence of a distinct ISO code before 2018 further complicates data collection in a specific variety, particularly Western Armenian, whose data are especially scarce and scattered.

Existing speech recognition models (MMS-1B-all, wav2vec2-large-xls-r-300m-armenian, and AssemblyAI) show limited performance on this language, with high character error rates. The contact of Western Armenian with other languages introduces pronunciation variations, making the task even more complex. To overcome these obstacles, a phonemic transcription approach was adopted, allowing better management of linguistic variation despite the limited amount of data available. A variant detection tool was also developed to create more homogeneous datasets.

Variant identification

A language identification (LID) model based on a Gaussian Mixture Model (GMM) (Sarmah & Bhattacharjee, 2014) was trained on various Eastern and Western Armenian audio data (podcasts, TV programs, audiobooks, etc.), which were not part of our training or test data for ASR model training.

This model computes the log-likelihood associated with the prediction for each audio sample. The gap between the log-likelihoods associated with the two standard variants (SWA and SEA) is measured and interpreted as a “degree of belonging” to one of the two language variants.

Phonetic transcription

To obtain Western Armenian data, we used testimonies from Syrian speakers of Armenian origin, totaling 8 hours of annotated recordings.²² The variant detector enabled the creation of five training and validation datasets of 2h47 each, as well as a 33-minute test set. The median score of the log-likelihood gaps (here, 4.4) was used to determine the detector’s confidence in variant identification (a score above 4.4 indicating that the detector is confident in its classification), thereby allowing the creation of more or less homogeneous but quantitatively equivalent datasets (see Table 1).

Three models were trained on these data: Wav2Vec2-XLS-R (300M and 1B parameters) (Babu et al., 2021) without fine-tuning, and Multipa²³ (Taguchi et al., 2023), optimized for phone recognition. The results obtained on the test set are presented in Table 2.

²² Data from the Rerooted project, available for use in academic research: <https://www.rerooted.org/>

²³ <https://github.com/ctaguchi/multipa/tree/main?tab=readme-ov-file>

Dataset	Detection score	Number of files
1	100% <i>score</i> $\geq 4,4$	4151
2	100% <i>score</i> $< 4,4$	4310
3	50% <i>score</i> $\geq 4,4$ and 50% <i>score</i> $< 4,4$	4214
4	75% <i>score</i> $\geq 4,4$ and 25% $2,6 \leq \text{score} < 4,4$	4164
5	75% <i>score</i> $\geq 4,4$ and 25% <i>score</i> $< 2,6$	4206
Test	Randomly selected	1168
SWA detected as SEA		310

Table 1: Overview of the datasets, with datasets 1 to 5 corresponding to the different training sets used to train the recognition models

Model	GPU use	Training set	PER (in %)
Multipa	20 to 22 Go	1	23,13
		2	23,59
		3	<u>22,25</u>
		4	22,68
		5	22,60
XLS-R-300M	7 to 9 Go	1	21,07
		2	21,33
		3	<u>20,19</u>
		4	20,58
		5	20,35
XLS-R-1B	24 to 26 Go	1	24,14
		2	24,44
		3	<u>23,00</u>
		4	23,78
		5	22,97

Table 2: Phone error rates of the different models on the test set, as a function of the training data

Despite the limited amount of data, we obtain encouraging results with restricted resources: a 16GB GPU and 2h45 of audio are sufficient to approach state-of-the-art performance for low-resource languages (Babu et al. report a PER of 17% for Hong Kong Cantonese (zh-HK) with 51 hours of audio for pretraining and 1 hour of annotated data). The data and scripts are available in the project repository.²⁴

²⁴ https://gitlab.huma-num.fr/ertim/dalih/asr_dalih/-/tree/main?ref_type=heads

Références

- Babu, A., Wang, C., Tjandra, A., Lakhotia, K., Xu, Q., Goyal, N., Singh, K., von Platen, P., Saraf, Y., Pino, J., Baevski, A., Conneau, A., & Auli, M. (2021). *XLS-R: Self-supervised Cross-lingual Speech Representation Learning at Scale* (No. arXiv:2111.09296). arXiv. <https://doi.org/10.48550/arXiv.2111.09296> (consulté le 17.07.2024)
- Dhawan, K., Rekesh, Kd., & Ginsburg, B. (2023). Unified Model for Code-Switching Speech Recognition and Language Identification Based on Concatenated Tokenizer. In G. Winata, S. Kar, M. Zhukova, T. Solorio, M. Diab, S. Sitaram, M. Choudhury, & K. Bali (Eds.), *Proceedings of the 6th Workshop on Computational Approaches to Linguistic Code-Switching* (pp. 74–82). Association for Computational Linguistics. <https://aclanthology.org/2023.calcs-1.7> (consulté le 05.06.2024)
- Li, H., Ma, B., & Lee, K. A. (2013). Spoken language recognition: from fundamentals to practice. *Proceedings of the IEEE*, 101(5), 1136-1159.
- Mortensen, D. R., Littell, P., Bharadwaj, A., Goyal, K., Dyer, C., & Levin, L. S. (2016). PanPhon: A Resource for Mapping IPA Segments to Articulatory Feature Vectors. *International Conference on Computational Linguistics*. <https://www.semanticscholar.org/paper/PanPhon%3A-A-Resource-for-Mapping-IPA-Segments-to-Mortensen-Littell/cef1b4c0f46d2c7b1568dec36b06e1319fdcfea1> (consulté le 17.06.2024)
- Sarmah, K., & Bhattacharjee, U. (2014). GMM based Language Identification using MFCC and SDC Features. *International Journal of Computer Applications*, 85(5).
- Taguchi, C., Sakai, Y., Haghani, P., & Chiang, D. (2023). Universal Automatic Phonetic Transcription into the International Phonetic Alphabet <https://doi.org/10.48550/arXiv.2308.03917> (consulté le 10.06.2024)

