

STUDIA UNIVERSITATIS

Babeş-Bolyai

HÉRITAGE ET
TRANSMISSION
DANS LES
LANGUES ROMANES



PHILOLOGIA

4/2025

STUDIA UNIVERSITATIS BABEŞ-BOLYAI

PHILOLOGIA

Volume 70 (LXX), 4/2025, December 2025

ISSN (print): 1220-0484; ISSN (online): 2065-9652; ISSN-L: 1220-0484
© STUDIA UBB PHILOLOGIA. Published by Babeş-Bolyai University

STUDIA UNIVERSITATIS BABEŞ-BOLYAI PHILOLOGIA

EDITORIAL OFFICE: 31 Horea St, Cluj-Napoca, Romania, Phone: +40 264 405300

ADVISORY BOARD:

Professor Rosalind Margaret Ballaster, University of Oxford, UK
Professor Bruno Blanckeman, Sorbonne Nouvelle University, France
Professor Emerita Maria Helena Araújo Carreira, University of Paris 8, France
Professor Cora Dietl, Justus Liebig University, Germany
Professor Isabel Margarida Duarte, University of Porto, Portugal
Professor Thomas Johnen, University of Zwickau, Germany
Professor Emeritus Declan Kiberd, University Notre Dame, USA
Professor Katalin É. Kiss, Pázmány Péter Catholic University, Hungary
Professor Emerita Judith Yaross Lee, Ohio University, USA
Professor Patrick McGuinness, University of Oxford, UK
Professor Christian Moraru, University of North Carolina, Greensboro, USA
Professor Eve Patten, Trinity College Dublin, Ireland
Professor Kerstin Schoor, Europa University Viadrina, Germany
Professor Silvia-Maria Chireac, University of Valencia, Spain
Associate Professor Jorge Figueroa Dorrego, University of Vigo, Spain
Associate Professor Elisabet Arnó Macià, Polytechnic University of Catalunya, Spain
Associate Professor Keith B. Mitchell, The University of Massachusetts Lowell, USA
Associate Professor John Style, Rovira i Virgili University, Spain
Associate Professor David Vichnar, Caroline University of Prague, Czech Republic
Associate Professor Annalisa Volpone, University of Perugia, Italy
Associate Professor Serenella Zanotti, University Roma Tre, Italy
Assistant Professor Margarida Vale de Gato, University of Lisbon, Portugal
Lecturer Emilia David, University of Pisa, Italy
Lecturer Ilaria Natali, University of Florence, Italy
Assistant Professor Nikolina Dobрева, Middlebury College, USA

EDITOR-IN-CHIEF:

Associate Professor Rareş Moldovan, Babeş-Bolyai University, Romania

EXECUTIVE EDITOR:

Associate Professor Carmen Borbely, Babeş-Bolyai University, Romania

EDITORIAL BOARD MEMBERS:

Professor Anna Branach-Kallas, Nicolaus Copernicus University, Toruń, Poland
Associate Professor Fiorenzo Fantaccini, University of Florence, Italy
Associate Professor Ágnes Zsófia Kovács, University of Szeged, Hungary
Associate Professor Daniela Vladu, Babeş-Bolyai University, Romania
Lecturer Andrei Lazar, Babeş-Bolyai University, Romania
Lecturer Veronica Manole, Babeş-Bolyai University, Romania
Lecturer Boglárka-Noémi Németh, Babeş-Bolyai University, Romania
Lecturer Elena Păcurar, Babeş-Bolyai University, Romania
Lecturer Anamaria Curea, Babeş-Bolyai University, Romania
Lecturer Petronia Popa Petrar, Babeş-Bolyai University, Romania
Lecturer Alexandru-Ioan Ciorogar, Babeş-Bolyai University, Romania
Assistant Professor Ema Vyroubalova, Trinity College Dublin, Ireland
Assistant Professor Mihaela Buzec, Babeş-Bolyai University, Romania

ASSISTANT EDITORS:

Assistant Professor Paul Mihai Paraschiv, Babeş-Bolyai University, Romania
MA student Robert-Giulian Andreescu, Babeş-Bolyai University, Romania

Studia UBB Philologia is a Category A scientific journal in the rating provided by Romania's National Council for Scientific Research.

As of 2017 *Studia UBB Philologia* has been selected for indexing in Clarivate Analytics' Emerging Sources Citation Index for the Arts and Humanities.

YEAR
MONTH
ISSUE

Volume 70 (LXX) 2025
DECEMBER
4

PUBLISHED ONLINE: 2025-12-12
PUBLISHED PRINT: 2025-12-30
ISSUE DOI:10.24193/subbphilo.2025.4

STUDIA UNIVERSITATIS BABEȘ-BOLYAI PHILOLOGIA

4

SPECIAL ISSUE: Héritage et transmission dans les langues romanes
*Guest Editors: Francisco Calvo del Olmo, Veronica Manole, Sanda Moraru, and
Anamaria Curea*

SUMAR – CONTENTS – INDICE – INHALT

Introduction – Introducción – Introduzione

FRANCISCO CALVO DEL OLMO, VERONICA MANOLE, SANDA MORARU, 9
and ANAMARIA CUREA

STUDIES

Per una storia dei nomi delle lingue romanze in italiano
For a History of the Names of Romance Languages in Italian
Pentru o istorie a numelor limbilor romanice în italiană

PAOLO D'ACHILLE 21

I latinismi integrali nella comunicazione medica scritta mediata tecnicamente:
un caso di studio

*Non-adapted Latinisms in Technically Mediated Written Medical Communication: A
Case Study*

Latinisme neadaptate în comunicarea medicală scrisă mediată tehnic: un studiu de caz

KEVIN DE VECCHIS 37

L'héritage lexical latin commun à l'ancien français et au roumain : mots commençant par les lettres A, B et C

The Latin Lexical Heritage Common to Old French and Romanian: Words Beginning with the Letters A, B and C

Moștenirea lexicală latinească comună vechii franceze și românei: cuvinte care încep cu literele A, B și C

ADRIAN CHIRCU

55

Sur les traces du lat. HABERE dans la morphosyntaxe verbale du roumain : polygrammaticalisation, universaux linguistiques et hypothèses étymologiques

On the Traces of the Lat. HABERE in the Verbal Morphosyntax of Romanian: Polygrammaticalization, Linguistic Universals, and Etymological Hypotheses

Pe urmele lui Lat. HABERE în morfosintaxa verbală a limbii române: poligramaticalizare, universalii lingvistice și ipoteze etimologice

CECILIA MIHAELA POPESCU

65

Herencia y transmisión de términos gastronómicos en Hispanoamérica

Inheritance and Transmission of Gastronomic Terms in Latin America

Moștenirea și transmiterea termenilor gastronomici în America Latină

OLIVIA NARCISA PETRESCU

83

L'heritage de la labiovelaire dans les langues romanes : le cas du lat. coquere 'cuire'

The Heritage of the Labiovelar in Romance Languages: The Case of the Lat. coquere 'to cook'

Moștenirea labiovelarei în limbile romanice: cazul lat. coquere 'a coace'

SIMONA GEORGESCU, THEODOR GEORGESCU

107

Le roumain langue d'héritage : le cas des étudiants francophones en médecine. Orientations et approches

Romanian as a Heritage Language: The Case of Francophone Medical Students. Directions and Approaches

Limba română moștenită. Cazul studenților mediciniști francofoni. Direcții și abordări

ALINA ANDREICA, NORA-SABINA NEAMȚ

121

Pérdida y sustitución de arabismos en español: recipientes y muebles de la cocina

Loss and Substitution of Arabisms in Spanish: Kitchen Containers and Furniture

Pierderea și înlocuirea arabismelor în spaniolă: recipiente și mobilier din bucătărie

MIHAI ENĂCHESCU

143

A Corpus-Based Analysis of the Event-Perception Constructions in Romanian

O analiză de corpus a construcțiilor de percepție evenimentială în limba română

MIHAELA GHEORGHE

161

Trabajo y travail: evolución semántico-discursiva en español y francés <i>Trabajo and Travail: A Semantic-Discursive Evolution in Spanish and French</i> <i>Trabajo și travail: evoluție semantico-discursivă în spaniolă și franceză</i>	181
TIMEA TOCALACHIS	

MISCELLANEA

Cultural Sensitivity in AI Translation of Chinese Language: Evaluating AI's Capability to Grasp Culturally Rich Texts <i>Sensibilitate culturală în traducerea asistată de IA a limbii chineze: evaluarea capacității de redare a textelor cu densitate culturală</i>	197
IULIA ELENA CÎNDEA, JIONG WANG	
Polarity in the Verbal Domain. Preliminary Results and Considerations <i>Polaritatea în domeniul verbal. Rezultate și considerații preliminare</i>	219
HAJNALKA DIMÉNY	
Literature's Response to Information Overload in Don DeLillo's <i>White Noise</i> <i>Răspunsul literaturii la supraîncărcarea informațională în White Noise de Don Delillo</i>	247
CRISTINA MIRELA NICOLAESCU	
Metaphorical Framing of Obesity in Online News Discourse <i>Reprezentări metaforice ale obezității în discursul știrilor online</i>	267
ANIȘOARA POP	

REVIEWS

Ágnes Zsófia Kovács, <i>The Memory of Architecture in Edith Wharton's Travel Writings</i> , New York: Routledge, 2025, 226 p.	283
DANIEL DARVAY	
Bianca-Livia Bartoș, <i>Eugène contre Ceașescu. Le cas de la Lettre à mon dictateur</i> , Paris : Hermann, 2024, 108 p.	287
ANDREEA GÎNGA	
Adrian Tudurachi (Ed.), <i>Dicționarul cronologic al romanului românesc de la origini până în 2000</i> , revised second edition, Cluj-Napoca: Presa Universitară Clujeană, 2023, 770p + 757p (2 volumes).	289
IRINA GORGAN	
Alex Ciorogar, <i>Ascensiunea autorului în epoca globalizării digitale</i> , Cluj-Napoca: Presa Universitară Clujeană, 2025, 418 p.	293
ANDREEA-DIANA HOSU	

Bárbara Marqueta Gracia y Fernando López García (coords.), *La divulgación lingüística: formación, información y desinformación*, 2025, Prensas de la Universidad de Zaragoza, 278 p.

GONZALO LLAMEDO-PANDIELLA 297

G. Herrera Peña, *Sinfonía de acentos. Rasgos de la variante guatemalteca del español*, Ciudad de Guatemala: Editorial La Pepita, 2025, 93 p.

JOSÉ LUIS RAMÍREZ LUENGO 301

Mary A. Kato, Ana Maria Martins, Jairo Nunes, *Português Brasileiro e Português Europeu: Sintaxe Comparada*, São Paulo: Contexto, 2023, 383 p.

VERONICA MANOLE 305

Anca Pârvulescu, *Face and Form: Physiognomy in Literary Modernism*, Cambridge: Cambridge University Press, 2025, 196 p.

PETRONIA POPA-PETRAR 309

Corina Croitoru, *Fronturi interioare. Poezia românească a celor două războaie mondiale*, Cluj-Napoca: Presa Universitară Clujeană, 2024, 273 p.

AMELIA PRECUP 313

INTRODUCTION – INTRODUCCIÓN – INTRODUZIONE

Le présent numéro de la revue *Studia Universitatis Babeş-Bolyai Philologia* est consacré à l'héritage et à la transmission au sein de la famille néolatine, depuis le latin vulgaire et la formation des langues romanes jusqu'à nos jours. Les dix articles qui le composent forment une mosaïque variée, tant par leurs objets d'étude que par leurs approches théoriques et méthodologiques, couvrant la linguistique historique, la linguistique de corpus, la lexicographie, la didactique et la dialectologie. En outre, les contributions sont rédigées en français (quatre articles), en espagnol (trois), en italien (deux) et en anglais (un) ; l'ensemble enrichit la nature polyphonique de la publication et l'inscrit dans la tradition plurilingue et intercompréhensive qui a caractérisé la linguistique romane tout au long de son parcours.

Le volume s'ouvre par un article de **Paolo D'Achille** consacré aux appellations que les principales langues romanes ont reçues en italien au fil des siècles. Cette question lui permet de discuter la manière et le moment où les différentes langues issues du latin ont été perçues, identifiées et dénommées. Le texte accorde une attention particulière aux glottonymes renvoyant aux peuples qui habitaient les diverses régions avant la latinisation, puis dresse, dans un second temps, la liste des différents vocables employés à des époques variées pour désigner la langue italienne, avant que cette dernière dénomination ne s'impose. Par une analyse de facture lexicale, l'étude nous donne à comprendre l'autoconscience et la perception métalinguistique des communautés romanophones.

Toujours dans l'aire italique, **Kevin De Vecchis** centre sa contribution sur le rapport entre le latin et les langues romanes, en mettant l'accent sur le domaine du lexique. Il propose une étude de cas consacrée aux latinismes non adaptés relevant du langage médical et à leur emploi dans la communication médiée à l'aide de la technologie entre médecins et patient·e·s. Il rappelle d'abord le rôle du latin dans la constitution du lexique italien et son importance dans les langages de spécialité, puis présente une méthodologie fondée sur un corpus de latinismes non adaptés. L'examen de ces mots, tant du point de vue morphologique que sémantique, prend en considération le genre, la catégorie grammaticale, les sens assumés en italien et leur typologie, tout en vérifiant leur présence dans la



lexicographie italienne. Ce cadre lui permet de formuler des réflexions pertinentes sur le statut du latin au sein du langage médical – y compris dans son rapport à l'anglais – et au sein du langage commun.

En passant du deuxième au troisième article, on change de langue d'étude : **Adrian Chircu** mène une étude comparative du lexique latin commun au français ancien et au roumain. L'auteur analyse une série de lexèmes commençant par les lettres A, B et C qui faisaient partie du vocabulaire fondamental du français ancien. Avec le temps, ces mots se sont réduits à des formes dialectales, ou ont même disparu, pour des raisons parfois difficiles à démêler, étant remplacés par des synonymes, des créations lexicales internes ou des emprunts. En rapprochant ces vocables du français ancien de leurs correspondants roumains, maintenus jusqu'à aujourd'hui, l'auteur montre combien ces deux langues romanes étaient beaucoup plus proches dans le passé et met en évidence le caractère plus conservateur du roumain. L'enquête reprend et actualise ainsi des procédés et des concepts de la tradition romaniste.

La quatrième contribution, signée par **Cecilia Mihaela Popescu**, partage avec les précédentes l'intérêt pour les études historiques ainsi que pour la relation entre le latin et les langues romanes actuelles. L'objectif est ici de retracer les empreintes du verbe HABERE dans la morphosyntaxe verbale du roumain. L'auteure considère que l'évolution s'est opérée à la fois « verticalement », par l'héritage du latin, et « horizontalement », par des innovations et des influences partagées au fil des siècles. Dans cette perspective, elle met l'accent sur le processus de polygrammaticalisation du verbe latin mentionné, selon différentes fonctions et dans divers paradigmes du verbe roumain. Pris dans leur ensemble, ces processus rendent compte de la manière dont ces morphèmes verbaux se rapportent aux universaux cognitifs et linguistiques. L'auteure fournit, en outre, des indices probants quant à l'origine de l'auxiliaire du conditionnel en roumain.

Avec l'article d'**Olivia Petrescu**, on passe du français comme langue de rédaction à l'espagnol, et de la *Romania continua* européenne aux vastes territoires de la *Romania nova*. Cette contribution s'intéresse au domaine gastronomique latino-américain sous un angle lexical, étymologique et culturel. L'auteure part de l'idée que la gastronomie est un art et une forme de communication culturelle : dans l'aire hispanophone, des ingrédients propres aux cultures précolombiennes se mêlent à ceux apportés par la colonisation espagnole et, plus tard, par de nouvelles vagues d'immigrants. Son enquête propose des classifications et des exemples du lexique gastronomique de cultures telles que la mexicaine, l'argentine, la péruvienne, la colombienne et la vénézuélienne, selon les termes hérités ou transmis qui renvoient principalement aux ingrédients, aux modes de préparation et aux dénominations des plats. Il en ressort une terminologie d'une grande variété et des combinaisons culturelles susceptibles de désigner un même concept.

Le lexique gastronomique hispano-américain constitue ainsi un témoignage des traditions, mais aussi des innovations, qui en font un objet d'attention privilégié en tant que modèle de métissage culturel.

Sans quitter entièrement le domaine culinaire, **Simona Georgescu** et **Theodor Georgescu** examinent l'héritage de la labio-vélaire latine dans les langues romanes à partir du verbe *coquere*, lequel aurait présenté en protoroman deux variantes : COQUERE et COCERE. Selon les auteurs, le français ancien serait la seule langue à avoir hérité des deux formes, assorties d'une distinction sémantique et d'une différenciation phonétique. Autrement dit, en discutant l'héritage de la consonne labio-vélaire /kw/ dans les langues romanes, l'étude envisage la possibilité qu'une différenciation sémantique s'accompagne d'une bipartition phonétique. Leur analyse minutieuse les conduit cependant à conclure que, dans ce cas, la reconstruction de deux prototypes n'est pas nécessaire : une forme unique, dépourvue de labio-vélaire, suffit à expliquer les descendants romans et leurs deux sens principaux.

Des reconstructions étymologiques, l'on passe à la réalité actuelle des langues romanes et de leurs locuteurs, puisque **Alina Andreica** et **Nora-Sabina Neamț** proposent une analyse sociolinguistique de la compétence orale en roumain comme langue d'héritage chez des étudiant-e-s francophones de deuxième génération qui retournent temporairement en Roumanie pour suivre des études supérieures. Adoptant pour cadre théorique le plurilinguisme en contexte de mobilité académique internationale, les auteures examinent tant des productions préparées que des interactions spontanées d'un groupe de participants, en identifiant des phénomènes tels que l'interférence linguistique, l'instabilité grammaticale et des traits phonétiques spécifiques, et en évaluant leur compétence plurilingue et de médiation. Leurs résultats révèlent des différences de niveau de compétence associées à l'exposition formelle au roumain, aux contextes d'usage et à l'implication familiale, soulignant le lien entre ces facteurs et l'acquisition de compétences communicatives dans une nouvelle langue.

Les études lexicales réapparaissent comme l'un des fils rouges de ce numéro avec l'article de **Mihai Enăchescu**. L'auteur aborde un thème d'histoire de la langue espagnole, en se concentrant sur le remplacement de certains arabismes au cours des derniers siècles du Moyen Âge. Il dresse l'inventaire de seize termes d'origine arabe désignant des récipients et des meubles utilisés dans les cuisines médiévales, et met en évidence l'évolution et l'histoire de ces vocables sur l'axe diachronique du castillan : certaines formes ont été conservées, d'autres ont disparu au profit de nouveaux termes, le plus souvent issus du latin.

L'étude de **Mihaela Gheorghe**, la seule rédigée en anglais, propose une analyse des Constructions de Perception d'Événements (CPE) en roumain, en centrant la recherche sur l'interaction entre le type de verbe de perception, la

classe sémantico-aspectuelle du prédicat et le choix du complément propositionnel. S'appuyant sur des données du corpus RoTenTen2021, elle montre que la distribution des types de compléments est affectée à la fois par le domaine sensoriel (par exemple, événements visuels, auditifs ou bimodaux) et par les propriétés aspectuelles. Ces éléments lui permettent d'identifier et de classer certains schémas et tendances propres à ces constructions.

Dans le dernier article, signé par **Timea Tocalachis**, les lecteurs retrouvent une autre étude d'étymologie comparée, cette fois consacrée aux changements sémantiques du terme « trabajo » en espagnol, mis en contraste avec son cognat français « travail ». Les deux langues remontent à l'étymon du latin vulgaire TRIPALIUM, instrument de torture. Le terme fut d'abord associé au registre de la souffrance et de la contrainte, avant de subir des processus de réinterprétation sémantique conduisant à désigner l'activité productive et l'effort humain dans divers domaines de la vie sociale. Selon l'exposé de l'auteure, l'évolution de ces deux mots ne reflète pas seulement la modification de leur signification sur l'axe diachronique ; elle témoigne aussi de la transformation des principes et de la vision de la société dans son ensemble quant à la valeur et au sens du travail dans les mondes hispanophone et francophone.

Le numéro continue avec la section *Varia* et se clôt sur une brève série de comptes rendus présentant quelques publications récentes consacrées à la linguistique appliquée aux langues romanes. La lecture de l'ensemble offre une vision plurielle et actualisée de la recherche dédiée à la famille néolatine et met en lumière l'actualité de la linguistique romane, carrefour d'études, de méthodes et de voix.

El presente número de la revista *Studia Universitatis Babeş-Bolyai Philologia* está consagrado a la herencia y la transmisión al interno de la familia neolatina, desde el latín vulgar y la formación de las lenguas románicas hasta la actualidad. Los diez artículos que lo integran componen un variado mosaico tanto por sus temas de estudio como por sus abordajes teóricos y metodológicos, que abarcan la lingüística histórica, la lingüística de corpus, la lexicografía, la didáctica y la dialectología. Además, las contribuciones están redactadas en francés (cuatro artículos), español (tres artículos), italiano (dos) así como en inglés (uno); todo ello enriquece la naturaleza polifónica de la publicación y la inscribe en la tradición plurilingüe e intercomprensiva que ha caracterizado a la lingüística románica a lo largo de toda su trayectoria.

Así pues, el volumen se abre con un artículo de **Paolo D'Achille** dedicado a los nombres que las principales lenguas románicas han recibido a lo largo de los siglos en italiano. Esta cuestión le permite debatir cómo y cuándo los diferentes idiomas derivados del latín fueron percibidos, identificados y denominados. El texto otorga especial atención a los glotónimos referidos a los pueblos que habitaban las diversas regiones antes de la latinización para, en un segundo momento, listar los variados vocablos que se utilizaron en épocas distintas para designar la lengua italiana antes de que se impusiera este último término. Es pues un trabajo que, mediante un análisis de cuño lexical, nos permite comprender la autoconciencia y la percepción metalingüística de las comunidades romanófonas.

Continuando en ámbito itálico, **Kevin De Vecchis** centra su trabajo en la relación entre el latín y las lenguas románicas, poniendo el foco en el campo del léxico para hacer un estudio de caso dedicado a los latinismos no adaptados pertenecientes al lenguaje médico y su uso en la comunicación mediada técnicamente entre doctores y pacientes. Así, en primer lugar, explica el papel del latín en la formación del léxico italiano y su importancia en los lenguajes con fines específicos para, acto seguido, presentar una metodología basada en un corpus de latinismos no adaptados. El examen de tales vocablos, tanto desde el punto de vista morfológico como semántico, toma en consideración el género, la categoría gramatical, los significados asumidos en italiano además de verificar su presencia en la lexicografía italiana. Todo ello le permite formular algunas reflexiones relevantes sobre el estatus del latín dentro del lenguaje médico, inclusive en relación con el inglés, y el lenguaje común.

Pasando del segundo al tercer artículo cambiamos de lengua de estudio ya que **Adrian Chircu** lleva a cabo un estudio comparado del léxico latino común al francés antiguo y al rumano. El autor analiza una serie de lexemas, comenzados por las letras A, B y C, que formaban parte del vocabulario fundamental en el francés antiguo. Con el paso del tiempo, estas palabras se redujeron a voces dialectales cuando no desaparecieron siendo sustituidas por sinónimos, creaciones léxicas internas o préstamos por razones que, a veces, resulta espinoso vislumbrar. Al relacionar estos vocablos del francés antiguo con sus correspondientes rumanos, mantenidos hasta la actualidad, el autor ilustra cómo las dos lenguas románicas estaban mucho más próximas en el pasado, al tiempo que pone de manifiesto el carácter más conservador del rumano. Una investigación que retoma y actualiza procedimientos y conceptos de la tradición romanista.

La cuarta contribución, firmada por **Cecilia Mihaela Popescu**, comparte con las anteriores el interés por los estudios históricos y por la relación entre el latín y las lenguas románicas actuales. En este caso, el objetivo es trazar las huellas de verbo HABERE en la morfosintaxis verbal de la lengua rumana. Para ello, toma en consideración que la evolución se produjo tanto «verticalmente»,

a través de la herencia tomada del latín, como «horizontalmente», a través de innovaciones e influencias compartidas a lo largo de los siglos. Asumiendo esta perspectiva, pone el foco en el proceso de poligramaticalización del mencionado verbo latino con diferentes funciones y en distintos paradigmas del verbo rumano. Tales procesos, considerados en su conjunto, dan buena cuenta de la manera en que dichos morfemas verbales se relacionan con los universales cognitivos y lingüísticos. Además, la autora proporciona indicios conclusivos sobre el origen del auxiliar del modo condicional en rumano.

Con el artículo de **Olivia Petrescu** saltamos del francés como lengua de redacción al español y de la *Romania Continua* de Europa a los vastos territorios de la *Romania Nova*. Esta contribución se interesa por el ámbito gastronómico latinoamericano desde una perspectiva léxica, etimológica y cultural. Así, la autora parte de la idea de que la gastronomía es un arte y una forma de comunicación cultural, en la que, en el ámbito hispanohablante, se fusionan ingredientes específicos de las culturas precolombinas con los traídos por la colonización española y, posteriormente, por nuevas levadas de inmigrantes. Su investigación incluye clasificaciones y ejemplos del léxico gastronómico de culturas como la mexicana, la argentina, la peruana, la colombiana y la venezolana, en función de los términos que se han heredado o transmitido y que se refieren principalmente a los ingredientes, la forma de preparar los platos y su denominación. Todo ello pone de manifiesto una terminología extremadamente variada y combinaciones culturales que pueden definir un mismo concepto. De este modo, el léxico gastronómico hispanoamericano constituye una prueba del reflejo de las tradiciones, pero también de las innovaciones, lo que lo hace objeto de especial cuidado y atención como modelo de mestizaje cultural.

Sin abandonar completamente el ámbito culinario, **Simona Georgescu** y **Theodor Georgescu** examinan la herencia de la labiovelar latina en las lenguas románicas mediante el análisis del verbo *coquere*; este tendría en protorromance dos variantes: COQUERE y COCERE. Según la exposición de los autores, el francés antiguo es la única lengua que heredó ambas formas, con una distinción semántica que estaría acompañada por una diferenciación fonética. En otras palabras, al debatir la herencia de la consonante labiovelar /kw/ en las lenguas románicas, el estudio analiza la posibilidad de que una diferenciación semántica pueda ir acompañada de una bipartición fonética. Sin embargo, la conclusión que presentan tras un minucioso análisis es que, en este caso, no es necesaria la reconstrucción de dos prototipos, sino que una única forma, sin la labiovelar, puede explicar los descendientes románicos y sus dos significados principales.

De las reconstrucciones etimológicas, pasamos a la realidad actual de las lenguas románicas y de sus usuarios ya que **Alina Andreica** y **Nora-Sabina Neamț** desarrollan un análisis sociolingüístico de la competencia oral en rumano

como lengua de herencia en estudiantes francófonos de segunda generación que regresan temporalmente a Rumanía para cursar estudios superiores. Asumiendo como perspectiva teórica el plurilingüismo en contextos de movilidad académica internacional, las autoras examinan tanto producciones preparadas como las interacciones espontáneas de un grupo de participantes, identificando fenómenos tales como interferencia lingüística, inestabilidad gramatical y rasgos fonéticos específicos, además de evaluar de forma general su competencia plurilingüe y de mediación. Sus resultados revelan diferencias en el nivel de competencia lingüística asociada a la exposición formal al rumano, los contextos de uso y la implicación familiar, subrayando la relación entre estos factores y la adquisición de competencias comunicativas en una nueva lengua.

Los estudios lexicales vuelven a aparecer como uno de los hilos rojos de este número con el artículo de **Mihai Enăchescu**. El autor aborda un tema de la historia de la lengua española, centrándose en la sustitución de algunos arabismos durante las últimas centurias de la Edad Media. En su investigación, inventaría dieciséis términos de origen árabe que hacían referencia a recipientes y muebles utilizados en las cocinas medievales y destaca la evolución y la historia de estos vocablos en el eje sincrónico del castellano que condujo a la conservación de algunas y a la desaparición de otras sustituidas por nuevos términos, sobre todo procedentes del latín.

El estudio de **Mihaela Gheorghe**, el único redactado en inglés, propone un análisis de las Construcciones de Percepción de Eventos (CPE) en rumano, centrando el foco de la investigación en la interacción entre el tipo de verbo de percepción, la clase semántico-aspectual del predicado y la elección del complemento oracional. Para ello, se apoya en datos del corpus RoTenTen2021, con lo que logra mostrar que la distribución de los tipos de complementos se ve afectada tanto por la prominencia sensorial (tales como eventos visuales, auditivos o de modalidad dual), como por las propiedades aspectuales. Estos elementos le permiten llegar a una cierta clasificación de patrones y tendencias en dichas construcciones.

En el último artículo, firmado por **Timea Tocalachis**, los lectores encuentran de nuevo un estudio etimológico comparado; en este caso, dedicado a los cambios semánticos del término “trabajo” en español en contraste con su cognado francés “travail”. Ambas lenguas parten del étimo del latín vulgar, donde TRIPALIUM se refería a un instrumento de tortura. Por ello, el término estuvo inicialmente anclado al ámbito del sufrimiento y la coacción, desde donde experimentó procesos de resignación hasta llegar a designar la actividad productiva y el esfuerzo humano en diversos ámbitos de la vida social. De acuerdo con la exposición de la autora, la evolución de los dos términos no solo refleja la alteración de su significado en el eje diacrónico, sino también la transformación de los principios

y la visión de la sociedad en su conjunto con respecto al valor y el significado del trabajo en el ámbito hispanohablante y francófono.

El número continúa con la sección *Varia* y se cierra con una breve serie de reseñas consagradas a presentar algunas publicaciones recientes cuyo tema común es la lingüística aplicada a las lenguas románicas. La lectura de la revista en su conjunto ofrece una visión plural y actualizada de la investigación dedicada a la familia neolatina y pone de manifiesto la vigencia de la lingüística románica, como encrucijada de estudios, métodos y voces.

Il presente numero della rivista *Studia Universitatis Babeş-Bolyai* è dedicato allo studio dell'eredità e della trasmissione, all'interno della famiglia neolatina, dal latino volgare e dalla formazione delle lingue romanze fino ad oggi. I dieci articoli che lo compongono costituiscono un variegato mosaico, tanto per i loro oggetti di studio quanto per i loro approcci teorici e metodologici, i quali spaziano dalla linguistica storica alla linguistica di corpus, dalla lessicografia alla didattica e alla dialettologia. Inoltre, i contributi sono redatti in francese (quattro articoli), spagnolo (tre), italiano (due) e inglese (uno); il tutto arricchisce la natura polifonica della pubblicazione, annoverandola all'interno della tradizione plurilingue e intercomprensiva che ha caratterizzato la linguistica romanza lungo l'intero suo percorso.

Il volume si apre con un articolo di **Paolo D'Achille** dedicato ai nomi che le principali lingue romanze hanno ricevuto nel corso dei secoli in italiano. Tale questione consente all'autore di discutere come e quando le diverse lingue derivate dal latino siano state percepite, identificate e denominate. Il testo presta particolare attenzione ai glottonimi riferiti ai popoli che abitavano le varie regioni prima della latinizzazione, per poi elencare, in un secondo momento, i vocaboli usati in epoche diverse per designare la lingua italiana, prima che prevalesse la denominazione attuale. Si tratta dunque di un lavoro che, attraverso un'analisi di tipo lessicale, permette di comprendere l'autocoscienza e la percezione metalinguistica delle comunità romanofone.

Restando in ambito italoico, **Kevin De Vecchis** concentra il proprio studio sul rapporto tra latino e lingue romanze, soffermandosi sul campo del lessico per realizzare un caso di studio dedicato ai latinismi non adattati appartenenti al linguaggio medico e al loro impiego nella comunicazione mediata tecnologicamente tra medici e pazienti. L'autore illustra dapprima il ruolo del latino nella formazione del lessico italiano e la sua importanza nei linguaggi settoriali, per poi presentare una metodologia fondata su un corpus di latinismi non adattati. L'analisi di tali vocaboli, dal punto di vista sia morfologico sia semantico, prende

in esame il genere, la categoria grammaticale, i significati assunti in italiano e la tipologia, oltre a verificarne la presenza nella lessicografia italiana. Tutto ciò consente di formulare alcune riflessioni rilevanti sullo statuto del latino nel linguaggio medico, anche in rapporto all'inglese e al linguaggio comune.

Dal secondo al terzo articolo cambia la lingua d'indagine, poiché **Adrian Chircu** conduce uno studio comparato del lessico latino comune all'antico francese e al rumeno. L'autore analizza una serie di lessemi — con iniziali A, B e C — che facevano parte del vocabolario fondamentale dell'antico francese. Col tempo, tali parole si ridussero a forme dialettali o scomparvero, sostituite da sinonimi, creazioni lessicali interne o prestiti, per ragioni talvolta difficili da individuare. Mettendo in relazione questi vocaboli con i corrispondenti rumeni, conservatisi fino ad oggi, l'autore mostra quanto le due lingue romanze fossero un tempo più vicine, evidenziando al contempo il carattere più conservativo del rumeno. Si tratta di una ricerca che riprende e aggiorna metodi e concetti della tradizione romanistica.

Il quarto contributo, firmato da **Cecilia Mihaela Popescu**, condivide con i precedenti l'interesse per gli studi storici e per il rapporto tra latino e lingue romanze attuali. L'obiettivo è quello di rilevare le tracce del verbo HABERE nella morfosintassi verbale del rumeno. L'autrice considera che l'evoluzione sia avvenuta tanto «verticalmente», per eredità diretta dal latino, quanto «orizzontalmente», per innovazioni e influenze condivise nei secoli. In questa prospettiva, mette in rilievo il processo di poligrammaticalizzazione del verbo latino menzionato, secondo funzioni diverse e in vari paradigmi del verbo rumeno. Nel loro insieme, tali processi spiegano il modo in cui i morfemi verbali si rapportano agli universali cognitivi e linguistici. Inoltre, l'autrice fornisce indizi conclusivi sull'origine dell'ausiliare del modo condizionale in rumeno.

Con l'articolo di **Olivia Petrescu** si passa dal francese come lingua di redazione allo spagnolo e dalla *Romania Continua* europea ai vasti territori della *Romania Nova*. Questo contributo affronta l'ambito gastronomico latinoamericano da una prospettiva lessicale, etimologica e culturale. L'autrice parte dall'idea che la gastronomia sia un'arte e una forma di comunicazione culturale; nell'area ispanofona, si fondono ingredienti propri delle culture precolombiane con altri introdotti dalla colonizzazione spagnola e, in seguito, da nuove ondate migratorie. La ricerca propone classificazioni ed esempi del lessico gastronomico di culture quali quella messicana, argentina, peruviana, colombiana e venezuelana, in base ai termini ereditati o trasmessi, riferiti soprattutto agli ingredienti, ai modi di preparazione e alla denominazione dei piatti. Ne risulta una terminologia estremamente varia e una combinazione di elementi culturali in grado di esprimere un medesimo concetto. Il lessico gastronomico ispanoamericano costituisce così una testimonianza sia delle tradizioni sia delle innovazioni, rappresentando un modello esemplare di meticcio culturale.

Senza abbandonare del tutto il campo culinario, **Simona Georgescu e Theodor Georgescu** esaminano l'eredità della labiovelare latina nelle lingue romanze mediante l'analisi del verbo *coquere*, che in protoromanzo avrebbe avuto due varianti: COQUERE e COCERE. Secondo gli autori, l'antico francese è l'unica lingua ad aver ereditato entrambe le forme, accompagnate da una distinzione semantica e da una differenziazione fonetica. In altri termini, discutendo l'eredità della consonante labiovelare /kw/ nelle lingue romanze, lo studio valuta la possibilità che una distinzione semantica corrisponda a una bipartizione fonetica. La conclusione a cui giungono, dopo un'analisi accurata, è tuttavia che non sia necessario postulare due prototipi, poiché un'unica forma priva della labiovelare basta a spiegare i discendenti romanzi e i loro due principali significati.

Dalle ricostruzioni etimologiche si passa alla realtà attuale delle lingue romanze e dei loro parlanti, poiché **Alina Andreica e Nora-Sabina Neamț** sviluppano un'analisi sociolinguistica della competenza orale in rumeno come lingua d'eredità presso studenti francofoni di seconda generazione che rientrano temporaneamente in Romania per frequentare l'università. Adottando la prospettiva teorica del plurilinguismo in contesti di mobilità accademica internazionale, le autrici analizzano sia produzioni preparate sia interazioni spontanee di un gruppo di partecipanti, individuando fenomeni quali interferenza linguistica, instabilità grammaticale e tratti fonetici specifici, oltre a valutare in generale la competenza plurilingue e di mediazione. I risultati rivelano differenze di livello linguistico associate all'esposizione formale al rumeno, ai contesti d'uso e al coinvolgimento familiare, sottolineando la relazione tra tali fattori e l'acquisizione di competenze comunicative in una nuova lingua.

Gli studi lessicali riemergono come uno dei fili conduttori di questo numero con l'articolo di **Mihai Enăchescu**. L'autore affronta un tema di storia della lingua spagnola, concentrandosi sulla sostituzione di alcuni arabismi negli ultimi secoli del Medioevo. Nella sua ricerca elenca sedici termini di origine araba riferiti a recipienti e mobili usati nelle cucine medievali e mette in evidenza l'evoluzione e la storia di tali vocaboli lungo l'asse diacronico del castigliano, che portò alla conservazione di alcuni e alla scomparsa di altri, sostituiti da nuovi termini, per lo più di derivazione latina.

Lo studio di **Mihaela Gheorghe**, l'unico redatto in inglese, propone un'analisi delle Costruzioni di Percezione di Eventi (CPE) in rumeno, concentrando l'indagine sull'interazione tra tipo di verbo di percezione, classe semantico-aspettuale del predicato e scelta del complemento orazionale. Basandosi su dati del corpus RoTenTen2021, l'autrice mostra che la distribuzione dei tipi di complemento è influenzata sia dalla prominenza sensoriale (eventi visivi, uditivi o bimodali) sia dalle proprietà aspettuative, e arriva così a individuare schemi e tendenze caratteristici di tali costruzioni.

Nell'ultimo articolo, firmato da **Timea Tocalachis**, i lettori ritrovano un altro studio etimologico comparato, dedicato in questo caso ai cambiamenti semantici del termine « trabajo » in spagnolo, confrontato con il suo cognato francese « travail ». Entrambe le lingue risalgono all'etimo del latino volgare TRIPALIUM, strumento di tortura. Il termine era dunque inizialmente associato al dominio della sofferenza e della costrizione, da cui subì processi di reinterpretazione semantica che lo condussero a designare l'attività produttiva e lo sforzo umano in vari ambiti della vita sociale. Secondo l'esposizione dell'autrice, l'evoluzione dei due termini non riflette soltanto la modifica del loro significato lungo l'asse diacronico, ma anche la trasformazione dei principi e della visione della società nel suo insieme riguardo al valore e al senso del lavoro nel mondo ispanofono e francofono.

Il numero continua con la sezione *Varia* e si chiude con una breve sezione di recensioni dedicate alla presentazione di alcune pubblicazioni recenti incentrate sulla linguistica applicata alle lingue romanze. La lettura complessiva del volume offre una visione plurale e aggiornata della ricerca sulla famiglia neolatina e mette in risalto l'attualità della linguistica romanza quale crocevia di studi, metodi e voci.

Francisco CALVO DEL OLMO 

Ludwig Maximilians Universität, München,
francisco.calvo@romanistik.uni-muenchen.de

Veronica MANOLE 

Universităţii Babeş-Bolyai de Cluj-Napoca,
veronica.manole@ubbcluj.ro

Sanda MORARU 

Universităţii Babeş-Bolyai de Cluj-Napoca,
sanda.moraru@ubbcluj.ro

Anamaria CUREA 

Universităţii Babeş-Bolyai de Cluj-Napoca,
anamaria.curea@ubbcluj.ro

PER UNA STORIA DEI NOMI DELLE LINGUE ROMANZE IN ITALIANO

Paolo D'ACHILLE¹

*Article history: Received 14 March 2025; Revised 1 May 2025; Accepted 31 May 2025;
Available online 12 December 2025; Available print 30 December 2025*

©2025 Studia UBB Philologia. Published by Babeş-Bolyai University.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ABSTRACT. *For a History of the Names of Romance Languages in Italian.*

The study presents the names that the main Romance languages have had in the history of Italian, in order to examine how and when the different idioms derived from Latin were perceived, identified, and named. Particular attention will be given to the glottonyms referring to the peoples who inhabited the various regions before the Latinization. At the end of the essay, the various terms used over time to refer to Italian will be proposed, terms which, before this glottonym became dominant, did not distinguish it either from Florentine or Tuscan, nor from Latin, the mother tongue from which it is derived, and against which the Vulgar Latin originally stood.

Keywords: *glottonyms, Romance languages, Italian, lexicography*

REZUMAT. *Pentru o istorie a numelor limbilor romanice în italiană.* Studiul prezintă numele pe care principalele limbi romanice le-au avut în istoria limbii italiene, pentru a verifica cum și când diferitele limbi derivate din latina au fost percepute, identificate și numite. Se va pune accent în mod special pe glotonimele care se referă la popoarele care locuiau diversele regiuni înainte de latinizare. La sfârșitul eseului vor fi propuse diferitele termene care au fost folosite de-a lungul timpului pentru a desemna limba italiană, care, înainte ca acest glotonim să se impună, nu a fost diferențiată nici de florentină sau toscană, nici de latină, limba maternă din care provine, împotriva căreia vulgarul se opunea inițial.