Marinus Wiedner (University of Freiburg)

Développement d'un corpus annoté de l'occitan médiéval à l'aide d'intelligence artificielle

Mots-clés : Occitan médiéval ; Corpus ; annotation de parties de discours ; Large Language Models ; Handwritten text recognition

Résumé

L'occitan médiéval est l'une des premières langues étudiées dans le cadre de la philologie romane. Malgré la longue tradition de recherche sur l'occitan médiéval, certains aspects n'ont pas encore été étudiés en profondeur, par exemple le genre grammatical : Comment le neutre latin a-t-il évolué lors du passage du latin à l'occitan moderne ? Quel genre a été attribué aux néologismes, surtout concernant les substantifs dérivés d'un verbe à la suite d'une conversion ? Comment les doublons de genre se répartissent-ils ?

Le focus de cette communication sera mis sur la méthodologie pour créer un corpus constitué de textes médiévaux à l'aide d'intelligence artificielle. Les questions de recherche susmentionnées ne servent que comme exemples afin de démontrer les possibilités de ce corpus pour la recherche.

Pour répondre à ces questions, il est nécessaire de disposer d'un corpus assez large de textes en occitan médiéval. Comme la plupart des éditions ne correspondent pas à la lettre aux manuscrits (cf. Marzo/Wiedner à paraître) l'utilité de ces éditions, souvent créées au 19ème siècle dans des contextes historiographiques (cf. Selig 2005 : 156–257) reste limitée pour la recherche en linguistique. Pour aborder cet obstacle, j'ai décidé de créer un corpus moi-même. Ce corpus consiste en 17 textes du 13ème et 14ème siècle se répartissant entre l'aire linguistique du languedocien et du provençal. Ces manuscrits ont été transcrits à l'aide d'un modèle *Transkribus* (cf. Wiedner 2023) et ensuite corrigés manuellement. Ce corpus est constitué d'environ 900.000 *tokens*, réparti presque de manière équilibrée entre les deux régions.

Après la transcription, j'ai annoté une partie des textes. Le premier manuscrit a été pré-annoté à l'aide d'un tagger pour l'occitan moderne (cf. Poujade en cours) et corrigé manuellement. Le deuxième texte a été annoté à l'aide d'un *Large Language Model* (LLM) en suivant les propositions dans Schöffel et al. (2025) et ensuite de nouveau post-corrigé. Je dispose alors de 70.000 *tokens* avec une annotation de partie de discours vérifiée, et tout le reste de mon corpus est pré-annoté mais doit encore être vérifié. Le corpus est désormais publié sous le nom de COMETA (Wiedner 2025).

Dans ma communication, je vais présenter les démarches et les décisions nécessaires pour créer un corpus annoté d'une langue minorisée en commençant au niveau du manuscrit. En plus, je vais donner des pistes en ce qui concerne l'emploi de mon corpus pour la recherche en linguistique historique à l'exemple des questions du développement du genre grammatical.

Keywords: Medieval Occitan; Corpus; Part-of-speech annotation; Large Language Models; Handwritten Text Recognition

Abstract

Medieval Occitan is one of the first languages to have been studied within the field of Romance philology. Despite the long-standing research tradition on Medieval Occitan, some aspects have not yet been investigated in depth, for example grammatical gender: How did the Latin neuter evolve in the transition from Latin to modern Occitan? What gender was assigned to neologisms, especially to nouns derived from verbs through conversion? How are gender doublets distributed?

The focus of this paper will be on the methodology for creating a corpus of medieval texts with the help of artificial intelligence. The aforementioned research questions serve merely as examples to demonstrate the potential of such a corpus for linguistic research.

To address these questions, it is necessary to have access to a sufficiently large corpus of Medieval Occitan texts. Since most editions do not correspond verbatim to the manuscripts (cf. Marzo/Wiedner, forthcoming), their usefulness – often produced in the 19th century in historiographical contexts (cf. Selig 2005: 156–257) – remains limited for linguistic research. To overcome this obstacle, I decided to compile a corpus myself. This corpus consists of 17 texts from the 13th and 14th centuries, distributed across the Languedocian and Provençal areas. These manuscripts were transcribed using a Transkribus model (cf. Wiedner 2023) and subsequently corrected manually. The corpus comprises approximately 900,000 tokens, distributed almost evenly between the two regions.