Cuvinte-cheie: *glotonime, limbi romanice, italiană, lexicografie*

¹ **Paolo D'ACHILLE** è professore ordinario di Linguistica italiana presso l'Università Roma Tre e, dall'aprile 2023, presidente dell'Accademia della Crusca. I suoi principali interessi di ricerca riguardano i rapporti tra scritto e parlato nella storia linguistica italiana, il dialetto di Roma, le varietà regionali e popolari, la lingua delle scritture esposte, la lingua del melodramma, la storia del lessico. E-mail: paolo.dachille@uniroma3.it.

1. Premessa

In questo intervento² intendo riunire (senza pretesa di completezza) una serie di dati attinti a varie fonti, prevalentemente ma non esclusivamente lessicografiche, relativi ai nomi che in italiano (o diciamo meglio nel dominio italo-romanzo, visto che prenderemo in considerazione anche esempi da volgari antichi diversi da quelli toscani) hanno avuto le principali lingue romanze, italiano compreso. Le motivazioni della ricerca sono varie: da un lato, vorrei verificare come e quando i diversi idiomi derivati dal latino sono stati percepiti, individuati e nominati dagli italiani; da un altro lato, intendo documentare come certi cambiamenti siano dovuti al significato, anche politico che possono assumere i glottonimi, che solo di recente è stato messo in luce negli studi (rimando per questo a Papahagi 2021, che dà conto anche della bibliografia precedente) e che, nella fattispecie, si coglie nelle denominazioni che si riferiscono alle popolazioni prelatine che abitavano i territori in cui si parlano le lingue moderne; infine, sulla scia del volume dell'amico Lorenzo Tomasin (2011) sulla parola *Italiano*, desidero mostrare le ragioni per cui il termine che designa la nostra lingua ha tardato a imporsi, all'interno del nostro Paese ancor più che all'estero, ed è stato indicato con glottonimi che non lo hanno differenziato non solo dal fiorentino o dal toscano, ma neppure dalla lingua madre da cui la nostra è derivata, e cioè il latino.

Credo che siano opportune due brevi considerazioni preliminari. La prima è che in italiano, come in molte altre lingue romanze che presentano un'alternanza di genere grammaticale (almeno tra maschile e femminile) i glottonimi sono di genere maschile; fa eccezione, com'è noto, il romeno. Secondo Fornaciari (1881, 24) in italiano la cosa si spiega perché è "sottinteso il sostantivo *idioma, linguaggio o parlare*". Ma si tratta di una spiegazione poco convincente. Leggiamo infatti quanto dice al riguardo Anna M. Thornton (2004, 514-5), che ha trattato il problema all'interno del fenomeno della conversione da aggettivi a nomi:

Un [...] gruppo di nomi regolarmente omofoni di un aggettivo etnico sono i nomi di lingue e dialetti, che sono sempre maschili: *l'italiano, il francese, il veneziano, il torinese*. Il genere maschile non è facilmente spiegabile. L'ipotesi di un'origine ellittica da sintagmi del tipo *la lingua italiana* predirebbe per l'aggettivo sostantivato il genere femminile, contrariamente ai fatti. Altra ipotesi è che l'origine sia sì ellittica, ma a partire da contesti in cui il nome testa sia maschile, quali ad esempio *l'idioma italiano*. Questa

² Ringrazio i miei allievi, i dottorandi Matteo Agolini e Andrea Riga, per l'aiuto da loro ricevuto nella ricerca della documentazione e nella revisione finale del testo. Ringrazio anche i due anonimi revisori per i loro utili commenti, di cui ho naturalmente tenuto conto.

ipotesi non trova però un buon riscontro nei fatti. Innanzitutto, *lingua* e *idioma* fin dai primi secoli alternano liberamente in contesti analoghi, e *lingua* è usato più frequentemente di *idioma*. Contesti di gapping che potrebbero stare all'origine del tipo che ci interessa sono attestati per entrambi i nomi: *nel greco idioma o nel latino* (Bembo), *la lingua greca e la latina* (Foscolo), *l'idioma Tedesco e il greco* (Leopardi); tuttavia i primi esempi di uso assoluto precedono le prime attestazioni di contesti di gapping. [...] Questo insieme di dati fa supporre che l'uso di aggettivi sostantivati maschili con il valore di nomi di lingua non sia da ricondurre a fenomeni di ellissi, ma sia frutto di una regola di conversione indipendente [...] ed è più convincente l'ipotesi di una formazione per regola di conversione dall'etnico, con assegnazione del maschile in quanto genere di default.

Un'altra ipotesi sull'or[i]gine dei nomi di lingua maschili omofoni di aggettivi etnici, non incompatibile con quanto detto finora, è [...] che all'origine del tipo possano stare costrutti come *parlare italiano*, in cui *italiano* è ancora aggettivo o al massimo è un avverbio convertito da un aggettivo [...]. Una rianalisi di questi contesti avrebbe portato poi a interpretare *italiano* come oggetto di *parlare*, e dunque come nome. Il genere maschile di questi nomi si spiegherebbe allora come ereditato dal genere maschile dell'aggettivo [...], il quale riceve il genere maschile per default, perché modifica un verbo, cioè un elemento non capace di controllare l'accordo di genere e quindi di assegnare un genere per accordo.

La seconda questione, che propongo rapidamente a completamento di quanto già scritto da Papahagi (2021), è che i glottonimi – specie prima della loro istituzionalizzazione – non sempre corrispondono agli etnici e tanto meno, direi, ai coronimi dei Paesi in cui le lingue corrispondenti vengono parlate: basti citare, anche in diacronia, il caso di *tedesco* (richiamato giustamente dalla stessa Papahagi) o, in sincronia, quello di *inglese*, che è riferito anche alla lingua ufficiale e/o parlata dalla maggioranza della popolazione in Paesi come gli Stati Uniti – per i quali, al massimo, si usa (ma solo in contesti particolari) *anglo-americano* –, l'Australia, ecc. Su entrambi i temi avrò modo di tornare nel corso dell'intervento, presentando e commentando le attestazioni raccolte.

2. I dati lessicografici sui nomi delle principali lingue romanze

Inizio col proporre una tabella che riporta, in ordine alfabetico, i nomi che vengono oggi prevalentemente attribuiti (negli studi e non solo) alle principali lingue romanze (includendone anche alcuni su cui poi non tornerò), e le date delle prime attestazioni come sostantivi fornite dal *Deonomasticon Italicum* di

Wolfgang Schweickard (1997-2013; indicato con la sigla DI) e nel *Tesoro della lingua italiana delle origini* (TLIO), in rete, che documenta la situazione nei volgari antichi³; segnalo anche le varianti documentate anteriormente a quella oggi in uso (non quelle posteriori), nonché alcune retrodatazioni, che ho apportato grazie alla consultazione di Google Libri (GL):

Tabella 1. Le prime attestazioni in italiano dei nomi delle lingue romanze

glottonimo	DI	TLIO	retrodatazione
Catalano	1467ca.	-	-
Francese	dal 1480 ca. (1288 <i>francesco</i> ; 1389 ca. <i>fransos</i> ; 1395 <i>francioso</i> ; 1436 <i>franzese</i>)	(1288 <i>francesco</i> ; 1385-95 <i>francioso</i>)	-
Franco-provenzale	1873	-	-
Friulano	1824 (1347ca. <i>friolano</i> ⁴)	-	-
Italiano	1475	1322 ⁵	-
Ladino	1829	-	-
Portoghese	1524	-	-
Provenzale	1300ca. (1188 <i>provenzalesco</i>)	1315 (1188 <i>provenzalesco</i>)	-
Romancio	1729	-	1702 ⁶
Rumeno	1861	-	1853 ⁷

³ Riproduco, senza scioglierle, le sigle e le abbreviazioni con cui il TLIO e il DI citano testi e altre fonti, e mantengo anche le parole riportate in grassetto; ometto invece alcune sottolineature presenti nel TLIO.

⁴ Come vedremo più oltre, l'esempio a cui rimanda la data non si riferisce in realtà al glottonimo.

⁵ Nell'esempio compare come agg. pl. "[Con rif. alla lingua parlata in Italia]: [1] Alberto della Piagentina, 1322/32 (fior.), *Prologo*, pag. 8.16: Questo Boezio è commendato da Teodorigo re in una sua pistola, la quale nel libro di Cassiodoro si legge; nella quale li scrive così: «[[...]]. Nelle tue translazioni i pitagorici musici si leggono **italiani**; Nicomaco arismetico, Euclide geometro sono uditi romani...»". DI, II, 553, segnala un esempio di *loquela italiana* in Fazio degli Uberti tratto dal corpus OVI del 1367ca., anteriore di oltre un secolo a quella di *italiano* s.m. (Appiano A, Carteggi Sestan 1,522), datata 1475 (DI, II, 550).

⁶ "Ed alla Valle di Santa Maria, dove si parla Romancio, che è un parlar misto di molte lingue, mandò Padri della Provincia di Brescia" (F. Clemente da Brescia, *Istoria delle missioni de' frati minori capuccini* [sic] *della Provincia di Brescia nella Rezia*. Trento: Parone, 1702, p. 467; da GL).

⁷ "Il futuro nel Sardo, tanto nel Campidano quanto nel Logudoro, non è un tempo semplice come nella maggioranza delle lingue e dialetti romanzi, ma è composto, come nelle lingue teuto-gotiche, nel rumeno (valacco) e nel retico oberlandese" ([Giovenale] Vegezzi-Ruscalla, "Corrispondenze dell'annotatore friulano", *L'Annotatore Friulano. Giornale di agricoltura arti commercio e belle lettere* 1, 1853, p. 158; da GL). Per alcune attestazioni di rumano, vedi *infra*, nota 12.

glottonimo	DI	TLIO	retrodatazione
Sardo	1824	-	1792 ⁸
Spagnolo	1584 (1347ca. <i>ispagnuolo</i>)	- p.m. XIV (<i>ispagnuolo</i>)	1552 ⁹

Una prima osservazione (che non risulta peraltro dalla tabella, ma dagli spogli lessicografici): nella maggior parte dei casi la conversione dell'etnico da aggettivo a nome è posteriore, a volte anche di un secolo, agli esempi in cui viene usato come aggettivo, ma le occorrenze in cui si lega, al femminile, a parole come *lingua* (o *favella* o *loquela* o *parlatura*) sono più antiche e numerose di quelle in cui compare affiancato a *idioma* (o a *dialetto* o *linguaggio* o *volgare* o *parlare*) e quindi la proposta di Thornton di un fenomeno di conversione indipendente dall'ellissi sembra confermata; meno probabile, invece, è quella di una derivazione da espressioni con l'aggettivo maschile legato avverbialmente al verbo *parlare*, che il DI documenta solo eccezionalmente e tardivamente. Si potrebbe invece supporre che il sostantivo sottinteso fosse costantemente *volgare*, che (è vero) emerge solo in pochi casi, ma che potrebbe essere implicitamente contrapposto a *latino*. A proposito di latino, va ricordato che in questa lingua era possibile usare il sostantivato del neutro *Latinum* come equivalente a *lingua Latina*, anche qui, dunque, con un cambio di genere. Non mi risulta che questa particolarità sia stata oggetto di studi specifici da parte dei latinisti, i quali si sono occupati soprattutto del nome della lingua, che si richiama al Lazio e non a Roma (Poccetti, Poli, e Santini 1999): in effetti l'espressione *lingua romana* in latino è più tarda, e le sue occorrenze crescono proprio al momento del passaggio dal latino volgare ai volgari romanzi (a volte per indicare, specie con l'aggiunta di aggettivi come *rustica*, *vulgaris*, ecc., proprio questi ultimi), mentre in italiano lo specifico significato di 'latino volgare' o di 'volgare' dell'espressione (o di altre analoghe) si ha, secondo il DI (IV, 78), dal sec. XVIII in poi:

lingua romana f. 'latino volgare' (1775, 65 BettinelliRossi, GDLL: «fu la nuova lingua generale all'Europa del mezzodì ed ebbe un nome suo generale, cioè di «romana» a differenza del vero latino, onde fecesi poi «romanza»»; 1845, GalvaniOccitania 60s.: «codesta lingua «romana», la quale ebbe pur nome di «rustica»»; 1872ca., DeSanctisGallo, LIZ). – *lingua romana plebea* f. 'id.' (1708, GravinaQuondam, ib.: «alcuni poemi scritti in lingua provenzale,

⁸ “[...] mostrare che il Sardo non ha per fondo principale che il Latino, non già l’Italiano” (Matteo Madao, *Dissertazioni storiche apologetiche critiche delle sarde antichità*. Tomo I. Cagliari: Reale Stamperia, 1792, p. 108 in nota; nel testo ci sono vari altri esempi; da GL).

⁹ *Olivieri di Castiglia et Artus di Dalgarve*. Tradotto di Spagnolo in lingua Toscana per Francesco Portonari. Vinegia: appresso Francesco Portonari da Trino, 1552.

la quale [...] lingua romanza appellavasi dalla lingua romana plebea»), *lingua plebea romana* (1708, ib.). – *lingua romana rustica* f. 'id.' (dal 1821, LeopardiPacella, LIZ; 1841, BiondelliAtlante 1,104; 2010, CorpusWeb9). – *dialetto volgare romano* m. 'id.' (1821, LeopardiPacella, LIZ). – *favella romana rustica* f. 'id.' (1841, BiondelliAtlante 1,118). – *linguaggio romano rustico* m. 'id.' (1841, ib. 1,104). – Agg. sost.: *comune rustico romano* m. 'id.' (1823, LeopardiPacella, LIZ), *romano rustico* (1837, CattaneoBertani, GDLI).

Tornando al neutro *Latinum*, è vero che nel latino classico si tratta di un uso minoritario rispetto alle espressioni *lingua Latina* o *sermo Latinus*, ma l'amico Angelo Luceri mi segnala che tra le prime attestazioni di *Latinum* come glottonimo ce n'è una di Cicerone (*Tusculanae disputationes* 3, 14, 29: "licet enim, ut saepe facimus, in Latinum illa convertere"), il quale usa spesso anche l'espressione *Latinum verbum* 'parola latina'. Forse proprio il legame con *verbum* potrebbe aver determinato la scelta del neutro. In ogni caso, la sostituzione romanza del neutro col maschile e il riferimento a Cicerone potrebbero spiegare perché nella Romània occidentale ai nomi delle lingue e dei dialetti sia stato assegnato il genere grammaticale maschile, di norma succedaneo del neutro, prima alle lingue che dal *Latinum* erano derivate e poi a tutte le altre.

In secondo luogo, esaminando la tabella, si nota (e la cosa si spiega benissimo sul piano storico) che risalgono almeno al sec. XIV (con presenze già duecentesche nella forma aggettivale e addirittura una del 1188 già come nome maschile) le attestazioni dei tipi lessicali dei glottonimi della lingua parlata all'interno dei nostri confini (*italiano*) o quelli di lingue con cui l'italiano è entrato più presto in contatto: *francese* e *provenzale*, ma anche *spagnolo*, a proposito del quale l'assenza di riferimenti nel trattato dantesco *De vulgari eloquentia* costituisce tuttora una *crux* ed è stata variamente spiegata. Si datano all'epoca rinascimentale (secc. XV e XVI) le prime attestazioni di *catalano* e *portoghese* (affiancato, sempre nel Cinquecento, da *portogallese*), mentre sono settecentesche quelle di *romancio* (che peraltro non si è ancora del tutto stabilizzata) e di *sardo*; si collocano lungo l'Ottocento, in seguito ai progressi della linguistica romanza, che hanno portato alla "scoperta" e definizione dei vari idiomi, le prime attestazioni, nell'ordine, di *friulano* (l'attestazione trecentesca del DI non è attendibile perché *Friolano* nel passo citato, compreso nel corpus OVI, indica certamente il territorio)¹⁰, *ladino* e *franco-provenzale*, denominazione, questa, creata da Graziadio Isaia Ascoli

¹⁰ "Gabella per tutta Toscana. [...] Costuma per tutta l'isola d'Inghilterra. Fedo a Tunizi in Barberia. Munda in Friolano. Mangona e Talaoch in Ispagna. Tutti questi nomi voglio dire diritto che si paga di mercatantia e di merce" (Francesco Balducci Pegolotti, *La pratica della mercatura*, a cura di Allan Evans. Cambridge (MA): The Medieval Academy of America, 1936, p. 15).

(1873)¹¹. Risale al sec. XIX anche la prima attestazione in senso linguistico di *rumeno* (1853), a cui dal 1937 si affianca *romeno* (per le oscillazioni tra *ru-* e *ro-* rinvio a Valmarin 1989 e a Paoli 2013), perché in precedenza si parlava di *valacco*, termine documentato, stando al DI, dal sec. XVIII in diverse varianti (*valaco*, *valako*, *vallacco*, *walaco*, *walako*), riportato tra parentesi dopo *rumeno* nella prima attestazione di questo glottonimo e rimasto in uso fino al 1872, in Napoleone Caix (sarebbe questa, dunque, l'“ultima attestazione”, per riprendere l'espressione usata in D'Achille 2020)¹². In questo caso si trattò certo, come nota Papahagi 2021, di un cambio fatto per “souligner une filiation prestigieuse”, ma aggiungerei che il nuovo glottonimo si lega al ricongiungimento della cultura romena al mondo romanzo occidentale dopo la costituzione, nel 1859, della Romania come Stato indipendente grazie all'unione tra Valacchia e Moldavia dopo la guerra di Crimea (com'è noto, la Transilvania si sarebbe aggiunta, o ricongiunta, a seconda dei punti di vista, dopo la Prima guerra mondiale). Approfitto per segnalare che anche *moldavo* è documentato dal sec. XVIII, ma che solo dopo la Seconda guerra mondiale (con la nascita della sovietica Repubblica di Moldavia, indipendente dalla Romania), nel clima della “guerra fredda” e dell'Unione Sovietica, si tentò di usarlo per “promuovere” la varietà locale del romeno parlato in Moldavia a livello di lingua nazionale distinta (del dibattito al riguardo dà conto Tagliavini 1972).

Concludo questa parte dedicata ai glottonimi tuttora in uso per segnalare che, secondo il DI, *castigliano* in alternativa a *spagnolo* è documentato già dal 1341ca. (nella forma *castellano*, presente anche nel TLIO) e poi, dal 1650, nella forma attuale. Inoltre, accanto a *provenzale*, si parla anche di *occitano* (che per il DI, s.v. *Linguadòca*, è attestato solo dal 2000, ma che Google libri documenta

¹¹ Mantengo la data del DI, ma il vol. III dell'*Archivio glottologico italiano* (nella versione informatica visibile in GL) reca sul frontespizio la data 1878; tuttavia l'indice del volume, compreso il saggio ascoliano, è riportato nella rassegna dei periodici in *Rivista di filologia romanza* 2, 2, 1875, p. 124.

¹² Nella *Nuova enciclopedia popolare, ovvero Dizionario generale di scienze, lettere, arti, storia, geografia, ecc. ecc. opera compilata sulle migliori in tal genere, inglesi, tedesche e francesi coll'assistenza e col consiglio di scienziati e letterati italiani*, vol. XII, Torino Pomba, 1848, s.v. *Valachia*, si usa *rumano* per indicare lo Stato (“L'anno 1848 lasciò speranza di veder sorgere uno Stato rumano, come un'Italia, una Germania”, p. 440) ma ancora *valacco* per la lingua. Ma *rumano* per indicare la lingua è attestato come aggettivo (“idioma rumano”), in *La Ungheria antica e moderna: sua storia, arti, letteratura, monumenti* scritta da una società di letterati sotto la direzione di M. J. Boldenyi, Pistoia, a spese dell'editore, 1852, p. 75, e come nome (“avendo egli stesso ben appreso il rumano”) in Angelo De Gubernatis, *Dizionario biografico degli scrittori contemporanei*, Firenze, Le Monnier, 1879, p. 54 (nell'opera è usato anche come etnico). Lo stesso termine *rumano* indica invece il romancio in Felice Bähr, *Storia della letteratura romana*, tradotta sulla 3ª edizione tedesca da Tom. Mattei, Torino, Pomba, 1849, p. 16 (“il *rumano* de' Grigioni”), mentre per il romeno si usa *valacco*.

già nel 1968, in un articolo del periodico “Volontà”, in un altro del 1975 su “La ricerca dialettale” e poi in Harrieta 1976, un testo dedicato al franco-provenzale indicato come *arpitano*) oppure di *occitanico* (dal 1861).

Restando ai tipi lessicali che abbiamo individuato, non si può non rilevare come, prima di *francese*, si siano avuti altri glottonimi derivati da *Francia* (e non da *Franchi*), tra i quali spicca *francesco*, da cui anche l'avverbio dantesco (1319ca.) *francescamente* ‘in lingua francese’, con il suffisso in *-esco* allora produttivo per gli etnici e quindi per i glottonimi: si pensi a *romanesco*, che ha però spesso, *ab origine*, una connotazione negativa (cfr. Aprea 2008-2009; Schweickard 2010) o a *fiorentinesco* (D'Achille 2019). Ovviamente, il riferimento in questi esempi è al francese antico, la *lingua d'oïl*, la cui individuazione e il cui uso come glottonimo è molto antico: il DI (II, 106) cita “lat. mediev. *lingua francisca* f. ‘id.’ (secc. IX-XII, Müller, LRL 2/1,144). – Fr. *français* m. ‘id.’ (dal 1100ca. [*franceis*], ib. 2/1,145).

3. Altre denominazioni documentate nel corso dei secoli

Passiamo ora a esaminare possibili denominazioni alternative, anche se documentate da attestazioni isolate, che sotto vari aspetti risultano essere significative. E ripartiamo dal francese, servendoci del già ricordato contributo di Papahagi (2021), che nota come, in testi latini redatti tra il XII e il XVI secolo, si trovi l'espressione *lingua gallica*, che assume particolari connotazioni, sul piano dei rapporti, anche politici, tra il papato e la monarchia francese. Ebbene, la locuzione si incontra anche nelle note al Volgarizzamento toscano di Valerio Massimo, del 1346ca., dove peraltro potrebbe riferirsi agli antichi Galli (“Alpi in lingua gallica viene a dire ‘alti monti’”): l'esempio, tratto dal corpus OVI, non è segnalato né nel TLIO né nel DI. *Gallico* sostantivato risale solo al 1851 (primo e unico esempio riportato nel DI, II, p. 209, in Guerrazzi), ma espressioni come *lengua galica*, *lingua gallica*, *gallica favella*, *idioma gallico*, *linguaggio gallico*, *gallico sermone* sono documentate dal 1436 fino al 1843, in Gioberti: dunque, l'espressione in italiano ha avuto una sua vitalità, probabilmente collegata a sentimenti antifrancesi, specie sul piano linguistico: il termine “gallicismo” è frequente nell'uso puristico ottocentesco, per riferirsi a parole o costrutti modellati sul francese, indicato richiamando appositamente la popolazione della Gallia conquistata dai Romani prima ancora di esser cristianizzata dalla Chiesa di Roma.

Questo riferimento non si limita al francese, ma – se pure occasionalmente – riguarda anche altre lingue romanze. Presento le attestazioni che ho trovato nel DI seguendo l'ordine cronologico – sempre suscettibile di anticipazioni, o in alcuni casi (come si è visto per *friolano*) di posticipazioni – segnalando subito di aver

riportato esempi della locuzione con *lingua*, *idioma*, ecc., solo quando non ho reperito attestazioni dell'aggettivo maschile sostantivato:

ispano 'spagnolo' è documentato solo come aggettivo (*lingua hispana*, 1519 ca.; *hispano idioma* 1525; *hispano linguaggio* 1884);

ibero 'spagnolo' (*idioma ibero* 1687, Frugoni; *parlar ibero* 1802, Batacchi; DI, II, p. 414);

reto 'romancio' (*lingua reta* 1729, Da Sale; DI, IV, p. 34);

dacio 'romeno' (1838; DI, I, p. 623).

Tra il Cinquecento e l'Ottocento, dunque, anche per lo spagnolo c'è stato il riferimento al mondo prelatino, con la chiamata in causa dell'*Hispania* e dell'*Iberia*, così come, isolatamente, dei *Reti* per il romancio e dei *Daci* per il romeno. L'ottica di queste denominazioni mi sembra prevalentemente linguistica: non a caso *gallo-*, *ispano-*, *reto-*, e *daco-* (non *dacio-*, che si direbbe una retroformazione dal plurale) si usano tuttora come confissi, premessi a *romanzo*, negli studi linguistici, particolarmente in quelli che valorizzano molto il fenomeno del sostrato.

4. I vari nomi dell'italiano: italiano, lingua latina e lingua romana

In quest'ottica, eminentemente linguistica, che fa riferimento alla storia prelatina, è stato coinvolto anche l'italiano, soprattutto nel corso del Cinquecento, il momento in cui, nella questione della lingua, anche la denominazione da dare al nuovo volgare, ormai stabilizzatosi nelle scritture letterarie e che si avviava alla sua prima standardizzazione, era oggetto di acceso dibattito, specie dopo l'intervento del Trissino, per quanto riguarda la scelta tra *fiorentino*, *toscano* e *italiano* (rimando almeno a Giovanardi 1998 e a Tomasin 2011). Ebbene, il riferimento al sostrato spiega perché, come documenta il DI, verso la fine del sec. XVI il toscano venga talvolta definito *etrusco* (la stessa denominazione data, già dalla fine del secolo precedente, alla lingua degli antichi abitanti della regione, lingua che peraltro non era stata ancora decodificata) e che l'etichetta venga poi ripresa, nel sec. XVIII, da autori settentrionali (Parini, Baretti), probabilmente con una voluta accentuazione del carattere 'regionale' della parlata toscana:

lingua etrusca f. 'lingua toscana' (1562, Scroffa, CanticiTrifone 3; 1752, PariniMazzali 328); *lingua hetrusca* (1588, Pino, CanticiTrifone 177); *etrusca favella* f. 'id.' (1589, DellaPorta, CanticiTrifone 177); *linguaggio etrusco* m. 'id.' (1764, BarettiOpere 4,81).

Certo, non c'è allora da stupirsi se, in quest'ottica "sostratista", non si usi anche *latino* per indicare quello che oggi chiamiamo *italiano* (sebbene in questo caso giochino anche altri fattori: De Mauro 2000). Ma già dal Medioevo, sebbene *latino* si usasse allora normalmente per indicare la lingua di Cicerone, anche in quanto 'grammatica, lingua ordinata delle scritture', troviamo nel TLIO, s.v. *latino* (accezione 4) e s.v. *volgare* (accezione 3.1.2), alcuni esempi (uno, quello di Catenacci, riportato in entrambe le voci) in cui il *latino* si riferisce al volgare italo-romanzo:

4. Di lingua italiana (d'un volgare d'Italia, del toscano).

[1] *Ottimo, Inf.*, a. 1334 (fior.), c. 28, pag. 474.6: ogni lingua sarebbe insufficiente per lo difetto de' vocaboli **latini**, e della memoria, che non è tanto capace.

[2] Boccaccio, *Corbaccio*, 1354-55, parr. 441-50, pag. 120.6: Ma io così fidatamente ne favellava, per ciò che saper mi pareva, e so, che le sue orazioni e i suoi paternostri sono i romanzi franceschi e le canzoni **latine**, ne' quali ella legge di Lancelotto e di Ginevra e di Tristano e d'Isotta e le lor prodezze e i loro amori e le giostre e i torneamenti e l'assemblee...

4.1 Sost. Lingua volgare italiana. || *Volgare latino*: v. *volgare*.

[1] Bart. da San Concordio, 1302/08 (pis.>fior.), dist. 9, cap. 4, par. 6, pag. 169.20: Io recando questo libro in **latino**, abbo posto più intendimento per intendimento, che parola per parola.

[2] Catenacci, *Disticha Catonis*, XIII/XIV (anagn.), Proemio, 6, pag. 117: Lu Cato ch'è de gran doctrina plino / translateraiu p(er) vulgar(e) **latino**.

[3] Fr. Grioni, *Santo Stady*, a. 1321 (venez.), 4767, pag. 174: Signori, eo ve don ben a saver, / A tuti quanti rie e bony, / Che Franceschin dito Griony / Questa leçenda conçità / Siando in grande oschurità, / E lla rismà cossì in **latin** / Per la gràcia dello re divin...

[4] Ciampolo di Meo Ugurgieri, a. 1340 (sen.), L. 7 argomento, pag. 217.7: E questo nome tragedia è composto di due nomi greci, cioè *tràgos* e *oda*; e in **latino** tanto è a dire quanto laude o canto di becco...

[5] Boccaccio, *Decameron*, c. 1370, V, 2, pag. 344.35: la quale essalei che forte dormiva chiamò molte volte e, alla fine fattala risentire e all'abito conosciutala che cristiana era, parlando **latino** la dimandò come fosse che ella quivi in quella barca così soletta fosse arrivata.

3.1.2 *Volgare latino*: lingua popolare d'Italia (quella dello scrivente).

[1] a Catenacci, *Disticha Catonis*, XIII/XIV (anagn.), Proemio, 6, pag. 117: Lu Cato ch'è de gran doctrina plino / translateraiu p(er) **vulgar(e) latino**.

[2] Anonimo Genovese (ed. Cocito), a. 1311, 146.15, pag. 641: Donde, per mejo deszhairar / de questo fatto lo tenor / en lo nostro **latin vulgar**, / ò chi notao li nomi lor: / frai Venardí e Carlevar...

[3] Boccaccio, *Teseida*, 1339-41 (?), *Dedica*, pag. 246.30: trovata una antichissima istoria e alle più delle genti non manifesta [...], in latino **volgare** e per rima [...] ho ridotta.

[4] Marchionne, *Cronaca fior.*, 1378-85, Preambolo, pag. 1.18: e perchè così il laico come il litterato di ciò possa prendere frutto in **volgare** latino scriveremo.

Il DI (II, 646), da parte sua, data al 1501 un esempio di *latino* aggettivo riferito alla lingua italiana (che verosimilmente non sarà l'unico): "*latino* agg. 'in lingua italiana; relativo alla lingua italiana' (1501, Trevisan A, Nuovo Mondo Berchet 1,52)".

Anche *lazio* (come aggettivo, di ascendenza dantesca) si riferisce "alla lingua popolare d'Italia" già in Boccaccio (TLIO, s.v.):

[1] Boccaccio, *Teseida*, 1339-41 (?), L. 12, ott. 84.8, pag. 661: tu, o libro, primo a lor cantare / di Marte fai gli affanni sostenuti, / nel volgar **lazio** più mai non veduti.

[2] **GI** Boccaccio, *Chiose Teseida*, 1339/75, L. 12, 84.8, pag. 661.7: [**lazio**]: latino. - [*id.*]: lazio s'intende qui largamente per tutta Italia.

Tale significato viene ripreso occasionalmente tra Quattro e Cinquecento, come risulta nel DI (II, 673):

idioma lazio m. '(ling.) italiano' (1582, Bruno, *OpereGuzzo/Amerio* 62).
materno latio m. 'lingua materna italiana, toscana' (1443ca., Prodenzani, Ugolini, *ContrDialUmbra* 1.1,59: «Dissi ben che 'l vocabulo e 'l profatio, / del Patrimonio nel paese expericho, / non è accepto nel materno latio»).

D'altra parte, ancora il DI (II, 553, 567) riporta singoli esempi in cui le espressioni *lingua italiana* o *itala lingua* si riferiscono al latino, il che completa il cortocircuito tra i due etnici usati come glottonimi:

lingua italiana 'lingua latina' (1550 LeoneAfricano, *Navigationi Ramusio* 1,3 r).

itala lingua f. 'lingua latina' (1793, Pindemonte, *LiriciMaier* 1065) [si noti che subito prima si citano *lingua itala* f. 'lingua italiana' (1551, GelliMaestri 471); *itala favella* f. 'id.' (1802, CastiPedroia, LIZ); *italo sermone* m. 'id.' (1843, DeStefanisStampaMil 1,1367)].

Il problema, come si è appena visto, si pone anche per *italo*, ma non per *italico*, perché, fin dal Duecento, per *volgare italico* si intende soltanto quello che Dante chiama il “volgare di sì”, come risulta dalla voce *italico* del TLIO:

1.1.1 [Con rif. alla lingua:] italiano.

[1] Andrea da Grosseto (ed. Selmi), 1268 (tosca.), L. 3, cap. 12, pag. 223.18: E Tullio disse: la ragion dei cini[ci] è tutta di gittar via: et è a- ddire *cinis* in lingua greca quanto che in volgare **italico** è a- ddire cane; et indi sono detti cini[ci], cioè cani. || Cfr. Albertano, *Ars loquendi*, III: «*Cinos* graece, latine dicitur *canis*. Inde *cinici* dicuntur latrantes ut canes...».

[2] Dante, *Convivio*, 1304-7, I, cap. 6, pag. 25.16: Ma questo non è: ché uno abituato di latino non distingue, s'elli è d'Italia, lo volgare [inglese] dallo tedesco; né lo tedesco, lo volgare **italico** dal provinciale. Onde è manifesto che lo latino non è conoscente dello volgare.

[3] Dante, *Convivio*, 1304-7, I, cap. 9, pag. 37.16: Non avrebbe lo latino così servito a molti: ché se noi riducemo a memoria quello che di sopra è ragionato, li litterati fuori di lingua **italica** non avrebbon potuto avere questo servizio...

Qualche incertezza nell'individuazione del referente si ha anche nel caso dell'etnico derivato da *Roma*: vero è che, come segnala il DI, IV, 78-79, l'aggettivo *romano*, anche con riferimento alla lingua, ha una gamma di usi piuttosto estesa. A volte, infatti, equivale a *romanzo* (e anzi lo anticipa) per indicare, al plurale, le lingue neolatine:

lingue romane f.pl. 'lingue neolatine, romanze' (dal 1817, Londonio, Bellorini Romanticismo, GDL; 1819, PellicoScotti, ib.; 1841, 90 Gioberti BelloCastelli, ib.; 1855, BonghiVilla ib.; 1872, CaixSaggio IX; 2010, CorpusWeb10);

oppure indica singole lingue romanze, come il romancio:

lingua romana f. 'dialetto romancio del cantone dei Grigioni' (1645, NicolosiHilgers 88: «[...] mà di più ogni Valle hà Lingua diversa dall'altra nel parlare, nello scrivere, et nella stampa. Chiamano questo miscuglio di parole, Lingua Romana, et Grisa»). – *dialetto romano* m. 'id.' (1856, FaurielDante 2,198: «dialetto romano della Svizzera»; 1864, Zuccagni OrlandiniRaccolta 93: «L'idioma «retico» è volgarmente detto dai Grigioni «romencio», ma dividesi in due dialetti principali, l'uno dei quali è chiamato dal P. De Specha «romano», e l'altro «ladino»);

oppure ancora il provenzale:

romano m. 'lingua provenzale' (1879, CarducciOpere, GDLI: «una nuova lingua, il «provenzale» o «romano»»). – *romano francese* m. 'lingua francese' (1839, CattaneoSalvemini/Sestan, GDLI) o il francese antico [ma l'esempio è contemporaneo, dal CorpusWeb].

In ogni caso, il DI (IV, 79) ci offre anche un'attestazione trecentesca di *romana loquela* in cui l'aggettivo non si riferisce al latino (come avviene nella maggioranza degli esempi ivi registrati) e neppure al dialetto della capitale (l'espressione *lingua romana* è usata, a partire dal 1529, da Trissino, il quale adopera invece *romanesco* per riferirsi all'italiano parlato nello Stato della Chiesa; cfr. Castelveccchi 1986): si tratta di un esempio in Boccaccio: "*romana loquela* f. 'lingua italiana' (1336ca., BoccaccioOpereBranca, OVI: «Giulia, udendo la romana loquela, la quale Ascalion, lungamente dimorato a Roma, impresa avea, alzò il viso verso lui»)".

5. Da fiorentino e toscano a italiano

Sulla lunga "lotta" tra *italiano* e *fiorentino* o *toscano* per avere il primato come glottonimo molto è stato detto e non voglio ripeterlo. Restano ancora fondamentali le pagine scritte da Tomasin (2011), circa l'esistenza dell'etnico, in latino, già nel *Tresor* di Brunetto Latini e, in volgare, in Boccaccio, e sul fatto che, nonostante l'attestazione nella "terza corona", la voce non sia stata accolta dalle prime due edizioni del *Vocabolario degli Accademici della Crusca*, che del resto preferirono non nominare la lingua di cui raccoglievano le parole (Muljačić 1997; De Mauro 2000; D'Achille e Proietti 2008) sebbene la loro intenzione, come ha chiarito lo stesso Tomasin, fosse quella di scegliere il termine *toscano*. Al riguardo, devo anche segnalare un limite (peraltro comprensibile) del DI, che, mentre distingue, pur se forse in modo non proprio convincente, tra *lingua fiorentina* 'dialetto di Firenze' (1328ca., DellaLanaBiagi, OVI) e *fiorentino* 'la parlata di Firenze in uso tra le persone di elevata cultura, norma ideale, secondo una lunga tradizione, dell'uso italiano colto' (dal 1341ca., LibriAlfonsoKnecht 179), s.v. *Firenze* (DI, II, 71, 76–7), non opera la stessa distinzione, forse a causa della documentazione, tra *favellatura toscana* (1288, ColonnaVolgCorazzini, OVI: «ma la favellatura, qual sia o tedesca o francesca o toscana [...]») e *toscano* (1288, ColonnaVolgCorazzini, OVI: «El quale libro, maestro, per lo comandamento del nobile re di Francia, coll'aiuto di Dio à

traslatato di latino in francesco, ed io di francesco in toscano»), s.v. *Toscàna* (DI, IV, 611–2), dando questa definizione: “*favellatura toscana* f. l’insieme delle parlate della Toscana dalle quali (spec. dal fiorentino scritto del Trecento) ha tratto origine l’italiano”.

Di certo, il definitivo accoglimento di *italiano* come glottonimo si ebbe, come è noto, solo nel Settecento, dopo la sua accettazione anche all’interno della Crusca con Anton Maria Salvini. Dal punto di vista della lessicografia, resta tutt’oggi importante la dichiarazione dell’Alberti di Villanova, varie volte citata:

Prima però di proceder oltre, ragion vuole, ch’io accenni i motivi per cui, anziché *Dizionario della lingua Toscana*, siami piaciuto intitolarlo della *Lingua Italiana*. Senza voler decidere se *Italiana*, o *Cortigiana*, o *Toscana* debba chiamarsi la nostra favella, osserverò, che siccome *lingua Francese, Inglese, Spagnuola, Tedesca* si chiama la lingua di queste Nazioni, non perché in una sola provincia o nelle sole città capitali più purgata favellisi; ma perché in tutta l’estensione di quelle provincie e di que’ regni si parla, e si scrive dalle persone colte, ed eziandio perché, malgrado la diversità de’ dialetti, s’incontrano in ogni lingua molti vocaboli, li quali, o schietti, o poco alterati, sono comuni all’intera nazione (D’AlbVill 1797, prefazione).

Sono degne, dunque, di un certo interesse le ultime, forse un po’ polemiche occorrenze di *fiorentino* e di *toscano* nel senso di *italiano* segnalate da D’Achille e Proietti (2011), tra le quali spicca la loro provocatoria combinazione in un testo di Pietro Fanfani: *La Paolina. Novella Scritta in Lingua Fiorentina Italiana* (1868 e 1870) e poi, con inversione dei termini, in *Lingua Italiana Fiorentina* (1880), in accesa polemica con la Relazione di Manzoni. Nello stesso 1868, del resto, l’autore scrisse che «La lingua italiana c’è stata, c’è e si muove», aveva già in precedenza pubblicato un *Vocabolario dell’uso toscano* (1863) e avrebbe poi compilato un *Vocabolario della lingua italiana* (1882).

6. Conclusioni

Nel corso del tempo, le lingue romanze sono state indicate in italiano con nomi diversi: ne abbiamo analizzati alcuni, per mostrare come l’introduzione di nuovi glottonimi si leghi da un lato all’individuazione di specifiche varietà linguistiche in precedenza sconosciute, dall’altro alle scelte effettuate all’estero, a cui l’italiano si è tendenzialmente uniformato. Il dato più interessante, in rapporto alla varietà delle denominazioni, riguarda però proprio la scelta del glottonimo riferito alla lingua di Dante.

Infatti, prima che si stabilizzasse l'uso di *italiano* (che nel corso del Novecento è stato adoperato anche al plurale, a documentarne la variabilità interna: D'Achille e Proietti 2011), i glottonimi usati nel corso dei secoli non sempre hanno differenziato quella che possiamo definire come lingua nazionale dal fiorentino o dal toscano, che di certo ne hanno costituito la base, e neppure dalla lingua madre da cui l'italiano è derivato, e cioè il latino (specie se denominato *romano*).

In questo hanno avuto certamente un peso la polimorfia lessicale e il carattere pletorico del lessico italiano, ma il dato documenta anche un problema concreto per uno Stato policentrico, la cui unità linguistica, raggiunta prima nello scritto e molto più tardi, e non definitivamente, nel parlato, era stata individuata da Dante solo sulla base della comune modalità di affermare.