After transcription, I annotated part of the texts. The first manuscript was pre-annotated with a tagger for modern Occitan (cf. Poujade, in progress) and then manually corrected. The second text was annotated using a Large Language Model (LLM), following the approach of Schöffel et al. (2025), and subsequently post-corrected. As a result, I now have 70,000 tokens with verified part-of-speech annotation, while the remainder of the corpus is pre-annotated but still requires verification. The corpus has now been published under the name COMETA (Wiedner 2025).

In this presentation, I will discuss the steps and decisions involved in creating an annotated corpus of a minoritized language, starting at the manuscript level. In addition, I will outline possible applications of my corpus for research in historical linguistics, using the example of questions relating to the development of grammatical gender.

Références

Marzo, D. & Wiedner, M. (à paraître) : « Remarks on Grammatical Gender in Romance ».

Poujade, C. (en cours) : *La linguistique outillée à l'épreuve de la variation : Ressources pour l'analyse de parlars occitans de l'Ariège*. Thèse de doctorat en cours à l'Université de Toulouse.

Selig, M. (2005) : « Schreiberprofile und Sprachstandardisierung. Bemerkungen zur mediävistischen Korpuslinguistik. » in : Schrott, A. & Völker, H. (éds.). *Historische Pragmatik und Historische Varietätenlinguistik*. Göttingen: Universitätsverlag Göttingen, 255–268.

Schöffel, M. & Wiedner, M. & Garcias Arias, E. & Ruppert, P. & Heumann, C. / Aßenmacher, M. (2025) : « Modern Models, Medieval Texts: A POS Tagging Study of Old Occitan ». In: *Proceedings of the 5th International Conference on Natural Language Processing for Digital Humanities*, 334–349.

Wiedner, M. (2023): *OldOccitanHandwriting*, (Modell-Nr. 52822, CER=3,51 %), *PyLaia-model for handwritten Occitan, 13th and 14th century*. <https://readcoop.eu/de/modelle/old-occitan-handwriting/> (dernière vue 16.06.2025).

Wiedner, M. (2025): COMETA = Corpus de l'occitan médiéval comparatif et annoté: Provence et Languedoc. <https://zenodo.org/records/15300719> (dernière vue 16.06.2025).

Word order consistency of object pronouns: A typological asymmetry

Keywords: word order; asymmetry; pronoun; reflexives; reciprocals

Abstract

Typological studies show that the order of object and verb has strong correlations with other elements in the sentence (Dryer, 1992; Kalyan & Donohue, 2023). In these studies, the prototypical object is understood as a lexical and nonpronominal noun phrase, based on the assumption that pronominal objects behave differently from prototypical objects. For example, French has SVO order (1a), but when the object is a pronoun, it takes preverbal position (1b). The reflexive pronoun aligns with the personal pronoun in this case (1c). This study examines the word orders of single-object constructions with a nominal phrase or a pronoun—personal, reflexive or reciprocal.

(1) French (Guentchéva & Rivière, 2008)

- a. Les parents lavent l'enfant.
ART parents wash.PRES ART.enfant
'The parents are washing the child.'
- b. Les parents le lavent.
ART parents PRON wash.PRES
'The parents are washing him.'
- c. Les parents se lavent.
ART parents REFL wash.PRES
'The parents are washing themselves.'

Our data survey 156 languages which have the feature of pronominal type coded in the Universal Dependencies project (UD v2.15; Zeman et al., 2024). Nominal and pronominal phrases are always coded in UD, but only a subset of languages specifically code personal pronouns and other pronominal types. Instances with exactly one object in the verb phrase are extracted. The matched instances are then categorised into exclusive groups of nominal noun phrases (1a), personal pronouns (1b; excluding reflexives and reciprocals), reflexives (1c), and reciprocals.

Figure 1 shows the proportion of VO order as opposed to OV order in percentage, when the object belongs to NOUN, personal pronoun (PRON), reflexive (REFL), or reciprocal (RECP). A dot plotted on the rightmost means that all instances in that category have VO order. The y-axis is sorted by descending percentages of NOUN in VO order (red dots). The smooth transition from right to left signifies word order flexibility in general. An asymmetry surfaces when PRON often appears in the preverbal domain even for highly consistent and rigid VO languages, so that [V NP & Pn V] is common, but the mirror situation [NP V & V Pn] occurs much less commonly in OV languages.

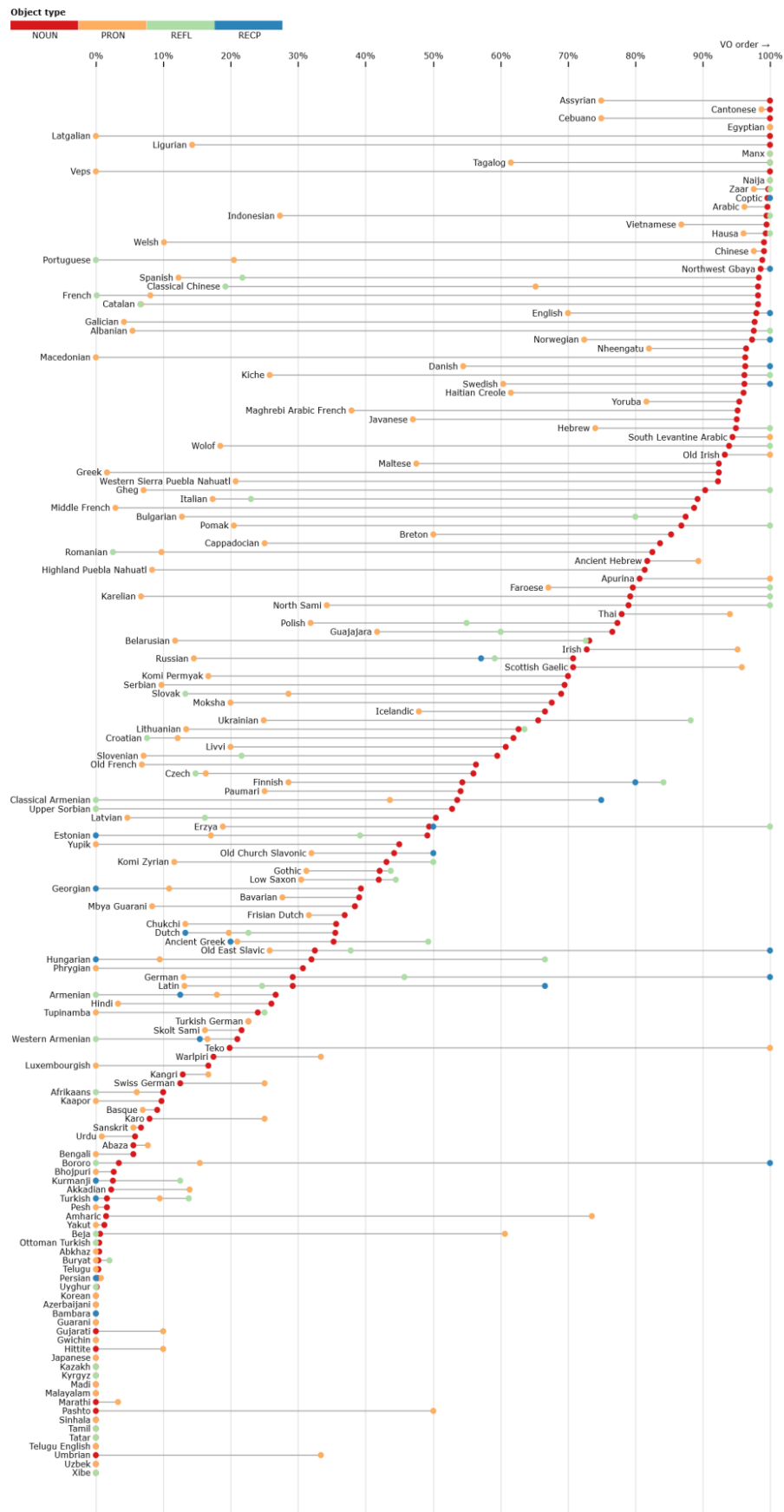
We then identify word order preference by an arbitrary dividing line of 50% corpus occurrence, and define word order consistency as whether NOUN has the same preference as PRON, REFL, or RECP. Table 1 summarises the number of languages with consistent word order in languages

with rigid VO (NOUN VO order > 90%), non-rigid VO (NOUN VO order > 50%), non-rigid OV (NOUN VO order < 50%), and rigid OV (NOUN VO order < 10%). Rigid and non-rigid OV languages are strongly consistent, VO LLcD 2025 counterparts much less so. However, the low consistency in the non-rigid VO class cannot be explained solely by a preference to place PRON early in the sentence, as rigid VO languages have a near 50% chance to prefer having PRON on the same side as NOUN. Thus the asymmetry remains. Some possible efficiency-based explanations are discussed based on Hawkins (2014).

Table 1: Proportion of languages having consistent word order between nominal object and personal pronoun (PRON), reflexive (REFL), or reciprocal (RECP) respectively

Classe	PRON	%	REFL	%	RECP	%
1 Rigid VO	25/45	56%	14/19	74%	6/6	100%
2 Non-rigid VO	6/37	16%	12/21	57%	3/3	100%
3 Non-rigid OV	30/31	97%	12/16	75%	7/12	58%
4 Rigid OV	40/43	93%	13/13	100%	4/5	80%
Total	101/156	65%	51/69	74%	20/26	77%

Fig. 1: Percentage of VO constructions with NOUN, personal pronoun (PRON), reflexive (REFL), or reciprocal (RECP) as the direct object



Mots-clés : ordre des mots ; asymétrie ; pronom ; réfléchis ; réciproques

Résumé

Les études typologiques montrent que l'ordre de l'objet et du verbe a de fortes corrélations avec d'autres éléments de la phrase (Dryer, 1992 ; Kalyan & Donohue, 2023). Dans ces études, l'objet prototypique est compris comme un syntagme nominal lexical et non pronominal, basé sur l'hypothèse que les objets pronominaux se comportent différemment des objets prototypiques. Par exemple, le français a un ordre SVO (1a), mais quand l'objet est un pronom, il prend une position préverbale (1b). Le pronom réfléchi s'aligne avec le pronom personnel dans ce cas (1c). Cette étude examine les ordres des mots des constructions à objet unique avec un syntagme nominal ou un pronom — personnel, réfléchi ou réciproque.