BIBLIOGRAFIA

- Aprea, Fabio. 2008-2009. "Per la storia del glottonimo *romanesco*". *Contributi di filologia dell'Italia mediana* 22: 219-50; 23: 81-99.
- Ascoli, Graziadio Isaia. 1873. "Schizzi franco-provenzali". *Archivio glottologico italiano* 3: 61-120.
- Castelvecchi, Alberto. 1986. "Introduzione". In Giovan Giorgio Trissino, *Scritti linguistici*, a cura di Alberto Castelvecchi, XI-LXVI. Roma: Salerno editrice.
- D'Achille, Paolo. 2019. "Gli 'errori' dei fiorentini". In: *Firenze e la lingua italiana* Atti della Piazza delle Lingue, 8 (Firenze, 29 settembre-2 ottobre 2016), a cura di Claudio Marazzini e Annalisa Nesi, 85-101. Firenze: Accademia della Crusca.
- D'Achille, Paolo. 2020. "«A te l'estremo addio»? Il problema dell'ultima attestazione nella linguistica e nella lessicografia italiana". *Studi di lessicografia italiana* 37: 333-56.
- D'Achille, Paolo, e Domenico Proietti. 2008. "Toponimi ed etnici nel *Vocabolario degli Accademici della Crusca*". In *Lessicografia e onomastica 2*. Atti delle giornate internazionali di Studio (Università degli Studi Roma Tre, 14-16 febbraio 2008), a cura di Paolo D'Achille ed Enzo Caffarelli, 375-92. Roma: Società Editrice Romana.
- D'Achille, Paolo, e Domenico Proietti. 2011. "Articolazioni e determinazioni nella definizione della lingua nazionale: l'italiano con aggettivi' dall'Unità a oggi". In *Storia della lingua italiana e storia dell'Italia unita. L'italiano e lo stato nazionale*. Atti del IX Convegno ASLI Associazione per la Storia della Lingua Italiana (Firenze, 2-4 dicembre 2010), a cura di Annalisa Nesi, Silvia Morgana e Nicoletta Maraschio, 215-30. Firenze: Franco Cesati.

- De Mauro, Tullio. 2000. "Stratificazioni sociolinguistiche dell'eredità latina nel lessico italiano". In *Linguistica storica e sociolinguistica*. Atti del Convegno della Società italiana di glottologia (Roma, 22-24 ottobre 1998), a cura di Palmira Cipriano, Rita d'Avino e Paolo Di Giovine, 163-88. Roma: il Calamo (rist. in Idem. 2005. *La fabbrica delle parole. Il lessico e problemi di lessicologia*, 193-218. Torino: Utet Libreria).
- DI = Wolfgang Schweickard, *Deonomasticon Italicum. Dizionario storico dei derivati da nomi geografici e da nomi di persona*, voll. I-IV, *Derivati da nomi geografici*. Tübingen: Niemeyer; poi Berlin-Boston: de Gruyter, 1997 [vol. I pubbl. a fascicoli dal 1997 al 2002]-2013.
- Fornaciari, Raffaello. 1881. *Sintassi italiana dell'uso moderno*. Firenze: Sansoni.
- Giovanardi, Claudio. 1998. *La teoria cortigiana e il dibattito linguistico nel primo Cinquecento*. Roma: Bulzoni.
- Harrieta, Joze. 1976. *La lingua arpitana (francoprovenzale). Con particolare riferimento alla lingua della Val di Aosta*. Romano Canavese: Tipografia Ferrero.
- Muljačić, Žarko. 1997. "Perché i glottonimi *linguaggio italiano*, *lingua italiana* (e sim.) appaiono per indicare «oggetti» reali e non soltanto auspicati molto più tardi di altri termini analoghi che si riferiscono a varie lingue gallo e iberoromanze?". *Cuadernos de Filología Italiana* 4: 253-64.
- OVI = *Corpus OVI dell'Italiano antico*, in rete all'indirizzo <<http://gattoweb.o-vi.cnr.it>>.
- Paoli, Matilde. 2013. "Si dice *romeno* o *rumeno*", in rete all'indirizzo <<https://accademiadellacrusca.it/it/consulenza/si-dice-romeno-o-rumeno/821>>; rist. *Italiano digitale* 17, 2021: 96-9.
- Papahagi, Cristiana. 2021. "Nommer la langue française au Moyen Âge. Un enjeu politique". In *Langages du pouvoir au Moyen Âge et au début de la modernité*, a cura di Catalina Girbea, 71-95. Paris: Garnier.
- Poccetti, Paolo, Diego Poli, e Carlo Santini. 1999. *Una storia della lingua latina. Formazione, usi, comunicazione*. Roma: Carocci.
- Schweickard, Wolfgang. 2010. "I glottonimi *romano* e *romanesco* nella storia dell'italiano". *Studi italiani di linguistica teorica e applicata* 39: 103-20.
- Thornton, Anna M. 2004. "Conversione". In *La formazione delle parole in italiano*, a cura di Maria Grossmann e Franz Rainer, 499-553. Tübingen: Niemeyer.
- Tagliavini, Carlo. 1972. *Le origini delle lingue neolatine. Introduzione alla filologia romanza*. 6ª ed. Bologna: Patron.
- TLIO = *Tesoro della Lingua Italiana delle Origini*, in rete all'indirizzo <<http://tlcio.o-vi.cnr.it/TLIO/>>.
- Tomasin, Lorenzo. 2011. *Italiano. Storia di una parola*. Roma: Carocci.
- Valmarin, Luisa. 1989. "La guerra del *ru-* e del *ro-*". In *Miscellanea di studi in onore di Aurelio Roncaglia a cinquant'anni dalla sua laurea*, IV: 1385-409. Modena: Mucchi.

I LATINISMI INTEGRALI NELLA COMUNICAZIONE MEDICA SCRITTA MEDIATA TECNICAMENTE: UN CASO DI STUDIO

Kevin De VECCHIS¹ 

Article history: Received 20 February 2025; Revised 5 May 2025; Accepted 25 May 2025;
Available online 12 December 2025; Available print 30 December 2025

©2025 Studia UBB Philologia. Published by Babeş-Bolyai University.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ABSTRACT. *Non-adapted Latinisms in Technically Mediated Written Medical Communication: A Case Study.* The study focuses on the analysis of non-adapted Latinisms belonging to the medical language and their use in Technically Mediated Communication between physicians and patients. The introduction illustrates the role of Latin within the Italian lexicon and its importance in languages for specific purposes and in particular in the medical language. § 2 is dedicated to the methodology: the *corpus* of non-adapted Latinisms was set up through research conducted within the Medicialia website. In § 3 we report the results of the analysis that examined non-adapted Latinisms both from a morphological and semantic point of view, considering gender, grammatical category, meanings assumed in Italian and typology, as well as verifying their presence in the Italian lexicography (§ 3.1.). Some uses of non-adapted Latinisms by doctors or patients are then examined (§ 3.2.). In the conclusions some reflections are given concerning the *status* of Latin within the medical language (also in relation to English) and the common language.

Keywords: *Italian lexicon, Latinisms, language for specific purposes, medical language, Computer-Mediated Communication*

REZUMAT. *Latinisme neadaptate în comunicarea medicală scrisă mediată tehnic: un studiu de caz.* Studiul se concentrează pe analiza latinismelor neadaptate din limbajul medical și utilizarea acestora în comunicarea mediată tehnic între medici și pacienți. Introducerea ilustrează rolul limbii latine în

¹ Kevin DE VECCHIS è ricercatore di Linguistica italiana presso l'Università di Pisa. I suoi principali interessi di ricerca riguardano la dialettologia, con particolare riferimento al dialetto di Roma, l'italiano contemporaneo, la lessicografia e il linguaggio della medicina. E-mail: kevin.devecchis@unipi.it.

lexicul italian și importanța acesteia în limbile specializate, în special în limbajul medical. § 2 este dedicat metodologiei: corpusul de latinisme neadaptate a fost creat printr-o cercetare efectuată pe site-ul Medicialia. În § 3 prezentăm rezultatele analizei care a examinat latinismele neadaptate atât din punct de vedere morfologic, cât și semantic, luând în considerare genul, categoria gramaticală, semnificațiile asumate în italiană și tipologia, precum și verificând prezența lor în lexicografia italiană (§ 3.1.). Sunt apoi examinate unele utilizări ale latinizărilor neadaptate de către medici sau pacienți (§ 3.2.). În concluzii sunt prezentate câteva reflecții privind statutul limbii latine în limbajul medical (inclusiv în ce privește limba engleză) și în limbajul comun.

Cuvinte-cheie: *Lexicon italian, latinisme, limbaj specializat, limbaj medical, comunicare mediată de computer*

1. Introduzione

È fatto noto che il latino costituisce la base su cui si è formato nel corso dei secoli il patrimonio linguistico italiano:

At the heart of the Italian lexicon are words that do not, strictly speaking, “derive from” but “continue” Latin, as they passed from the spoken Latin of the late imperial era (referred to by scholars as Vulgar Latin) into Italian (with phonetic, morphological, and semantic shifts), often (although not always) remaining in use until this day. In Italian lexicography, entries inherited directly are preceded by the tag “Lat[in],” to be distinguished from the tag “from Lat[in],” the latter being used for Latinisms (D’Achille 2023, 7).

Nel lessico della lingua italiana si distinguono così le parole che hanno continuato il latino volgare, ossia il latino parlato nella tarda età imperiale, dai cosiddetti latinismi o parole dotte (o ancora *cultismi*)², termini del latino classico o di derivazione latina ripresi in epoche diverse e destinati, almeno in un primo momento, all’uso scritto. Tra questi ultimi si possono distinguere³ i latinismi adattati alla veste fonomorfológica italiana (es. *escandescenza* ‘impeto improvviso d’ira accompagnato da atti e parole violente’, dal lat. *excandescētia(m)*, Zingarelli 2025, d’ora in poi Z); i latinismi parzialmente adattati alla struttura grafico-fonetica (es. *auscultare* ‘eseguire un’auscultazione’, dal lat. *auscultāre* ‘ascoltare

² Il termine *cultismo* ha invero un significato più ampio: ‘parola o espressione del linguaggio letterario o colto’ e appare per la prima volta nel 1936 (Z).

³ La distinzione è in Mazzini (2015, 115).

con attenzione', Z)⁴; i latinismi integrali⁵ (es. *flatus vocis* 'parola o discorso privi d'importanza e consistenza', Z), anche detti, fin dall'Ottocento, "crudi" o "nudi e crudi" (Trifone 2024, 103), vale a dire parole o espressioni che hanno mantenuto la forma originale latina⁶. I latinismi integrali (d'ora in avanti LI) si riscontrano maggiormente negli usi colti della lingua e nei linguaggi specialistici (cf. Dardano 1994; Gualdo e Telve 2011), mentre sembrano soltanto lambire la lingua comune, sottraendosi dagli usi più informali.

Proprio su quest'ultima tipologia si concentra lo studio qui condotto, che si propone di raccogliere i LI presenti nel linguaggio della medicina e analizzare l'uso che ne viene fatto nella *Comunicazione Mediata Tecnicamente* (CMT)⁷ tra medici e pazienti (cf. Orletti 2000; Orletti e Iovino 2018). I LI sono, inoltre, qui esaminati da un punto di vista morfologico (il genere e la categoria grammaticale che hanno assunto in italiano) e semantico (con quale significato sono utilizzati) e ne viene verificata la presenza all'interno della lessicografia italiana (Z). Sia per la costituzione del *corpus* sia per la verifica dell'uso nella CMT si è scelto di fare affidamento sul sito Medicitalia (§ 2). Si tratta di testi che, come si dirà, presentano gran parte delle caratteristiche relative alla categoria dello scritto trasmesso (D'Achille 2019).

La scelta è ricaduta sul linguaggio della medicina (cf. Serianni 2005; Gualdo e Telve 2011; Bellina 2011; Orletti e Iovino 2018; Piro 2022) per varie ragioni. Innanzitutto, tra tutti i lessici specialistici quello medico "vanta la presenza più alta nei dizionari" (Gualdo 2009, 397)⁸ e perché, come ha scritto Serianni, "nessuna scienza come la medicina coinvolge nella stessa misura il pubblico dei profani" (Serianni 2005, 239). Questa maggiore pervasività del linguaggio della medicina comporta, almeno in potenza, una più alta probabilità di

⁴ Mazzini fa riferimento a quelle forme che hanno perso i morfemi dei casi ma hanno conservato intatta la struttura fonetico-grafica interna: "forme generalmente non [...] isolate" che "compaiono in composti: così ad es. *auricolo-tomia*, da *auricola* 'orecchio'; *medullo-blastoma*, da *medulla* 'midollo'; *oculo-rinite*, da *oculus* 'occhio', ecc." (Mazzini 2015, 115).

⁵ La locuzione *latinismo integrale* è già in Sabatini (1962, 20).

⁶ Orletti e Iovino (2018, 45) ricordano anche che "la persistenza del latino nella lingua medica italiana risulta evidente [...] nella possibilità di avere parole e composti di creazione recente con materiale latino".

⁷ L'espressione è ripresa da Prada (2016, 335) e sostituisce *Computer-Mediated Communication*, in italiano *comunicazione mediata da computer* (CMC), perché l'accesso alla rete da diversi anni non avviene più soltanto tramite computer.

⁸ Il lessico della medicina risulta essere circa "il 13,70% del complessivo 52,50% del patrimonio linguistico tecnico specialistico del vocabolario italiano, secondo il GRADIT" (Gualdo e Telve 2011, 286), mentre la sua presenza nei dizionari raggiunge circa il 5,45% (Serianni 2005, 115). Si compone di circa 160.000 termini, di cui molte migliaia sono di origine greco-latina (Mazzini 2015, 114). Una breve ricerca sullo Z riporta i seguenti risultati (tra parentesi il numero di lemmi per ciascun ambito della medicina): med. (18022); anat. (1230); chir. (197); fisiol. (180); farm. (173), veter. (99).

sconfinamento di LI di àmbito medico⁹ nella lingua comune e, di conseguenza, la possibilità di verificare gli usi di tali termini in conversazioni asimmetriche tra medici e pazienti.

Bisogna considerare, inoltre, che nella struttura del linguaggio della medicina si può individuare una maggioranza di termini di derivazione latina (e greca) che “convivono con pari dignità scientifica” (Gualdo e Telve 2011, 287) al fianco di voci derivanti da tutte quelle altre lingue che nel corso dei secoli hanno potuto, per diverse ragioni storico-culturali, lasciarvi traccia¹⁰. A livello storico è noto che il latino, così come in parte anche il greco, non solo rimase la lingua originale dei testi di studio dal Medioevo fino almeno al Seicento, ma fu per secoli la lingua di riferimento per la produzione scientifica (cf. Mazzini 2015, 118; Orletti e Iovino 2018, 44). Il latino è stato poi parte del “bagaglio obbligato della formazione del medico fino al Novecento”, tanto da rimanere “in diverse denominazioni ufficiali di uso internazionale (*areola*, *angina pectoris*, ecc.)” (Serianni 2005, 171). Al successo del latino ha contribuito, inoltre, il lavoro di standardizzazione terminologica e nomenclatoria compiuto dall'*International Federation of Associations of Anatomists* (IFAA), che lo ha posto come lingua di riferimento¹¹, considerandolo una sorta di lingua franca, ormai incapace di evolversi e quindi potenzialmente globale (Gualdo e Telve 2011, 298; Carli 2006, 181)¹².

⁹ Infatti “la lingua specialistica di ambito medico, tra le lingue specialistiche dell’italiano, è tra le più caratterizzate da una massiccia presenza del latino, oltre che dalla ben nota componente greca” (Orletti e Iovino 2018, 41).

¹⁰ È nota la stratificazione diacronica del linguaggio medico, composto, oltre che dalle lingue classiche, in gran parte dall’arabo, dal francese e, recentemente, dall’inglese, il cui influsso, non solo nel linguaggio della medicina, è ogni giorno più invasivo (cf. Serianni 2005).

¹¹ Come ricordano Gualdo e Telve (2011, 297), l’IFAA ha coordinato i lavori del *Basilensia Nomina Anatomica* (1895), il primo vocabolario anatomico, sostituito dalla *Terminologia Anatomica* (Stuttgart: Thieme, 1998; 2^a ed., 2019: <https://fipat.library.dal.ca/TA2/>). L’IFAA ha pubblicato poi *Terminologia Histologica* (Philadelphia: Lippincott Williams & Wilkins, 2008), *Terminologia Embryologica* (2017: <https://FIPAT.library.dal.ca>) e *Terminologia Neuroanatomica* (2017: <https://FIPAT.library.dal.ca>).

¹² Sul rapporto tra latino e lingua medica cf. anche Mazzini (2015, 115), che tra i LI individua “per lo più [...] espressioni [...] di tradizione accademica: *cutis laxa* ‘pelle flaccida’, *per vias naturales* ‘attraverso le vie naturali’ (riferito alla somministrazione di medicinali), *ab ingestis* ‘in conseguenza degli alimenti ingeriti’, *bacillus coli* ‘colibacillo’, *bacillus rigidus* ‘bacillo rigido’, *treponema pallidum* ‘treponema pallido’ (il batterio della sifilide), ecc.”, appartenenti al lessico anatomico, patologico e biologico; Orletti e Iovino (2018), che hanno studiato i latinismi utilizzati come tecnicismi specifici e collaterali all’interno di un *corpus* allestito tramite interviste a medici specialisti, cartelle cliniche di area chirurgica, schede tecniche di farmaci in commercio e manualistica universitaria; Dionigi (2023), professore emerito di Chirurgia, che ha raccolto in un glossario le voci latine generiche e le voci settoriali, tutte di esclusiva pertinenza medico-chirurgica; il glossario *Le Parole della Salute* (PDS), nato in collaborazione tra Accademia della Crusca e Regione Toscana (cf. De Vecchis 2023), che si è proposto di raccogliere e semplificare tecnicismi d’ambito medico e sanitario (<https://leparoledellasalute.federsanitatoscana.it/>).

2. Metodologia della ricerca

La ricerca dei LI è stata condotta all'interno del sito Medicialia¹³. Si tratta di una "piattaforma web gratuita di consulti medici online e di divulgazione medico-scientifica", attiva dal 2000, con l'obiettivo di "avvicinare il medico e il paziente abbattendo le barriere socio culturali; aumentare la consapevolezza rispetto alle scelte sulla propria salute; promuovere la cultura medica per evitare l'autodiagnosi e l'autocura". All'interno del sito gli utenti, previa registrazione, possono aprire uno spazio di discussione e chiedere un consulto specialistico ai medici iscritti al sito¹⁴. Ai soli specialisti è permessa, inoltre, la pubblicazione di articoli divulgativi di vario genere inerenti all'ambito della salute (solitamente si configurano come "una indagine e rielaborazione di pubblicazioni scientifiche a livello nazionale ed internazionale e di contenuti prodotti dagli stessi professionisti iscritti [...] con l'assistenza tecnica del team di Medicialia")¹⁵. I consulti e gli articoli, corredati sempre di firma e di data, sono catalogati all'interno di sezioni tematiche che facilitano l'interrogazione (es. "donna", "uomo", "bambini", "sistema nervoso", "occhi", "orecchie e naso", ecc.)¹⁶.

Oltre ai consulti e agli articoli scientifici, è presente anche un *Dizionario medico*, di cui pure si è tenuto conto per la presente indagine. Il lemmario consiste di circa 4300 lemmi¹⁷, redatti da professionisti iscritti alla piattaforma, e può essere interrogato per lettera iniziale, per parola o per specialità clinica (tra quelle che contengono più termini [> 100] risultano in ordine alfabetico: andrologia [140], chirurgia generale [161], ginecologia e ostetricia [489], medicina generale [828], oculistica [371], odontoiatria e odontostomatologia [157], psichiatria

¹³ <https://www.medicialia.it/> (ultima consultazione 31/1/2025: tutti i risultati discussi nel presente lavoro sono aggiornati a questa data). Le citazioni all'interno del presente paragrafo che riguardano Medicialia sono tratte dalla pagina stessa.

¹⁴ Al momento dell'iscrizione sul sito i medici devono certificare "la loro identità e la loro professione". Ciò avviene attraverso l'invio da parte degli specialisti di un documento che attesti la loro iscrizione all'albo professionale.

¹⁵ Gli articoli devono sempre riportare "il riferimento all'autore, al revisore o alla fonte" e rispettare le linee guida del sito (a tal fine vi è una redazione scientifica che effettua operazioni di controllo e moderazione). Nella seguente dichiarazione relativa alle linee guida degli articoli "tutti i contributi scientifici sono editati da professionisti (medici specialisti, odontoiatri e psicologi iscritti), da Medical Writer o curati da Medicialia con il supporto di Opinion Leader esterni e supervisionati dalla Redazione Scientifica" appare interessante notare l'uso del termine *medical writer*. Si tratta di una locuzione sostantivale presa dall'inglese, ancora non registrata dalla lessicografia italiana, che indica una figura professionale che si occupa di redigere testi di ambito medico.

¹⁶ La sezione dei consulti, in particolare, si configura come una sorta di forum, tipologia di piattaforma digitale che generalmente viene fatta rientrare nella categoria della cosiddetta chat asincrona (Prada 2016, 361).

¹⁷ Il conteggio è stato fatto manualmente il 16/10/2024.

[367], psicologia [512], psicoterapia [102] e urologia [123]). La costituzione stessa del lemmario e le definizioni non rispettano i criteri scientifici della lessicografia. Basti pensare che “ogni termine medico inserito nel vocabolario può essere più volte ripetuto ed avere più definizioni”, non è mai esplicitata la categoria grammaticale, la definizione è a volte soltanto discorsiva (es. “*enucleazione* ‘il bulbo oculare viene completamente asportato per un tumore presente all'interno dell'occhio. Il volume orbitario viene ricostruito per non avere la caduta e l'infossamento delle palpebre’”) e il lemmario contiene anche voci non propriamente d'ambito medico (es. *facebook*, *salumi*, ecc.).

Al 31/1/2025 il sito vanta 478.374 utenti registrati (ma è possibile leggere e consultare le risposte anche senza registrazione), 9.831 specialisti certificati, 2.826.576 consulti e 7.608 articoli. Come solitamente avviene sui forum, blog o social networks, i medici con più recensioni, contenuti pubblicati e consulti erogati (nonché con più visualizzazioni) ottengono un simbolo di riconoscimento (si tratta di una coccarda arancione posta vicino al nome utente), una sorta di certificazione che garantisce la loro professionalità. I più meritevoli sono promossi poi a referenti scientifici, titolo che ha “carattere di onorificenza [...] funzionale alla gestione del sito ed allo sviluppo della comunità”.

Il sito non è strutturato come un *corpus* destinato a ricerche linguistiche e, dunque, non consente l'estrazione automatica dei termini e soprattutto non assicura la conservazione delle discussioni presenti, che possono essere in qualsiasi momento cancellate. Per la presente indagine si è fatto, dunque, ricorso alla barra di ricerca, dotata di poche funzioni utili al raffinamento¹⁸, che offre la possibilità di ricercare una o più parole all'interno dei consulti e degli articoli (la barra riporta la frase “Ricerca sintomo o malattia”). I risultati che si ottengono sul sito vengono mostrati sotto forma di lista, in ordine casuale, suddivisi in pagine, accompagnati da una breve anteprima. Il totale delle occorrenze è sempre riportato in alto.

L'indagine è stata così condotta: è stata ricercata prima la presenza nel sito di LI reperiti da altri studi sulla lingua della medicina¹⁹; sono stati poi inseriti nella barra di ricerca sostantivi e aggettivi latini indicanti parti del corpo, organi e malattie, declinati nei vari casi (es. per il sostantivo *piede* sono state immesse le forme *pes*, *pedis*, *pedi*, *pedem*, ecc.; per l'aggettivo *grave* le forme *gravis*, *grave*, *gravem*, ecc.); in aggiunta, sono state digitate nella barra di ricerca tutte le combinazioni possibili formate dal carattere jolly * seguito da una consonante o una vocale e dalla desinenza latina di ciascuna declinazione (es. **aae*, **bae*, **zae*; **aam*, **bam*, ... **zam*, ecc.)²⁰.

¹⁸ Si può ricercare per “sezione”, “area”, “genere”, “età” e per anno.

¹⁹ Si tratta degli studi citati nel § precedente.

²⁰ Per il nominativo di terza declinazione sono state cercate le varie uscite: es. *-bs/-ps*, *-x*, *-s*, *-ut*, *-m/-n*, *-or*, ecc. Questa modalità di ricerca, che al momento risulta disabilitata, non si è rivelata efficace per alcune desinenze latine (es. nominativo singolare di prima declinazione).

Tra i grandi limiti di questa modalità di ricerca rientra l'impossibilità di filtrare i risultati ed escludere alcuni termini (italiani, stranieri o sigle) che presentano terminazioni uguali alle parole latine (es. **ata* o **ato* mostra per lo più parole italiane così terminanti; **ris* ha un alto numero di occorrenze perché vi è *ris*, abbreviazione di *risonanza*). La ricerca, inoltre, non è *case sensitive*, non distingue l'accento e non dà modo di avere sequenze di parole tra loro consecutive: la combinazione di due parole rileva tutte le loro occorrenze a qualsiasi distanza nel testo (es. se si cerca *in vivo*, si troveranno tutte le occorrenze di *in* e di *vivo* sia consecutive, e quindi *in vivo*, sia lontane nel testo, es. *in questo modo non vivo più*).

Date queste premesse e la grande mole di consulti e articoli presenti nel sito, nei casi in cui un termine o una combinazione ha riportato molte occorrenze (es. la sequenza **tum* conta 3.868 risultati in 194 pagine), sono state spogliate soltanto le prime dieci pagine di risultati (in ogni pagina ci sono in media 22 risultati).

Nel *corpus* della ricerca sono state incluse le frasi e le locuzioni latine (es. *ipse dixit, senectus ipse/-a morbus*), che hanno costituito una singola entrata. Sono stati esclusi, invece, i latinismi parzialmente adattati alla struttura grafico-fonetica (es. *acne*, derivato dal lat. scient. *acne*)²¹; i nomi di medicinali, che sono spesso falsi latinismi e che possono accompagnarsi a termini inglesi nella formazione di una locuzione (es. *allergicum defence, arvenum, compendium, dicloreum, fastum gel, fluxum, imodium, nexium, prometrium, prostatricum, sibelium, sursum, valium*, ecc.); i nomi scientifici di funghi e alghe (*candida albicans, fucus, ganoderma lucidum*), di batteri e virus (es. *bifidobacterium longum, clostridium difficile, enterococcus, proteus mirabilis, staphylococcus aureus, treponema pallidum*) e di piante (es. *aconitum, chelidonium, gelsemium, ledum palustre, psyllium, pygeum africanum, viscum album*); i nomi di rimedi o composti omeopatici (es. *antimonium crudum, argentum nitricum, arsenicum album, calcium carbonicum, natrum muriaticum, nitricum acidum, oscillococcinum, plumbum aceticum, zincum metallicum*).

3. Analisi

Sono stati raccolti 319 LI per un totale di circa 121 mila occorrenze²². I LI sono stati analizzati in base al numero di attestazioni, alla categoria grammaticale,

²¹ Se il sostantivo compare all'interno di loc. sost. in cui l'agg. è un LI, allora si considera l'intera loc. come un LI e viene inclusa nel *corpus* (es. *acne aestivalis, acne vulgaris, alopecia barbae, angina pectoris* 'dolore parossistico del torace', *causa mortis*, ecc.). Questo non vale, naturalmente, per quei casi in cui il sost., seppur seguito da un agg. latino, non può essere considerato un LI, es. *diabete mellitus*.

²² Come specificato nel paragrafo precedente, il numero di occorrenze non è affidabile soprattutto nel caso delle locuzioni. Inoltre, non è possibile stabilire quante occorrenze contenga l'intero sito: per avere un'idea dell'ordine di grandezza, la parola grammaticale *per* ha 729.290 occorrenze, mentre una parola piena come *medico* 285.591.

alla declinazione originaria (nel caso dei nomi), all'ambito di appartenenza (patologia, anatomia, ecc.) e alla loro presenza in Z. Sono stati poi suddivisi in quattro categorie: 1. latinismi propri anche di altri ambiti o tratti da essi; 2. latinismi della medicina (o propri anche di altri ambiti o tratti da essi) entrati nella lingua comune e/o la cui presenza è attestata in Z; 3. latinismi tecnici collaterali; 4. latinismi tecnici specialistici.

In ultimo si riportano alcune considerazioni sui contesti d'uso in cui i LI appaiono nella comunicazione tra medici e pazienti, prendendo in esame la tipologia e la forma di LI utilizzata, lo scrivente e la finalità²³.

3.1. Caratteristiche dei LI

I LI più frequenti, ossia con un numero di occorrenze superiore a 500, sono *iter* (9.339), *die* 'al giorno' [PDS] (7.722)²⁴, *ictus* (6.470), *plus* (4.020)²⁵, *visus* 'capacità visiva dell'occhio' [Z] (2.716), *ratio* 'rapporto' (2.004)²⁶, *liquor* 'liquido, in particolare una soluzione acquosa che contiene una sostanza medicinale', 'termine riferito a determini liquidi corporei' (1.302), *lupus* 'membro di un gruppo di malattie cutanee in cui le lesioni sono caratteristicamente ulcerate' (1.191), *anulus* 'anello o struttura circolare' (1.089), *videat* 'esame o visita specialistica' [PDS] (697) e *angulus* 'punto in cui due confini o superfici intersecanti convergono' (634)²⁷. Tra i meno attestati, vale a dire con un numero di occorrenze inferiore alle 50, vi sono *facies* 'volto', 'superficie' (45), *quoad vitam* 'espressione usata nella formulazione di prognosi mediche, che fa riferimento alla sola sopravvivenza, senza pronunciarsi in merito al riacquisto della salute o a un recupero funzionale' [VT] (44), *morsus* 'morso della lingua provocato dall'epilessia' [def. propria] (36), *pancreas divisum* 'anomalia congenita del pancreas' [def. propria] (31), *defluvium capillorum* 'caduta dei capelli' [def. propria] (21), *caput medusae* 'vene cutanee dilatate attorno all'ombelico' (19), *substantia nigra* 'lo strato di sostanza grigia che separa il

²³ Il significato dei LI riportati nell'articolo verrà esplicitato solo se il termine risulta poco trasparente. In questi casi il significato deve considerarsi tratto dal DMD. Qualora il termine non dovesse essere presente nel DMD, verrà esplicitata la fonte da cui si prenderà la definizione tra parentesi quadre (PDS, VT, Z o "def. propria").

²⁴ Attestato anche nella forma *pro die* 'al giorno, ogni giorno, quasi esclusivamente nella posologia farmacologica e medica' [Z].

²⁵ Sul problema della pronuncia "all'inglese" cf. Gualdo (2022).

²⁶ Il termine ha anche un valore più tecnico nella medicina in quanto indica il rapporto "tra la capacità di coagulazione del sangue del paziente e la capacità di coagulazione di un sangue standard usato come riferimento" (Orletti e Iovino 2018, 48).

²⁷ A rigore rientrerebbero nella lista anche le loc. *lichen sclerosus* 'patologia cronica atrofica della cute' (970), *herpes simplex* (857), *angina pectoris* (704), *post partum* (667) e *fundus oculi* (645), ma i totali potrebbero essere sovrastimati per via delle problematiche legate, come già detto, al conteggio delle occorrenze nei casi delle locuzioni.

teg[u]mento del mesencefalo dal peduncolo cerebrale' (16), *coxa valga* 'deformità dell'anca' (14), *morsicatio buccarum* 'morsicatura della mucosa delle guance' [def. propria] (8), *os peroneum* 'piccolo osso accessorio del piede e della caviglia' [def. propria] (6), *pulvinar* 'parte prominente mediale dell'estremità posteriore del talamo' (5), *rigor mortis* (4), *pars tensa* 'parte del timpano' (3), *tinea barbae* (3), *foetor hepaticus* 'alito cattivo dovuto a una malattia del fegato' [def. propria] (2), *limen nasi* 'cresta che segna il confine tra il vestibolo del naso e la cavità nasale' (2) e molti LI con una sola occorrenza, es. *xantelasma palpebrarum* 'xantoma planare che colpisce le palpebre' (1).

I dati consentono alcune considerazioni: non stupisce la presenza tra le forme con più occorrenze di LI d'ambito non strettamente specialistico, la cui circolazione è attestata anche negli scritti di media-alta formalità (es. *iter*, *plus*), e di LI d'ambito medico ma considerabili di basso e medio specialismo (sulla classificazione si rimanda a Serianni 2008), es. *ictus*, *visus*, *liquor*, *lupus* (registrati nello Z) e *anulus*, *videat* e *angulus* (assenti in Z). I termini medici più frequenti riguardano malattie e disturbi avvertiti dai pazienti come potenzialmente pericolosi (di conseguenza l'utente cerca ulteriori pareri in rete oltre a quelli ricevuti dal proprio medico curante) o ritenuti imbarazzanti (in questo caso l'anonimato aiuta i pazienti a esporsi sul sito).

Tra le parole con meno occorrenze, invece, ricorrono sia termini di malattie o disturbi generalmente lievi per cui i pazienti che ne soffrono evitano di rivolgersi agli specialisti (es. *xantelasma palpebrarum*), sia termini il cui concorrente in un'altra lingua è più frequente (es. *anxietas tibiatarum* ha una sola occorrenza, *sindrome delle gambe senza riposo* circa 490).

Passando, invece, alle categorie grammaticali, risultano maggioritari le loc. sost. (50,47%), di cui il 47,2% di genere femminile e il 52,8% di genere maschile, e i sost. (26,01%), di cui il 30,12% di genere femminile e il 69,88% di genere maschile. Tra le prime si segnalano alcuni LI attestati soltanto al plurale (9 in totale), come ad es. *striae albae* 'smagliature bianche' [def. propria] e *striae rubrae* 'smagliature rosse' [def. propria]. Tra i secondi si segnalano gli pseudo-latinismi (almeno sul piano morfologico e semantico) *deficit*, propriamente terza pers. del pres. indic. di *deficere* 'venir meno, mancare', e *videat*, terza pers. del pres. cong. di *vidēre* 'vedere'²⁸. Piuttosto marginali sono le loc. agg. (7,84%), es. *prurito sine materia* 'senza causa' [def. propria], e le loc. avv. (6,58%), es. *per os* 'per bocca' [def. propria], a cui si sommano sei esempi di LI impiegati sia come loc. agg. che loc. avv. (es. *in utero*), un esempio di LI impiegato sia come loc. agg. che loc. sost. (*ab ingestis* 'a causa di materiali ingeriti', def. propria) e

²⁸ Si tratta di un tipo di formazione molto frequente nella lingua medica, che porta alla creazione dei cosiddetti "nomi cartellino" (Migliorini 1975; Thornton 2004). Su *videat* si rimanda a Orletti e Iovino 2018, 58; De Vecchis 2021).

un esempio di LI impiegato sia come loc. avv. che loc. sost. (*extrema ratio*). Si conteggiano, infine, 12 frasi (es. *primum non nocere* ‘per prima cosa non nuocere’, trad. propria), 2 avv. (*conquibus* ‘in soldoni’, def. propria, e *plus*, usato anche come sost.), 4 agg. (*mellitus* ‘mellito’, riferito a *diabete*, *motorium* e *vegetativum*, entrambi attestati in un referto di un’elettroencefalografia e probabilmente riferiti a *sistema*, *praeter* da *preternaturale* ‘artificiale’, Z), 3 prep. (*sine*, *sub* e *versus*) e un’abbreviazione (*gtt* per *guttae* ‘gocce’)²⁹.

Per quanto riguarda la declinazione di appartenenza dei sostantivi e delle locuzioni sostantivali (per cui è stata considerata la declinazione del sostantivo), dal cui conteggio si escludono gli pseudolatinismi *deficit* e *videat*, si ha il 19,92% dalla I declinazione (es. *coxa* ‘anca’); il 27,11% dalla II declinazione (es. *lupus*, *defluvium*); il 44,07% dalla III declinazione (es. *hallux*, *rigor*, *vis* ‘forza’); l’8,05% dalla IV declinazione (es. *aditus* ‘apertura o ingresso’, *habitus* ‘atteggiamento’, ‘fisico’, *genu* ‘ginocchio’, *veru* ‘collicolo seminale’); lo 0,42% dalla V declinazione (*die* e *facies*).

In merito all’ambito di appartenenza, la maggior parte dei LI designa malattie, disturbi o condizioni del fisico e della mente (33,23%), es. *hallux limitus* ‘alluce rigido’ [def. propria], o indica parti anatomiche (31,66%), es. *ostium primum* ‘apertura nella parte inferiore della membrana che divide il cuore embrionale nei lati destro e sinistro’³⁰. Segue poi l’ambito medico *lato sensu* (19,12%), in cui rientrano termini non riconducibili a una specializzazione specifica (es. *guttae*, *per os*, *quoad vitam*), a cui si può aggiungere un 11,91% di termini generici condivisi anche con altri ambiti (es. *facies*, *habitus*, *sine causa*). Sono stati conteggiati a parte tutti quei termini circostanziali e sentenze che fanno parte della lingua comune (13,48%).

La presenza dei LI nella lessicografia italiana, qui rappresentata da Z, raggiunge il 19,75% (63 termini). Si riscontrano sia termini in realtà appartenenti e circolanti in contesti non obbligatoriamente specialistici (es. *coram populo*, *in extremis*, ecc.), sia termini del linguaggio medico di basso o medio specialismo (es. *rigor mortis*, *ictus*, ecc.).

Si passa, infine, a esaminare la tipologia di latinismi riscontrati³¹. Questa la classificazione: [1] latinismi propri anche di altri ambiti o tratti da essi; si

²⁹ Il DMD registra anche *guttat.* per *guttatim* ‘goccia a goccia’.

³⁰ Piuttosto marginali, invece, ambiti quali la sessuologia (1,25%), es. *ante portam* ‘prima della penetrazione’, e la chirurgia (1,25%), es. *exenteratio orbitae* ‘rimozione dell’intero contenuto dell’orbita’ (s.v. *exenterazione*).

³¹ In Orletti e Iovino (2018, 48) l’analisi dei tecnicismi specifici riscontrati nel loro *corpus* ha permesso di documentare anche “alcuni dei procedimenti alla base della formazione del lessico tecnico”, come quello metaforico (es. *fovea*, che in latino significava ‘fossa’, è passato a indicare nel linguaggio medico la ‘depressione della retina’ [Z]).

tratta di 31 termini (9,75%), es. *primum movens* ‘causa primaria’, termine da ricondurre alla filosofia aristotelica, penetrato poi in altri àmbiti specialistici o in testi di alta formalità; [2] latinismi della medicina (o propri anche di altri àmbiti o tratti da essi) entrati nella lingua comune e/o la cui presenza è attestata in Z. Si tratta di 62 termini, corrispondenti al 19,50%, es. *ictus* ‘qualsiasi manifestazione morbosa che si manifesta in modo improvviso’ [Z]; [3] latinismi tecnici collaterali³²; sono 52 termini, ossia il 16,35%, es. *ante mortem* ‘che si verifica prima della morte’ o *per venam* ‘per via venosa’; [4] latinismi tecnici specialistici³³, in cui si distinguono due sottocategorie: [4a] latinismi tecnici specialistici facilmente sostituibili (per cui, cioè, l’alternativa italiana è ampiamente attestata): es. *claudicatio* ‘claudicazione, zoppia’. Sono il 31,76%; [4b] latinismi tecnici specialistici difficilmente sostituibili perché il termine latino ha un uso internazionale, è di alto o altissimo specialismo e non ha una traduzione italiana diffusa. Sono il 22,33%, es. *cauda equina* ‘insieme delle radici spinali che discendono dalla parte inferiore del midollo spinale e che occupano il canale vertebrale al di sotto del midollo’.

Nei gruppi [1] e [2] sono presenti sia latinismi circostanziali (es. *ad hoc*) sia latinismi comuni (es. *optimum* o lo pseudolatinismo *vademecum*), ossia quei termini che Zolli (1989, 43) considerava già parte del linguaggio comune. Serianni (2011, 138), più convincentemente, parlava di una lingua “ormai distante, semplicemente orecchiata” e non realmente padroneggiata. A questi si aggiungono sentenze, frasi e locuzioni cristallizzate che hanno lo scopo di innalzare il registro stilistico (*causa mali est ipse clamabit, conditio sine qua non, cum grano salis, hic et nunc*, ecc.)³⁴.

Nel gruppo [3] si trovano una serie di termini utilizzati a fini eufemistici, che sostanzialmente tabuizzano concetti legati alla morte (*exitus, obitus, senectus*) o a comportamenti e atti socialmente immorali o privati (*ante portam, potus* ‘alcolismo’). In Orletti e Iovino (2018, 57), ma anche in Gualdo e Telve (2011) e in Serianni (2005), viene più volte messa in rilievo la funzione dei “latinismi come strategia di sostituzione eufemistica per assolvere efficacemente a una funzione etica”.

Nel gruppo [4] la distinzione dei LI in due categorie impone alcune riflessioni. Nella prima [4a] la sostituzione del termine italiano a quello latino

³² Mette conto sottolineare la difficoltà, a volte, di collocare i LI nella categoria di tecnicismo collaterale [3] o di tecnicismo specifico [4a]. Spesso, infatti, un termine latino concorrente all’italiano, seppur dotato di un certo rigore terminologico, può essere utilizzato con il solo scopo di innalzare il registro stilistico: es. *cerebellum* vs. *cervelletto*.

³³ Orletti e Iovino (2018, 50) individuano in alcuni latinismi tecnici specifici “una certa plasticità funzionale e categoriale”, ad es. *fundus* “può essere utilizzato sia da solo sia in unione al genitivo che lo specifica (*fundus oculi*), in riferimento tanto all’anatomia [...] quanto alla procedura di analisi”.

³⁴ Nel gruppo [2] si trovano, inoltre, anche tecnicismi specifici, che si sono diffusi nella lingua comune e nella lessicografia: es. *lupus*.

non comporterebbe nessuna perdita di significato, data anche la vicinanza strutturale tra le due lingue, anzi abbasserebbe “l’aura di prestigio” senza discapito per l’elemento denotativo scientifico (Gualdo 2023, 15). Al contrario, in [4b], la sostituzione con termini italiani non circolanti nella comunità scientifica potrebbe ostacolare la comunicazione specialistica. Occorre infatti precisare che alcuni latinismi in [4a] e [4b], seppur sostituibili, rispondono a usi ed esigenze del settore medico e hanno una loro circolazione consolidata anche in inglese.