(1) Français (Guentchéva & Rivière, 2008)

- a. Les parents lavent l'enfant.
ART parents laver.PRÉS ART.enfant
« Les parents lavent l'enfant. »
- b. Les parents le lavent.
ART parents PRON lavent.PRÉS
« Les parents le lavent. »
- c. Les parents se lavent.
ART parents RÉFL lavent.PRÉS
« Les parents se lavent. »

Notre enquête de données porte sur 156 langues qui ont la caractéristique du type pronominal codée dans le projet Universal Dependencies (UD v2.15 ; Zeman et al., 2024). Les syntagmes nominaux et pronominaux sont toujours codés dans UD, mais seul un sous-ensemble de langues code spécifiquement les pronoms personnels et autres types pronominaux. Les instances avec exactement un objet dans le syntagme verbal sont extraites. Les instances correspondantes sont ensuite catégorisées en groupes exclusifs de syntagmes nominaux (1a), pronoms personnels (1b ; excluant les réfléchis et réciproques), réfléchis (1c), et réciproques.

La Figure 1 montre la proportion d'ordre VO par opposition à l'ordre OV en pourcentage, quand l'objet appartient à NOM, pronom personnel (PRON), réfléchi (RÉFL), ou réciproque (RÉCP). Un point tracé à l'extrême droite signifie que toutes les instances de cette catégorie ont l'ordre VO. L'axe des y est trié par pourcentages décroissants de NOM en ordre VO (points rouges). La transition fluide de droite à gauche signifie une flexibilité générale de l'ordre des mots. Une asymétrie apparaît quand PRON apparaît souvent dans le domaine préverbal même pour les langues VO hautement cohérentes et rigides, de sorte que [V NP & Pn V] est commun, mais la situation miroir [NP V & V Pn] se produit beaucoup moins couramment dans les langues OV.

Nous identifions ensuite la préférence d'ordre des mots par une ligne de division arbitraire de 50% d'occurrence dans le corpus, et définissons la cohérence de l'ordre des mots comme le fait que NOM ait la même préférence que PRON, RÉFL, ou RÉCP. Le Tableau 1 résume le nombre de langues avec un ordre des mots cohérent dans les langues avec VO rigide (ordre VO NOM >

90%), VO non-rigide (ordre VO NOM > 50%), OV non-rigide (ordre VO NOM < 50%), et OV rigide (ordre VO NOM < 10%). Les langues OV rigides et non-rigides sont fortement cohérentes, les contreparties VO beaucoup moins. Cependant, la faible cohérence dans la classe VO non-rigide ne peut pas être expliquée uniquement par une préférence à placer PRON tôt dans la phrase, car les langues VO rigides ont près de 50% de chance de préférer avoir PRON du même côté que NOM. Ainsi l'asymétrie demeure. Quelques explications possibles basées sur l'efficacité sont discutées basées sur Hawkins (2014).

Tableau 1 : Proportion de langues ayant un ordre des mots cohérent entre objet nominal et pronom personnel (PRON), réfléchi (RÉFL), ou réciproque (RÉCP) respectivement

Classe	PRON	%	RÉFL	%	RÉCP	%
1 VO rigide	25/45	56%	14/19	74%	6/6	100%
2 VO non-rigide	6/37	16%	12/21	57%	3/3	100%
3 OV non-rigide	30/31	97%	12/16	75%	7/12	58%
4 OV rigide	40/43	93%	13/13	100%	4/5	80%
Total	101/156	65%	51/69	74%	20/26	77%

Références

- Dryer, M. S. (1992). The Greenbergian word order correlations. *Language*, 68(1), 81–138.
- Guentchéva, Z., & Rivière, N. (2008). Reciprocal and reflexive constructions in French. In *Reciprocal constructions* (pp. 561–607). John Benjamins Publishing Company.
- 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.
- Zeman, D., Nivre, J., Abrams, M., Ackermann, E., Aepli, N., Aghaei, H., ... Znotiņš, A. (2024). *Universal dependencies 2.15*. Retrieved from <http://hdl.handle.net/11234/1-5787> (LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University)

Qiujun Zhang (Umeå University)

*Revisit the connection between progressive aspect and motion event construal:
Evidence from Mandarin and Swedish*

Keywords: motion event construal; grammatical aspect; path; endpoint; eye-tracking

Abstract

The domain of motion has been a central topic for investigating Whorfian effects over the past two decades. The present study focuses on the temporal properties of motion events and tests whether grammatical aspect may influence the extent to which speakers attend to the goal of motion events. Previous research has found the cross-linguistic effect of grammatical aspect among speakers of Afrikaans, Arabic, Dutch, English, German, Spanish, and Swedish. Specifically, speakers of [-aspect] languages are more likely to include endpoints in linguistic descriptions of motion events and to allocate more attention to potential endpoints during verbalization (e.g., von Stutterheim et al., 2012). However, more recent studies failed to replicate the above cross-linguistic effects on endpoint behavior when a new language pair is investigated (Liao et al., 2020) or when a new experimental setting is introduced (Misersky et al., 2022). In light of the mixed findings, the study re-examines the relationship between progressive aspect and motion event cognition by exploring Mandarin and Swedish, a language pair that has been understudied in this area. Using a linguistic description task and a memory-based triads-matching task (Athanasopoulos and Bylund, 2013), Zhang's (2024) study establishes a baseline of endpoint behavior in [+aspect] Mandarin speakers. They mentioned event endpoints at a similar frequency to [+aspect] English speakers but significantly less often than [-aspect] Swedish and Afrikaans speakers (Bylund et al., 2013). Results from the categorization task also showed that Mandarin speakers, similar to English speakers, exhibited lower percentages of selecting reached endpoints as the criterion for categorizing goal-oriented motion events as compared to speakers of non-aspect languages.