3.2. *Usi dei LI*

Tralasciando le caratteristiche generali della varietà di lingua utilizzata sul sito³⁵, in questo paragrafo si prendono in considerazione soltanto alcune modalità d’uso riguardanti i LI da parte degli utenti e degli specialisti.

La premessa obbligatoria a questo tipo di analisi riguarda il tipo di conversazione che si instaura tra medici e pazienti. Come ha messo ben in luce Orletti (2000, 2015, 2019; Orletti e Iovino 2018, 25)³⁶, questa relazione è un’“interazione asimmetrica istituzionale” che si caratterizza per “un notevole squilibrio” dovuto principalmente alla disparità sociale ed epistemica. In particolare, nella CMT si parla di “un contatto tra medico e paziente che passa attraverso la presenza di «strumenti di mediazione»”, che fungono da “terzo interlocutore” (Orletti e Iovino 2018, 31). Questo permette e facilita la comunicazione di argomenti che riguardano malattie e disturbi gravi o problemi che causano imbarazzo se affrontati di persona. D’altra parte, la CMT scritta consente agli stessi pazienti di articolare un messaggio più preciso e più approfondito, data anche la possibilità di riportare per iscritto referti o analisi.

È chiaro che l’asimmetria tra medici e pazienti è dovuta soprattutto alla scelta della lingua. A tal riguardo, si possono notare alcuni usi dei LI³⁷.

Il primo è un uso riportato dei LI e avviene per mano dei pazienti. Quando questi si rivolgono agli specialisti per avere un consulto, presentano il

³⁵ In questi generi di siti, simili ai forum online, solitamente si riscontra una lingua “poliedrica, fortemente legata all’argomento e alla comunità che si raduna a discutere intorno a un tema” (Biffi 2016, 59). In questo caso, trattandosi di un forum d’ambito medico in cui gli utenti si rivolgono agli specialisti, si ha un italiano neostandard che non di rado tenta alcune “risalite sulla china della diafasia” (Prada 2016, 345) attraverso l’inserimento di termini appartenenti a un registro più formale e ricercato. Non mancano, tuttavia, gli elementi tipici delle scritture digitali, come disgrafismi (uso dell’apostrofo in sostituzione dell’accento), tachigrafie (es. contrazioni, troncamenti), realizzazioni grafonetiche espressive (es. l’uso di <k>), logogrammi (es. x, +), ecc (cf. Prada 2016).

³⁶ Per ulteriore bibliografia si rimanda a Orletti e Iovino (2018, 25).

³⁷ Non sono stati presi in considerazione altri usi riguardanti il trattamento dei LI (es. riformulazioni, glosse esplicative, ecc.).

loro quadro clinico³⁸ o trascrivendo il referto medico o riportando quanto detto da altri specialisti. Questa modalità li costringe a inserire LI letti (es. *addendum* ‘aggiunta’ che si trova scritto in alcuni referti) o “orecchiati” (cf. Serianni 2011, 138). Tuttavia la scarsa padronanza con il linguaggio medico specialistico, sommata a un’inadeguata competenza nella lingua latina, porta a una serie di errori³⁹. Sul piano grafico si possono osservare grafie insolite, es. *spekulum* per *speculum*, in cui l’uso del grafema <k> per <c> (con valore di /k/), ampiamente usato nello scritto trasmesso giovanile (e non solo)⁴⁰, si potrebbe spiegare come un’interpretazione della parola come anglismo o, addirittura, grecismo⁴¹; mancate percezioni dei confini di parola, es. *lanulus* per *l’anulus* con concrezione dell’articolo, *ob torto collo* per *obtorto collo*, in cui una simile separazione può nascere dal fatto che è stato isolato *torto* da *obtorto*, part. perfetto di *obtorquēo*, in quanto unica parola italiana riconoscibile (si esclude un’interpretazione che tenga conto di un riconoscimento della preposizione latina *ob*); errori di accentazione: *diè* per *die* ‘al giorno’; semplificazioni dei dittonghi, sequenze vocaliche o nessi consonantici complessi o estranei all’italiano: *-ae > -e* (es. *gutte*), *oe > e* (es. *locus ceruleus*) *quoad > quod* o *quaod*, *iusta* per *iuxta*, *interposizio* per *interpositio*, *addomen* per *abdomen*, ecc. (riguardano anche la fonetica i casi di semplificazioni vocaliche in direzione dell’italianizzazione, es. *fondus* per *fundus*, *carenatum* per *carinatum*); scambi di consonanti, es. *contromalacia* per *condromalacia* (dal gr. χόνδρος ‘granello, cartilagine’ e μαλακία ‘mollezza’), in cui l’elemento formativo *condro-*, evidentemente sconosciuto, è stato rianalizzato sull’italiano *contro-* (se il parlante fosse meridionale, non si escluderebbe anche l’azione di spinte ipercorrettistiche dovute al dialetto soggiacente, in particolare al fenomeno della lenizione postnasale).

Per quanto riguarda la morfologia, si assiste a casi di cambi di declinazione: il fenomeno riguarda principalmente la sostituzione di desinenze poco note con altre più frequenti (es. passaggio dalla III in *-or* alla II in *-us*: *pavus nocturnus* per *pavor nocturnus*) oppure la confusione tra desinenze simili di declinazioni diverse (es. *-us* della IV e *-um* della II: *morsum* per *morsus*; *-u* neutro di IV e *-um* della II: *genum* per *genu*); cambi di genere, per cui *-o* del nominativo singolare di alcuni sostantivi femminili di III è interpretato come morfema grammaticale

³⁸ Sul cosiddetto “problema del paziente” e sulla co-costruzione dell’asimmetria in questo tipo di relazioni si rimanda a Orletti e Iovino (2018, 32–5).

³⁹ Alcuni errori si trovano anche nell’analisi di Orletti e Iovino (2018, 50–1).

⁴⁰ Prada (2016, 342); D’Achille (2019, 92).

⁴¹ Forse si può pensare a una semplice associazione di “lettera esotica” a parola altrettanto “esotica”, cioè con cui non si ha familiarità nella sua forma scritta. Nel resto del messaggio il fonema /k/ è reso con <c> o <ch> (es. la congiunzione *che*).

del maschile singolare, es. *il portio, il ratio, il fundoplicatio, il claudicatio*, ecc.⁴²; casi di invariabilità (es. *corpus* per *corpora*, *foramen* per *foramina*).

A livello lessicale si hanno scambi di termini tra lat. e lat. (es. *plenum* per *planum*; *microplasma* per *micoplasma*) e tra lat. e gr. (es. il lat. *algor* ‘freddo’ viene scambiato per il gr. *algos* ‘dolore’); latinizzazione impropria di termini greci: *lobus* invece del gr. *lobos* ‘lobo’, *calatio* invece dell’it. *calazio*, dal gr. *khalázion*, propr. dim. di *khálaza* ‘granello’; slittamenti di significato: es. *conquibus* propriamente ‘denaro’ (scherz. *cum quibus nummis*) utilizzato col significato di ‘in soldoni, dunque’ (es. “in conquibus le aritmia riscontrate sono [...]”, “e veniamo al conquibus”); malapropismi: es. *angelus* per *angulus* ‘parte anatomica dell’intestino’⁴³.

Il secondo uso riscontrato è l’uso citazionale dei LI da parte dei medici. Questo ricorre maggiormente nelle risposte degli specialisti e contribuisce ad affermare una certa superiorità epistemologica. La presenza di espressioni e locuzioni quali *post hoc ergo propter hoc*, *testis unus testis nullus* comprovano lo *status symbol* degli specialisti (Mazzini 2015, 118).

A questa finalità contribuisce naturalmente anche l’uso di tecnicismi collaterali e di tecnicismi specifici (Orletti e Iovino 2018, 53), che appaiono cospicui anche in un sito divulgativo come quello qui esaminato.

4. Conclusioni

Lo studio qui condotto ha mostrato alcune caratteristiche e alcuni usi dei LI all’interno della CMT, rappresentata nel dettaglio dai consulti e dagli articoli presenti sul sito Meditalia.

Si può innanzitutto dire che il latinismo “nudo e crudo” gode ancora di discreta salute in ambito medico. Il numero di LI appare, infatti, non del tutto trascurabile (si compone sia di LI strettamente specialistici sia di LI circostanziali o comuni) e la loro presenza riesce a infiltrarsi nella rete anche negli usi dei non specialisti. Gualdo e Telve (2011, 298) parlano opportunamente di una certa “pervasività” e “radicamento storico del latino nella lingua scientifica (anche inglese)”, che potrebbe “indurre qualcuno non solo a chiedersi se la lingua medica possa essere saldamente dominata senza la conoscenza del latino e del greco, ma anche a ritenere che l’attuale pressione dell’inglese potrà risultare un domani solo un *historical interlude*”. Probabilmente i tempi non sono ancora maturi per valutare l’entità della pressione angloamericana nel campo medico,

⁴² Cf. anche Orletti e Iovino (2018, 49).

⁴³ Alcuni di questi tratti potrebbero essere anche semplici refusi (es. *plenum* per *planum* o *angelus* per *angulus*). Si può tendenzialmente escludere questa ipotesi però nei casi in cui il termine errato è ripetuto più volte all’interno del messaggio.

ma alcuni cambiamenti in atto a favore dell'inglese e a discapito del latino si sono già verificati (si pensi ad esempio ai risvolti linguistici dovuti alla pandemia da COVID-19, su cui cf. almeno Pietrini 2021). Secondo Mazzini (2015, 115), infatti, si va incontro a un rischio reale: dato il costante calo nei manuali universitari e nelle pubblicazioni scientifiche dei LI dovuto anche alla perdita di iscritti negli studi classici delle scuole secondarie, il latino potrebbe assumere "quasi esclusivamente, una funzione ornamentale; [...] una sorta di soprammobile prezioso, eredità di famiglia, che abbellisce il mobilio moderno di una casa di antiche tradizioni" (Mazzini 1999, 292).

Per quanto riguarda le caratteristiche dei LI, i dati relativi alle categorie grammaticali, alle declinazioni o agli ambiti di appartenenza non sorprendono di certo. Si può, però, riflettere sulle tipologie di LI individuate: in particolare, il gruppo [3] dei tecnicismi collaterali e il gruppo [4a] dei tecnicismi specifici traducibili in italiano potrebbero rappresentare una categoria unica se l'uso che si fa dei LI del gruppo [4a] fosse finalizzato esclusivamente all'innalzamento del registro stilistico.

Per quanto riguarda gli usi dei LI, si concorda con quanto rilevato da Orletti e Iovino (2018, 50): "il lessico latino sembra [...] essere utilizzato, da un lato, in maniera meccanica, senza una precisa cognizione che si tratti di 'vero' latino; dall'altro lato, con una consapevolezza metalinguistica quasi eccessiva, che induce i medici a dare al lessico medico una patina latina non prevista". Si può aggiungere che vi è una scarsa padronanza della lingua latina da parte dei pazienti, soprattutto negli usi riportati e "orecchiati": in questi casi gli utenti commettono errori per lo più dovuti all'adattamento del latino al sistema linguistico italiano. Questo potrebbe comportare un cambiamento nel latino "medico" o in direzione di un'italianizzazione di termini latini (si ritorna al gruppo [4a]) o verso la sostituzione dei LI con termini provenienti da altre lingue di prestigio, come l'inglese, che, in alcuni casi, potrebbero risultare più facili da adoperare rispetto al latino (o all'italiano stesso?).

BIBLIOGRAFIA

- Bellina, Massimo. 2011. "medicina, lingua della". In *Enciclopedia dell'Italiano*, II, a cura di Raffaele Simone, 862–5. Roma: Istituto della Enciclopedia Italiana.
- Biffi, Marco. 2016. *Le parole nella Rete*. Roma: Gruppo Editoriale L'Espresso (*L'Italiano. Conoscere e usare una lingua formidabile*, VII, a cura dell'Accademia della Crusca e Repubblica).
- Carli, Augusto. 2006. "Una scienza – una lingua? I conflitti linguistici nella comunicazione scientifica". In *Zhì. Scritti in onore di Emanuele Banfi*, a cura di Gabriele Iannaccaro e Nicola Grandi, 177–88. Roma: Caissa Linguistica.

- D'Achille, Paolo. 2019. *L'italiano contemporaneo*. 4^a ed. Bologna: il Mulino.
- D'Achille, Paolo. 2023. "History of the Italian Lexicon". In *Oxford Encyclopedia of Romance Linguistics*, a cura di Michele Loporcaro, 1–27. Oxford: Oxford University Press.
- Dardano, Maurizio. 1994. "I linguaggi scientifici". In *Storia della lingua italiana*, II. *Scritto e parlato*, a cura di Luca Serianni e Pietro Trifone, 497–551. Torino: Einaudi.
- De Vecchis, Kevin. 2021. "Un *videat* linguistico". *Italiano digitale* 19, n. 4 (ottobre-dicembre): 6–8.
- De Vecchis, Kevin. 2023. "Le Parole della Salute. Un glossario in fieri". In *Lingua e patologia. Parole dentro parole fuori*, a cura di Francesca M. Dovetto, 471–80. Roma: Aracne.
- Dionigi, Renzo. 2023. *Le parole in chirurgia: dal Medioevo al SARS-CoV-2*. Milano, Udine: Mimesis.
- DMD = *Dizionario Medico Dorland. Edizione tascabile*. 27^a ed. Milano: Elsevier, 2006.
- Gualdo, Riccardo. 2009. "Linguaggi specialistici". In *XXI Secolo. Enciclopedia italiana*, 395–405. Roma: Istituto della Enciclopedia Italiana.
- Gualdo, Riccardo. 2022. "«Non se ne può plus!»: i (neo)latinismi e la loro pronuncia". *Italiano digitale* 20, n. 1 (gennaio-marzo): 110–3.
- Gualdo, Riccardo. 2023. "Prefazione linguistica". In *Le parole in chirurgia: dal Medioevo al SARS-CoV-2*, di Renzo Dionigi, 14–7. Milano, Udine: Mimesis.
- Gualdo, Riccardo, e Stefano Telve. 2011. *Linguaggi specialistici dell'italiano*. Roma: Carocci.
- Mazzini, Innocenzo. 1999. "Appunti per una storia del latino dei medici dal Rinascimento ai nostri giorni. Saggio di indagine". *Annali della Facoltà di Lettere e Filosofia dell'Università di Macerata*, 32: 289–315.
- Mazzini, Innocenzo. 2015. "Greco-latino e inglese nella lingua medica italiana contemporanea: convivenza pacifica o sopraffazione?". *Lingue antiche e moderne*, 4: 113–31.
- Migliorini, Bruno. 1975. "I nomi-cartellino". *Archivio Glottologico Italiano*, 60: 222–7.
- Orletti, Franca. 2000. *La conversazione diseguale. Potere e interazione*. Roma: Carocci.
- Orletti, Franca. 2015. "Le forme dell'autodiagnosi: negoziazioni dell'asimmetria di conoscenze fra medico e paziente". In *Tra linguistica medica e linguistica clinica. Il ruolo del linguista*, a cura di Franca Orletti, Anna Cardinaletti e Francesca M. Dovetto. *Studi Italiani di Linguistica Teorica e Applicata* 44: 546–60.
- Orletti, Franca. 2019. "Latin as a tool for social differentiation and epistemic asymmetry". *Language and Dialogue* 9, n. 1: 106–24.
- Orletti, Franca, e Rossella Iovino. 2018. *Il parlar chiaro nella comunicazione medica. Tra etica e linguistica*. Roma: Carocci.
- Pietrini, Daniela. 2021. *La lingua infetta: l'italiano della pandemia*. Roma: Treccani.
- Piro, Rosa. 2022. *L'italiano della medicina*. Roma: Carocci.
- Prada, Massimo. 2016. "Lingua e Internet". In *La lingua italiana e i mass media*, a cura di Ilaria Bonomi e Silvia Morgana, 333–84. Roma: Carocci.
- Sabatini, Francesco. 1962. "Una scritta in volgare amalfitano del secolo XIII". *Studi di filologia italiana* 20: 13–30.
- Serianni, Luca. 2005. *Un treno di sintomi. I medici e le parole: percorsi linguistici nel passato e nel presente*. Milano: Garzanti.
- Serianni, Luca. 2008. "Terminologia medica: qualche considerazione tra italiano, francese e spagnolo". *Studi di lessicografia italiana* 25: 254–69.

- Serianni, Luca. 2011. "Il latino nella scuola e nella società oggi. Riflessioni di uno storico della lingua italiana". In *Latinum est, et legitur... Metodi e temi dello studio dei testi latini*. Atti del Convegno Arcavacata (Rende, 4-6 novembre 2009), a cura di Raffaele Perrelli e Paolo Mastandrea, 137-50. Amsterdam: Hakkert.
- Thornton, Anna Maria. 2004. "Conversione". In *La formazione delle parole in italiano*, a cura di Maria Grossmann e Franz Rainer, 499-533. Tübingen: Niemeyer.
- Trifone, Maurizio. 2024. *L'eredità latina nel lessico italiano*. Firenze: Franco Cesati.
- VT = *Vocabolario Treccani*, <http://www.treccani.it/vocabolario/> (consultato tra il 10/01/2025 e il 31/01/2025).
- Z = *Lo Zingarelli 2025. Vocabolario della lingua italiana di Nicola Zingarelli*, rist. della 12^a ed., a cura di Mario Cannella, Beata Lazzarini e Andrea Zaninello. Bologna: Zanichelli, 2024 (con DVD).
- Zolli, Paolo. 1989. *Come nascono le parole dialettali*. Milano: Rizzoli.

L'HÉRITAGE LEXICAL LATIN COMMUN À L'ANCIEN FRANÇAIS ET AU ROUMAIN : MOTS COMMENÇANT PAR LES LETTRES A, B ET C

Adrian CHIRCU¹ 

Article history: Received 30 January 2025; Revised 20 February 2025; Accepted 15 March 2025; Available online 12 December 2025; Available print 30 December 2025

©2025 Studia UBB Philologia. Published by Babeş-Bolyai University.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ABSTRACT. *The Latin Lexical Heritage Common to Old French and Romanian: Words Beginning with the Letters A, B and C.* In our study, we propose to discuss some words of Old French (lexemes beginning with the letters A, B, and C) that have disappeared over time, being replaced by synonyms, internal lexical forms, or borrowings. These words were inherited from Latin, but for various reasons, sometimes difficult to discern, they have been lost or have become dialectal. The lexical items in question were mainly part of the fundamental vocabulary of French. After their identification, the words attested in French will be related to those passed down to Romanian. This fact illustrates that the two Romance languages were much closer in the past, with Romanian remaining much more conservative even today. To support our investigation, we relied heavily on the language facts inventoried by d'Algirdas Julien Greimas (DAF) regarding French and on some representative dictionaries (DLR, PEW, DELR) regarding Romanian. Such types of analyses help us to better understand the evolution of the Romance languages, as well as the lexical unit of the Romance languages.

Keywords: *Latin, Old (French), Romanian, comparative analysis, lexis, dictionaries, lexical unit*

¹ **Adrian CHIRCU** est Professeur des universités HDR en linguistiques roumaine et romane et il assure, depuis 1996, des cours à l'Université « Babeş-Bolyai » de Cluj-Napoca (Roumanie). Entre les années 2000-2004, il a été lecteur de roumain à l'Université de Provence, Aix-Marseille I (France), devenue depuis 2012 partie intégrante d'Aix-Marseille Université. Ses travaux de recherche portent principalement sur différents aspects de la linguistique, synchroniques, diachroniques et comparatifs (langues romanes). Courriel : adrian.chircu@ubbcluj.ro.

REZUMAT. Moștenirea lexicală latinească comună vechii franceze și românei: cuvinte care încep cu literele A, B și C. În studiul nostru, ne propunem să discutăm pe marginea unor unități lexicale prezente în vechea franceză (lexeme ce încep cu literele A, B și C), care, de-a lungul timpului, au dispărut, fiind înlocuite prin sinonime, prin creații lexicale interne ori prin împrumuturi. Cuvintele avute în vedere au fost moștenite din latină, însă, din diverse motive, uneori greu de întrevăzut, s-au pierdut sau au devenit dialectale. Este vorba despre elemente lexicale care, în principal, au făcut parte din vocabularul fundamental al francezei. După identificarea lor, cuvintele atestate în franceză vor fi puse în legătură cu cele transmise limbii române, ilustrând faptul că cele două limbi romanice erau mult mai apropiate în trecut, româna rămânând mult mai conservatoare chiar și în zilele noastre. În sprijinul investigației noastre, pentru limba franceză, ne-am bazat în bună măsură pe faptele de limbă inventariate de către d'Algirdas Julien Greimas (DAF), iar, în ceea ce privește româna, pe unele dicționare reprezentative (DLR, PEW, DELR). Asemenea analize, precum cea de față, ne ajută să înțelegem mult mai bine evoluția limbilor romanice, ca, de altfel, și unitatea lexicală de odinioară a acestora.

Cuvinte-cheie: *limba latină, limba franceză (veche), limba română, studiu comparativ, lexic, dicționare, unitate lexicală*

0. Préliminaires

Nous assistons ces dernières années, en linguistique à un revirement concernant l'étude systématique des anciens faits de langue. Ceux-ci sont souvent susceptibles de confirmer ou d'infirmer les changements survenus le long des siècles, ainsi que d'illustrer des tendances évolutives moins perceptibles au premier abord.

1. Points de vue sur le français et sur le roumain

1.1. Compte tenu de ce repositionnement, nous avons considéré qu'il est opportun d'aborder certains aspects (d'ordres lexico-sémantique et étymologique) assez rarement pris en compte par les spécialistes des langues romanes.

Ainsi, nous pouvons mieux comprendre les similitudes entre l'ancien français et le roumain, deux idiomes romans particuliers, qui, selon Michel Banniard, sont « des langues d'évolution forte » (Banniard 1997, 39) et sont cataloguées par Martin-Dietrich Glessegen comme « langues par écart et par élaboration », à côté de l'italien, de l'espagnol et du portugais, vu leur tradition d'écriture.

1.2. Si le français a été soumis aux emprunts latins savants majeurs et à des normalisations successives (Bertrand 2011, 71–109), le roumain a connu un nombre important d'influences étrangères (y compris française) par l'intermédiaire desquelles il a enrichi son vocabulaire de façon significative (Sala 1999, 99–115).

1.3. Malgré cela, le roumain reste une langue romane isolée du continuum roman, ce qui fait qu'il n'est pas un idiome qui a beaucoup évolué, mais plutôt une langue assez conservatrice par rapport au latin, préservant un nombre important d'unités lexicales (Pușcariu 1937, 33–39).

Selon le linguiste clujois,

« notre langue [le roumain] apparaît comme plus conservatrice que les autres langues romanes : elle conserve des éléments qui ailleurs ont disparu et ont été remplacés par d'autres. D'autre part, les innovations qui, lors de la séparation des deux groupes, existaient en germe mais ne s'étaient pas encore développées, se trouvent chez nous comme chez les Romains d'Occident, mais avec un aspect différent » (Pușcariu 1937, 28).

1.4. Cette spécificité du roumain a été d'ailleurs observée il y a presque un siècle par A. Iacobeanu qui soutenait que

« le fait que beaucoup de mots du latin populaire, conservés presque intouchés en roumain, existants aussi en ancien français, ne se trouvent plus en français d'aujourd'hui met encore plus en évidence le conservatisme et le traditionalisme de la langue roumaine, qui a préservé intouché l'ancien trésor lexical latin, en opposition avec le français, qui, menant une vie plus intense, plus vive, tant à travers la langue parlée que la langue écrite, a vite évolué, ayant subi beaucoup de changements » (Iacobeanu 1932, 3).²

1.5. Les auteurs des études comparatives franco-roumaines ont plutôt décrit les stades actuels de langue, même si les spécialistes auraient pu tirer profit des données linguistiques offertes par les anciens textes et la comparaison pourrait être plus profitable.

À ce propos, Ovide Densusianu affirmait que « plus nous descendons profondément dans le passé, plus les ressemblances deviennent plus nombreuses »

² « Faptul că multe cuvinte vechi din latina populară, păstrate aproape neatins în limba română, existente și în franceza veche nu se mai găsesc în franceza de astăzi, pune mai mult în evidență conservatismul și tradiționalismul limbii române, care a păstrat mai neatins tezaurul lexical vechi latin, în opoziție cu limba franceză, care, ducând un trai mai intens, mai viu, atât prin grai, cât și prin scris, a evoluat cu mare iuteală, suferind mai multe schimbări » (notre traduction).

(Densusianu 1902, 23) et A. Iacobeanu observait qu'« en lisant des textes français datant du siècle de milieu, nous sommes frappés par les similitudes de nature morphologique, syntaxique et lexicale existant entre le français de l'époque et le roumain d'aujourd'hui » (Iacobeanu 1932, 3).³

2. L'héritage lexical latin

2.1. Tenant compte de ces remarques préliminaires, nous nous proposons de nous attarder dans cette étude sur quelques mots (commençant par les lettres A, B et C) présents en ancien français et qui, au fil des siècles, ont disparu pour diverses raisons.⁴

Ceux-ci seront mis en relation avec ceux répertoriés en roumain, afin d'illustrer le parcours évolutif commun qui caractérisait les deux langues il y a quelques siècles. Même si une telle démarche a ses limites, dues au décalage temporel, nous considérons qu'elle est pertinente pour illustrer le patrimoine lexical commun.

Les mots français concernés ont été hérités du latin mais, pour diverses raisons, parfois difficiles à entrevoir, ils sont sortis d'usage. Il s'agit principalement de mots faisant partie du vocabulaire fondamental mais qui ont été remplacés par des mots empruntés ou par des mots formés au sein de la langue française.

À ce propos, Édouard Bourciez remarque que « la langue française du Moyen Âge connaissait encore beaucoup de mots d'origine latine, tombés plus tard en désuétude, et c'est là ce qui fait en partie la richesse de sa nomenclature » (Bourciez 1967 : 315). À son tour, Gaston Zink tient à préciser que « le vocabulaire de l'ancien français comprend, à la fin du XIII^e siècle, toutes régions et échantillons de langue confondus, entre 20 000 et 25 000 mots [...], tirés du latin à 95 % » (Zink 2000 : 102).

2.2. À l'appui de notre investigation, pour le français, nous avons fait principalement appel au dictionnaire d'Algirdas Julien Greimas, *Dictionnaire de l'ancien français* (DAF), qui contient un inventaire assez riche et donc adéquat pour une telle entreprise et, pour le roumain, nous avons consulté des ouvrages lexicographiques spécialisés (DLR, PEW et DELR). Dans des situations particulières, nous nous sommes aussi reportés à certains dictionnaires de référence pour l'étude des langues romanes (DILF ou REW).

³ « Citind texte franceze din veacul de mijloc, suntem izbiți de asemănările de natură morfologică, sintactică și lexicală dintre limba franceză din acea epocă și limba română de astăzi » (notre traduction).

⁴ Pour plus de détails concernant la naissance et l'évolution du lexique français, voir Alin Ray *et alii* (2011, *passim*) et Quentel (2018, *passim*).

Vu l'ampleur d'une telle recherche, nous avons limité la description des faits de langue aux mots commençant par les lettres A, B et C, en nous proposant de poursuivre l'analyse dans d'autres études de la même facture.

3. Anciens mots français et leurs correspondants roumains

Pour notre analyse, en premier lieu, nous avons attentivement parcouru les colonnes de l'ouvrage lexicographique antérieurement mentionné (DAF), en dressant une liste des mots susceptibles d'être d'intérêt pour notre démarche analytique. Les unités lexicales retenues pour notre analyse seront discutées ci-dessous.

3.1. La présentation de mots concernés débute avec l'adjectif *albe/aube* (DAF s.v.), dont l'étymon est ALBUS 'blanc' (fém. ALBA). En roumain, sa présence est constante et commune, *alb* (DER s.v.). En français, celui-ci a été remplacé par *blanc*, dont les origines remontent au germ. *BLANK 'brillant'.

3.2. *Almaille* 'le gros bétail, une bête à cornes' (DAF s.v.; VAF s.v.) renvoie sans doute au pluriel latin ANIMALIA et doit être mis en relation avec le mot roumain dialectal (Banat) *nămaie* 'vită cornută/bête à cornes' (PEW 1151; REW 476; DLR s.v.; « bêtes [...] à cornes, appelées aussi ailleurs *nămaie* »⁵ – Pușcariu 1940, 351).

3.3. *Altrier* (DAF s.v.) est un composé adverbial ayant le sens 'avant-hier' et qui est présent aussi en roumain *alaltăieri* (DLR s.v. – un composé *alalt*, -ă + *ieri*; DELR s.v. – *ILLUM/ILLE + ILLUM/ILLE + ALTER) et en italien *altrièri/ieri* *l'altro*, ce qui nous fait penser à un syntagme déjà présent en latin. Celui-ci connaît d'ailleurs en français une forme articulée, *l'altrier* 'l'autre jour, naguère' (VAF s.v.).

3.4. *Amortir*, -er (DAF s.v.) renvoie au verbe latin *ADMORTIRE, un dérivé du nom latin *mors*, -is 'mort' (DAF s.v.) et nous devons le mettre en relation avec le verbe roumain *a amorți* (dans PEW 83, est proposé comme étymon *AMMORTIO, -IRE).

3.5. *Antain* et *ante* 'tante' sont deux formes nominales (cas régime et cas sujet) – DAF s.v.; VAF s.v.. Elles doivent être rapportées au lat. AMITA qui est aussi à l'origine du nom de parenté roumain *mătușă* 'tante', qui est, en fait, une forme dérivée (PEW 1054) à l'aide d'un suffixe (-ușă) provenant du substrat traco-dace.

⁵ « vitele [...] cornute, numite pe alocuri și *nămaie* » (notre traduction).

3.6. L'origine de *aprof*, var. *apruef* et *aproef* (DAF s.v.), est sans doute latine (AD + PROPE), les formes ayant des valeurs sémantiques diverses en ancien français (valeur adverbiale temporelle 'ensuite, après' et prépositive 'après' (VAF s.v.): *Aprof le vendredi Fut faiz samedi* – Ph. de Thaun). Nous rencontrons en roumain un adverbe similaire (*aproape* – PEW 101 ; REW 6781 ; DLR s.v. ; DELR s.v.) qui a lui aussi des valeurs modales et de lieu (*Costă aproape zece lei* 'il coûte presque dix lei' ou *Locuiește aproape* 'il habite à proximité').

3.7. Intéressant reste pour notre étude *arain* (var. *airain*) provenant du lat. *ARAMEN < AERAMEN (DAF s.v.), qui a été concurrencé, en français, par le terme *bronze*. En roumain, nous avons *aramă* (PEW 107 ; REW 242 ; DELR s.v.) qui est généralement répandu.

3.8. Un autre mot qui a retenu notre attention est *araine* 'sable, rivage de la mer' (< lat. ARENA 'sable') que nous devons comparer avec le roum. *arină* (PEW 120 ; REW 630 ; DELR s.v.), qui a une signification similaire et qui est signalé en Crișana (TDR 1984, 312). Nous mentionnons l'adjectif *arenos* 'sablonneux' qui a lui aussi un correspondant en roumain, *arinos* 'nisipos' < lat. ARENOSUS, -A, -UM (PEW 122).

3.9. *Ardoir* et *ardre* 'brûler' (*Ardeir en feu ou en aive neier* – Cour. Louis ; DAF s.v.) renvoient au lat. ARDERE et ont comme correspondant en roumain le verbe *a arde* (PEW 114, DELR s.v.), qui connaît un usage généralisé (cf. afr. *arsure* 'brûlure' – VAF s.v.).

3.10. Comme le verbe roumain *a ațâta* (PEW 163 ; DELR s.v.), le verbe *aticer* 'attiser ; exciter, provoquer' (DAF s.v.) revoie au lat. *ATTITIARE et a des significations semblables.

3.11. Un verbe qui a retenu notre attention est *auner* 'assembler, réunir ; entasser ; unir, marier' (DAF s.v.; *La gens d'Orléans avec lui s'aūna*. – Aliscans – VAF s.v.) qui a un correspondant en roum. *a aduna* 'ramasser, récolter, cueillir, recueillir, etc.' (PEW 31 ; DLR s.v.), les deux descendant du lat. *ADUNIRE ou de ADUNARE (un composé de *unus*).

3.12. Le nom de parenté *ave* 'aïeul, grand-père' (DAF s.v.) a ses origines dans lat. AVUS ou lat. AVIUM (REW 839), dont une trace se trouve aussi en roumain dialectal (Oltenie) *auș* 'vieillard' (**au* + *-uș*), aroum. et méglénoroumain (PEW 173, DLR s.v. ; DELR s.v.).

3.13. L'adjectif français *balçant*, -cens, -sain 'blanc et nor, tacheté, pie' (*Cheval pie*) a ses origines dans lat. *BALTEANUM 'garni d'une ceinture rayée', un dérivé de BALTEUS 'ceinture, bande' (REW 919 ; DAF s.v.). En roumain, nous avons *bălțat* ' (en parlant d'un animal) bigarré, bariolé, tacheté, rayé' qui renvoie au lat. BALTEATUS, -A, -UM 'centuré' (DLR s.v. ; DELR s.v. *balț*).

3.14. *Blandir* est un verbe qui signifiait en ancien français 'caresser ; flatter' (*Elle beisa lo cent foiz et cent Molt lo blandit et losanja* – Eneas ; DAF s.v.) et qui est hérité du lat. *BLANDIO, -IRE = BLANDIOR, -IRI 'flatter, caresser, charmer' (cf. aussi VAF s.v.). Celui-ci a comme correspondent en roumain *a îmblânzi* 'caresser, apaiser, calmer ; dresser, apprivoiser' (PEW 779 ; REW 1149 ; DLR s.v.) qui est d'usage à nos jours. En ancien français, il existait aussi l'adjectif *bland* (cf. roum. *blând*).

3.15. Le français *bot* 'outre ; grosse bouteille ; vase pour servir à table les liquides' (DAF s.v. ; 'nabot' – VAF s.v.). Le mot a un correspondant en roumain : *bute* 'tonneau, botte, boute', les deux renvoyant au lat. pop. *BUTTEM (DAF s.v.) / BUTTIS (PEW 241 ; REW 1427 ; DELR s.v. ; DLR s.v.).

3.16. L'adverbe français *buer*, var. *boer*, *bor*, *bur* (DAF s.v. ; VAF s.v.), signifiait 'bien, heureusement ; avec raison, à propos' et il est à l'origine un composé (< lat. BONA HORA ?) qui a un correspondant en roumain, respectivement *bunăoară* 'par exemple, disons' (DLR s.v.).

3.17. Le verbe *cener* 'souper' (DAF s.v. ; REW 1808) doit être mis en relation avec le verbe roumain *a cina* 'souper, donner à souper' (PEW 369 ; DELR s.v. ; DLR s.v.), les deux ayant comme étymon le lat. CENO, -AVI/CENO, -ARE 'souper'.

3.18. Intéressant, surtout pour le roumain, est le nom français *cercel/cercele* 'cercle ; chandelier ; tout ce qui a la forme d'un cercle' (DAF s.v.), qui a comme étymon lat. CIRCELLUS 'petit cercle ; boudin', un dérivé de CIRCUS 'cercle'. Son équivalent roumain est *cercel* 'boucle d'oreille ; anneau, chaîne' (PEW 343 ; REW 1939 ; DLR s.v. ; DELR s.v.).

3.19. Un autre mot qui a retenu notre attention est *chael*, *cheel* 'petit chien, petit d'un animal en général' (DAF s.v. ; 'enfant' – VAF s.v.) ; voir aussi *chaele* 'petite chienne' (lat. CATELLA). L'étymon latin est CATELLUS (CATULUS) 'petit chien'. Ce mot latin s'est aussi transmis au roumain où il connaît la forme *cățel* 'petit chien ; petit d'un animal sauvage'. Par rapport au français, le mot roumain a acquis de nombreuses significations 'nom d'amitié donné à un nourrisson qui

mord la mamelle en tétant ; flatteur intrigant, etc.’ (voir PEW 314 ; REW 1763 ; DLR s.v.. DELR s.v.). Quant à la forme de féminin, en roumain elle est *cățea* ‘petite chienne, chienne, etc.’ Vu que les significations s’entremêlaient et prêtaient à confusions, en roumain a été créée une autre forme diminutive *cățelușă*.

3.20. *Chalchier* (var. *cauchier*) ‘fouler aux pieds, pressurer ; marcher, piétiner ; couvrir la femelle’ est aussi répertorié dans le dictionnaire de Greimas (*chaucier sur serpenz et sur escorpiuns* – Garn. ; DAF s.v.). L’étymon latin est sans doute CALCARE ‘fouler, marcher sur qqch ; piétiner, comprimer en foulant [la terre] ; (fig.) fouler aux pieds’. Un descendant de celui-ci se trouve aussi en roumain. Il s’agit de *a călca* (PEW 254 ; REW 1491 ; DELR s.v.) qui a une multitude de significations ‘mettre le pied (sur) ; passer (sur) ; mettre les pieds, entrer ; passer (par) ; fouler, écraser (en marchant ou en passant sur...), etc.’ (DLR s.v.).

3.21. Le nom féminin français pluriel *chanes* (var. *chaines*) avait la signification ‘cheveux blancs’ (DAF s.v.) et, de point de vue étymologique, il est un continuateur du lat. CANAS, formé à partir de CANUS ‘blanc, d’un blanc brillant, argenté [en parl. des choses], blanc [en parl. des cheveux, de la barbe]’. En roumain, nous trouvons l’adjectif *cărunt* ‘à cheveux gris, chenu’ (< lat. CANUTUS ; PEW 301 ; DLR s.v. ; DELR s.v.) et en aroum. *cănut*, les deux adjectifs correspondant au fr. *chanu* ‘qui a des cheveux blancs’. Il faut aussi mentionner le verbe *chanuer/chanuir* ‘avoir des cheveux blancs’ qui a, en roumain, un correspondant *a (în)cărunți* ‘grisonner’ (DLR s.v.).

3.22. Le mot français *cive* ‘ciboulette’ (DAF s.v.) provient du lat. CAEPA (CEPA) ‘oignon’ (REW 1817) qui connaît un descendant aussi en roumain, respectivement *ceapă* (PEW 327 ; DLR s.v. ; DELR s.v.).

3.23. L’ancien français *corb* (var. *corp*, *corf*) ‘corbeau’ (DAF s.v. ; REW 2269) est plus près formellement du nom roumain *corb* < lat. CORVUS ‘corbeau’ (PEW 399 ; DLR s.v. ; DELR s.v.) que l’est son dérivé *corbeau*.

3.24. *Cuidier* ‘penser ; croire, s’imaginer ; prétendre, avoir de la présomption ; être sur le point de, manquer de’ (DAF s.v. ; VAF s.v.) doit être mis en relation avec le verbe roumain *a cugeta* ‘penser, réfléchir ; penser à ; réfléchir, hésiter ; penser, se dire que...’ (PEW 431 ; DLR s.v. ; DELR s.v.), les deux mots se trouvant en étroit lien avec le lat. COGITARE ‘remuer dans son esprit ; penser, songer, se représenter par esprit ; méditer, projeter ; avoir des pensées, des intentions bonnes, mauvaises à l’égard de qqn’ (DILF s.v.).

4. Conclusions

4.1. Une telle analyse comparative des lexèmes que nous avons menée nous a permis de mieux comprendre l'évolution du vocabulaire non seulement de deux langues romanes mais aussi du latin. Les ressemblances relevées témoignent de l'unité lexicale qui caractérisait les divers stades de langue et du fait que le vocabulaire roumain s'avère très conservateur par rapport à celui du français, malgré le passage des siècles.

4.2. Ainsi, nous avons pu constater que, dans bon nombre de cas, les différences formelles, sémantiques et lexicales ne sont pas significatives. Elles illustrent le fait que le roumain a préservé la plupart des mots hérités du latin jusqu'à nos jours et que, malgré les influences subies, il a gardé mieux sa latinité. Par contre, le français de nos jours semble être plus évolué, en innovant et en trouvant d'autres solutions pour compenser les disparitions lexicales répertoriées en ancien français.

BIBLIOGRAPHIE

- Banniard, Michel. 1997. *Du latin aux langues romanes*. Paris : Éditions Nathan.
- Bertrand, Olivier. 2011. *Histoire du vocabulaire français. Origines, emprunts et création lexicale*. Paris : Éditions de l'École Polytechnique.
- Bourciez, Édouard. 1967. *Éléments de linguistique romane*, cinquième édition révisée par l'auteur et par les soins de Jean Bourciez. Paris : Éditions Klincksieck.
- DAF = Algirdas Julien Greimas. 2001 [1979]. *Dictionnaire de l'ancien français*. Paris : Éditions Larousse-Bordas/HER.
- DELR = Marius Sala, Andrei Avram, Ion Giurgea et Cristian Moroianu (coord.). 2012-2024-2015-2018. *Dicționarul etimologic al limbii române*, vol. I. A-B, vol. I. B (serie nouă), vol. II. C₁ (Partea 1. Ca-Cizmă) – C₂ (Clac-Cyborg). București : Editura Academiei Române.
- Densusianu, Ovid. 1902. *Filologia romanică în universitatea noastră. Lecțiuni de deschidere*. București : J.V. Soccec, 1902.
- DILF = Felix Gaffiot. 1967[1934]. *Dictionnaire illustré latin-français*. Paris : Librairie Hachette.
- DLR = Marius Sala, Gheorghe Mihăilă et Monica Busuioc (coord.). 2010 [1913]. *Dicționarul limbii române*, ediție anastatică (DA-DLR), tomurile I. A-B, vol. II-C. București : Editura Academiei Române.
- Glessegen, Martin-Dietrich. 2007. *Linguistique romane. Domaines et méthodes en linguistique française et romane*. Paris : Éditions Armand Colin.

- Iacobeanu, A. 1932. *Tradiționalismul limbii române – comparație cu limba franceză*, partea I-a. *Lexicul*. Botoșani : Tipografia B. Saidman.
- PEW = Sextil Pușcariu. 1975 [1905]. *Etymologisches Wörterbuch der rumänischen Sprache*, zweite. unveränderte Auflage. Heidelberg : Carl Winter Universitätsverlag.
- Pușcariu, Sextil. 1937. *Études de linguistique roumaine*. Cluj–București : Monitorul Oficial și Imprimeriile Statului & Imprimeria Națională.
- Pușcariu, Sextil. 1940. *Limba română*, vol. I. *Privire generală*. București : Fundația pentru Literatură și Artă « Regele Carol II ».
- Quentel, Gilles. 2018. *La genèse du lexique français. La formation du lexique français des origines au Moyen-Âge*. Berlin : Peter Lang GmbH.
- Rey, Alain, Frédéric Duval et Gilles Siouffi. 2011. *Mille ans de langue française, histoire d'une passion*, vol. I. *Des origines au français moderne*. Paris : Éditions Perrin.
- REW = W. Meyer-Lübke. 2009 [1911]. *Romanisches Etymologisches Wörterbuch*, 7. unveränderte Auflage. Heidelberg : Universitätsverlag Winter.
- Sala, Marius. 1999. *Du latin au roumain*. Traduction de Claude Dignoise. Paris–Bucarest : Éditions L'Harmattan & Univers Enciclopedic.
- TDR = Valeriu Rusu (coord.). 1984. *Tratat de dialectologie românească*. Craiova : Editura Scrisul Românesc.
- VAF = Magali Rouquier. 1992. *Vocabulaire de l'ancien français*. Paris : Éditions Nathan.
- Zink, Gaston. 2000. *L'ancien français*, cinquième édition mise à jour. Paris : Presses Universitaires de France.

SUR LES TRACES DU LAT. HABERE DANS LA MORPHOSYNTAXE VERBALE DU ROUMAIN : POLYGRAMMATICALISATION, UNIVERSAUX LINGUISTIQUES ET HYPOTHÈSES ÉTYMOLOGIQUES

Cecilia Mihaela POPESCU¹ 

*Article history: Received 30 April 2025; Revised 15 June 2025; Accepted 2 July 2025;
Available online 12 December 2025; Available print 30 December 2025*

©2025 Studia UBB Philologia. Published by Babeş-Bolyai University.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ABSTRACT. *On the Traces of the Lat. HABERE in the Verbal Morphosyntax of Romanian: Polygrammaticalization, Linguistic Universals, and Etymological Hypotheses.* Inheritance and transmission in Romance languages are better understood by considering that their evolution occurred both “vertically” through a shared Latin heritage and “horizontally” through common innovations and influences over time (Iliescu 2013 [2008]: 101). In accordance with this perspective, our approach traces the polygrammaticalization of the Latin verb HABERE > Rom. (a) *avea* as a (i) temporal morpheme: (a) in the compound past structure (e.g., *am cântat* “I have sung”), (b) in a future construction expressing lower probability (e.g., *am să cânt* “I am going to sing”), and (c) in the future-in-the-past paradigm (e.g., *aveam să cânt* “I would have sung”), but also (ii) as a modal morpheme of the conditional mood (e.g., *aş cânta* “I would sing”). These processes viewed together can reveal how these verbal morphemes relate to cognitive and linguistic universals and could also offer insights into the origin of the Romanian conditional mood auxiliary.

Key-words: *Lat. HABERE, Rom. (a) avea, (poly)pragmaticalization, cognitive and linguistic universals, Romanian verbal system*

¹ **Cecilia-Mihaela POPESCU** est Professeur des universités en Linguistique romane et Pragmatique et elle assure, depuis 1999, des cours à la Faculté des Lettres de l'Université de Craiova (Roumanie). Elle travaille aussi comme chercheur principal à l'Institut de Linguistique « Iorgu Iordan – Alexandru Rosetti » de l'Académie Roumaine, au Département de linguistique romane. Son domaine d'intérêt porte sur trois axes de recherche majeurs, à savoir : a) le comportement sémantique et fonctionnel de certains tiroirs verbaux des langues romanes (le futur, le subjonctif, le conditionnel, le présomptif), analysé du point de vue synchronique, diachronique ou bien en contrastivité, b) l'emprunt lexical du roumain au français et c) le fonctionnement et la signification des marqueurs discursifs des langues romanes. Courriel électronique : cecilia.popescu@edu.ucv.ro

REZUMAT. Pe urmele lui Lat. HABERE în morfosintaxa verbală a limbii române: poligramaticalizare, universalii lingvistice și ipoteze etimologice. Moștenirea și transmiterea în limbile romanice sunt mai bine înțelese dacă luăm în considerare faptul că evoluția acestora s-a realizat atât „vertical”, printr-o moștenire comună din limba latină, cât și „orizontal”, prin inovații și influențe comune de-a lungul timpului (Iliescu 2013 [2008]: 101). În acord cu această perspectivă, abordarea noastră urmărește procesul de poligramaticalizare a verbului latin HABERE > Rom. (a) avea ca (i) morfem temporal: (a) în structura formelor de perfect compus (**am cântat**), (b) într-o construcție de viitor care exprimă o probabilitate slabă (**am să cânt**), și (c) în paradigma viitorului în trecut (**aveam să cânt**), dar și (ii) ca morfem modal în construcția formelor condiționalului analitic (**aș cânta**). Aceste procese privesc laolaltă pot da seamă despre modul în care aceste morfeme verbale se raportează la universaliiile cognitive și lingvistice și pot, totodată, oferi o serie de indicii privind originea auxiliarului modului condițional din limba română.

Cuvinte-cheie: Lat. HABERE, Rom. (a) avea, (poly)pragmaticalizare, universalii cognitive și lingvistice, sistemul verbal al limbii române

1. Introduction

L'héritage et la transmission dans les langues romanes pourraient être mieux compris si l'on considère que l'évolution de ces systèmes linguistiques s'est réalisée tant « verticalement », par leur héritage commun du latin, aussi bien qu'« horizontalement », par les « innovations communes et les éventuelles influences communes subies au long du temps » (v. Iliescu 2013 [2008]: 101). Dans cette perspective, notre approche se propose de faire une synthèse sur le processus de polygrammaticalisation du verbe latin HABERE > roum. (a) avea (analysé, le plus souvent, de manière séquentielle), comme morphème temporel: (i) dans la construction du passé composé (en abrégé : PC) (**am, ai, a, am, ați, au_{IND.PRÉS.}** + le participe passé du verbe lexical : **am cântat** « j'ai chanté »), (ii) dans la construction d'un certain type de FUT exprimant une probabilité moins forte, notamment dans le registre parlé (**am, ai, are, avem, aveți, au_{IND.PRÉS.}** + le subjonctif (en abrégé : SUBJ) du verbe lexical : **am să cânt** « je vais chanter »), (iii) dans la construction d'une structure analytique actualisant le futur dans le passé (en abrégé: FUT) (**aveam, aveai, avea, aveam, aveati, aveau_{IND.IMP.}** + le SUBJ du verbe lexical : **aveam să cânt** « j'aurais chanté ») et, probablement, comme morphème modal, dans la construction du conditionnel (présent et passé) (en abrégé : COND), dont l'origine de son auxiliaire reste encore incertaine, **aș, ai, ar, am, ați, ar** étant des formes dérivées soit de (a) avea, soit de (a) vrea, soit de

(a) *fi* : *aș zice* « je dirais ». Une synthèse des données grammaticales vues en synchronie et présentées *supra* se retrouve dans le tableau ci-dessous :

Tableau 1 . L'ensemble des paradigmes verbaux construits avec les cognats de HABERE en roumain

A avea vb. lexical PRÉS. / IMP.	Passé composé	Futur périphrastiques, Type - 3	Périphrase prospective	Conditionnel
<i>am / aveam</i>	<i>am</i>	<i>am</i> _{PRÉS.} + <i>să</i> SUBJ	<i>aveam</i> _{IMP.} + <i>să</i> SUBJ	<i>aș</i>
<i>ai / aveai</i>	<i>ai</i>	<i>ai</i> _{PRÉS.} + <i>să</i> SUBJ	<i>aveai</i> _{IMP.} + <i>să</i> SUBJ	<i>ai</i>
<i>are / avea</i>	<i>a</i>	<i>are</i> _{PRÉS.} + <i>să</i> SUBJ	<i>avea</i> _{IMP.} + <i>să</i> SUBJ	<i>ar</i>
<i>avem / aveam</i>	<i>am</i>	<i>avem</i> _{PRÉS.} + <i>să</i> SUBJ	<i>aveam</i> _{IMP.} + <i>să</i> SUBJ	<i>am</i>
<i>aveți / aveati</i>	<i>ați</i>	<i>aveți</i> _{PRÉS.} + <i>să</i> SUBJ	<i>aveati</i> _{IMP.} + <i>să</i> SUBJ	<i>ați</i>
<i>au / aveau</i>	<i>au</i>	<i>au</i> _{PRÉS.} + <i>să</i> SUBJ	<i>aveau</i> _{IMP.} + <i>să</i> SUBJ	<i>ar</i>

2. Grammaticalisation, regrammatisation et polygrammaticalisation – esquisse du cadre théorique

L'histoire de la formation du PC analytique, du FUT et du COND roumain commence en latin tardif, mais elle se continue aussi au sein des systèmes linguistiques romans, de sorte que les formes considérées *supra* sont le résultat d'un processus cognitif abstrait et complexe de *grammaticalisation* (impliquant toute sorte d'autres processus cognitifs, tels que *la réanalyse* (voir De Mulder 2001 : 16), *la désémantisation* etc.) et même de *regrammatisation*.

Le processus de *grammaticalisation*, défini dans la manière canonique proposée par Meillet [1912 (1928/1958): 130-148], comme évolution d'une unité lexicale au statut de morphème grammatical, se rencontre dans la création du PC, du FUT et du COND des langues romanes, dans le passage de la synthèse à l'analyse. De l'autre côté, la *regrammatisation* se retrouve plutôt dans la création du COND roman, en général, puisque les différentes formes d'expression fortement ancrées initialement (en latin tardif) dans le domaine de l'actualisation [prospective] en référence au [passé] viennent grammaticaliser initialement ce trait originel inhérent (au sens où elles deviennent des marqueurs du *futur dans le passé*), mais, ensuite, à travers le temps, elles perdant (aux degrés divers) la possibilité de se replier dans le laps de temps susmentionné et elles vont développer (également aux degrés divers) des compétences modales. La *regrammatisation* est donc un processus ultérieur à la grammaticalisation, qui entraîne des modifications et des réorganisations au niveau du système de la langue, « [...] un processus qui accorde un statut grammatical à des éléments ou

à des systèmes qui ne l'avaient pas autrefois ou *qui changent les systèmes grammaticaux existants* [s.n.] » (Lindschouw 2011 : 59).

La polygrammaticalisation, initialement définie par Craig (1991) et reprise plus tard par Dostie (2001 : 62 ; 2004 : 34), Buchi (2004 : 251-264), Robert (2004), et, dans la linguistique roumaine, par C. Papahagi (2014), désigne le processus par lequel une même unité lexicale génère plusieurs unités relevantes au niveau grammatical: « one lexical item is the source of several grammaticalization clines, and ultimately of different grams » (Papahagi 2014 : 317).

En résumé, ce processus évolutif concerne, dans notre cas, le verbe latin HABERE, qui à partir de son statut de lexème plein, syntaxiquement transitif et avec le sens premier de [possession], entre dans diverses constructions grammaticales, associant un complément infinitif (en abrégé : INF) ou même une structure au SUBJ, et finit par perdre progressivement son sens originel, réduisant du même coup son corps phonétique et son rôle morphosyntaxique. En d'autres termes, le verbe latin, HABERE, sera grammaticalisé et deviendra un morphème dans la construction de certaines périphrases de PC, de FUT et, probablement, par association, également de COND.

Toutes ces modifications traduisent, en fait, des universaux cognitifs et/ou linguistiques, comme, par exemple, le besoin de *l'analytisme, la proéminence (à savoir, la préférence et la focalisation au niveau de l'expression) du temporel par rapport à la modalité ou vice-versa*. De sorte, la création du FUT et du COND roman souligne la nature profondément cognitive qui détermine tout processus de *grammaticalisation*, car, par exemple, l'apparition de ces deux morphèmes verbaux dans tout le domaine verbal roman signifie, en fait, le passage :

« *de una temporalidad egocéntrica* [s. n.], que llega al ser humano y que él no puede modificar, a significar *una temporalidad egodeíctica, modelada y construida en gran medida por el ser humano* [s. n.] » (Company Company 2006: 353-354).

3. Les cognats du lat. HABERE et leur rôle dans la morphosyntaxe du Roumain

3.1. La périphrase de Passé Composé

En roumain actuel, le PC est formé avec l'auxiliaire *a avea* < lat. HABERE, dont les formes *am, ai, a(u), am, ați, au* sont suivies ou, plus rarement, précédées par le participe passé (en abrégé : PART._{passé}) (une forme non-finie) du verbe lexical (*au mers, zis-au*):

- (1) « HABEO > *aibu* > *amu* > *am* (analogie avec la forme de la 1^e personne du pluriel)
 HABES > *ae* > *ai*
 HABET > *ae(t)* > *a*
 HABEMUS > *aemu* > *amu* > *am*
 HABETIS > *aeți* > *ați*
 *HABUNT > *au* » (Nicolae 2020 : 400).

En ancien roumain, au 16^e siècle, le PC était déjà très fréquent et, à cette époque-là, il semble être entièrement grammaticalisé. Pour ce qui est de son origine, il est à noter que le PC est une forme verbale romane obtenue à la suite de la priorisation de l'opposition temporelle (*présent* – *passé* – *futur*) en défaveur de l'opposition aspectuelle *infectum* – *perfectum*, aussi bien qu'à la suite de la tendance vers l'analytisme des systèmes linguistiques romans, en général. La seule différence par rapport à l'usage moderne concerne sa fréquence réduite par rapport au passé simple (en abrégé : PS), une forme entièrement héritée du latin et très bien représentée en ancien roumain, notamment dans les textes savants.

De plus, l'opposition entre le PS et le PC repose aussi sur leur sémantisme, le premier étant une forme véritable de prétérît (c'est-à-dire une forme de passée perfective, ponctuelle) et le second fonctionnant tantôt comme une *forme de passé accompli*, tantôt comme *un véritable passé simple* (v. Dragomirescu et al. 2022: 246).

Il est à noter aussi que la forme de PC a survécu dans le système verbal du roumain contemporain à la suite de la concurrence non seulement avec le PS, mais aussi avec sa propre structure surcomposée du type: *am fost* PART._{passé}. de *a fi* « être » *dat* PART._{passé}. du verbe lexical² (actualisant le plus souvent une action accomplie, mais ponctuelle) - aussi plus proche de la signification du PS – v. Frâncu 1983-1984 : 27), ce qui a contribué à l'augmentation de son caractère perfectif.

Dans la langue actuelle, le PC est la seule forme de passé ayant la structure HABERE + PART._{passé}, ce qui la distingue des formes perfectives du SUBJ, du FUT et du COND réalisées avec le cognat de ESSE suivi du PART._{passé} du verbe lexical (cf. *să fi cântat* – *aș fi cântat* – *voi fi cântat* vs. *am cântat*).

3.2. Les formes du futur construites avec l'auxiliaire *a avea*

Pour ce qui est de la création des formes du FUT, il est à noter que la langue roumaine a sélectionné deux auxiliaires :

² Il convient de noter qu'en roumain ancien, les formes verbales surcomposées apparaissent fréquemment dès le XVI^e siècle, dans tous les types de textes, qu'ils soient traduits ou originaux (v. Călărășu 1987; Frâncu 2009). En effet, ces formes contribuent à la mise en place d'un système temporel, aspectuel et modal complexe, fondé sur deux types de constructions, selon que l'auxiliaire est suivi d'une forme au PART._{passé} ou au PART._{PRÉS} (gérondif) du verbe lexical.

(i) d'un côté et contrairement aux langues de l'Occident roman, le verbe VOLERE (> VELLE) « vouloir » (sous 2), un verbe originellement volitif, qui se trouve tant à la base du FUT – Type 1 (le FUT canonique) : *voi cânta*, aussi bien qu'à la base de deux autres structures prospectives périphrastiques, le FUT – Type 2 : *oi cânta* (aux. + INF) et le FUT – Type 4 : *o să cânt* (aux. + SUBJ), une forme verbale fortement deictique.

(2) « **velit faciem tangere** » (Mihăescu 1960 : 143)
« il aimerait toucher son visage »,

(ii) de l'autre côté et dans la direction ouverte par les langues romanes occidentales, HABERE « avoir », qui se trouve dans la construction du FUT – Type 3, *am să cânt* (aux. + SUBJ).

Ce choix s'expliquait cognitivement et typologiquement (voir Fleischmann 1982 ou Bybee *et al.* 1994) par le fait que le *prospectif* est une notion cognitive, universelle, catégorisée par les sèmes : [+nécessité], [+volonté], [+intention].

Il est donc à souligner que ces deux auxiliaires pouvaient se combiner soit avec l'infinitif (« the inherited latin infinitive subordination » – Papahagi 2014 : 331), soit avec le SUBJ du verbe lexical (« the Balkanic subjunctive subordination » – Papahagi 2014 : 331), mais aussi avec l'INF du verbe *être* accompagné du PART. PRÉS/PASSE. du verbe lexical. Ces possibilités combinatoires ont donné naissance à un paradigme riche de formes verbales à signification prospective, synthétisée dans le tableau suivant (v. aussi Popescu 2013) :

Tableau 2. Le paradigme des formes de FUT en ancien roumain

Siècle	Le 16e siècle			Le 17e siècle		
Etimologie	+ INF	+ SUBJ	+ GER	+ INF	+ SUBJ	+ GER
VOLERE	Le Type 1 canonique : <i>voi afla,</i> <i>vei ceti</i> <i>audzi-veri,</i> <i>sluji-veri</i>	Le Type 1a : <i>va să judece,</i> <i>va să rugăm</i>	Le Type 2 : <i>veri fi aducând</i>	Le Type 1 canonique : <i>voi afla,</i> <i>vei ceti,</i> <i>audzi-veri,</i> <i>sluji-veri</i>	Le Type 1a : <i>voi să cânt,</i> <i>va să rugăm</i>	Le Type 2 : <i>voi fi cântând</i>
				Le Type 1b : <i>oi cânta</i>	Le Type 1c : <i>oi să cânt</i>	

Siècle	Le 16e siècle			Le 17e siècle		
HABERE	Le Type 3 : <i>am a</i> <i>trimite,</i> <i>are a opri,</i> <i>are a vedea</i>			Le Type 3 : <i>am a</i> <i>trimite,</i> <i>are a opri,</i> <i>are a vedea</i>	Le Type 3c : <i>avem să</i> <i>vă facem</i>	
				Le Type 3a : <i>voi avea a</i> <i>cânta</i>		
				Le Type 3b : <i>să aibă a</i> <i>să crede</i>		

Par conséquent, très caractéristique pour la fin du 16e siècle et le début du 17e siècle est le FUT construit avec l’auxiliaire *avea + a + INF* [le FUT – Type 3] (< lat. HABEO AD CANTARE) : *am a trimite, are a opri, are a vedea* etc., dont la valeur est à cette époque, comme dans le cas du FUT – Type 2, une valeur déictique, étroitement liée au sens déontique originel de l’auxiliaire : « un futur qui exprime plutôt la nécessité »³ (Guțu-Romalo 1968 : 427-432, cité par Iliescu 2000 : 433) : *eu mai mult am a te sluji*. Ce type se retrouve surtout dans les textes religieux et il est peu représenté dans les écrits à caractère administratif et dans ceux qui montrent une influence prononcée de la langue parlée.

« Apparaissant dans des propositions principales, en particulier à la fin des lettres, pour formuler l’engagement d’une obligation, la construction est synonyme de *voiu, vei, va + avea + a + infinitif*, exprimant, en plus de la valeur temporelle du futur, une signification modale spéciale, celle de l’action présentée comme nécessaire »⁴ (Călărașu 1987 : 179).

Pour le FUT – Type 3, les histoires de la langue roumaine enregistrent également l’invariant [le Type – 3a] : *voi, vei, va + avea + a + infinitif*, avec une diffusion plus large dans la langue parlée, par rapport à la fréquence de son type de base (le Type 3). C’est pourquoi, au 18e siècle, cette structure apparaît dans

³ «un viitor care exprimă mai ales necesitatea» (Guțu-Romalo 1968 : 427-432, cité par Iliescu 2000: 433).

⁴ «Apărând în propoziții principale, în special, în finalul scrisorilor, pentru a formula asumarea unei obligații, formația este sinonimă cu *voiu, vei, va + avea + a + infinitivul*, exprimând alături de valoarea temporală de viitor, o semnificație modală specială de acțiune prezentată ca necesară» (Călărașu 1987 : 179).

des textes avec une valeur purement prospective, ce qui a ensuite contribué à sa restriction et puis à sa disparition de la langue.

Aussi, une autre invariante [= Type – 3b], avec l’auxiliaire au SUBJ : *să am, ai (aibi), aibă + a + infinitif*, apparaît-elle avec une diffusion toutefois limitée – aux 17^e et 18^e siècles –, surtout dans les lettres et les documents, où elle est fréquemment enregistrée dans l’apodose du système conditionnel du type : *si p, q*, devenant ainsi un synonyme du SUBJ présent à valeur iussive (v. *infra* l’ex. 3):

- (3) « *Și de să vor găsi și nescar dirése **să nu aibă a să crede*** » (*Docum. Hist.*, A IXI, 136, cité par Călărașu 1987 : 184).

« Et s’ils trouvent aussi des actes seigneuriaux (des documents officiels) afin que l’on ne croie pas ».

En même temps, l’auxiliaire *am, ai, are*, etc. apparaît accompagné non seulement de l’INF, mais aussi du SUBJ du verbe lexical [le Type – 3c]. Dans les textes législatifs du 17^e siècle, cette structure exprime l’idée de nécessité dans le futur (voir l’ex. 4), mais à partir du siècle suivant, cette valeur modale spécifique va se perdre, le sous-Type – 3c étant utilisé majoritairement dans les subordonnés :

- (4) « *Noi **avem să vă facem** legē cu nușii să nu vă pără bucatele dōmnilor voastre* » (LRB, 64, cité par Călărașu 1987 : 187).

« Nous allons vous faire une loi avec eux pour que vous ne perdiez pas vos biens ».

A la fin du 18^e siècle, le FUT – Type 3 (*am a cânta*) sera remplacé par cette dernière variante, à savoir le Type – 3c (avec l’auxiliaire suivi du SUBJ du verbe lexical). Dès lors, la structure *am a cânta* n’appartiendra plus au paradigme de FUT, compte tenu de la forte valeur déontique qui la régit.

3.3. La périphrase *avea să* désignant le Futur dans le Passé (FdP)

Nous avons souligné maintes fois (v., par exemple, Popescu 2013 : 60, 255) que, parmi les traits spécifiques du roumain actuel se trouve l’impossibilité du COND (présent et passé) d’exprimer le FdP, ce paradigme verbal n’ayant dans la langue actuelle que des valeurs modales. Pourtant, cette situation semble ne pas être caractéristique à l’ancien roumain, où, dans de nombreux textes rédigés entre les 16^e et 17^e siècles (v. *Palia de la Orăștie*, le *Noul Testament de la Bălgrad*, le *Sicriul de Aur*, la *Biblia de la București*, ou bien chez N. Costin, D. Cantemir et Gh. Șincai), ce tiroir verbal apparaît fréquemment soit comme

substitut du FUT déictique, soit avec une valeur d'IMP, de PC ou de plus-que-parfait de l'IND (voir Avram 1976 : 353–358). On l'y rencontre surtout dans des circonstancielles temporelles, mais également dans les complétives.

Il résulte que, pour exprimer le prospectif dans la tranche de passé, le roumain recourt (v. Popescu 2013 : 157) :

- (i) au FUT littéraire (le FUT – Type 1), *va veni* « il viendra » (5a),
 - (ii) au FUT – Type 3 : *are să vină* (5b) et au FUT – Type 4 : *o să vină* (5c) « il viendra »,
 - (iii) à une construction périphrastique spécialisée, mais pas encore grammaticalisée, construite avec l'auxiliaire *a avea*_{IMP} « avoir » et le SUBJ: *avea să vină* (« il allait venir ») (v. 5d) :
- (5) a. « Paul *a spus* (ieri / acum o lună) că **va veni** (*ieri / azi / mâine / peste două săptămâni). »
 c. « Paul *a spus* (ieri / acum o lună) că **are să vină** (*ieri / mâine / peste două săptămâni). »
 b. « Paul *a spus* (ieri / acum o lună) că **o să vină** (*ieri / azi / mâine / peste două săptămâni). »
 d. « Paul *a spus* (ieri / acum o lună) că **avea să vină** (ieri/azi/mâine / peste două săptămâni) » (Popescu 2013 : 157).

La construction *avea* + le SUBJ se retrouve dans divers types d'énoncés (complétives, relatives, circonstancielles) employés en discours indirect ; elle tend à être particulièrement utilisée aujourd'hui pour marquer les « rétrospectives historiques » (GALR, I, 2005 : 442), notamment, dans la presse écrite :

- (6) « Nici un semn nu mă pregătise pentru descoperirea pe care **aveam s-o fac** la amiază » (*Transylvania Journal*, cité par Uricaru 2003, 190).
 « Aucun indice ne m'avait préparé pour la découverte que j'*allais faire* à midi ».
- (7) « Abia atunci am realizat eu că acest cântec **avea să-mi schimbe** oarecum viața » (*Formula AS*, 2003, cité par GALR, I, 2005 : 442).
 « C'est seulement à ce moment-là que je me suis rendu compte que cette chanson *changerait* / *allait changer* en quelque sorte ma vie ».

Par son comportement sémantique et fonctionnel, *avea* + le SUBJ semble être équivalent aux structures du type : (es.) *iba a* + INF; (fr.) *avait à* ou *allait* ou *devait* + INF ; (it.) *doveva* + INF (Popescu 2014), réalisées donc avec l'IMP d'un verbe modal ou d'un verbe de mouvement et suivi de l'INF, constructions utilisées dans la Romania Occidentalis pour exprimer le FdP. Mises en parallèle avec leur formes correspondantes de COND, ces constructions sont tenues pour dénoter le FdP « objectif » (v. Popescu 2013, 2014 ; Korzen, Nølke (2001, 130–131) ; Nilsson-Ehle (1973, 183)).

En revanche, les formes déictiques de FUT (la forme littéraire aussi bien que *o să vină* et *are să vină*) peuvent fonctionner dans des énoncés semblables à (6) et (7), régies par une forme verbale au passé :

- (8) « Actuala putere anunța în 1996 că **va aduce** la guvernare 15000 de specialiști. S-a văzut că promisiunea nu avea acoperire » (Uricaru 2003, 190).
« Le pouvoir actuel annonçait en 1996 qu'elle *installerait* au pouvoir 15000 spécialistes. On a vu que cette promesse n'avait aucun fondement ».

Du point de vue de la fréquence et de la distribution des occurrences, les formes de FUT à valeur de FdP sont préférées surtout dans les cadres discursifs canoniques des structures explicites (v. 8 et 5d).

La différence subtile entre toutes ces formes consiste dans le fait que, par rapport aux formes de FUT, la structure périphrastique a un fonctionnement exclusivement anaphorique, se rapportant, tout comme le COND français, par exemple, à un ensemble référentiel qui se place dans le passé. La capacité de *avea* + le SUBJ de désigner un procès placé dans la tranche temporelle de passé, de même que son fonctionnement anaphorique, tout cela détermine son emploi fréquent de nos jours, pour l'actualisation des « rétrospectives historiques », dans la presse écrite. Aussi, *avea* + le SUBJ se combine-t-il aisément avec des items lexicaux ou des expressions exprimant un positionnement antérieur, simultané ou postérieur par rapport au *nunc* énonciatif. À l'inverse, le FUT déictique (la forme littéraire et celles populaires) – en tant que FdP – constitue une manière « subjective » de se rapporter à un référent passé, distinct de T₀, projetant ainsi dans la narration « un effet de contemporanéité des événements » (v. Popescu 2013, 2014).

Il est à noter aussi que les formes de FUT utilisées en roumain actuel pour exprimer le FdP diffèrent :

(i) en rapport avec la variation diatopique/diamésique : la forme canonique étant fréquemment inscrite dans la langue littéraire, les autres appartenant au registre familier, populaire ou dialectal.

(ii) par leur comportement syntaxique : si la forme canonique de FUT peut apparaître dans des structures indépendantes aussi bien que dans des subordonnées, les deux autres types, construits avec le SUBJ du verbe lexical, s'enregistrent au sens de FdP, seulement au sein de structures explicites.

(iii) au niveau morphosyntaxique : le fonctionnement dichotomique *discours* vs. *récit / histoire* est relevant uniquement pour le couple : FUT – Type 3 vs. la périphrase avec l'auxiliaire *avea*_{IMP} et le SUBJ : *are să vină* / *avea să vină*.

(iv) du point de vue de certains effets de sens contextuels : la forme canonique de FUT est non-marqué par rapport à ces traits, tandis que les formes

périphrastiques de FUT sont parfois chargées contextuellement de la nuance *d'incertitude, de probabilité, d'imminence, de proximité* ou *d'atténuation illocutoire*.

3.4. Le conditionnel

Le COND (présent) du roumain est une forme analytique avec la structure morphématique *auxiliaire + infinitif court du verbe lexical* (eu *aș cânta* « je chanterais », el *ar cânta* « il chanterait », etc.), enregistrée pour la première fois dans des textes du 16^e siècle, non pas avec les valeurs qui se retrouvent dans la langue actuelle, mais comme une forme verbale prospective semblable au FUT déictique ou bien comme une forme rétrospective traduisant le passé.

Ce type de COND analytique n'est pas spécifique au roumain commun, puisque pour la référence au présent, l'ancien roumain a connu aussi une forme synthétique⁵ de COND, entre le XIII^e et le XV^e siècle et même dans les textes du XVI^e siècle, avec un nombre d'occurrences extrêmement réduit.

Pour ce qui est de la zone du passé, le roumain a enregistré dès le 16^e siècle, seulement des structures périphrastiques non-homogènes :

- certaines construites avec l'auxiliaire *a vrea* « vouloir » (employé au présent, mais le plus souvent à l'IMP, au PC ou parfois même au FUT de l'IND) suivi de l'INF ou bien de + *fi* « INF du verbe *être* » + PART_{passé} / le participe présent PART_{présent} du verbe lexical [= Type 1] : « *ascunde-mevrea, vreai spune, scula-se-vrea* ou bien *au vrut putea, se-au vrut părea, înghiți au vrut, nu s-au vrut crede* ou bien *vrea fi ajutat, se vrea fi făgăduitu-se, fi-vrea împărțit, fi-se-vrea părut, fi-vreați crezut* ou bien *ai vrut fi pus, am vrut fi furăți, au vrut fi* ou bien *fi-vreați știind* » (Densușianu 1961, II: 148).

⁵ En fait, pour cette période, le COND synthétique est la seule « réalité morphologique » ayant, *grosso modo*, la valeur du COND_{présent} de la langue actuelle. Malgré son caractère déjà archaïque, vieilli au 16^e siècle, le COND synthétique restera dans le système verbal du roumain jusqu'au début du 17^e siècle. Pour ce qui est de sa structure morphématique (réconstituée soit par la comparaison avec les dialectes historiques du roumain – l'aroumain et l'istroroumain, soit avec les patois du daco-roumain, à savoir le patois parlé au sud-ouest de la Roumanie, dans la région du Banat), celle-ci a la structure suivante : *Racine + suffixe identique à celui du parfait (spécifique selon la classe de conjugaison) + suffixe spécifique -RE- (généralement valable) + désinences* (v. ILR, II 1969: 270).

En revanche, pour ce qui est de son origine, il n'y a pas un point de vue unanimement accepté dans la littérature de spécialité. On considère (v. Tasmowski et Bourova 2005), que le COND synthétique est soit hérité du SUBJ parfait latin (CANTAVÉRIM), soit du FUT parfait latin (CANTAVERO), soit il s'agit d'une forme mixte obtenue à la suite de la confusion morphologie et sémantique entre ces deux tiroirs du latin (Dimitrescu *et al.* 1978 : 321).

- d'autres, avec les formes de l'auxiliaire identiques à celles du COND_{PRÉS.} : *aș(i), ai, are (ară), am, ați, are (ară) + fi* « INF du verbe *être* » + PART_{passé} / PART_{présent} du verbe lexical [= Type 2] : « *ară fi lăsat, ară fi venită* ou bien *ai fi știind, fire-ași ascultându* » (v. Dimitrescu et al. 1978 : 322).

3.4.1. Pour ce qui est des *hypothèses étymologiques* enregistrées dans la littérature par rapport à l'origine des formes de l'auxiliaire du COND analytique du roumain, il y en a trois.

La première hypothèse (lancée en 1896 par Gustave Weigand) considère comme étymon de l'auxiliaire du COND roumain, le verbe lat. VOLERE (employé à l'IMP) : CANTARE VOLEBAM > *aș(i) cânta*, etc. L'argument fondamental réside dans le fait qu'un paradigme transparent, quasi semblable à celle du COND littéraire, à savoir : *vreaș, vreai, vream, vreați, vrea* s'enregistre dans certaines variétés régionales actuelles du daco-roumain (le Banat et le nord de l'Olténie) et, aussi, en istro-roumain. Malgré tout cela, toute une série de difficultés ressort lorsqu'il faut démontrer l'évolution des formes de la 1ère personne et de la 3e personne du singulier : (*aș(i)*, respectivement, *ară (are)/ar*). Plusieurs contre-arguments à cette théorie se retrouvent chez Tiktin (1904) et, plus tard, chez Coene et Tasmowski (2006) et Zafiu (2017 : 1–20).

D'ailleurs, pour cette hypothèse qui semble être construite de manière évidente comme un modèle mimétique par rapport à celui formulé pour ce qui est de l'origine du COND des langues romanes occidentales (cf. HABEO CANTARE – HABEBAM CANTARE vs. VOLO CANTARE – VOLEBAM CANTARE), un autre contre-argument très important semble être l'existence même dans l'histoire du roumain (entre le 17e siècle et le 18e siècle) de certaines tournures construites toujours avec l'auxiliaire *a vrea* « vouloir » à l'IMP, mais dont le sens était celui d'un COND_{passé} (v. *supra*).

La deuxième hypothèse établit pour étymon de l'auxiliaire du COND roumain, le verbe HABERE. Formulée pour la première fois par Hariton Tiktin (1904), cette direction sera développée plus tard par Alexandru Rosetti (1978), qui considère les formes de l'IND_{PRÉS} du verbe latin HABERE en tant que base étymologique pour le COND analytique du roumain. Mais, cette théorie, elle ne se justifie non plus puisque la provenance de la forme *aș(i)* de la 1ère personne du singulier est difficile de prouver (??HABEO > *aș(i)*). Comme explication, Rosetti fait appel à la forme de plus-que-parfait du SUBJ latin, HABUISSEM (> *aș(i)*)⁶.

Cette hypothèse qui considère le verbe latin HABERE comme étymon du COND du roumain, a été reprise par Mark Elson (1992 : 360–375), qui considère

⁶ Il y a aussi des orientations (comme, par exemple, celle d'Alf Lombard) qui considèrent qu'à la base de l'auxiliaire du COND roumain se trouve un paradigme mixte, obtenue de l'immixtion des auxiliaires *a vrea* « vouloir » et *a avea* « avoir ».

que le paradigme temporel qui se trouve à l'origine des formes *aș(i)*, *ai*, etc. n'est pas celle de l'IND_{PRÉS}, mais celle de l'IND parfait (latin) du verbe HABERE (HABUI, HABUISTI, etc., comme en italien moderne). Dans le cas de *aș(i)*, Elson (1992 : 360–375) propose le découpage morphématique suivant : [*a_{AUX}* + une marque de personne : *a_{AUX}* + -*ș(i)*].

La troisième hypothèse (Coene et Tasmowski 2006) est constituée autour de l'auxiliaire *a fi* « être » suivi de l'infinitif long du verbe lexical (par ex.: *era cântare*) et elle prend en compte les arguments suivants :

(a) l'appartenance au domaine du passé (ou, plus précisément, à la zone aspectuelle [+perfectif]) de tous les auxiliaires qui se trouvent dans la structure morphématique de chaque paradigme du COND, tant dans les langues romanes, que dans l'aire des langues balkaniques ;

(b) la grammaticalisation en roumain (tout comme dans les autres langues slaves voisines) du verbe *a fi* « être » pour l'expression de la tranche temporelle de passé, pour la formation des temps composés du FUT (*voi fi cântat* « j'aurai chanté »), du SUBJ (*să fi cântat* « que j'aie chanté »), du COND (*aș fi cântat* « j'aurais chanté ») et de l'INF. (*a fi cântat* « avoir chanté »).

4. Considérations finales

Tout en suivant les particularités de la grammaticalisation et du fonctionnement des formes verbales présentées *supra*, notre démarche essaie de souligner, autour du cadre théorique de la (poly)grammaticalisation et de la typologie romane, que la création de ces morphèmes verbaux dérivés de HABERE est fortement liée aux universaux cognitifs et linguistiques. Cet ensemble des similitudes typologiques peut être interprété comme un argument en faveur de l'existence des matrices (frames) partagées dans la représentation linguistique de certaines notions universelles.

De plus, cette approche pourrait nous fournir des données supplémentaires concernant l'origine de l'auxiliaire du COND roumain. Même si aucune solution ne s'impose avec évidence, on peut néanmoins envisager comme origine une forme verbale du passé – peut-être « an imperfect, probably contaminated by the simple past » (Dragomirescu et al. 2022 : 22 ; Zafiu 2017) de *a vrea* ou bien le PS de *a avea* (Elson 1992 ; dans ce dernier cas, cf. la situation du COND de type *cantarebbe*, de l'italien contemporain). De l'autre côté, la création de la périphrase *avea să* pour l'expression du FdP, dont l'auxiliaire *a avea* (< lat. HABERE) se trouve à l'IMP (cf. le COND de l'espagnol, du français et de l'ancien italien qui fonctionne toujours comme un FdP), serait un autre argument en faveur d'une hypothèse ou de l'autre.

En sélectionnant des formes spécifiques de l'auxiliaire pour le FUT et le COND, le roumain confirme sa propension à désambigüiser et à créer des paradigmes spécialisés, une caractéristique récurrente de son système verbal, comme en témoigne la distinction entre le lexème *avoir*₁ et le morphème *avoir*₂ dans les structures fortement grammaticalisées telles que le PC et le COND.

Il est à noter pourtant que, pour le choix d'un paradigme ou de l'autre, la langue a payé attention en diachronie à d'autres paramètres, dont on mentionne :

- la *variation de registre* bien évidente, par exemple, dans le parcours évolutif de la structure prospective du FUT –Type 3 (*am să cânt*), enregistrée en roumain à la fin du 18e siècle, initialement dans la langue écrite (où se trouvait notamment dans les subordonnées) avec une forte nuance modale déontique.
- la *proéminence* de certaines catégories sémantiques et cognitives, telles que la *modalité* ou bien la *temporalité* est un autre facteur d'autant plus important (v. *infra* Pîrvu 1998 : 675 et aussi Nicolae 2020 : 395–422) pour la polygrammaticalisation de HABERE et, notamment, pour la regrammatisation de ses cognats romans, un facteur qui mériterait un développement tout particulier.

« La differenza tra il romeno e le altre lingue romanze consiste quindi nel fatto che mentre nel primo gli ausiliari si sono specializzati per indicare *realtà* o *possibilità*, nelle seconde esprimono la transitività o l'intransitività dell'azione verbale » (Pîrvu 1998 : 675).

À ce propos, nous sommes d'avis que la proéminence de la modalité est beaucoup plus saillante en roumain par rapport aux autres langues romanes, et cette particularité dérive, en tout premier lieu, de la possibilité de l'auxiliaire de se combiner soit avec une forme verbale non-finie (infinitif/participe passé/participe présent) soit avec une forme verbale finie (le SUBJ) du verbe lexical :

« *Romanian privileges mood: auxiliaries systematically grammaticalize as exponents of mood, and this has consequences for their internal structure and for the periphrases they are part of. Periphrases with auxiliaries whose feature matrix is richer and also includes tense features fail to enter the periphrastic system of standard Romanian (and, in some cases, give way to functionally equivalent synthetic formations)* » (Dragomirescu et al. 2022 : 25).