This ongoing study builds on previous findings and further explores the factors beyond grammatical aspect that may affect the conceptualization of path in motion events. The path of motion is the core schema of motion events (Talmy, 2000), and different path elements such as trajectory, ground elements, as well as endpoints might all play a part in constituting speakers' mental representations of motion events. Moreover, according to cultural psychology, Easterners tend to think holistically and pay attention to context or background, while Westerners think analytically and focus on central objects (Miyamoto & Wilken, 2013). Considering the role of culture, culturally relevant stimuli depicting motion events are created with half set in Chinese contexts and the other half in Swedish contexts. Additionally, the stimuli are manipulated according to trajectory length: (i) a short trajectory with a highly evident but unreached endpoint; (ii) a long trajectory with a potential endpoint in the distance. During the verbalization task, participants' eye gaze towards various components of motion events will be tracked to examine their attention allocation at different stages of thinking-for-speaking. Mandarin speakers are predicted to distribute their attention not only to the focal areas such as the moving entity and ground, but also to the background less relevant to a motion event (e.g., the color of a walking person's clothes, a large forest by the roadside). As to Swedish speakers, they are predicted to focus on the moving entity and its movement towards a goal. Additionally, compared to Swedish

speakers, Mandarin speakers are more likely to focus on the ongoing phase and therefore mention an entity's location in the middle of the trajectory rather than its goal due to the frequent use of progressive markers.

Mots-clés : construction d'événements de mouvement ; aspect grammatical ; trajectoire ; point final ; oculométrie

Résumé

Le domaine du mouvement a été un sujet central pour étudier les effets *whorfiens* au cours des deux dernières décennies. La présente étude se concentre sur les propriétés temporelles des événements de mouvement et teste si l'aspect grammatical peut influencer la mesure dans laquelle les locuteurs portent attention au but des événements de mouvement. Des recherches antérieures ont trouvé l'effet translinguistique de l'aspect grammatical parmi les locuteurs d'afrikaans, d'arabe, de néerlandais, d'anglais, d'allemand, d'espagnol et de suédois. Spécifiquement, les locuteurs de langues aspectuelles sont plus susceptibles d'inclure des points finaux dans les descriptions linguistiques d'événements de mouvement et d'allouer plus d'attention aux points finaux potentiels pendant la verbalisation (par ex., von Stutterheim et al., 2012). Cependant, des études plus récentes ont échoué à répliquer les effets translinguistiques mentionnés ci-dessus sur le comportement de point final lorsqu'une nouvelle paire de langues est étudiée (Liao et al., 2020) ou lorsqu'un nouveau cadre expérimental est introduit (Misersky et al., 2022). À la lumière de ces résultats mitigés, l'étude réexamine la relation entre l'aspect progressif et la cognition d'événements de mouvement en explorant le mandarin et le suédois, une paire de langues qui a été sous-étudiée dans ce domaine. En utilisant une tâche de description linguistique et une tâche d'appariement de triades basée sur la mémoire (Athanasopoulos et Bylund, 2013), l'étude de Zhang (2024) établit une ligne de base du comportement de point final chez les locuteurs de mandarin [+aspect]. Ils ont mentionné les points finaux d'événements à une fréquence similaire à celle des locuteurs anglais [+aspect] mais significativement moins souvent que les locuteurs suédois [-aspect] et afrikaans (Bylund et al., 2013). Les résultats de la tâche de catégorisation ont également montré que les locuteurs de mandarin, similairement aux locuteurs anglais, présentaient des pourcentages plus faibles de sélection de points finaux atteints comme critère de catégorisation d'événements de mouvement orientés vers un but comparativement aux locuteurs de langues sans aspect.

Cette étude en cours s'appuie sur les résultats précédents et explore davantage les facteurs au-delà de l'aspect grammatical qui peuvent affecter la conceptualisation de la trajectoire dans les événements de mouvement. La trajectoire de mouvement est le schéma central des événements de mouvement (Talmy, 2000), et différents éléments de trajectoire tels que la trajectoire proprement dite, les éléments de fond, ainsi que les points finaux peuvent tous jouer un rôle dans la constitution des représentations mentales des événements de mouvement des locuteurs. De plus, selon la psychologie culturelle, les Orientaux tendent à penser de manière holistique et à porter attention au contexte ou à l'arrière-plan, tandis que les Occidentaux pensent de manière analytique et se concentrent sur les objets centraux (Miyamoto & Wilken, 2013). Considérant le rôle de la culture, des stimuli culturellement pertinents représentant des événements de mouvement sont créés avec la moitié située dans des contextes chinois et l'autre moitié dans des contextes anglais. De plus, les stimuli sont manipulés selon la longueur de trajectoire : (i) une