BIBLIOGRAPHIE SÉLECTIVE

- Andersen, Henning (éd.). 2001. *Actualization : Linguistic Change in Progress*, Amsterdam and Philadelphia : John Benjamins, 21-57.
- Andersen, Henning. 2006. « Grammatication, regrammatication and degrammatication: tense loss in Russian ». *Diachronica*, 23, 2, 231-258.
- Avram, Mioara. 1976. « Condiționalul cu valoare de indicativ trecut în texte vechi românești ». *Studii și Cercetări Lingvistice*, 4-27, 353-358.
- Buchi, Eva. 2004. « *Approche diachronique de la (poly)pragmaticalisation de français déjà* » (article « Quand le grammème est-il devenu pragmatème, déjà ? ». In Trotter, David (éd.). *Actes du XXIV Congrès International de Linguistique et de Philologie Romanes*, Tome III. Max Niemeyer Verlag, 251-264, 2007) communication présentée au XXIVe CILPR, Aberystwyth, 2004, en ligne sur: <https://shs.hal.science/halshs-00149275/document>.
- Bybee, Joan, William Perkins, and J.-R. Pagliuca. 1994. *The Evolution of Grammar. Tense, Aspect and Modality in the Languages of the World*. Chicago and London : The University of Chicago Press.
- Bugeanu, Dan. 1970. « Formarea condiționalului în limba română ». *Studii și cercetări lingvistice*, 21, 543-563.
- Călărașu, Cristina. 1987. *Timp, mod, aspect în limba română în secolele al XVI-lea – al XVIII-lea*. București : Tipografia Universității din București.
- Closs-Traugott, Elizabeth. 1989. « On the Rise of Epistemic Meanings in English: An Example of Subjectivation in Semantic Change ». *Language*, 65, 1, 31-66.
- Closs-Traugott, Elizabeth, and Bernd Heine (éds.). 1991. *Approaches to grammaticalization*. Amsterdam : John Benjamins Publishing Company.
- Closs-Traugott, Elizabeth, and Richard Dasher. 2002. *Regularity in Semantic Change*. Cambridge, Cambridge University Press.
- Coene, Martine, and Liliane Tasmowski. 2006. « On the Balkan-Slavic origins of the Romanian conditional ». *Revue Roumaine de Linguistique* 51, 2, 321-340.
- Company Company, Concepción. 1985-1986. « Los futuros en el español medieval: sus orígenes y su evolución ». *Nueva Revista de Filología Hispánica*, 34, 1, 48-107.
- Company Company, Concepción. 2006. « Tiempos de formación romance. II. Los futuros y los condicionales ». Company Company, Concepción (éd.), *Sintaxis histórica de la lengua española: La frase verbal*, vol. I. México : FCE, UNAM, 349-422.
- Coteanu, Ion. 1981. *Structura și evoluția limbii române (De la origini până la 1860)*. București : Editura Academiei Republicii Socialiste România.
- de Mulder, Walter. 2001. « La linguistique diachronique, les études sur la grammaticalisation et la sémantique du prototype: présentation ». *Langue française*, 130, 8-32, en ligne sur : https://www.persee.fr/doc/lfr_0023-8368_2001_num_130_1_1024.
- Densusianu, Ovid. 1961. *Istoria limbii române*, 2 voll. București : Editura Științifică.
- Dietrich, Wolf. 1984. « La grammaticalisation du passé composé ». In Buridan, Claude (éd.), *Romanistique – germanistique. Une confrontation. Actes du Colloque de Strasbourg organisé par le Centre de Philologie Romane près de l'Université de Strasbourg*, 23-24 mars, 1984, 81-95.

- Dimitrescu, Florica et al. 1978. *Istoria limbii române. Fonetică, Morfosintaxă, Lexic*. București : Editura Didactică și Pedagogică.
- Dostie, Gaétane. 2001. « La gradation du sens et ses traces morphologiques et syntaxiques. Considérations sur la (poly)pragmaticalisation ». In Col, Gilles, and Daniel Roulland (éds.). *Travaux linguistiques du CERLICO 14, Grammaticalisation 2. Concepts et cas*. Rennes : Presses Universitaires de Rennes, 61-91.
- Dostie, Gaétane. 2004. *Pragmaticalisation et marqueurs discursifs : Analyse sémantique et traitement lexicographique*. Bruxelles : Duculot / De Boeck.
- Dragomirescu, Adina, Alexandru Nicolae, and Rodica Zafiu. 2022. « The loss of analyticity in the history of Romanian verbal morphology ». In Ledgeway, Adam, John Charles Smith, and Nigel Vincent. *Periphrasis and Inflexion in Diachrony: A View from Romance*. Oxford : Oxford University Press, 241-271.
- Elson, Mark. 1992. « A Latin Source for the Conditional Auxiliary in Romanian ». *Zeitschrift für romanische Philologie*, 108, 5-6, 360-375.
- Fleischman, Suzanne. 1982. *The Future in Thought and Language. Diachronic Evidence from Romance*. Cambridge : Cambridge University Press.
- Frâncu, Constantin. 1983-1984. « Geneza și evoluția timpurilor verbale supracompuse în limba română ». *Anuar de lingvistică și istorie literară*, XXIX, A, 23-62.
- Frâncu, Constantin. 2009. *Gramatica limbii române vechi (1521 – 1780)*. Iași : Casa Editorială „Demiurg”.
- GA/GLR – Academia Română. 1963². *Gramatica limbii române*, vol. II. București : Editura Academiei Republicii Populare Române.
- GALR – Academia Română, Institutul de Lingvistică „Iorgu Iordan – Al. Rosetti”, Guțu-Romalo, Valeria (coord.). 2008 [2005]. *Gramatica Limbii Române*, vol. I : *Cuvântul*, vol. II : *Enunțul*. București : Editura Academiei Române.
- GBLR – Pană-Dindelegan, Gabriela (coord.). 2016. *Gramatica de bază a limbii române*. București : Editura Univers Enciclopedic.
- GR – Pană-Dindelegan, Gabriela (coord.). 2013. *The Grammar of Romanian*. Oxford : Oxford University Press.
- Guțu-Romalo, Valeria. 1968. *Morfologia structurală a limbii române*. București : Editura Academiei Române.
- Heine, Bernd. 1995. « Agent-oriented vs. epistemic modality. Some observations in German modals ». In Bybee, Joan, and Suzanne Fleischman (éds.), *Modality in grammar and discourse*. Amsterdam and Philadelphia : John Benjamins Publishing, 17-53.
- Heine, Bernd. 1997. « Grammaticalization and Language Universals ». *Mémoires de la Société de Linguistique de Paris*, nouvelle série, 5: *Grammaticalisation et reconstruction*. Paris : Librairie Klincksieck, 11-23.
- Hopper, Paul J., and Elizabeth Closs-Traugott. 2003 [1997]. *Grammaticalization*. Cambridge : Cambridge University Press.
- Iliescu, Maria. 2000. « Grammaticalisation et modalités en roumain : le futur déictique et épistémique ». In Coene, Martine et al. (éds.), *Traiani Augusti Vestigia Pressa Sequamur. Studia Linguistica in Honorem Lilianae Tasmowski*. Padova : Unipress, 429-441.

- Iliescu, Maria. 2013 [2008]. « Phénomènes de convergence dans la Romania : Morphosyntaxe et Syntaxe ». In Ernst, Gerhard et al. (éds.), *Histoire linguistique de la Romania*, Tome 3, Berlin-New York : Walter de Gruyter, 3266–3281, article repris dans Iliescu, Maria, 2013, *Varia Romanica. Universaux linguistiques, analyse du discours et caractère variationnel de la Romania*, Berlin, Frank und Timme, 101-130.
- ILR II – Academia Română, Al. Rosetti (rédacteur responsable). 1969. *Istoria limbii române*, vol. II : A. *Latina dunăreană*, B. *Româna comună*, C. *Influențe*. București : Editura Academiei Republicii Populare Române.
- Korzen, Hanne, and Henning Nølke. 2001. « Le conditionnel : niveaux de modalisation ». In Dendale, Patrick, and Liliane Tasmowski (éds.), *Le conditionnel en français*. Paris : Librairie Klincksieck, 125-167 ; repris Nølke, Henning (éd.). 2001. *Le regard du locuteur. 2. Pour une linguistique des traces énonciatives*. Paris : Editions Kimé, 35-58.
- La Fauci, Nunzio. 2005. « Il fattore *HABEO*. Prolegomeni a una nuova considerazione delle genesi del perfetto e del futuro romanzo ». In Kiss, Sándor, Luca Mondin, and Giampaolo Salvi (éds.), *Latin et Langues Romanes. Études de Linguistique Offertes à József Herman à l'Occasion de son 80ème Anniversaire*. Tübingen : Max Niemeyer, 441-451.
- La Fauci, Nunzio. 2006. « Dinamiche sistematiche, perifrasi perfettive e futuro sintetico: dal latino al romanzo ». In Oniga, Renato, and Luigi Zennaro (éds.), *Atti della Giornata di Linguistica Latina. Venezia, 7 maggio 2004*. Venezia : Università Ca' Foscari, 101–132.
- Lamiroy, Béatrice. 1999. « Auxiliaires, langues romanes et grammaticalisation ». In Bat-Zeev-Shylskrot, H. (éd.), *Langages. Les auxiliaires : délimitation, grammaticalisation et analyse*. 135, 33-45.
- Lamiroy, Béatrice. 2007. « Typologie et grammaticalisation ». In Bertrand, Olivier et al. (eds.), *Discours, diachronie, stylistique du français. Hommages à Bernard Combettes*. Berne : Peter Lang, 141-157.
- Lindschouw, Jan. 2011. « L'évolution du système du futur du moyen français au français moderne : la réorganisation comme un cas de régrammation ». *Revue de Linguistique Romane*, 75, 297-298, 51-98.
- Lindschouw, Jan. 2013. « Passé simple et passé composé dans l'histoire du français. Changement paradigmatique, réorganisation et régrammation ». *Revue de Linguistique Romane*, 77, 305-306, 87-120.
- Meillet, Antoine. 1912 [1928/1958]. « L'évolution des formes grammaticales ». In Meillet, Antoine. *Linguistique historique et linguistique générale*. Paris : Honoré Champion, 130-148.
- Mihăescu, Haralambie. 1960. *Limba latină în provinciile dunărene al Imperiului Roman*. București : Editura Academiei Republicii Populare Romîne.
- Nicolae, Alexandru. 2020. « Grammaticalization as pattern formation : Romanian auxiliaries from a diachronic Romance perspective ». *Revue Roumaine de Linguistique*, 65, 4, 395-422.

- Nilsson-Ehle, Hans. 1973. « Sur le conditionnel – temps en italien ». *Revue Romane*, 8, 178-184, en ligne sur : http://img.kb.dk/tidsskriftdk/pdf/rro/rro_0008-PDF/rro_0008_94411.pdf.
- Nocentini, Alberto. 2001. « La genesi del futuro e del condizionale sintetico romanzo ». *Zeitschrift für romanische Philologie*, 117, 3, 367-401.
- Pagliuca, William. 1994. *Perspectives on Grammaticalization*. Amsterdam and Philadelphia : John Benjamins Publishing Company.
- Pîrvu, Elena. 1998. « Sulle cause della diversa distribuzione degli ausiliari temporali in italiano e in romeno ». In Ruffino, Giovanni (éd.), *Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza*, vol. 2. Tübingen : Niemeyer, 673-678.
- Pană Dindelegan, Gabriela, and Martin Maiden (coord.). 2021. *O gramatică a limbii române altfel*. București : Univers Enciclopedic.
- Papahagi, Cristiana, 2014, « Convergent grammaticalization of some Romance auxiliaries », in *Revue Roumaine de Linguistique*, 59, 4, 317-334.
- Popescu, Cecilia Mihaela. 2013. *Viitorul și condiționalul în limbile romanice. Abordare morfosintactică și categorizare semantică din perspectivă diacronică*. Craiova : Editura Universitaria.
- Popescu, Cecilia Mihaela. 2014. « <Viitorul în trecut> în limba română contemporană. Un punct de vedere semantic-pragmatic ». *Revista de Filologie Română*, 31, 1, Madrid : Ediciones Universidad Complutense de Madrid, 111-125.
- Popescu, Cecilia Mihaela. 2015. *Exprimarea 'potențialului' și a 'irealului' în latină, franceză și română*. Craiova : Editura Universității din Craiova.
- Robert, Stéphane. 2004. « The challenge of polygrammaticalization for linguistic theory: fractal grammar and transcategorial functioning ». In Frajzyngier, Z., A. Hodges, and D. S. Rood (éds.), *Linguistic Diversity and Language Theories*. Amsterdam and Philadelphia : John Benjamin, 119-142.
- Roberts, Ian. 1992. « A formal account of grammaticalisation in the history of Romance futures ». *Folia linguistica historica* 13, 1-2, 219-258.
- Rosetti, Alexandru. 1978. *Istoria limbii române*. București : Editura Științifică și Enciclopedică.
- Tasmowski, Liliane, and Viara Bourova. 2005. « La formation du conditionnel en ancien roumain ». In M. Coene (éd.), *On space and time in language*. Cluj : Editura Clusium, 333-352.
- Tiktin, Hariton. 1904. « Die Bildung des Rumänischen Konditionalis ». *Zeitschrift für romanische Philologie*, 28, 691-704.
- Titova, V. P. 1959. « O problemă litigioasă a morfologiei istorice românești (Originea condiționalului) ». *Studii și cercetări lingvistice*, 10, 561-571.
- Uricaru, Lucia. 2003. *Temporalitate și limbaj*. București : Editura ALLFA.
- Zafiu, Rodica. 2017. « The auxiliary of the Romanian conditional: semantic and functional arguments concerning the reconstruction of a disputed grammaticalization process ». *Diacronia*, 6, 1-20.

HERENCIA Y TRANSMISIÓN DE TÉRMINOS GASTRONÓMICOS EN HISPANOAMÉRICA

Olivia Narcisa PETRESCU¹ 

*Article history: Received 30 May 2025; Revised 2 October 2025; Accepted 3 October 2025;
Available online 12 December 2025; Available print 30 December 2025*

©2025 Studia UBB Philologia. Published by Babeş-Bolyai University.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ABSTRACT. *Inheritance and Transmission of Gastronomic Terms in Latin America.* The present study focuses on the Latin American gastronomic field, from a lexical, etymological and cultural perspective, based on the enormous potential for exploration and development that gastronomy has, converted today into a science-art about cultures, people, places and the reciprocal influences that occur among them. We start from the idea that culinary heritage is a form of cultural communication, in which both traditions and modern hybridizations have determined enrichments, fusions and a profound change in the world's food and gastronomy.

Moreover, the article aims to approach, classify and exemplify the gastronomic lexicon of the most relevant Hispanic American culinary cultures, according to the terms, inherited or transmitted, used to designate ingredients, dishes and culinary techniques. Besides, we describe several of the most used lexicographic works, pointing out some challenges, and also, a few limitations within the Hispanic lexical resources in this field. Moreover, we draw attention to a rich variety of terms, cultural and technical combination of styles for the same concept, depicted from different ancient and modern periods, with special emphasis on the pre-Columbian indigenous heritage, but without neglecting the colonial and

¹ **Olivia Narcisa PETRESCU** es doctora *magna cum laude* en Literatura Comparada y licenciada por la Universidad “Babeş-Bolyai” de Cluj-Napoca en Letras y Derecho, profesora titular en el Departamento de Lenguas Modernas Aplicadas de la Facultad de Letras. En calidad de docente, investigadora, coordinadora de tesis y otros proyectos culturales, a la vez que traductora e intérprete jurada, sus intereses profesionales y áreas de investigación son la traducción especializada y los estudios culturales. Sus publicaciones más relevantes son: *Cultura y gastronomía en América Latina* (2024), *Laberintos posibles. Cultura y significados hispánicos* (2008), *La significación mítica del espacio en la literatura iberoamericana e inglesa del siglo XX* (2014), *Estudios de traducción y lenguajes de especialidad* (2014) y varias traducciones literarias. olivia.petrescu@ubbcluj.ro

the contemporary cultures. From Mexican and Guatemalan *tamales* to Peruvian or Ecuadorian *ceviche*, Colombian and Venezuelan *arepas* and Argentinean *asado*, Latin America demonstrates an authentic richness and diversity in its gastronomic lexicon, living sample of the stages of history, geography, fauna and flora, economy or traditions, that deserve to be inventoried and preserved as a valuable piece of the cultural heritage of humanity.

Keywords: *gastronomy and culture, inheritance and transmission, culinary terms, lexicography, etymology.*

REZUMAT. Moștenirea și transmiterea termenilor gastronomici în America Latină. Studiul de față se concentrează asupra domeniului gastronomic latino-american, dintr-o perspectivă lexicală, etimologică și culturală, pornind de la potențialul enorm de explorare și dezvoltare pe care îl are gastronomia, transformată astăzi într-o știință-artă despre culturi, oameni, locuri și influențele reciproce care se produc între acestea. Pornim de la ideea că patrimoniul culinar este o formă de comunicare culturală, în care atât tradițiile, cât și hibridările moderne au determinat dezvoltări reciproce, fuziuni și o schimbare profundă în alimentația și gastronomia mondială.

Mai mult, articolul își propune să abordeze, să clasifice și să exemplifice lexicul gastronomic al celor mai relevante culturi culinare hispano-americane, în funcție de termenii, moșteniți sau transmiși, utilizați pentru a desemna ingredientele, mâncărurile și tehnicile culinare. Totodată, vom descrie câteva dintre cele mai utilizate lucrări lexicografice, subliniind unele provocări și, de asemenea, câteva limitări în cadrul resurselor lexicale hispanice din acest domeniu. În plus, evidențiem bogăția și varietatea termenilor, combinațiile culturale sau tehnico-stilistice pentru același concept, așa cum se regăsesc în diferite perioade antice sau moderne, cu precădere în patrimoniul indigen precolumbian, dar fără a neglija culturile coloniale și contemporane. De la *tamales* mexicane și guatemaleze la *ceviche* peruvian sau ecuadorian la *arepas* columbiene și venezuelene la *asado* argentinian, America Latină demonstrează o bogăție și o varietate autentică în lexicul său gastronomic, mostră vie a istoriei, geografiei, faunei și florei, economiei și tradițiilor, care merită să fie inventariat și conservat ca o parte importantă a patrimoniului cultural al umanității.

Cuvinte-cheie: *gastronomie și cultură, moștenire și transmitere, termeni culinari, lexicografie, etimologie.*

1. Significados culturales de la gastronomía, entre tradición y modernidad

Ante todo, cabe aclarar el significado del término, su herencia etimológica y su transmisión, en función de dos acepciones particulares, que

conlleven tradición, por un lado, y modernidad, por el otro. De hecho, en Francia, país cuna de semejante arte, se encuentra el primer escrito europeo que alude a *gastronomía*. Se trata de un poema en cuatro cantos firmado por Joseph Berchoux que data del año 1805, traducido posteriormente a diversos idiomas, incluso al castellano por José de Urcullu. Además, el término comenzó a divulgarse por tierras británicas a principios del siglo XIX en textos ingleses (Pearsall y Hanks, 2005²) y, después, con más frecuencia, en documentos españoles a partir de 1871 (Corominas y Pascual, 1980, 123).

Cierto es que en Francia se plantea la primera vez el enfoque científico al acto de comer en uno de los textos del año 1825, escrito por Jean Anthelme Brillat-Savarin, titulado *Fisiología del gusto*. El libro es considerado fundamental para la ciencia gastronómica, además abunda en muchos aforismos, entre los que está el conocido refrán “dime qué comes y te diré quién eres”, retomado incluso por Balzac o Feuerbach, entre otros, al debatir literariamente sobre el arte del buen comer. En efecto, en el apartado del libro titulado *Definición de gastronomía*, Brillat-Savarin procura instituir una ciencia gastronómica que define como un conocimiento inteligente de todo lo que concierne la alimentación del ser humano. Al mismo tiempo, el autor subraya que la gastronomía es parte de la historia, economía política y colonizaciones históricas, por su aporte y clasificación de las sustancias alimentarias acorde con los principales momentos de evolución de la humanidad; de la física, por su observación de la composición y calidades alimenticias; de la química, por sus análisis a los que someten los ingredientes en la elaboración de platos y procesos, y, no, en último lugar, del arte, por saber crear y adaptar los platos de comida con el fin de hacerlos placenteros al gusto en función de las clases sociales. Como conclusión, “[la gastronomía] gobierna toda nuestra vida” (Brillat-Savarin 1825/1949, 51). La visión de Brillat-Savarin ha sido divulgada por muchos escritores de la época, incluso Alejandro Dumas hijo (1873), al reivindicar “los términos *gastrónomo* y *gourmand* para designar a individuos que no cocinan ni comen por necesidad ni por glotonería, sino por el placer de consumir la comida en formas racionales y legisladas por el saber científico de su tiempo, así como por el goce que derrocha un buen plato” (Teuteberg 2007, 234).

Desde ya, se concreta lo que llamamos la acepción tradicional que incumbe la herencia universal del término. Como argumento contemporáneo adicional, el historiador mexicano José Iturriaga expone con respecto a la gastronomía que “[...] dentro de dicho término debe entenderse [...] a todas las expresiones culinarias de las diversas regiones y de los distintos estratos

² Pearsall, J.& Hanks, P. ed. 2005. *Oxford Dictionary of English*. 2nd, revised ed. Oxford: Oxford University Press.

sociales” (Iturriaga 2015, 15), haciendo hincapié en el *Patrimonio Inmaterial del mundo* (2008), según la UNESCO³ que introdujo el concepto desde 2010.

Desde esta perspectiva, la *gastronomía* llega a ser un símbolo territorial, una muestra de culturas y naturalezas que relaciona a seres humanos con raíces e influencias en un determinado espacio y tiempo. De todo ello, deducimos que se puede acuñar el binomio gastronomía & cultura, asociado a cada contexto específico y reflejado en la recolección y en el cultivo de determinadas plantas, flores, semillas, sustancias y en la cría o caza de ciertos animales, peces, aves e insectos, que llegan a incorporarse en platos originales y señas de la identidad cultural de cada pueblo, tal como lo demuestran las expresiones culinarias de las diversas regiones y de los distintos estratos sociales (UNESCO, 2010). En el mismo sentido, Fusté-Forné (2016) sostiene que el *patrimonio culinario* es una forma de comunicación y herencia cultural, en la que las tradiciones, normalmente rurales, de esas culturas, pero también, añadimos nosotros, los mestizajes y las innovaciones han venido transmitiendo y enriqueciendo culturalmente como resultado de múltiples factores políticos, económicos, sociales, educacionales, geográficos o climáticos.

Llegamos así a la segunda significación, la actual, de hecho la más difundida, que se asienta en conceptos como transmisión, hibridación, fusión y creación (Ayora Díaz 2018, 6-14). Ciertamente es que el término *gastronomía* se ha solapado a lo largo del tiempo con *cocina* o *cuisine*, la variante francesa. Por ejemplo, desde la Antigüedad clásica, las clases altas europeas consumían platos codificados, es decir muy refinados o sofisticados, que formaban parte de una *alta cocina* u *haute cuisine*. Hoy día, la transmisión del término gastronomía se usa fundamentalmente para designar los nuevos códigos culinarios desarrollados en la modernidad, y, sobre todo, a partir de las normas y los procedimientos de la cocina francesa, exportados al mundo como estética gastronómica por los *chefs* de los restaurantes con estrellas Michelin. Sin embargo, aunque la contribución francesa a la gastronomía es muy importante, tal visión exclusivista nos parece, a nuestro juicio, algo limitativa.

En lo que nos concierne, insistimos en el hecho de que la gastronomía abarca una gran variedad de culturas, incluidas las patrimoniales o colonizadoras, afronegristas e indígenas, y sus respectivas técnicas, ingredientes y platos, apuntando hacia su eclecticismo e interferencias culturales, aspectos que nos proponemos resaltar en la selección subjetiva de vocabulario gastronómico extraído de los idiomas autóctonos arahuaco, caribe, taíno, náhuatl, quechua,

³ UNESCO (s.n.) *Inintangible Cultural Heritage*. <https://ich.unesco.org/> y <https://www.unesco.org/archives/multimedia/subject/13/intangible+heritage/> (consultado el 10 de octubre de 2024)

guaraní, mapuche o aimara, entre otros. Aun más, aprovecharemos los criterios histórico-culturales y de variación lingüística retomados de los diccionarios de lengua española para analizar algunos vocablos del léxico culinario, en clave de herencia y transmisión entre esas lenguas autóctonas y el español peninsular convertido con el tiempo en el español atlántico o de ultramar. En este sentido, nosotros recalcamos en su origen, difusión y persistencia en los territorios hispanos, como por ejemplo las técnicas de preparación del ceviche o la incorporación del pavo/guajolote en la comida rural. Además, la elección de los términos comparados y de los recursos lexicográficos ha sido en función de su legado lingüístico, corroborado con la perspectiva de los significados culturales que intervinieron: uniformización o diferenciación, procesos aglutinadores, añadidos o hibridados de índole regional, nacional o global, tal como observaremos en la tabla comparativa entre algunos de los platos hispanos actuales más populares que se preparan en Venezuela y Colombia (arepas), España, Argentina, Chile y Brasil (empanadas) y México y Guatemala (quesadillas y tamales). La metodología que aplicaremos consistirá en inspeccionar algunas fuentes lexicográficas, etimológicas y enciclopédicas del vocabulario culinario y su relevancia funcional, en términos de formación, difusión y reconocimiento actual, aunque sin referirnos a un *corpus* culinario específico. Al mismo tiempo, compararemos tanto las definiciones y técnicas de preparación, como las etimologías y referencias culturales, con miras a encontrar elementos comunes o diferenciadores, tal como préstamos, adaptaciones, paralelismos y, sobre todo, cambios semánticos e innovaciones léxicas. Por ende, subrayaremos la gran unidad en diversidad de América Latina, denominación cuyo uso “resurgió con gran fuerza a fines del siglo XIX y principios del XX, aunque ahora con una definida connotación antinorteamericana, cuando algunos pensadores [...] esgrimían el legado de la tradición latina (Ariel) para contraponerlo al brutal expansionismo anglosajón (Calibán)” (Guerra Vilaboy 1997, 9). Sin embargo, nuestro estudio abarcará únicamente los territorios de habla hispana, a pesar de que el nombre designa a “todos los países ubicados del río Bravo a la Patagonia, incluyendo al Caribe de lengua inglesa y holandesa, Brasil, las antiguas colonias francesas y los grandes conglomerados indígenas” (Guerra Vilaboy *ibíd.*). En realidad, muchos vocablos de las lenguas indígenas han pasado a través del español a otras regiones e idiomas no hispanohablantes, difundiendo lo que, desde los primeros tiempos, los cronistas y colonizadores españoles habían descrito. El gran encuentro con una fauna y una flora tan diferentes en comparación con las de sus lugares de origen, se ha convertido en modos distintos de comer y cocinar, cuyas raíces exploraremos a continuación.

2. Préstamos, mestizaje e innovación gastronómica. Una mirada a los indigenismos culinarios

Hispanoamérica formó su identidad cultural y culinaria sobre un sustrato indígena, recogiendo de los españoles lo que éstos ya habían perfeccionado en la Edad Media y el Renacimiento. A estas influencias se añadieron, en menor medida, notas africanas y asiáticas, debido a la importación masiva de mano de obra esclava y a las oleadas migratorias procedentes de China, Japón o los países hispanos vecinos, en función de las etapas de progreso económico o de crisis.

Si tuviéramos que repasar brevemente la herencia culinaria transmitida a América, después de la conquista y la colonización española a partir de finales del siglo XV, sería preciso observar y analizar la evolución de la Península Ibérica de los últimos dos mil años. En concreto, a través de todas las influencias aportadas por las culturas gastronómicas más relevantes, como la fenicia, la hebrea, la griega, la romana y la árabe, la cultura española del Renacimiento y Barroco, ya formada y enriquecida, logró exportar a América un contenido cultural altamente desarrollado. El descubrimiento del nuevo continente en 1492 y las expediciones emprendidas por Cristóbal Colón y otros exploradores han constituido un verdadero detonante en la historia de España y América Latina. Desde ese momento ha habido un constante intercambio de productos entre lo que llamamos el Nuevo Mundo (el territorio de las Américas, y más tarde Oceanía y Australia) y el Viejo Mundo (la Península Ibérica en nuestro estudio) y, por extrapolación, toda Europa. Por defecto, la gastronomía española ha evolucionado mucho, al beneficiarse de hortalizas como patatas, pimientos, habichuelas-frijoles, tomates, calabaza, yuca/mandioca, el girasol, junto a muchas frutas exóticas: aguacate, piña, níspero, guayaba, pitayo, a la vez que importantes cereales como el maíz, especias y productos exquisitos: chile-guindilla, vainilla o incluso aves y animales: guajolote, alpaca, guanaco, ñandú etc. Uno de los productos estrella, el cacao en grano, ingrediente adorado por los mayas y los mexicas con el que preparaban el famoso *chocolatl*, deleitó a los españoles como una trufa, que inicialmente sólo la aristocracia podía degustar. Como señala Gutiérrez de Alva en *Historia de la gastronomía* (2012, 91-92), todo el siglo XVI y los dos siguientes representaron una revolución alimentaria en España y, por extensión, en todo el continente europeo. Los intercambios fueron afortunadamente recíprocos, ya que españoles y el resto de colonizadores aportaron una diversidad de proteínas a la comida usual mesoamericana, con alimentos, incluyendo el indispensable arroz desde Asia, pero también los cultivos más familiares del Viejo Mundo, entre los cuales estaban los de caña de azúcar, café, cebada, limones y naranjas, mangos, olivos, plátanos y animales como burros, caballos, cabras u ovejas. Por consiguiente, la alimentación conoció unos cambios definitivos y siguió

refinándose a medida que se ensanchaban los territorios conquistados y se consolidaban las vías de comunicación y comercio entre la madre patria y las colonias. No obstante, algunas plantas no fueron muy bien recibidas por los españoles, siendo tachadas de tóxicas, o simplemente no aceptadas en la cultura ibérica, como, por ejemplo, insectos, algas, perros *xoloitzcuintli* o variantes terribles de tarántulas o carne humana, que solían integrar la cocina de algunas tribus amerindias.

Al abarcar las principales etapas de la conquista hispánica de América, Antonio Alatorre proporcionaba su acertada visión comparatista, poniendo de relieve la rápida propagación y asimilación de la lengua castellana, a la par con su cultura y la religión cristiana:

“La hispanización del Nuevo Mundo ofrece ciertas semejanzas con la romanización de Hispania y con la arabización de España. Al igual que los romanos y los árabes (a diferencia no sólo de visigodos, sino también de los ingleses, franceses y holandeses que colonizaron otras regiones de América), los conquistadores y los pobladores españoles se mezclaron racialmente desde un principio con los conquistados, y este mestizaje de sangre fue, desde luego, el factor que más contribuyó a la lengua y cultura de España. Los romanos latinizaron con pasmosa rapidez toda la península⁴ (salvo el territorio vasco), y el latín de los escritores hispanos de los primeros siglos de nuestra era (...). Los moros arabizaron profundamente a España a partir del siglo VIII (...). En la historia americana, (...) el primer siglo de conquista (había) escritores de sangre americana como los mexicanos Hernando de Alvarado Tezozómoc y Fernando de Alva Ixtlilxóhtitl y los peruanos Felipe Guamán Poma de Ayala y Garcilaso Inca de la Vega. A fines del siglo XVI ya estaban echadas en todo el continente las raíces de la lengua nacional de los países hispanoamericanos de hoy.” (Alatorre 1979-1991, 2015-216).

En este sentido, tal como aseveraron varios estudiosos, las hibridaciones y las nuevas creaciones fueron modestas al principio, aunque el español trasplantado a ultramar, ya con sus variantes dialectales heterogéneas, empezó paulatinamente a producir variantes locales, otorgando nombres “a muchas cosas, animales, plantas, costumbres... Unas veces se adoptaban indigenismos que estaban a mano, otras, se recurrió a palabras españolas que no significaban exactamente lo mismo y (...) a medida que todo se iba conociendo mejor (...) comenzaron las innovaciones locales.” (López Morales 1992, 281). Dicho de otra manera, asistimos a una importante denominación de la nueva realidad, que iba al compás con las regionalizaciones lingüísticas, las modalidades socio-culturales y geográficas,

⁴ Se trata de la Península Ibérica (n. n).

que no solamente se extendería a los *criollos*, es decir a los hijos de europeos ya nacidos en territorios americanos, sino que sería un rasgo distintivo de su personalidad, según J. A. Frago y M. Franco Figueroa (2001: 300).

De lo antes mencionado, desprendemos como característica general del léxico americano, el hecho de que refleja, sin duda, amplias variedades dialectales, determinadas por áreas regionales muy extendidas y realidades étnicas y antropológicas diferentes, con porcentajes fluctuantes de lenguas y poblaciones amerindias, cuyos índices de vitalidad lingüística son bastante inestables sino débiles. Aún más, un inmenso acervo léxico americano, ha dado lugar a cuatro grupos fundamentales, de los cuales, los más extendidos y complejos por analizar, a nuestro juicio, serían los primeros dos:

- *Léxico patrimonial*, adaptado a las realidades americanas y fuente de creaciones con la observación de que pueden ser “términos creados en América por derivación, composición o acortamiento sobre bases españolas, sea palabras que tienen el mismo significante en España, pero han alterado el significado en América” (Vaquero de Ramírez 2003, 39); también incluye el grupo de los *-ismos*: regionalismos, marinerismos, arcaísmos, andalucismos, canarismos;
- *Léxico autóctono* o *indígena*, prestado de las lenguas indígenas cuya difusión varía mucho, tal como ejemplificaremos en el presente análisis del vocabulario culinario;
- *Léxico africano* o *afroamericano*, presentes en determinadas zonas, especialmente las Antillas, a raíz de la importación de los esclavos africanos, aunque en la actualidad experimenta una caída notable;
- *Léxico extranjerizante* (calcos y préstamos tomados de otros idiomas, en función de las migraciones pasadas o actuales) entre los cuales se destacan: los anglicismos (México y Estados Unidos), los italianismos (zona Río de la Plata y Argentina), los lusismos y los germanismos, en zonas más restringidas (Torres Torres 2000, 9).

Los cuatro bloques enumerados acuñaron variantes combinatorias exóticas y mestizadas, acordes con cada territorio (véase, por ejemplo, las peculiaridades dialectales que se dan entre los espacios urbanos y rurales), y, asimismo, crearon:

“(...) un rico mosaico léxico, de base hispánica, con zonas donde puede destacarse el componente indígena (andina, centroamericana o yucateca, zonas donde lo africano es evidente (Caribe) y zonas en que lo patrimonial acoge y asimila otras influencias (países del Sur). Cuando se habla del mestizaje como elemento caracterizado de la cultura hispanoamericana, se alude a este proceso de fusiones étnico-culturales, reflejado en el léxico.” (Vaquero de Ramírez 2003, 40).

Tras este enfoque general, matizamos que trataremos los términos analizados como americanismos léxicos, unos de los conceptos bastante discutibles en la lingüística hispánica, que se refiere a cualquier unidad léxica o valor semántico utilizados en alguna parte de Hispanoamérica. Sin embargo, al analizar el concepto, hay que tener en cuenta dos criterios distintos: por un lado, el histórico-etimológico, basado en el origen del término, y, por el otro, el sincrónico (contrastivo entre todos los países y España o no contrastivo) que aborda usos y su difusión actual. De todas maneras, en la práctica lexicográfica, J. P. Rona (1969, 135-140) observa la compleja distinción entre los americanismos de origen y americanismos de uso, que finalmente ha sido resuelta, prescindiendo de todos los *-ismos*, indicando meramente la extensión geográfica general o particular, lo que también aplicaremos a nuestra metodología.

*

De hecho, las áreas léxicas donde se ha podido estudiar mayor presencia de indigenismos han resultado la flora, la fauna, la alimentación, seguidas por la organización social, los objetos del hogar, la vestimenta y la toponimia. Por tanto, si consideramos las primeras tres para nuestra ejemplificación, llegamos a la cocina tradicional indígena, cuyos términos, cierto es, se ven más implantados en las hablas locales. Aún más, señalamos que tal como ocurrió en la mezcla étnico-racial, la lengua y su vocabulario tuvieron que adaptarse para reflejar la nueva realidad, a través de procedimientos muy diferentes: revitalizando lexemas obsoletos en España, mezclando o inventando otros. A continuación, pasaremos revista a las más destacadas clases de indigenismos y algunos términos relacionados al acto de comer, a los objetos, ingredientes, técnicas históricas o utensilios tradicionales de la cocina. Recordamos que en un principio, fueron las abuelas y mujeres, cocineras en su mayoría, quienes conservaron los saberes y sabores de las tradiciones culinarias, gracias a las recetas transmitidas de generación en generación. Creemos que todos esos alimentos y prácticas repetidas dan identidad a cada región o pueblo, al relacionar saberes ancestrales con expresiones vivas de la gastronomía actual. Por consiguiente, la terminología, incluso los indigenismos, enriquecidos gracias al mestizaje, se convierten en instrumentos de aprendizaje cultural y lingüístico, testimonio por la unidad en diversidad entre todos los países del continente americano.

- a) *Léxico arahuaco*⁵. El arahuaco es una de las lenguas de las Antillas, una de las fuentes más antiguas de indigenismos. A consecuencia de la invasión de las Antillas por los *caribes* antes de la llegada de los españoles, a veces resulta difícil diferenciar entre el a) y c).

⁵ La mayoría de las referencias y ejemplos citados han sido extraídos de Antonio Torres Torres 2016. *El Español de América*, Universitat de Barcelona, pp. 76-84.

Ejemplos: *jaiba* (especie de cangrejo), *guanajo* (pavo), *guaicán* (pez), *canoas*⁶ etc.

- b) *Léxico arahuaco taíno*. El arahuaco insular que se localiza en Haití, República Dominicana, Cuba, Puerto Rico, Jamaica y otras islas se ha denominado dialecto taíno. La mayoría de los taínismos se usan poco en España, aunque se encontraron desde las Crónicas de Indias.

Ejemplos: *yuca*, *maíz*, *barbacoa* (parrilla para asar la carne), *budare* (un tipo de cazuela utilizada para cocer el pan de maíz), *cebucán/sebucán/cibucán/sibucán* (colador para la yuca rallada), *ají* (guindilla picante), *tuna* (higo chumbo), *dahao* (pez comestible), *maguey* (agave).

- c) *Léxico caribe*. Aunque son complicadas las diferencias del arahuaco, se atribuyen los siguientes: *caníbal*, *caribito* (pez piraña), *caimán*.

- d) *Léxico caribe-cumanagoto*. Se observan bastantes préstamos cumanagotos en el español de América, mientras que algunos son conocidos en el español ibérico.

Ejemplos: *manare* o *mapire* (especie de cesta), voces muy difundidas por América: *totuma* (vasija hecha con la mitad de una calabaza), *arepa* (torta redonda de maíz), *masato* (bebida de maíz, plátano o yuca fermentados), *guacharaca* (un ave parecida a la gallina), *patilla* (clase de sandía), *ayama* (calabaza cuyo fruto es el mayor de los vegetales).

- e) *Léxico nahua o náhuatl*. La lengua corriente del Imperio azteca ha tenido tanta importancia que algunas voces se transmitieron al español general: *tequila*, *chicle*, *coyote*, *tomate*, *caco*, *chocolate*, *cacahuete* (cuya forma originaria *cacahuate* permanece en México), *aguacate*, *metate* (piedra para de moler), *comal* (disco de barro para cocer o tostar), *tecomate* (vasija de barro).

Otros ejemplos que aluden a la alimentación: *pulque* (bebida alcohólica, blanca y espesa, obtenida tras la fermentación del maguey), *pinole* (harina tostada de maíz), *pozole* (cierto guisado), *totopo/totoposte* (rosquilla de harina de maíz), *chancaca* (masa preparada con azúcar o miel), *camote* (su área de difusión es mayor que el taíno *batata/ boniato* o *patata dulce*).

Una de las historias curiosas que solapa aspectos en cuanto a herencia y transmisión lingüística, es la del origen del pavo, ave de corral domesticada y conocida en todo el mundo (*Numida meleagris*). Según los investigadores (Gutiérrez de Alva 2012, 92-93), se rastrearon dos teorías válidas: una aboga por una especie de gallina europea, mientras otra argumenta que el pavo no llegó al Viejo Continente hasta el siglo XVI. La mayoría de los estudiosos opina

⁶ Algunos aparecen en el *Diario de navegación* de Cristóbal Colón y se recogen en el *Diccionario* de Antonio de Nebrija que se publicó en Salamanca en 1492.

que el primer ejemplar domesticado procedía de América, a través de la ruta hispana. Así, de *huexolotl* en lengua *náhuatl* proviene el nombre del actual *guajolote* mexicano, un pavo silvestre, originario de la Península Yucatán desde la antigüedad precolombina. Parece que criar pavos era una antigua costumbre indígena, tal como resaltó Bernardino de Sahagún en su *Códice florentino*, el estudio etnográfico bilingüe del siglo XVI sobre Mesoamérica, acompañando por un dibujo de mujeres engordando un pavo. Sea cual sea la perspectiva, el pavo fue un éxito rotundo en Europa, cuando se presentó como exquisitez en los banquetes de Enrique VIII de Inglaterra y en la boda de Carlos IX de Francia en 1570, la primera vez que lo habían mencionado las fuentes escritas (Petrescu, 2024, 46-47).

- f) *Léxico maya*. Las voces mayas se observan casi exclusivamente en la Península Yucatán y se notaron también préstamos mayas al taíno. Ejemplos: *cenote* (depósito de agua dulce de gran profundidad), *chichibé* (planta cuyas fibras se utilizan para confeccionar escobas y hamacas). Como elemento incierto, “las más antiguas representaciones de hombres fumando se encuentran en códices mayas; los cronistas no citan esta voz [el cigarro], sino *cañutos de humo o de tabaco*” (Torres Torres 2000, 80).
- g) *Léxico quechua*. Solía ser el idioma hablado en el Imperio Inca y en el presente representa la variedad con el mayor número de hablantes (unos diez millones de hablantes) y la más extendida difusión americana, principalmente en el Perú. Ejemplos: *cáncana* (asador), *callana* (cierto plato de barro), *cocaví* (provisión de víveres para un viaje), *chupe* (especie de guisado con patatas), *mote* (maíz granado hervido con sal/postre de maíz y trigo caliente), *sanco/sango* (gachas de maíz o yuca), *carapulcra* (guisado de carnes y patatas deshidratadas). A nivel del español general, se conoce el *pisco* (cierto aguardiente de uva) y la literatura diseminó la infusión de yerba *mate* (significado original: calabaza vaciada para tomar hierba mate), *papa* (patata), *coca* (con su propósito medicinal de aliviar dolores), *choclo/chócolo* (mazorca de maíz); animales como: *llama*, *vicuña*, *cuy* (conejillo de Indias, uno de los platos típicos de Bolivia), *suri* (avestruz, ñandú) etc.
- h) *Léxico aimara*. Una de las lenguas que comparte más de un cuarto de raíces lingüísticas con el *quechua*, por ende es bastante complicado delimitar muchas voces. Se extienden tanto por Perú como por Bolivia, ejemplos como *cañihua/cañahua* (planta y fruto que se fermenta para preparar el alcohol barato llamado *chicha*), *poroto* (alubias), *palta* (aguacate), *pacay* (guamo/guabo).
- i) *Léxico chibcha*. La difusión de las voces se delimita entre Colombia, Panamá, Ecuador, ya que pocos lexemas se transmitieron al español general.

Ejemplos: *chicha* (bebida alcohólica), *fica/jica* (hoja de maíz), *cura* (aguacate), *fique* (agave, pita).

- j) Léxico *araucano* o *mapuche*. Algunos araucanismos perduraron en Chile y Argentina, aunque hay pocos préstamos al español general.

Ejemplos: *curanto* (algas y mariscos cocidos en hoyos con piedras), *pilco/pirco* (guiso de alubias con maíz), plantas con propiedades medicinales: *cacha(n)lagua*, *palqui/palque* y *boldo*.

- k) *Léxico tupí-guaraní*. En realidad, se pueden distinguir dos variedades: el del sur, llamado *guaraní*, que ha dado lugar al *guaraní* moderno, lengua cooficial en Paraguay, y, respectivamente, el del norte, que ha originado el *tupí* moderno, presente también en otros territorios, incluso en Brasil. Ejemplos: *cobayo*, *tiburón*, *tapioca* (del portugués brasileño), *paca* (liebre), *pirarucú/pacú* (un tipo de pez grande), *yacaré* (caimán), *maracaná* (tipo de loro), *mandí* (especie de bagre) etc.

Vale la pena mencionar que, aparte del lexema *tapioca*, sobre todo en los países rioplatenses, se destacan unas voces que entraron en el español general debido al portugués brasileño, como *ananá(s)* (piña), *mandubí* (cacahuete), *guapurú* (árbol de cuyo fruto se produce vino) o *acajú* (caoba).

*

Más allá de los indigenismos procedentes de las lenguas autóctonas, cabe subrayar que en el español americano intervino léxico de las lenguas africanas, sobre todo en las áreas de las Antillas y el interior del Caribe, el litoral del Pacífico desde Panamá hasta Ecuador. De todas maneras, los africanos que habían llegado a América como esclavos o mano de obra barata, hablaban una multitud de lenguas, por lo cual representaron “un sustrato, una fuerza lingüística mínimamente uniforme” (Frago y Franco Figueroa 2001, 295). Ejemplos: *mondongo* (pedazo de estómago de cerdo, vacuno u ovino, destinado al consumo), *tafia* (líquido que sale de la caña de azúcar pasada por el trapiche), *cola* (semilla de un árbol ecuatorial apreciada por sus propiedades tónicas y vitamínicas).

Al mismo tiempo, quedó constancia de los vínculos entre la población afronegrista y los hispanohablantes, lo que estableció una buena integración lingüística de los primeros. El mestizaje, tal como lo hizo con las lenguas africanas, intervino en el caso del inglés, alemán, francés o italiano, contribuyendo, de esa manera, al mestizaje cultural y a las innovaciones lingüísticas. Ejemplos: del inglés: *panqueque* en vez de crepes (tortita de harina frita), *frízer* (Argentina, congelador), *lonchera* (México, cajita de merienda); del francés: *frigider* (Chile); del italiano: *grapa* (Argentina, aguardiente obtenido del orujo de la uva), *ricota* (requesón); del portugués: *feijoada* (Uruguay, guiso a base de frijoles negros), *caneca* (vasija), *cachaza* (aguardiente), *charuto* (cigarro puro) (según Torres Torres 2016, 86).

3. Diccionarios y otras fuentes de léxico gastronómico. Ejemplos comparados y su legado cultural

Recordamos una fecha significativa, marcada en el Siglo de las Luces, cuando se funda La Real Academia Española (RAE), más exactamente en 1713, mientras que la primera Academia Americana, la Colombiana, se inaugura en 1871, seguida por las de Ecuador, México, El Salvador, Venezuela, Chile, Perú, Guatemala y, así en adelante, hasta contar con veintitrés por todo el continente⁷.

La historia e incorporación de los americanismos en los diccionarios españoles generales representa un tema pendiente para una futura investigación en el campo de la lexicografía. De todas maneras, el *Nuevo Diccionario de Lengua Castellana*, realizado por Vicente Salvá en 1864, constituyó una obra pionera, al adelantarse incluso a la Real Academia Española con “la primera recogida sistemática de americanismos que se conoce en la historia del español, utilizando para ello un método riguroso que denota en el autor una concepción sorprendentemente moderna de las variedades diatópicas del español” (Azorín Fernández 2008, 13). Aunque, tal como observa la investigadora, hasta la edición décima quinta del DRAE (1925) la Academia no había incluido en sus entradas los dialectalismos ya comprobados como uso, la obra de Salvá determina un cambio radical de visión y de denominación para el *Diccionario de la lengua castellana* (desde 1726) que se convirtió en *Diccionario de la lengua española*. Además, según las afirmaciones de Salvá, citadas por Dolores Azorín Fernández, se ha arrojado luz en:

“[...] una casi total omisión [en el DRAE] de las voces que designan las producciones de las Indias [...] y de de los *provincialismos* –estos es, las verdaderas variantes diatópicas– de sus habitantes. Según el texto, el autor parece distinguir entre los vocablos que designan conceptos propios de la cultura americana, independientemente de que puedan ser conocidos o no fuera de América, y aquellos otros cuyo empleo se supone privativo de los habitantes del nuevo continente y que son desconocidos, por tanto, de los hablantes del español peninsular”. (Azorín Fernández 2008, 13)

En el siglo XX, más exactamente en el 1951 nace la Asociación de las Academias Americanas de Lengua Española (ASALE), que celebra reuniones

⁷ Los países que conforman la Asociación de Academias de la Lengua Española (ASALE) son: España, Colombia, Ecuador, México, El Salvador, Venezuela, Chile, Perú, Guatemala, Costa Rica, Filipinas, Panamá, Cuba, Paraguay, Bolivia, República Dominicana, Nicaragua, Argentina, Uruguay, Honduras, Puerto Rico, Estados Unidos y República de Guinea Ecuatorial. (n. n).

periódicas y coopera en varios ámbitos, unos de los más importantes siendo el lexicográfico (Torres Torres 2000, 17). En este campo, se han observado varios criterios como ejes ordenadores que seleccionan, definen, o clasifican los americanismos léxicos, entre los cuales siguen incidiendo el criterio del origen (*indigenismos*) y el criterio enciclopédico o el de los conceptos representativos de las realidades y culturas americanas, aunque esas pueden ser patrimonio compartido con el ibérico (*voces patrimoniales*). A todo ello, se ha sumado el criterio del uso actual en Hispanoamérica, lo que nos ha servido también para nuestro análisis comparado. Efectivamente, la continua incorporación de americanismos al *Diccionario de la Real Academia Española* (DRAE) responde a una política panhispánica actual que reconoce la unidad en la diversidad del español y documenta los usos léxicos propios de América en recursos complementarios como el *Diccionario de americanismos*.

En concreto, el DRAE ha agregado cada vez más las voces americanas de una forma sistémica. Si la edición de 1992 incluía 13.758 americanismos, traducido en 12% del total de 83.018 entradas, la última impresa, es decir la vigesimotercera del 2014⁸ cuenta con 93.111 entradas, de las que casi 19.000 americanismos⁹, porcentaje equivalente a más de 20%. Como colofón, la RAE presentó en diciembre de 2024 la actualización con el número 23.8 de su *Diccionario de la Lengua Española*, en colaboración con ASALE y con la adición de 4.074 nuevos términos, expresiones y acepciones, muchas prestadas de otras lenguas, además de una imagen *online* revitalizada. Santiago Muñoz Machado, presidente de ASALE, adelantó que “esta sería la última actualización de la edición número 23 del diccionario, cuya 24 versión se prevé para 2026, donde la RAE intentará términos que han faltado, en especial americanos”, efecto que podrá duplicar la cantidad de lexemas.

Con el mismo patrocinio y atenta supervisión de ASALE, en 2009 se publicó el *Diccionario de americanismos*¹⁰, obra que constituye un repertorio léxico muy extendido con los términos específicos del español americano, y que recoge 70.000 voces, lexemas complejos, frases, locuciones y un total de 120.000

⁸ Se conoce como la edición del tricentenario de la Academia Española y representa la labor de colaboración de las veintitrés entidades integradas en ASALE. (n. n), URL: <https://dle.rae.es/docs/DLE-Edici%C3%B3n-del-Tricentenario.pdf> (consultado el 7 de marzo de 2025)

⁹ V. El nuevo *Diccionario de la RAE*, URL: <https://www.rtve.es/noticias/20141016/nuevo-diccionario-rae-mas-americano-mas-cheli-drae-real-academia-espanola/1031284.shtml> (consultado el 28 de marzo de 2025)

¹⁰ *Diccionario de americanismos*, URL: https://www.asale.org/sites/default/files/Prologo_Diccionario_de_americanismos.pdf (consultado el 2 de mayo de 2025)

acepciones. Ciertamente es que estos dos recursos lexicográficos no recogen únicamente léxico gastronómico, ni se concentran en una sola área geográfica, social o cultural.

Antes de refinar el campo lexicográfico a nivel gastronómico, aseveramos que hemos comprobado a lo largo de nuestras investigaciones en la última década unas actualizaciones de voces de América cada vez más numerosas en los diccionarios españoles impresos o digitalizados, de forma ordenada y metódica. Efectivamente, todo el caudal de americanismos debería ser revisado constantemente, ya que ello implica supresiones o enmiendas y adiciones, con criterios que, según muchos expertos, deberían atender más a la frecuencia que a las autoridades. Como corolario, se indaga sobre la importancia de los usos americanos tanto en las nuevas gramáticas académicas como en los trabajos en conjunto y por separado de las Academias americanas. Gracias a esa inmensa labor, en los últimos cincuenta años se han publicado diccionarios de peruanismos, argentinismos, mexicanismos, chilenismos, colombianismos etc. para cada una de las variantes diatópicas relevantes.

Sobre gastronomía se han realizado muchos estudios que describen la forma de preparar un plato o, la historia culinaria de una región, pero elaborar diccionarios que recogieran ese patrimonio cultural tan vivo en cada región no suele ser tan habitual, porque hay una predominancia de crear diccionarios nacionales o especializados en otros campos más complejos, por ejemplo, diccionarios económicos, jurídicos, técnicos o médicos. Como curiosidad, muchos americanismos se han incorporado últimamente como calcos al léxico culinario de otras lenguas y no necesitan traducirse a nivel mundial, entre ellos están los términos: *taco*, *burrito*, *nachos*, *ceviche* etc.

De todos modos, la iniciativa pionera del *Diccionario de gastronomía en España* data del año 2019 y reúne 7000 términos en castellano, acompañados por sus equivalencias en inglés y francés. Julio Valles, presidente de la Academia Castellana y Leonesa de Gastronomía, coordinó el proyecto y contó tanto con la participación de lingüistas, como de periodistas gastronómicos, cocineros de prestigio y otros académicos o expertos en la materia, “profundizando en sus aplicaciones gastronómicas e identificando los diferentes nombres con que se conoce cada concepto en las regiones y países iberoamericanas”.

Sin embargo, unas de las obras lexicográficas más importantes es el *Diccionario panhispánico de gastronomía* (2024) elaborado por ICPNA – La Academia Peruana de Lengua, proyecto presidido por Marco Martos. El diccionario ofrece un vasto repertorio de la gastronomía de cada uno de los países integrante de ASALE, con el fin de poner de relieve sus identidades dentro de la diversidad culinaria panhispánica. Cabe señalar que por cada región se presentan los patrones lingüísticos del léxico gastronómico más emblemático, entradas que oscilan entre cincuenta y doscientas por país. El

procedimiento detallado recurre a descripciones lingüísticas y enciclopédicas, e incluye, a veces, fotografías o ilustraciones¹¹.

Hemos elegido analizar un término muy popular y conocido bajo el nombre *ceviche* (mayoritariamente Perú), *cebiche* (mayoritariamente Ecuador) o *sebiche* (Cuba, Honduras, México) y sus distintas representaciones, de las cuales mostramos dos en la figura 1. El primer artículo pertenece al *Diccionario de la gastronomía* antes mencionado y es muy detallado, casi enciclopédico, indicando que el término se heredaría de la época preincaica o podría ser una simple hibridación del *escabeche* medio-oriental, técnica transmitida por los conquistadores españoles. El ingrediente principal es el pescado y su manera de marinarlo crudo, añadiendo algún cítrico, cebolla, guindilla y sal, y otros elementos que suelen variar. El segundo artículo es del DRAE y su descripción es mucho más sintética.

cebiche

Preparación a base de pescado crudo cortado en trozos y marinado con lima o naranja agria, ají amarillo u otros chiles, cebolla, sal y cilantro fresco. Hay cierta discusión sobre su origen, ya que hay quien considera que tiene su origen en la cultura prehispánica moche en Perú, pero según otros estudios esta preparación es una variación del escabeche, la forma de conservación mediterránea de inspiración medio-oriental llevada a las Américas por los españoles. Se trata de un plato muy difundido por todos los países de Latinoamérica, cada uno con sus diferentes versiones: en algunos países se le añade tomate, en otros aguacate, etc. En Perú se considera plato nacional y han sido precisamente los cocineros peruanos los que le han dado su actual fama internacional.

ceviche

Tb. **ceviche**.

Quizá del ár. hisp. *assukkabāj*, y este del ár. *sikbāj*.

Artículo

Sinónimos o afines

1. m. Plato propio de algunos países americanos, hecho a base de pescado o marisco crudo cortado en trozos pequeños y preparado en un adobo de jugo de limón o naranja agria, cebolla picada, sal y ají.

Sin.: *sebiche*.

Sinónimos o afines de «ceviche»

Fig. 1. Representación del término *ceviche/cebiche* según fuentes lexicográficas distintas

Según otras fuentes consultadas que detallamos en la parte de webgrafía, descubrimos que desde 2004, el *ceviche*, cuya etimología procedería del *quechua siwichi* y significa «pescado fresco» ha llegado a ser unos de los símbolos del patrimonio cultural de Perú, así como en uno de los platos más renombrados de toda la gastronomía de la costa del Pacífico. La variante ecuatoriana se diferencia de la versión peruana, por el hecho de que se decora con mucho cilantro y cebolla roja. Desde el punto de vista gastronómico, las técnicas utilizadas patentizan la preparación del *marinado*, a la vez una exquisita conservación, sin procesamiento térmico, de modo que las calidades organolépticas del pescado mejoren y exploten plenamente los sabores básicos: salado, ácido y picante.

¹¹ El *Diccionario panhispánico de gastronomía*, descripción del contenido del 2021, URL: <https://cultural.icpna.edu.pe/wp-content/uploads/2021/08/Marco-Antonio-Lovon-Cueva-El-diccionario-gastronomico-hispanoamericano.pdf> (consultado el 20 de mayo de 2025).

Además, en muchas zonas costeras, se utiliza el pescado fresco, combinado con gambas, calamares u otros moluscos; otro procedimiento esencial consiste en cómo filetear el pescado o el marisco, de tal manera que se mantengan las aromas de los ingredientes y que esos absorben toda la sal y el zumo de lima, limón, naranja ácida, maracuyá o carambola por todas partes.

Como se puede observar, hay cierta libertad al elegir el tipo de pescado o marisco que se emplea, llegándose incluso a pulpo o vieiras. La versión peruana actual que está más de moda es *el tiradito*, una síntesis peruano-japonesa, conocida por muchos platos estrella bajo el nombre de *cocina nikkei*. La técnica consiste en combinar ingredientes de la tierra peruana de primera calidad con métodos culinarios japoneses, una costumbre introducida por la primera oleada de inmigrantes japoneses, los cuales se establecieron hacia 1899 en el puerto del Callao, cerca de Lima. En aquella época, los asiáticos se afincaban en el Perú, huyendo de las malas condiciones económicas de su país y de algunas enfermedades graves. Desde trabajos temporales como agricultores, hasta reconocidos *chefs* de cocina, en el presente hay más de medio millón de peruanos de ascendencia japonesa, mientras Lima se ha consolidado, desde unos años, como la capital gastronómica de Hispanoamérica.

A continuación, en la figura 2, presentamos una comparación entre unos de los platos hispanos más notorios y, a la vez, hibridados, que se dan en Venezuela y Colombia (las *arepas*), en México y Guatemala (*quesadillas* y *tamales*), en España y casi toda Hispanoamérica (*empanadas*), junto a su definición, técnicas de preparación, etimología y referencias culturales, antes de proceder a las conclusiones:

Nombres y denominaciones locales	Definición y preparación	Etimología y referencias culturales
AREPAS (Colombia, Venezuela)	Especie de pan de forma redonda, hecho con harina maíz ablandada a fuego lento o con harina de maíz tratado o precocido (desde 1950 se llama <i>maíz nixtamalizado</i>). Abiertas y con varios rellenos: mayonesa, queso, huevos cebolla, tomate, aguacate, chicharrones, tiernos o crujientes. Tradicionalmente, las indígenas remojaban los granos de maíz y los	“El lema <i>arepa</i> proviene del vocablo erepa ‘maíz’, préstamo fonético derivado del idioma indígena <i>cumanagoto</i> ”. ¹² La arepa guarda mucha similitud con “la <i>tortilla</i> panameña, el <i>talo</i> vasco y con la <i>manaba</i> ecuatoriana; además, la arepa rellena es similar a la <i>pupusa</i> de

¹² Picón Salas, M. 1962. *Pequeña historia de la arepa* en Picón Salas, *obras selectas* (2.ª edición). Caracas y Madrid: Edime. pp. 295-300 (consultado el 30 de marzo de 2025).

Nombres y denominaciones locales	Definición y preparación	Etimología y referencias culturales
	molían para hacer harina. Después, con la masa, formaban bolas que aplanaban. Utilizaban <i>batanes</i> (piedra para moler) y <i>budares</i> (plancha de barro para cocinarlas). La cocción es en seco: asadas u horneadas.	América Central y a la <i>gordita mexicana</i> ¹³ . El lexema se registró en 1884 por la Real Academia Española. Francisco Javier Pérez en el <i>Diccionario histórico del español de Venezuela</i> (2012/2016) rastrea “el primer testimonio escrito de este vocablo y sus variantes que se remonta entre 1539 y 1553 en el libro <i>Viaje y descripción de las Indias de Galeotto Ceí</i> ” ¹⁴ .
EMPANADAS (España, Hispanoamérica, Filipinas).	Masa de pan, quebrada u hojaldre, rellena de carne, pescado, verdura, aceitunas, fruta o queso. Se emplea harina de trigo, menos de maíz, se amasa y se corta en discos para el relleno. Normalmente se hornean, aunque raras veces se fríen en grasa.	La tradición cuenta la costumbre rústica de rellenar rebanadas de pan con carnes o vegetales, una comida muy práctica para pastores o viajeros. Paulatinamente, la masa de pan y el relleno empezaron a cocerse en el horno, dando origen a los <i>pâtés</i> franceses, el <i>calzone</i> italianos, la <i>empanada gallega</i> o a los <i>Cornish pastry</i> británicos. En la Península Ibérica, Ruperto de Nola ¹⁵ mencionó la primera vez “las empanadas rellenas de atún,

¹³ Univision 2016. *Clásicos de la comida latina: arepas, pupusas y gorditas* 14.01.2016. <https://www.univision.com/estilo-de-vida/tipos-de-comida/clasicos-de-la-comida-latina-arepas-pupusas-y-gorditas> (consultado el 20 de abril de 2025).

¹⁴ Cey, G., Lovera J. R. 1994. *Viaje y descripción de las Indias*, 1539-1553. Caracas: Fundación Banco Venezolano de Crédito. <https://search.worldcat.org/title/viaje-y-descripcion-de-las-indias-1539-1553/oclc/909795847> (consultado el 15 de abril de 2025).

¹⁵ Citado por Long, J. 2003. *Conquista y comida: consecuencias del encuentro de dos mundos*. México: UNAM, p. 8.

Nombres y denominaciones locales	Definición y preparación	Etimología y referencias culturales
		anguilas, trucha y congri, en un recetario catalán publicado en Barcelona en 1520”¹⁶. Así, las empanadas hispanoamericanas derivarían de tales empanadas europeas y de las de Oriente Próximo (<i>fatay</i> y <i>sfihas</i>) transmitidas por los españoles.
QUESADILLAS (México)	Tortilla de maíz (México central y sureño) o de trigo (la capital y México norteño), doblada por la mitad, rellena, a veces de queso, y, a veces, de otros ingredientes, propia de la cocina mexicana. Se cocinan en sartén hasta que la tortilla esté ligeramente dorada y el queso derretido. Abiertas de una orilla, al relleno se añade lechuga, queso fresco, flor de calabaza, <i>huitlacoche</i> (hongo del maíz) y una buena salsa.	En español, las palabras terminadas en <i>-illa</i> se consideran diminutivos desde el siglo XV. Por tanto, una quesada sería un plato o postre que incorpora queso como ingrediente base en su elaboración. Tras esta lógica, el término es un diminutivo de origen español ibérico y no proviene de la lengua <i>náhuatl</i>.
TAMALES (México, Guatemala)	Especie de empanadas de harina de maíz o arroz rellenas de carne, queso, fruta (guayaba, piña y zarzamora) o verduras, envueltas en hojas de plátano, maguey, aguacate o mazorca de maíz. Cocidos al vapor o en agua. Tienen un formato rectangular, hay una gran variedad de sabores, rellenos y salsas. En la era prehispánica, los mexicas comían tamales con muchos ingredientes: pavo, flamenco, rana, ajo lote, conejo, pescado, huevos de	El tamal proviene de la lengua náhuatl, <i>tamalli</i>, al ser un alimento precolombino, propio de algunas gastronomías mesoamericanas. Hay evidencia arqueológica sobre los tamales utilizados en rituales religiosos, en ofrendas y tumbas. En México, Fray Bernardino de Sahagún en <i>Historia general de las cosas de Nueva España</i> (siglo XVI), mejor conocido como

¹⁶ V. de Nola, R. *Llibre de doctrina per a ben servir, de tallar y del art de coch*. Alacant: Biblioteca Virtual Cervantes. <https://www.cervantesvirtual.com/obra/libre-de-doctrina-per-a-ben-servir-de-tallar-y-del-art-de-coch--1/> (consultado el 3 de mayo de 2025)

Nombres y denominaciones locales	Definición y preparación	Etimología y referencias culturales
	pavo, miel, frutas, calabaza y frijoles.	<i>The Florentine Codex</i> ¹⁷ , describió detalladamente los tamales. En Venezuela representan un tipo de comida navideña que se llama <i>hallaca/hayaca</i> , mientras que los tamales peruanos llevan el nombre <i>humitas</i> .

Fig. 2. Análisis y comparación entre variedades de platos difundidos en Hispanoamérica

En primer lugar, cabe mencionar que hemos elegido adrede unos platos muy populares, relacionados con la manera de vivir y de alimentarse de las poblaciones hispanas, con una amplia difusión geográfica, cultural y social y una rica tradición gastronómica. Mientras unos se remontan cultural o etimológicamente a sus civilizaciones y lenguas indígenas (el *cumanagoto* en el caso de las arepas y el *náhuatl* para los tamales), otros remiten al patrimonio heredado por los ibéricos (quesadilla) o incluso por los árabes (empanadas). En cuanto a su definición y preparación, todos suponen previamente el cultivo y la cosecha local de los ingredientes, y luego, una preparación cocida al vapor, frita o asada en el horno, en función de las preferencias culinarias del lugar, como, por ejemplo, el maíz en todas sus formas, incluso con un hongo asociado que se llaman *huiltlacoche*, el uso de flores de calabaza, las hojas de plátano y de maguey como envoltorio y rellenos con frijoles, pescado, piña, aguacates, tomates, huevos, carnes, chicharrones, cebolla, fruta, y verdura.

Al echar un vistazo a las referencias culturales, compartimos la opinión de Cristina Nunes dos Santos, de que “el gusto y las preferencias culinarias son formados culturalmente y controlados socialmente” (Nunes dos Santos 2007, 234-235). En otras palabras, los platos culinarios, sus sabores y su consumo más o menos frecuente son influenciados por criterios diferenciadores como la geografía, el clima, la raza, la clase social, a los que se añadirían la edad y la salud, la educación y la situación social, a la vez que otras costumbres o rituales asociados (véase la comida navideña etc.).

¹⁷ V. Sahagún B. fr. *Historia general de las cosas de Nueva España* s. XVI/ 1938, México: P. Robredo. Ed. digitalizada: https://ia601506.us.archive.org/24/items/b29827620_0001/b29827620_0001.pdf (consultado el 20 de mayo de 2025)

Hoy día, la cultura urbana global y la vida moderna demandan desgraciadamente cada vez más una comida rápida, promocionada por tantas cadenas de *fast-food*, sin señas de identidad alguna, pero ello no constituye el objeto de nuestro análisis, por tanto nos referiremos únicamente a los métodos de preparación y uso tradicional.

De todas maneras, la idea general que se desprende de nuestro estudio es que la manera de cómo se cocina y en qué modo se come define la identidad cultural de los pueblos y sus hábitos alimentarios. En realidad, en todas las culturas se encuentran preferencias y restricciones de ingredientes, a raíz de la religión, de la temporada, de la clase social o de otras cuestiones contextuales. Todas representan características de cada cultura que pueden llevar diferentes nombres, como en el caso de las arepas o las tortillas, para que al final demuestren su faceta ecléctica y multicultural. Cada gastronomía en sí podrá verse exquisita para personas pertenecientes a otras culturas y tradiciones distintas, pero en la actualidad, debido a tantas fusiones, hibridaciones, intercambios sociales o emigraciones a otras tierras, nos hallamos ante un panorama gastronómico hispanoamericano muy diversificado y difundido por todo el mundo, revelándose repleto de colores y sabores.

De hecho, cualquier identidad cultural-gastronómica no logra abarcar todos los elementos que han moldeado su herencia, formación y transmisión, pero cuenta con aspectos claves que se manifiestan a lo largo de la vida de esa identidad. Como valor añadido, todos los significados que se atribuyen a los alimentos y a los hábitos sociales por un determinado tiempo, acaban acomodando cada cultura en el mundo, y, de vez en cuando, salen a la luz en formas sorprendentes como la cocina de fusión, una de las más refinadas e innovadoras. Independientemente si se trata de Colombia, Venezuela, Perú, Ecuador, Argentina, México o Guatemala, cada país hispano y cada una de sus regiones manifiestan estas interferencias culinarias y semánticas en su marca cultural-gastronómica, socialmente formada a medida que su cultura incorpora los cambios a su carácter.

4. Conclusiones

Tras el breve análisis emprendido por el campo del léxico gastronómico, se puede afirmar que su herencia y transmisión atestiguan una parte importante de la cultura de un país o región que merece ser inventariada debidamente. Partimos de la idea de que los términos utilizados para describir los ingredientes, los platos de comida y las técnicas de elaboración reflejan la geografía, la historia, la economía, la cultura o los rituales de esos lugares. Al adentrarnos en la búsqueda de varias voces americanas relacionadas al acto de comer, sea de herencia indígena o transmisión colonial, sea de mestizaje y/o de innovación léxica, hemos podido comprobar una serie de aspectos complicados con respecto a las obras lexicográficas y su clasificación.

En primer lugar, el campo gastronómico en sí representa un reto para la lexicografía, ya que no es una tarea sencilla representar términos y significados culinarios para zonas distintas, porque muchas de las concepciones, a pesar de las reivindicaciones nacionalistas o regionalistas, suelen ser muy parecidas o incluso se confunden. Además, tal argumento apunta hacia las interdependencias culturales y las variedades dialectales en sincronía en un proceso lingüístico diacrónico, hecho que complica su diferenciación y sistematización en diccionarios especializados.

En segundo lugar, toda clasificación de léxico gastronómico resulta limitativa en lo que concierne su selección, delimitación y referencias culturales, debido a la versatilidad de los conceptos (fondos culturales) y las múltiples representaciones en el habla (formas cultas, formas populares, generales y específicas). A pesar de ello, se ha puesto de relieve el vínculo muy estrecho entre las culturas indígenas, mestizadas o modernas y los términos que caracterizan esas identidades, definiendo un binomio gastronomía-cultura con el que se podrá operar en futuras investigaciones, quizás aplicadas directamente a un corpus como CORDE.

Por último, cuando estábamos proponiendo el resumen del presente estudio, hemos topado con el anuncio sobre un número temático de la Revista *Letras* no. 64 (105) del año 2024, titulado “Léxico y gastronomía”, coordinado por el Instituto Venezolano de Investigaciones Lingüísticas y Literarias Andrés Bello. Esa noticia demuestra la necesidad y vitalidad de los estudios pluridisciplinarios, ya que la gastronomía no solamente se refiere al buen comer, sino también al hecho de describir y documentar su léxico, ese enorme bagaje lingüístico y cultural. El camino de la investigación queda ancho y largo, pero, con cada estímulo a degustar y reflexionar sobre la comida como identidad, herencia y transmisión, contribuimos a una mayor comprensión de la riqueza léxica y gastronómica en español, con sus múltiples variedades americanas.

BIBLIOGRAFÍA

- Alatorre, Antonio. 1978/1991. *Los 1001 años de la lengua española*. México: FCE-Tezontle.
- Ayora Díaz, S. I. 2018. “Cocina, gastronomía y modernidad: hacia una definición de lo gastronómico”. *Diario de Campo* 4: 6-14.
<https://revistas.inah.gob.mx/index.php/diariodecampo/article/view/13334>
(consultado el 26.04. 2025).
- Azorín Fernández, Dolores. 2008. “Para la historia de los americanismos léxicos en los diccionarios del español”. *Revista de Investigación Lingüística* 11: 13 -43.
- Brillat-Savarin, Jean Anthelme. 1825/1949. *The Physiology of Taste or Meditations on Transcendental Gastronomy*. Nueva York: The Heritage Press.

- Corominas, Joan y José Antonio Pascual. 1980. *Diccionario crítico etimológico castellano e hispánico*. Barcelona: Gredos.
- de Nola, Rupert. 2000. *Llibre de doctrina per a ben servir, de tallar y del art de coch*. Alacant: Biblioteca Virtual Cervantes., ed. digital basada en C. Amorós 1520, Barcelona y ed. Facsímil de 1988: Borriana, Ediciones Histórico Artísticas <https://www.cervantesvirtual.com/obra/libre-de-doctrina-per-a-ben-servir-de-tallar-y-del-art-de-coch--1/> (consultado el 26.08. 2025)
- de Sahagún, Bernardino. S. XVI/ 1938, *The Florentine Codex o Historia general de las cosas de Nueva España*, ed. Digitalizada. México: Ed. Pedro Robredo.
- Frago, Juan Antonio y Mariano Franco Figueroa. 2001. *Historia del español de América*. Madrid/Cádiz: Servicio de Publicaciones de la Universidad.
- Freedman Paul (ed.). 2007. *Food. The History of Taste*. Berkeley: University of California Press.
- Fusté-Forné, Francesc. 2016. “Los paisajes de la cultura: la gastronomía y el patrimonio culinario”. *Dixit* vol. 24, 1. Uruguay: Montevideo. http://www.scielo.edu.uy/scielo.php?pid=S0797-36912016000100001&script=sci_abstract (consultado el 26.03. 2025)
- Guerra Vilaboy, Sergio. 1997. *Etapas y procesos de la historia de América Latina*. Argentina: Clacso, <http://biblioteca.clacso.edu.ar/> (consultado el 26.03. 2024)
- Gutiérrez de Alva, Cecilia Isabel. 2012. *Historia de la gastronomía*. México: Red Tercer Milenio, <https://gastronomiamia.jimdofree.com/lecturas/historia-de-la-gastronomia/> (consultado el 26.12. 2024).
- Iturriaga, José N. 2015. *Gastronomía*. México: Conaculta/Debate.
- Long, Janet. 2003. *Conquista y comida: consecuencias del encuentro de dos mundos*. México: UNAM.
- López Morales, Humberto. 1992. “Los primeros contactos lingüísticos con el español de América”. En *Historia y presente del español en América*, 281-293. Valladolid: Junta de Castilla y León, Pabecal.
- Nunes dos Santos, Cristiane. 2007. “Somos los que comemos. Identidad cultural, hábitos alimenticios y turismo”. *Estudios y Perspectivas en Turismo*, 16: 234-242.
- Pearsall, Judy y Patrick Hanks (ed.). 2005. *Oxford Dictionary of English*. 2nd revised ed. Oxford: Oxford University Press.
- Petrescu, Olivia. 2024. *Cultură și gastronomie în America Latină*. Cluj: Presa Universitară Clujeană.
- Rona, José Pedro. 1969. “¿Qué es un americanismo?”, in *Actas del Simposio*, 135-148. México: UNAM. https://ia601506.us.archive.org/24/items/b29827620_0001/b29827620_0001.pdf (consultado el 20 de mayo de 2025).
- Teuteberg, Hans Jürgen. 2007. “The Birth of the Modern Consumer Age. Food Innovations from 1800”. En *Food: The History of Taste*, editado por Paul Freedman, 234-261. Berkeley: University of California Press.
- Torres Torres, Antonio. 2000. *El español de América*. Barcelona: Ed. de la Universitat de Barcelona.
- Torres Torres, Antonio. 2016. *El español de América*. Barcelona: Ed. de la Universitat de Barcelona, II^{nda} edición.

- UNESCO (s.n.) *Intangible Cultural Heritage*, URL: <https://ich.unesco.org/>, & <https://www.unesco.org/archives/multimedia/subject/13/intangible+heritage/> (consultado el 26.07. 2024)
- Vaquero de Ramírez, Maria. 1996/2003. *El español de América II. Morfosintaxis y Léxico*. Madrid: Arco/Libros.
- VV.AA. 2018. *Cultura en el mundo hispanohablante*, A2-B1. enClave-ELE.

Webgrafía de términos gastronómicos generales:

Diccionario de americanismos:

https://www.asale.org/sites/default/files/Prologo_Diccionario_de_americanismos.pdf

Diccionario panhispánico de gastronomía: Enciclopedia en línea,

<https://www.enciclochef.com/>

<https://gustolatinogastronomia.com/gastro-diccionario-gastronario-de-comida-latina/>.

<https://prezi.com/p/9i8w-jewnntg/glosario-gastronomia-latinoamericana/>.

<https://saborygusto.com/glosario.html>.

<https://www.escuelamasterchef.com/diccionario/>. <https://www.petitchef.es/articulos/recetas/glosario-espanol-latinoamericano-aid-1912>

<https://www.pucheromix.com/terminos-culinarios/>

Otros diccionarios:

Rodríguez, Guille. *Diccionario culinario gastronómico:*

<https://pilpileando.com/diccionario-culinario-gastronomico-pilpileando-m/>

Ruiz del Castillo Pérez de Arenaza, Carlos. 2024. *Diccionario Enciclopédico de Cocina:*

<https://gastronomiaaz.com/>.

Vasco v. Baselli, 2023. *Diccionario Gastronómico Latinoamericano*. Staten House.

https://books.google.ro/books/about/Diccionario_Gastron%C3%B3mico_Latinoamerican.html?id=wcxZ0AEACAAJ&redir_esc=y.

Vivancos Sámper, Ginés. 2003. *Diccionario de alimentación, gastronomía y enología española y latinoamericana*. Madrid: Everest.

L'HERITAGE DE LA LABIOVELAIRE DANS LES LANGUES ROMANES : LE CAS DU LAT. *COQUERE* 'CUIRE'

Simona GEORGESCU¹ , Theodor GEORGESCU²

Article history: Received 30 May 2025; Revised 30 July 2025; Accepted 15 August 2025;
Available online 12 December 2025; Available print 30 December 2025

©2025 Studia UBB Philologia. Published by Babeş-Bolyai University.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ABSTRACT. *The Heritage of the Labiovelar in Romance Languages: The Case of Lat. coquere 'to Cook'.* Lat. COQUERE appears in REW₃ (n° 2212) with two variants: COQUERE and COCERE. According to REW₃ and FEW (2, 1167a), Old French is the only idiom to have inherited both forms, with a semantic distinction: the (indirect) descendant of COQUERE, *cuivre* s.m./f., has a figurative meaning, namely 'worry, torment', while the descendant of COCERE, *cuire*, has retained the concrete meaning 'to cook'. Our aim is to find out whether this semantic bipartition was coupled with phonetic differentiation in Protoromance, or whether it merely represents an idioromance development. We will apply the comparative grammar-reconstruction method, as it is used in the DÉRom project. This will enable us to establish, on the one hand, whether the */k^w-e-/ prototype is found in an area wider than the Gallo-Romance domain, and, on the other, whether the semantic reconstruction also indicates two distinct prototypes. The Late Latin and Romance data we collected suggest that, by the 3rd century at the latest, the verb had already lost the labial appendice of the labiovelar, and the form in circulation had a simple velar: */k^w-e-/. Bringing into question the inheritance of the labiovelar consonant /k^w/ in Romance languages, the study analyzes the possibility that a semantic differentiation

¹ **Simona GEORGESCU** est chargée de cours à l'Université de Bucarest, où elle enseigne la linguistique romane comparée et l'espagnol. Son dernier livre publié en tant qu'auteur unique est *La regularidad en el cambio semántico. Las onomatopeyas en cuanto centros de expansión en las lenguas románicas*, Strasbourg, ELiPhi, 2021 (simona.georgescu@lls.unibuc.ro);

² **Theodor GEORGESCU** est chargé de cours à l'Université de Bucarest, au Département de Philologie Classique, où il enseigne la langue latine et la langue et littérature grecque ancienne. Son dernier ouvrage en tant qu'auteur unique est *Poezia elenistică. Theocritus, Idile*, Editura Universităţii din Bucureşti, 2010. Il est actuellement en train d'éditer, avec Constantin Georgescu et Simona Georgescu, le *Dictionnaire Grec-Roumain* (DGR), dont six des douze volumes prévus ont été publiés (theodor.georgescu@lls.unibuc.ro).

may be accompanied by a phonetic bipartition. The conclusion is that, in this case, the reconstruction of two prototypes is not necessary, but that a single form, without the labiovelar, can explain the Romance descendants and their two main meanings.

Key-words: *labiovelar consonant, coquere, comparative grammar-reconstruction, Late Latin, semantic reconstruction.*

REZUMAT. Moștenirea labiovelarei în limbile romanice: cazul lat. *coquere* ‘a coace’. Lat. COQUERE figurează în REW₃ (n° 2212) cu două variante: COQUERE și COCERE. Conform REW₃ și FEW (2, 1167a), franceza veche este singurul idiom care a moștenit ambele forme, cu o distincție semantică: descendentul (indirect) al lui COQUERE, *cuivre* s.m./f., are un sens figurat, și anume ‘îngrijorare, chin’, în timp ce descendentul lui COCERE, *cuire*, a păstrat sensul concret de ‘a coace’. Scopul nostru este să aflăm dacă această bipartiție semantică a fost dublată de diferențierea fonetică în protoromanică sau dacă reprezintă doar o evoluție idioromanică. Vom aplica metoda gramaticii comparate-reconstrucție, așa cum este utilizată în cadrul proiectului DÉRom. Aceasta ne va permite să stabilim, pe de o parte, dacă prototipul */k^w-e-/ se poate identifica într-o zonă mai largă decât domeniul galoromanic și, pe de altă parte, dacă reconstrucția semantică indică, de asemenea, două prototipuri distincte. Datele oferite de latina târzie, precum și cele extrase din limbile romanice, sugerează că, cel târziu în secolul al III-lea, verbul pierduse deja apendicele labial al labiovelarei, iar forma în circulație avea o velară simplă: */k^h-e-/. Punând în discuție moștenirea consoanei labiovelare /k^w/ în limbile romanice, studiul analizează posibilitatea ca o diferențiere semantică să fie însoțită de o bipartiție fonetică. Concluzia este că, în acest caz, reconstrucția a două prototipuri nu este necesară, ci o singură formă, fără labiovelară, poate explica descendenții romanici și cele două sensuri principale răspândite în majoritatea limbilor romanice.

Cuvinte-cheie: *consoană labiovelară, coquere, gramatică comparată-reconstrucție, latină târzie, reconstrucție semantică.*

1. Introduction : le problème

Le verbe *coquere*, d’origine indo-européenne, documenté depuis les plus anciens textes littéraires (Plaute, cf. TLL 4, 925 sqq.), a une descendance panromane.

Selon Meyer-Lübke (REW₃), dans le latin vulgaire circulaient deux formes phonétiques, *coquere* et *cocere*, dont la dernière a perduré dans toutes les langues romanes tandis que la première s’est maintenue uniquement en français. La même perspective est adoptée dans le FEW (2, 1167a) par Wartburg, qui insiste sur la seule continuation du prototype *coquere* en français.