trajectoire courte avec un point final hautement évident mais non atteint ; (ii) une trajectoire longue avec un point final potentiel au loin. Pendant la tâche de verbalisation, le regard des participants vers divers composants des événements de mouvement sera suivi pour examiner leur allocation d'attention à différentes étapes du penser-pour-parler. Les locuteurs de mandarin sont prédits distribuer leur attention non seulement aux zones focales telles que l'entité en mouvement et le fond, mais aussi à l'arrière-plan moins pertinent pour un événement de mouvement (par ex., la couleur des vêtements d'une personne qui marche, une grande forêt au bord de la route). Quant aux locuteurs suédois, ils sont prédits se concentrer sur l'entité en mouvement et son mouvement vers un but. De plus, comparativement aux locuteurs suédois, les locuteurs de mandarin sont plus susceptibles de se concentrer sur la phase en cours et donc de mentionner la localisation d'une entité au milieu de la trajectoire plutôt que son but en raison de l'usage fréquent de marqueurs progressifs.

Références

- Athanasopoulos, P., & Bylund, E. (2013). Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science*, 37(2), 286–309. <https://doi.org/10.1111/cogs.12006>
- Bylund, E., Athanasopoulos, P., & Oostendorp, M. (2013). Motion event cognition and grammatical aspect: Evidence from Afrikaans. *Linguistics*, 51(5), 929–955. <https://doi.org/10.1515/ling-2013-0037>
- Liao, Y., Flecken, M., Dijkstra, K., & Zwaan, R. A. (2020). Going places in Dutch and Mandarin Chinese: Conceptualizing the path of motion cross-linguistically. *Language, Cognition and Neuroscience*, 35(4), 498–520. <https://doi.org/10.1080/23273798.2019.1679123>
- Misersky, J., Peeters, D., & Flecken, M. (2022). The potential of immersive virtual reality for the study of event perception. *Frontiers in Virtual Reality*, 3, 697934. <https://doi.org/10.3389/frvir.2022.69793>
- Miyamoto, Y., & Wilken, B. (2013). Cultural differences and their mechanisms. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 970–985). Oxford University Press.
- Talmy, L. (2000). *Toward a cognitive semantics*: Volume I. MIT Press.
- Von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., & Schmiedtová, B. (2012). How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analyses, eye tracking data, and memory performance. *Linguistics*, 50(4), 833–867. <https://doi.org/10.1515/ling-2012-0026>
- Zhang, Q. J. (2024). Revisiting the connection between progressive aspect and motion event cognition: Evidence from L1 Mandarin Chinese and L2 Swedish. *Lingua*, 311, 103827.

Maria Zielenbach (Vrije Universiteit Amsterdam)

How the Historical Comparative Method can reveal language history beyond genetic affiliation

Keywords: Historical Comparative Method; Language relatedness; Contact languages; Papuan languages; North Halmahera languages; Austronesian

Abstract

When linguists describe languages as ‘related,’ they typically mean genetic affiliation according to the Tree Model (i.e., descent from a common ancestor). Establishing this kind of descent is commonly thought to presuppose a successful application of the Historical Comparative Method (HCM, cf. Campbell 2013: 107). If the HCM cannot be applied successfully to a group of potentially related languages, this opens the door to debate (cf. Miceli & Round 2022). Contact languages (pidgins, creoles, ...), derived from multiple parents, are said to defy genetic classification in the first place (Thomason & Kaufman 1988: 3).

Taking West Makian [mq̄s] as an example, I will argue that a thorough application of the HCM can reveal crucial aspects of the history of a language beyond the binary assessment of ‘related’ vs. ‘unrelated’. Since the 1970s, West Makian has been regarded as a primary branch of the North Halmahera (NH) language family, based on similarities in the lexicon and lexicostatistic evaluations (Watusake (1976), Voorhoeve (1988), Grimes & Grimes (1984)). Applying the HCM, I found that only about 8% of the total West Makian vocabulary attested in Voorhoeve (1982) shows regular sound correspondences with the NH languages.²⁵ Some of these correspondences have unique realizations in West Makian (e.g. West Makian /f/ vs. Ternate, Tidore, Galela /h/, other languages ∅) and cannot be attributed to phonological integration of loan words. About 12% of the remaining lexicon appears to be of NH origin as well but does not show regular correspondences. This suggests at least two layers of NH vocabulary in West Makian: one inherited from an otherwise unattested NH language, displaying regular sound correspondences, and another acquired through borrowing. About half of the attested vocabulary is unique to West Makian and another 23% is of Austronesian origin. Grammatical items are likewise of NH or Austronesian origin, or unique to West Makian.

The regular sound correspondences found in West Makian could be interpreted as evidence of its genetic affiliation with the NH family. However, the HCM also revealed that inherited NH vocabulary in West Makian is relatively scarce, particularly in contrast to the 49% of vocabulary unique to the language —a fact that would be overlooked if West Makian were simply classified as an NH language. Based on these findings, I will argue instead that West Makian is a contact language that consists of multiple linguistic components, one being an otherwise unattested NH language, and another being a language isolate. I will introduce the term ‘genetic outlier’ for a language that only partly belongs to a language family. This new term stresses the importance of the other linguistic components dominating West Makian in comparison to the NH component without negating the existence of regular sound correspondences. The (partly successful) application of the HCM highlights its value in uncovering the intricate history of West Makian.

²⁵ These are very closely related and the HCM can be applied to them without any issues.

Mots-clés: méthode comparative historique ; parenté linguistique ; langues de contact ; langues papoues ; langues du Halmahera du Nord ; langues austronésiennes

Résumé

Lorsque les linguistes décrivent des langues comme « apparentées », ils font généralement référence à une filiation génétique selon le modèle de l'arbre généalogique (c'est-à-dire une descendance à partir d'un ancêtre commun). Établir ce type de filiation est généralement considéré comme présupposant une application réussie de la méthode comparative historique (MCH, cf. Campbell 2013 : 107). Si la MCH ne peut être appliquée avec succès à un groupe de langues potentiellement apparentées, cela ouvre la porte au débat (cf. Miceli et Round 2022). Les langues de contact (pidgins, créoles, etc.), issues de parents multiples, sont réputées échapper à toute classification génétique (Thomason et Kaufman 1988 : 3).

Prenant l'exemple du makien occidental [mq̄s], je soutiendrai qu'une application approfondie du MCH peut révéler des aspects cruciaux de l'histoire d'une langue, au-delà de l'évaluation binaire « apparenté » et « non apparenté ». Depuis les années 1970, le makien occidental est considéré comme une branche principale de la famille des langues du Halmahera du Nord (NH), en raison de similitudes lexicales et d'évaluations lexicostatistiques (Watusseke (1976), Voorhoeve (1988), Grimes et Grimes (1984)). En appliquant le MCH, j'ai constaté que seulement 8 % environ du vocabulaire makien occidental total attesté par Voorhoeve (1982) présente des correspondances phonétiques régulières avec les langues NH.²⁶ Certaines de ces correspondances présentent des réalisations uniques en makien occidental (par exemple, /f/ en makien occidental contre /h/ en ternate, tidore, galela, ø dans d'autres langues) et ne peuvent être attribuées à l'intégration phonologique de mots empruntés. Environ 12 % du lexique restant semble également d'origine NH, mais ne présente pas de correspondances phonétiques régulières. Cela suggère l'existence d'au moins deux couches de vocabulaire NH en makien occidental : l'une héritée d'une langue NH non attestée, présentant des correspondances phonétiques régulières, et l'autre acquise par emprunt. Environ la moitié du vocabulaire attesté est propre au makien occidental et 23 % est d'origine austronésienne. Les éléments grammaticaux sont également d'origine NH ou austronésienne, ou propres au makien occidental.

Les correspondances phonétiques régulières trouvées en makien occidental pourraient être interprétées comme une preuve de son affiliation génétique à la famille NH. Cependant, le MCH a également révélé que le vocabulaire NH hérité en makien occidental est relativement rare, notamment en comparaison des 49 % de vocabulaire propre à cette langue – un fait qui serait négligé si le makien occidental était simplement classé comme une langue NH.

Sur la base de ces résultats, je soutiendrai plutôt que le makien occidental est une langue de contact composée de multiples composantes linguistiques, l'une étant une langue NH non attestée par ailleurs, et l'autre étant une langue isolée. J'introduirai le terme « aberration génétique » pour une langue n'appartenant qu'en partie à une famille linguistique. Ce nouveau terme souligne l'importance des autres composantes linguistiques dominantes en makien occidental par rapport à la composante NH, sans pour autant nier l'existence de correspondances

²⁶ Ces deux concepts sont étroitement liés et le modèle MCH peut leur être appliqué sans aucun problème.

phonétiques régulières. L'application (en partie réussie) du MCH souligne son intérêt pour la découverte de l'histoire complexe du makien occidental.

Références

- Campbell, Lyle. 2013 3 . *Historical linguistics: An Introduction*. Edinburgh: Edinburgh Univ. Press.
- Grimes, Charles E. & Barbara D. Grimes. 1984. Languages of the North Moluccas. In E. K. M. Masinam-bow (ed.), *Maluku dan Irian Jaya* (Buletin LEKNAS: Terbitan Khusus, III, 1), 35–63. Jakarta: Lembaga Ekonomi dan Kemasyarakatan Nasional, Lembaga Ilmu Pengetahuan Indonesia.
- Miceli, Luisa & Erich Round. 2022. Where have all the sound changes gone? Examining the scarcity of evidence for regular sound change in Australian languages. *Linguistics Vanguard* 8(s5). 509–518.
- Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- Voorhoeve, C. L. (ed.) 1982. *The Makian languages and their neighbours* (Pacific Linguistics, Series D, 46). Canberra: Dept. of Linguistics, Research School of Pacific Studies, Australian National University.
- Voorhoeve, C. L. 1988. The languages of the North Halmaheran stock. *Papers in New Guinea linguistics* 26. 181–209.
- Watuseke, F. S. 1976. West Makian, a Language of the North-Halmahéra Group of the West-Irian Phylum. *Anthropological Linguistics* 18(6). 274–285.