Une telle situation, où l'on a deux prototypes du même étymon en latin vulgaire hérités dans les langues romanes, ne serait pas unique : le DÉRom a identifié, parmi les 262 étymons protoromans étudiés jusqu'à présent (janvier 2025), plus de 15% qui imposent la reconstruction d'un double – sinon triple ou quadruple – prototype phonétique, par exemple : **/askul't-a-re/* ~ **/eskol't-a-re/* (Schmidt/Schweickard 2010–2020 in DÉRom s.v. **/as'kult-a-/*; Schmidt/Schweickard 2010–2020 in DÉRom s.v. **/es'kolt-a-/*) ; **/kas'tani-a/* ~ **/kas'tni-a/* (Medori 2010–2020 in DÉRom s.v. **/kas'tani-a/* ~ **/kas'tni-a/*) ; **/la'brusk-/* ~ **/la'brusk-/* ~ **/lam'brusk-/* ~ **/lam'brusk-/* (Reinhardt 2011–2021 in DÉRom s.v. **/la'brusk-/* ~ **/la'brusk-/*) ; **/re'tund-u/* ~ **/ro'tund-u/* ~ **/to'rund-u/* ~ **/'tund-u/* (Hegner 2011–2020 in DÉRom s.v. **/re'tund-u/*), etc. Cela ne peut que montrer une significative variété diatopique, diachronique et diastratique du latin parlé, 15% représentant un taux considérable.

Les questions qui dirigeront notre recherche sont les suivantes : (a) faut-il vraiment reconstruire deux prototypes ? (b) est-ce que l'éventuelle différenciation est seulement phonétique, ou se base aussi sur des critères sémantiques ? (c) est-ce que ce possible double développement est identifiable aussi dans d'autres langues romanes ? c'est-à-dire, le prototype *coquere* pourrait-il être documenté aussi en dehors du français ?³

Dans cette enquête, on va appliquer la méthode « grammaire comparée-reconstruction » (cf. Chambon 2007 ; 2014), dans son format le plus récent, tel que proposé par le projet DÉRom (cf. Buchi/Schweickard 2014 ; DÉRom). Cela nous permettra d'établir si le prototype **/kɔkw-e-/* se retrouve seulement en français ou dans une aire plus étendue et si la reconstruction sémantique indique aussi deux prototypes distincts.

2. *Coquere* vs *cocere* ou le destin de la labiovélaire : point de vue théorique

Il convient tout d'abord d'examiner la question de la labiovélaire */kʷ/*. L'existence de ce phonème en protoroman a été mise en doute, pour des raisons d'économie phonologique (cf. Gouvert 2014), par Hall (1976, 155), qui interprète **[kw]* comme la réalisation de la séquence **/ku/* en hiatus. Adoptant cette position théorique, le DÉRom a également traité le phonème traditionnellement interprété comme labiovélaire en termes d'une simple séquence phonologique

³ Il ne faut pas oublier que depuis le REW, la quantité de nouveaux documents découverts et données recueillies pour la documentation des langues romanes (voire les atlas linguistiques) n'est pas négligeable, on pourrait donc s'attendre à de nouvelles informations à ajouter au REW (et c'est, en effet, l'une des raisons pour lesquelles une entreprise comme le DÉRom a été lancée, cf. Schweickard 2022).

*/ku/ (cf. Greub 2025 in DÉRom, s.v. */'akuil-a/ ; Maggiore 2020 in DÉRom, s.v. */'kuer-e/). Cependant, déjà dans le premier volume publié de DÉRom, Gouvert (2014) plaide en faveur de la reconstruction de ce phonème latin, à l'unisson de la perspective traditionnelle et majoritaire (Meillet et Vendryes 1960, 58–59 § 80 ; Weiss 2009, 66). Gouvert (2014, 100) invoque le contraste */kw/ ~ */ku/ qui peut être clairement illustré par « la distinction importante entre protorom. */'kwi/ (pron. rel./interrog. 'qui') et */'kui/ (adv. 'ici'), perpétuée dans it. /'ki/ ~ /'kwi/, frioul. /'kwi/ ~ /'ki/. »

L'auteur cité établit ce phonème dans des mots tels que :

protorom. */'kwattor/ ~ */'kwattro/ 'quatre' > sard. *battor*, dacoroum. *patru*, vegl. *kʷatro*, it. *quattro*, frioul. *cuatri*, romanch. *quatter*, fr. *quatre*, occit. *quatre*, cat. *quatre*, esp. *cuatro*, port. *quatro* (REW3 s.v. *quattuor*) ; protorom. */'kwetu/ 'calme' > dacoroum. *cet*, it. *cheto*, frioul. *cet*, fr. *coi*, occit. *quet*, cat. *quet*, esp. *quedo*, port. *quedo* (REW3 s.v. *quiētus/quētus*) ; protorom. */'kwɪ/ 'quoi' > dacoroum. *ce*, vegl. *ke*, it. *che*, frioul. *ce*, romanch. *tge*, fr. *quoi*, occit. *que*, cat. *que*, esp. *que*, port. *que* (REW3 s.v. *qui/quēm/quam/quīd*) (Gouvert 2014).

En revanche, la séquence /ku-/ doit être reconstruite dans des lexèmes tels que :

protorom. */'kuindiki/ 'quinze' > sard. *bindigi*, vegl. *čonk*, it. *quindici*, romanch. *quindesch*, fr. *quinze*, occit. *quinze*, cat. *quinze*, esp. *quince*, port. *quinze* (REW3 s.v. *quīndēcim*) ; protorom. */(ad) 'kui/ 'ici' > it. *qui*, frioul. *chi*, occit. *aqui*, cat. *aqui*, esp. *aqui*, port. *aqui* (REW3 s.v. *hīc/*hīcce*) ; protorom. */(ad) 'kuistu/ 'celui-ci' > dacoroum. *acest*, it. *questo*, frioul. *chest*, romanch. *quest*, occit. *aquest*, cat. *aquest*, esp. *aqueste*, port. *aqueste* (REW3 s.v. *īste*) (Gouvert 2014).

Ces exemples et d'autres encore obligent à reconnaître la labiovélaire */k^w/ ⁴ comme un phonème à part entière, avec des résultats différents dans les langues romanes de ceux de la séquence */ku-/.

Pour ce qui est du verbe latin *coquo*, son prototype indo-européen se reconstruit comme *pek^wō 'cuire', qui subit une première assimilation consonantique régressive *quequo et une deuxième assimilation vocalique, aussi régressive, *quoquo. La troisième étape suppose la délabialisation de la première labiovélaire, suite à une dissimilation, ayant le résultat *coquo* (cf. gr. πέσσω [< *pek^wyō], scr. *pācati*) (IEW. s.v. *pek^w-* ; Leumann 1977, 137).⁵

⁴ Il en va de même pour sa version sonore, */g^w/, que nous ne décrirons pas ici.

⁵ La même dissimilation se trouve en *quinque* > *cinque* ou *quinquaginta* > *cinquaginta*, cf. Väänänen, 1981 : 52 ; Lausberg 1970 : 339 ; Costa, 2003 : 73.

Ce qui nous intéresse est donc l'évolution romane de la labiovélaire posttonique. Il faut signaler que dans les analyses de l'évolution phonétique on prend d'habitude comme base la forme infinitive, mais en effet cette forme n'est pas tout à fait usitée dans la langue parlée : les formes les plus fréquentes sont les conjuguées à la première personne du singulier et la troisième du singulier et pluriel. Donc, il faudrait plutôt envisager l'évolution de */'kɔk^w-o/, */'kɔk^w-i-/ (lat. cls. *coquit*) et */'kɔk^w-u-/ (lat. cls. *coquant*).⁶

Pour ce qui est de */'kɔk^w-o/ et */'kɔk^w-u-/, la question semble assez claire : la simplification de la labiovélaire en vélaire devant voyelle homorgane, /o/ et /u/, est bien représentée en latin vulgaire (Lausberg 1970, 335 ; Leumann 1977, 137).⁷ Le verbe même */'kɔk^w-e-re/ est très tôt réduit à */'kɔk^w-e-re/, comme, en latin classique, **loquutus* > *locutus*, **sequondus* > *secundus*, ou *quottidie* > *cottidie* (cf. Väänänen 1981, 51). Dans l'évolution romane, un exemple significatif est */'k^wo-modo/, dont l'issue sarde est *comente* (influencé par les adverbes en *-mente*) et *co*, le descendant roumain est *cum*, l'italien *come*, le fr. *comme*, etc., ce qui indique la réduction de la labiovélaire en vélaire en protoroman (cf. Lausberg 1970, 335).⁸

Väänänen (1981, 52) explique la double variante 'coquere' / 'cocere' (*/'kɔk^w-e-re/ ~ */'kɔk-e-re/) comme résultat de l'analogie avec la forme nominale double 'coquus' / 'cocus'.⁹ On pourrait alors présumer comme suffisante l'analogie pour expliquer la délabialisation dans tout le paradigme. Sinon, il faut analyser les possibles évolutions d'une labiovélaire devant une voyelle palatale (en considérant les formes */'kɔk^w-i-/ et */'kɔk^w-e-/).

Dans ce cas, on envisage deux possibilités. Selon Lausberg (1970, 340), « la conservación en sardo y rumano del elemento labial representa la auténtica evolución, originándose de la sucesión de la oclusiva palatal sorda y la fricativa labial sonora – por asimilación recíproca – una oclusiva labial (sonora en sardo y sorda en rumano) ». ¹⁰ Ce processus a eu lieu non seulement devant *-a-* (comme semblent l'indiquer les correspondances lat. *quattuor* – sard. *báttoro*, dacoroum. *patru*, lat. *quattuordecim* – sard. *battórdighi*), mais aussi devant une voyelle

⁶ On ne peut pas reconstruire les désinences de la troisième personne singulier et pluriel pour la réalité orale du latin.

⁷ Parfois, la forme originaire est refaite, cf. *sequor*, *linquant*, cf. Costa (2003, 73) ; les formes pronominales du type 'quo' sont aussi restituées selon *qui*, *quis qualis*, cf. Leumann (1977, 137).

⁸ En effet, cette réduction est attestée par l'épigraphie pompéienne, voire les formes *comodo* et *como* (cf. Väänänen 1981, 51).

⁹ L'explication est reprise par Corominas et Pascual (DCECH 2, 109), qui considèrent la forme 'vulgaire' *cocere* comme « debida a extensión analógica de la simplificación fonética de *coquo* y *coquant* en *coco*, *cocunt* ».

¹⁰ Le changement s'explique de la manière suivante : « la serie ku se desredondeó en fecha temprana, dentro de este dominio lingüístico, en *[kβ], naciendo después [b] de *[gβ] y [p] de [kφ] » (Lausberg 1970, 340).

palatale, voire lat. *quindecim* qui correspond au sard. *bíndighi*. La conservation en sarde, italien et sursilvan de l'élément labial représente, selon Lausberg (1970, 340) la prononciation originaire. Donc le protoroman */'kwin-deke/ (lat. class. *quindecim*), donnant lieu en sarde à *bíndighi*, a comme résultat en it. *quíndici* [kw-] et en sursilv. *quéndisch* [kw-] ; lat. *quercea* évolue en it. à *quercia*, et l'issue sursilvane de *qu(i)escere* est *quescher* 'se taire'.

De l'autre côté, y compris dans les régions archaïques citées (sarde, italien, sursilvan), on trouve assez de cas de réduction de la labiovélaire à la vélaire devant une voyelle palatale : lat. *qui* > sard. [ki], it. *chi* ; lat. *quid* – sard. [ki], it. *che* ; lat. *quaerere* > sard. *kerrere*, it. *chiedere* ; lat. *qu(i)etu* > it. *cheto*.¹¹

D'un point de vue de la chronologie relative, dans tous les cas cités, la réduction de */k^w/ à */k/ a eu lieu assez tard pour que la nouvelle */k/ ne participe pas au « premier impulse palatalisateur », selon Lausberg (1970, 337 ; cf. Väänänen 1981, 52) (lat. c+i,e > é > ts). C'est pour cela que lat. *qui* devient en it. *chi*, en fr. *qui*, mais *civitate* a comme résultat en it. *città*, en fr. *cité*.

Vu que, dans le cas de *coquere*, la labiovélaire a participé à ce que Lausberg (1970, 337) appelle « le premier impulse palatalisateur », en ayant le même résultat qu'une vélaire simple devant e/i en Italo-romania, Galo-romania et Ibero-romania, on peut déduire que la perte de l'élément labial a eu lieu très tôt. La présence en sarde d'une issue délabialisée (sard. *kòkere*, log. *kògere*, cf. DES), dont l'évolution est identique à celle des lexèmes avec vélaire simple devant une voyelle palatale, témoigne de la circulation de la protoforme */'kɔk-e-re/ dans la phase antérieure à l'individuation du protosarde (2^e m. 2^e s. [?], Straka 1956, 25 ; Dardel 1985, 268 ; Stefanelli 1996, 84). Cette protoforme s'est diffusée, sans aucun doute, dans toute la Romania.¹²

La même réduction */'kok^w/ > */'kok-/ est opérante aussi dans les mots apparentés come lat. *coquina*, *coquinare* et *coquus*, hérités à partir des protoformes */'ko'k-in-a/, */'kok-i'n-a-re/, */'koku-/ (cf. LEI, s.v. *coquere* [ms.]).¹³

Pourrait-on supposer qu'en protoroman continental circulaient néanmoins deux formes, l'une avec labiovélaire, l'autre délabialisée ?

Pour le savoir, on dispose de deux types de ressources : premièrement, il faut dépouiller les sources latines tardives qui offrent des indices sur cet aspect – des indications données par les grammairiens, concernant la graphie ou la prononciation, représentant une aide significative pour cette enquête.

¹¹ Pour une discussion plus détaillée et une liste extensive d'exemples, voir Wolf (2012). Pour une synthèse, voir Repetti (2016, 664).

¹² Il en va de même pour le verbe lat. *torquere*, dont la forme reconstruite est */'tɔrk-e-re/, aucune des issues romanes ne conduisant à une labiovélaire originelle (cf. REW n° 8798). De plus, la graphie attestée *extorquet* est significative (TLL 2047, 48).

¹³ En effet, Väänänen (2006, 52) explique les formes *cocere* pour *coquere* ou *cocina* pour *coquina* comme « des cas d'analogie ... d'après *coquus*, *cocus* ».

Deuxièmement, mais plus significatif d'un point de vue de la grammaire comparée–reconstruction, on doit inventorier toutes les formes romanes issues de l'étymon */'kək^w-e-re/ ~ */'kək-e-re/.

3. Ce que nous enseignent les grammairiens

Une analyse des témoignages fournis par les grammairiens peut nous offrir une image de la manière dont on percevait, dans la basse latinité, la différence entre vélaire et labiovélaire. Nous voulons savoir si cette opposition était vraiment valide dans la langue parlée, réelle du point de vue du locuteur, et, surtout, fonctionnelle.

On ne manque pas de témoignages qui semblent indiquer une perception toujours présente, au 5^e siècle ap. J.C., de l'opposition phonétique /k^w/ – /k/ :

PAPIRIAN. Cassiod. gramm. VII 164, 20 (ap. TLL 4, 925, 33-36), *coqui [...]* per 'c' primam syllabam, secundam per 'q' scribendam putant. **non enim dicimus cocere, sed coquere.**

(« on considère qu'on doit écrire la première syllabe de *coqus* avec 'c' et la seconde avec 'q'. Car on ne dit pas *cocere*, mais *coquere* »).

Toutefois, en employant le syntagme "scribendam putant", le grammairien se démarque de cette affirmation, en laissant planer le doute sur la validité de la corrélation entre la graphie et la prononciation.

Un autre témoignage appartient à Pseudo Probus, qui, dans son œuvre *Instituta artium* (rédigée au 4^e siècle ap. J.C.), signale pour sa part un certain doute concernant la prononciation :

PS. PROB. inst. gramm. IV 182, 21 (ap. TLL 4, 925, 37-38) *quaeritur qua de causa 'coquo' et non 'coco' dicatur.*

(« on se demande pourquoi on dit *coquo* et non *coco* »)

De l'autre côté, on trouve des graphies hypercorrectes qui semblent indiquer la neutralisation de l'opposition /k^w/ – /k/ : Velius Longus (117-138 ap. J.C.), dans son traité *De orthographia*, note l'observation suivante :

VEL. gramm. VII 79,7 (ap. TLL 4, 925, 28-33), *cocum nonnulli in utraque syllaba per 'q' scribunt, nonnulli et inserta <u>. in verbo etiam 'quoquere' per 'quo'. Nisus censet ubique 'c' litteram ponendam tam in nomine quam in verbo, quod mihi nimium videtur exile, nam sicut non est prima syllaba oneranda, sic sequens videtur explenda.*

(« quelques-uns écrivent *cocum* avec 'q' dans les deux syllabes, certains même avec 'u' inséré. Y compris dans le verbe ils écrivent *quoquere*, avec *quo*. Nisus considère qu'il faut mettre partout 'c', tant dans le nom que dans le verbe, ce qui me paraît trop faible, car, autant la première syllabe ne doit pas être chargée, autant la syllabe suivante me paraît devoir être remplie »)

Cette incertitude dans l'écriture du verbe relève du fait que pendant la première moitié du 2^e siècle ap. J.C. la prononciation était déjà indistincte entre la première vélaire et la seconde : si l'on pouvait écrire aussi bien *quoquere* et *quoquum*, que *cocere* et *cocum*, respectivement *coquere* et *coquum* sans aucun critère spécifique (sauf *nimum videtur exile* et *videtur explenda*), il résulte soit qu'il n'y avait plus de différence entre la prononciation d'une vélaire et d'une labiovélaire, soit que tant le nom que le verbe avaient évolué à *cocum*, *cocere*, tandis que la norme en vigueur imposait encore la prononciation avec labiovélaire.

Certaines remarques des grammairiens sur des lexèmes contenant une vélaire simple font ressortir l'oscillation entre les graphies 'c-' et 'qu-', apparemment sans influencer la prononciation, par exemple :

PRISC. gramm. II 36, 14 (ap. TLL 9, 441, 56-57, s.v. *oculus*) *apud antiquos frequentissime loco 'cu' syllabae 'quu' ponebatur ... ut oquulus pro ...oculus*. (« chez les anciens on mettait souvent 'quu' au lieu de 'cu' dans *locus*, de même qu'<on écrivait> *oquulus* au lieu d'*oculus*. »)

À remarquer la mention « *apud antiquos* », qui relève de la période précoce de la neutralisation.

La même incertitude pour ce qui est de la graphie est visible dans des témoignages comme :

PRISC. gramm. (sec. VI) II 504, 1 (ap. TLL 4, 925, 36) *coquo vel, ut alii, 'coco, coxi'*. (« *coquo* ou, comme d'autres <le disent/écrivent>, *coco, coxi* »)

Le doute exprimé par Priscius est résolu par son disciple Eutyches, (EUTYCH. gramm. V 468, 28, ap. TLL 4, 925, 36-37), qui recommande la graphie *coco, cocis*.

Les recommandations de Probus ou Pseudo Probus dans *Instituta artium* (Ps. Prob. inst. gramm. IV 126, 29) *coqus non cocus*, ou, dans la liste connue comme *Appendix Probi* (Ps. PROB. app. gramm. IV 197, 30, 4^e – 6^e s. (?)), *coquens, non cocens*, sont un indicateur du destin roman de ces mots, vu que, d'habitude, c'est la

forme marquée comme erronée qui est susceptible de se conserver dans les langues romanes.

La même *Appendix Probi* signale : (PROBI app. gramm. IV, 192, 2, TLL 2, 177, 22-24, s.v. *antiquus*) *per q, non per c litteram scribuntur equus coqus iniquus antiquus* (« on écrit *equus, coqus, iniquus, antiquus* avec *q*, non avec *c* »), ce qui ne représente qu'une autre preuve de la neutralisation entre les deux phonèmes, qui, à cette époque-là, ne semblent être plus que deux graphèmes.

Selon TLL, cette graphie ne manque pas chez Caton (*234, † 149 av. J.C.), Varron (*116, † 27 av. J.C.), Pline (*23, † 79), Martiale († ca 103), ainsi que dans les textes tardifs et dans les inscriptions (Corp. IV 2095). De même, on trouve la graphie hypercorrecte (ou originaire ?) *quoquo* chez Plaute (ca 254–184 av. J.C.), Lucrèce (ca 97, † 55 av. J.C.) etc. Excluant les fautes de manuscrit, cette variation graphique relève de l'évolution précoce */k^w/ (+o/i, e) > */k-/ (+o/i, e).

Les remarques des grammairiens ne semblent donc être que des tentatives visant à rétablir la norme au détriment de l'usage.

4. Ce que montrent les langues romanes

Toutes les branches romanes conduisent à la reconstruction du prototype */kək-e-re/ : dacorum. *coace* (Tiktin), veigl. *kukro* (Bartoli 2000 [1906], 413), engad. *couscher* (DRG 4,171sq.), afrioul. *chog[er]* (14^e s., TLIO), *quey*, frioul. *cuéi* (DESF), fr. *cuire* (dp. 881, Eulalie, sous la forme *coist* 'brûler (d'un corps humain)', FEW 2,1162b-1167b), occit. *cozer* (DOMel), cat.a. *coure* (13^e s., DELCat 2,1019sq.), esp. *cocer* (DCECH 1, 828-829), port. *cozer* (dp. 13^e siècle, DELP₃).

Selon Meyer-Lübke (REW₃) et Wartburg (FEW 2, 1167b), la protoforme */kək^w-e-re/ serait toutefois continuée dans le fr. *cuivre*, s.m. Ce n'est pas, cependant, la seule interprétation possible de l'origine de ce lexème.

Dans la *Grammaire des langues romanes* §494, Meyer-Lübke signalait lui-même que le groupe -kr- intervocalique est prédisposé à une labialisation, comme suit : *socru* 'beau-père' > a.fr. *soigvre* > *soivre* (cf. FEW 12, 15b, s.v. *socer* : afr. *sevre*, *suivre*) ; germ. *kokar* (reconstruit comme **kukur* dans le FEW 16, 429b) > **kukru* a.fr. *coigvre* > *coivre* > *cuivre* 'carquois'. De la même manière, la forme fr. *cuivre* pourrait être expliquée simplement comme une issue de *cocere*, syncopé > **cocre*, labialisé > **coigvre* simplifié > *coivre* > *cuivre*.

Il en résulte que, d'un point de vue phonétique, il n'est pas nécessaire de reconstruire deux prototypes, vu que */kək-e-re/ peut justifier y compris la forme traditionnellement considérée comme provenant de */kək^w-e-re/. La différence entre la forme proparoxytone et la variante syncopée pourrait être le reflet d'une séparation phonétique à base morphologique, plutôt que sémantique : le nom doit avoir subi la syncope avant le verbe, qui, pour sa part, a permis l'évolution régulière de la vélaire devant une voyelle palatale.

5. Différentiation sémantique?

Il nous reste toujours à vérifier la possibilité qu'une différenciation sémantique protoromane soit assortie à une bipartition phonétique. Cela impliquerait que le sens figuré de 'tourmenter' se soit séparé du sens propre 'cuire', en s'associant avec une forme phonétique légèrement différente de la plus commune : */'kək-e/ 'cuire' (sens propre), */'kək^w-e/ 'tourmenter' sens figuré. Ce type de divergence sémantique qui attire une bipartition phonétique n'est pas inconnu : juste pour donner un exemple, pensons-nous au lat. *sibilare* 'siffler', avec une variante tardive *sifilare*, qui a donné lieu à trois lexèmes en espagnol, à savoir *silbar* 'siffler', *chiflar* 'se moquer' y *chillar* 'pousser un cri' (cf. Malkiel 1990, 25).

Dans le cas de *cocere*, cette hypothèse n'est toutefois pas soutenue par les données latines et romanes : les deux sens se retrouvent, tant en latin que dans toutes les branches romanes, en relation polysémique, donc sous la même forme phonétique, sans exiger une bipartition lexicale.

Le latin écrit enregistre abondamment le sens figuré: (TLL 4, 928: *de affectibus excitandis, maxime de dolore cogitando accepto i.q. urere, macerare, opprimere, affligere*). Voici quelques exemples :

ENN. ann. 336 *quae cura te coquit et versat in pectore fixa*.
(« quel souci niché dans ta poitrine te cuit et te tord ? »)

PLAVT. Trin. 225 *egomet me coquo et macero et defetigo*.
(« Moi-même, je me cuit, je me tourmente et je m'épuise. »)

CATVLL. 83, 6 *irata est: hoc est, uritur et coquitur*.
(« elle est en colère : ça veut dire, elle se brûle et se cuit. »)

Pour ce qui est des langues romanes, à part le signifié 'cuire', un sens figuré coexiste dans les cas suivants : dacoroum. *a coace* 'tourmenter' (DLR), sard. *kòkere* 'harceler, malmener, affliger' (DES, NVLS), it.a. *cocersi* v.rifl. 'être emporté par les passions ; se tourmenter, être troublé, souffrir' (attesté avant 1321, TLIO; LEI)¹⁴, a.occ. *coire* 'être désagréable', a.esp. *cozer* 'affliger' (Kasten/Cody),¹⁵ gall. *cozer* 'causer un douleur, tourmenter' (DDGM), ptg. *cozer*

¹⁴ Les dialectes d'Italie en offrent des nombreux exemples : tic.alp.occ. (Vairano) *còs[es]* 'preoccuparsi, crucciarsi', mil. *caèuses* v.rifl. 'fremere, ribollire (per rabbia)', it. *cuocere* v.tr. 'provocare (in q.) sentimenti intensi e dolorosi; sconvolgere l'animo; far innamorare; addolorare, far soffrire; molestare, tormentare'¹⁴ (1335-36 (?)) ; piem. *cheuit* adj. 'gravemente danneggiato; perduto, rovinato', novar. (borgom.) *koécu* adj. 'triste, melanconico', apulo-bar. *cùette* adj. 'macerato da fatiche e dolori', umbro-merid.-or. *k¹ttu k¹ttu* agg. 'mortificato, avvilito, demoralizzato' (LEI).

¹⁵ De même, en espagnol moderne, *cocer* peut signifier 'padecer intensamente y por largo tiempo un dolor o incomodidad' (DRAE_{23.6}).

'tourmenter'. En plus, même le fr. *cuire* manifeste la polysémie latine et romane: à côté du sens propre 'rendre propre à l'alimentation', le signifié de 'ressentir un chagrin à propos de qch.' est attesté depuis le 12^e siècle) (FEW 2, 1166b).

La présence du sens figuré de 'provoquer un chagrin, une douleur physique ou psychique' en latin classique et dans toutes les branches romanes peut être interprétée de deux façons. D'une part, on peut supposer une évolution parallèle induite par des facteurs cognitifs. Ainsi, on a constaté la tendance, observée au niveau universel, à représenter les émotions en termes de fluides en ébullition ou sous pression, compte tenu de la perception du corps humain comme un récipient (cf. Kövecses 1986, 1990, 2000 ; Georgescu 2021). Pour donner un exemple parallèle en latin, *macerare* 'rendre doux, amollir en humectant, faire macérer' est attesté avec des sens figurés similaires à *coquere*, à savoir 'consumer, miner' [le corps] et 'tourmenter [l'esprit]' cf. TLL 8, 8-10. D'autre part, l'existence de ce sens figuré chez la plupart des descendants de */kək-e-re/ conduit à la reconstruction de ce signifié en protoroman. En plus, étant donné que le verbe français, à côté des autres langues romanes, a continué les deux signifiés sous une forme unique *cuire* (fr. *cuire* 'produire une douleur analogue à celle que cause l'action du feu' dp. *Eulalie*, cf. FEW 2, 1166a), il n'est pas justifié de supposer une bipartition phonétique à base sémantique dans le protoroman : un seul lexème était suffisant pour transmettre les deux sens, perçus comme propre et métaphorique.

6. Conclusion

La question de départ était de savoir s'il fallait reconstruire, comme le proposait Meyer-Lübke, deux prototypes pour le verbe latin *coquere*, qui puissent rendre compte des descendants romans : l'un avec labiovélaire */kək^w-e-re/ et l'autre qui atteste la réduction de la labiovélaire en vélaire simple devant la voyelle antérieure, */kək-e-re/. Une analyse détaillée des variantes graphiques du latin tardif attestées chez les grammairiens latins, ainsi qu'un examen des descendants romans à partir desquels nous avons appliqué la méthode grammairienne comparée-reconstruction, nous ont conduit à établir un seul prototype en latin vulgaire, à savoir */kək-e-re/ avec deux sens principaux, 'rendre propre à l'alimentation par l'action de la chaleur, du feu' et 'provoquer une douleur physique ou psychique, tourmenter'. Les deux sens sont continués dans la majorité des idiomes romans.

L'hypothèse d'un prototype double, */kək-e-re/ et */kək^w-e-re/, dont le second aurait perduré uniquement en français, devient futile dans la lumière des données qui concernent l'évolution phonétique romane et la répartition des sens. Le fr. *cuivre* s'explique simplement comme issu du même prototype */kək-e-re/ syncopé, */kək-re/, lexicalisé comme substantif en opposition avec le verbe non syncopé.

BIBLIOGRAPHIE

- Bartoli, Matteo Giulio. 2000 [original allemand : 1906]. *Il dalmatico. Resti di un'antica lingua romanza parlata da Veglia e Ragusa e sua collocazione nella Romània appennino-balcanica*. Rome : Istituto della Enciclopedia Italiana.
- Chambon, Jean-Pierre. 2007. « Remarques sur la grammaire comparée-reconstruction en linguistique romane (situation, perspectives) ». *Mémoires de la Société de linguistique de Paris*, n. s., 15, 57-72 (*Tradition et rupture dans les grammaires comparées de différentes familles de langues*). Louvain : Peeters.
- Chambon, Jean-Pierre. 2014. « Réflexions sur la reconstruction comparative en étymologie romane : entre Meillet et Herman ». In *Étymologie romane : objets, méthodes et perspectives* édité par Martin Glessgen et Wolfgang Schweickard, 141-159. Strasbourg : ELiPhi.
- Costa, Ioana. 2003. *Fonetică istorică latină*. București: Editura Universității din București.
- Dardel, Robert de. 1985. « Le sarde représente-t-il un état précoce du roman commun ? ». *Revue de linguistique romane* 49, 263-269.
- DCECH = Corominas, Joan et José Antonio Pascual. 1980-1991. *Diccionario crítico etimológico castellano e hispánico*, 6 vol. Madrid : Gredos.
- DDGM = González Seoane, Ernesto, María Álvarez de la Granja et Ana Isabel Boullón Agrelo. 2006. *Dicionario de dicionarios do galego medieval* (cédérom). Saint-Jacques-de-Compostelle : Universidade de Santiago de Compostela.
- DECat = Coromines, Joan. 1980-2001. *Diccionari etimològic i complementari de la llengua catalana*, 10 vol. Barcelone : Curial.
- DELP₃ = Machado, José Pedro. 1977 [1952-1959]. *Dicionário etimológico da língua portuguesa com a mais antiga documentação escrita e conhecida de muitos dos vocábulos estudados*, 5 vol. Lisbonne : Horizonte.
- DÉRom = Buchi, Éva et Wolfgang Schweickard (dirs.). 2008-. *Dictionnaire Étymologique Roman (DÉRom)*. Nancy, ATILF, <http://www.atilf.fr/DERom>.
- DES = Wagner, Max Leopold. 1960-1964. *Dizionario etimologico sardo*, 3 vol. Heidelberg: Winter.
- DESF = Zamboni, Albert, Manlio Cortelazzo, Giovan Battista Pellegrini, Paola Benincà, Laura Vanelli Renzi, Giuseppe Francescato, Franco Crevatin, Giovanni Frau, Mario Doria, Carla Marcato, Piera Rizzolatti et Marcello Marinucci. 1984-1987. *Dizionario etimologico storico friulano*, 2 vol. Udine : Casamassima.
- DLR = Iordan, Iorgu, et al. 1965-2010. *Dicționarul limbii române. Serie nouă*. Bucarest : Academia Română/Editura Academiei Române.
- DOMe1 = Stempel, Wolf-Dieter (éd.). 2014-. *Dictionnaire de l'occitan médiéval*. Munich : Bayerische Akademie der Wissenschaften. <http://www.dom-en-ligne.de>.
- DRAE_{23.6} = Real Academia Española. 2014^{23.6} [1780¹]. *Diccionario de la lengua española. Versión 23.6 en línea*. Madrid : Real Academia Española, <https://dle.rae.es/>.
- DRG = Planta, Robert de, Florian Melcher, Chasper Pult, Andrea Schorta, Alexi Decurtins, Felix Giger, Carli Tomaschett et Ursin Lutz (éds.). 1939-. *Dicziunari Rumantsch Grischun*. Coire/Winterthour, Bischofberger/Fabag : Winterthur/Institut dal Dicziunari Rumantsch Grischun.

- FEW = Wartburg, Walther von, et al. 1922–2002. *Französisches Etymologisches Wörterbuch. Eine darstellung des galloromanischen sprachschatzes*, 25 vol., Bonn/Heidelberg/Leipzig-Berlin/Bâle : Klopp/Winter/Teubner/Zbinden.
- Georgescu, Simona. 2021. « Conceptual metaphors leading to some names of anger in the Indo-European languages (with focus on the Romance languages) ». *Revue roumaine de linguistique*, 65, no. 2 : 189-197.
- Gouvert, Xavier. 2014. « Reconstruction phonologique ». In *Dictionnaire Étymologique Roman (DÉRom). Genèse, méthodes et résultats*, édité par Éva Buchi et Wolfgang Schweickard, 61-128. Berlin/Munich/Boston : De Gruyter.
- IEW = Pokorny, Julius. 1959. *Indogermanisches Etymologisches Wörterbuch*. Bern. (confronté avec Indo-European Language Revival Association. 2007. *Proto-Indo-European Etymological Dictionary – A Revised Edition of Julius Pokorny's Indogermanisches Etymologisches Wörterbuch*. <http://dnghu.org/>).
- Kasten/Cody = Kasten, Lloyd A. et Florian J. Cody. 2001² [1946¹]. *Tentative dictionary of medieval Spanish*. New York : Hispanic Seminary of Medieval Studies.
- LEI = Pfister, Max (fond.), Wolfgang Schweickard et Elton Prifti (éds.). 1979–. *Lessico Etimologico Italiano*. Wiesbaden : Reichert.
- Leumann, Manu. 1977. *Lateinische Laut-und Formenlehre*. München: Beck.
- Kövecses, Zoltán. 1986. *Metaphors of Anger, Pride and Love. A Lexical Approach to the Structure of Concepts*. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Kövecses, Zoltán. 1990. *Emotion concepts*. New York-Berlin: Springer.
- Kövecses, Zoltán. 2000. *Metaphor and Emotion: Language, Culture, and Body in Human Feeling*. Cambridge: Cambridge University Press.
- Lausberg, Heinrich. 1965/1966 [original allemand : 1957–1962¹]. *Lingüística románica*, 2 vol. Madrid : Gredos.
- Malkiel, Yakov. 1990. « Integration of phonosymbolism with other categories of language change ». In *Diachronic problems in Phonosymbolism. Edita and inedita, 1979-1988*, vol. I, 10-42. Amsterdam/Philadelphia, John Benjamins.
- Meillet, Antoine et Joseph Vendryes. ³1960 [1948]. *Traite de grammaire comparée des langues classiques*. Paris : Champion.
- Meyer-Lübke, Wilhelm. 1890–1906 [original allemand : 1890–1902]. *Grammaire des langues romanes*, 4 vol. Paris: Welter.
- NVLS = Pittau, Massimo. 2014. *Nuovo vocabolario della lingua sarda. Fraseologico ed etimologico*, 2 vol. Sestu : Domus de Janas.
- Repetti, Lori. 2016. « Palatalization ». In *The Oxford Guide to the Romance Languages*, édité par Adam Ledgeway et Martin Maiden, 658-668. Oxford University Press.
- REW₃ = Meyer-Lübke, Wilhelm. 1930–1935³ [1911–1920¹]. *Romanisches Etymologisches Wörterbuch*. Heidelberg : Winter.
- Schweickard, Wolfgang. 2022. « Il Dictionnaire Étymologique Roman (DÉRom) e la ricostruzione del protoromanzo ». *Medioevo romanzo* 46, 42-56.
- Stefenelli, Arnulf. 1996. « Thesen zur Entstehung und Ausgliederung der romanischen Sprachen ». In *Lexikon der Romanistischen Linguistik (LRL)*, édité par Günter Holtus, Michael Metzeltin et Christian Schmitt, vol. 2/1, 73-90. Tübingen : Niemeyer.

- Straka, Georges. 1956. « La dislocation linguistique de la Romania et la formation des langues romanes à la lumière de la chronologie relative des changements phonétiques ». *Revue de linguistique romane* 20 : 249-267.
- Tiktin = Tiktin, Hariton, Paul Miron et Elsa Lüder. 2001–2005³ [1903–1924¹]. *Rumänisch-deutsches Wörterbuch*, 3 vol. Wiesbaden : Harrassowitz.
- TLIO = Beltrami, Pietro G. (fond.), Lino Leonardi et Paolo Squillacioti (éds.). 1998–. *Tesoro della Lingua Italiana delle Origini*. Florence : CNR, <http://tlio.ovi.cnr.it/TLIO>.
- TLL = 1900–. *Thesaurus Linguae Latinae*. Leipzig/Stuttgart/Berlin/New York : Teubner/Saur/De Gruyter.
- Väänänen, Veikko. 1981³ [1963¹]. *Introduction au latin vulgaire*. Paris : Klincksieck.
- Weiss, Michael. 2009. *Outline of the Historical and Comparative Grammar of Latin*. Ann Arbor/New York : Beech Stave Press.
- Wolf, Heinz Jürgen. 2012. « La “palatalisation secondaire” romane, aussi en Sardaigne ». *Revue roumaine de linguistique*, 57, no 4 : 355-375.

LE ROUMAIN LANGUE D'HÉRITAGE : LE CAS DES ÉTUDIANTS FRANCOPHONES EN MÉDECINE. ORIENTATIONS ET APPROCHES

Alina ANDREICA¹ , Nora-Sabina NEAMȚ² 

Article history: Received 6 June 2025; Revised 25 October 2025; Accepted 30 October 2025;
Available online 12 December 2025; Available print 30 December 2025

©2025 Studia UBB Philologia. Published by Babeș-Bolyai University.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ABSTRACT. *Romanian as a Heritage Language: The Case of Francophone Medical Students. Directions and Approaches.* This study offers a sociolinguistic analysis of oral proficiency in Romanian as a heritage language among second-generation francophone students coming back to study in Romania. Focused on a small group of students, within the framework of the student scientific circle *Cultures and Heritage Languages*, the study analyzed both prepared monologic discourse and spontaneous interactions to highlight phenomena such as linguistic interference, grammatical instability, and phonetic particularities,

-
- ¹ **Alina ANDREICA** est maître assistant à l'Université de Médecine et Pharmacie « Iuliu-Hațieganu » de Cluj-Napoca, où elle enseigne le roumain langue étrangère aux étudiants francophones et anglophones. Ses recherches portent sur l'enseignement du roumain langue étrangère et sur objectifs spécifiques, le roumain langue héritée, l'immersion linguistique, le tandem linguistique et l'interprétation médicale. Elle est autrice de ressources pédagogiques et coordinatrice du cercle scientifique étudiant *Cultures et langues héritée* depuis 2025. Elle a également dirigé le projet international *Research & Action and Training in Medical Interpreting* (2019–2022). Courriel : alina.andreica@hotmail.com.
- ² **Nora-Sabina NEAMȚ** est maître assistant, docteur, membre de la Discipline des Langues Modernes à l'Université de Médecine et Pharmacie « Iuliu Hațieganu » de Cluj-Napoca, spécialisée dans l'enseignement du roumain langue étrangère / pour objectifs spécifiques et le développement de supports didactiques. De 2017 à 2021, elle a enseigné le roumain et la communication interculturelle à l'Université de Strasbourg, dans le cadre des projets IdEx et ARTLINGO. Ses intérêts de recherche portent sur l'enseignement des langues modernes, le RLE, la traduction, la linguistique appliquée, la pragmatique, l'argumentation, l'intercompréhension et la communication interculturelle. Elle est formatrice certifiée en accompagnement interculturel et coordonnatrice du cercle scientifique étudiant *Cultures et langues héritées* depuis 2025. Courriel : noraneamt@gmail.com.

as well as, more schematically, their levels of plurilingual and mediation competence. The study is grounded in theoretical perspectives on plurilingualism in the context of international academic mobility, with an emphasis on how the heritage language is maintained, used, and valued in real educational and social settings. The results reveal variations in linguistic competence correlated with the degree of formal exposure to Romanian, the context of language use, the level of proficiency attained, and the family's involvement in maintaining the heritage language. The findings point out a direct relationship between formal exposure, language acquisition, and proficiency level. The research supports theoretical trends in European heritage language studies while bringing into focus the characteristics of Romanian as a heritage language in the context of a temporary return to the language's original linguistic environment.

Keywords: *Bilingualism, Romanian as a Heritage language, Linguistic competence, Interference, Sociolinguistics*

REZUMAT. *Limba română moștenită. Cazul studenților mediciști francofoni.*

Directii și abordări. Lucrarea analizează dintr-o perspectivă sociolingvistică competențele de exprimare orală în limba română moștenită în rândul studenților francofoni din a doua generație, care se întorc în România pentru studii. În cadrul cercului științific studentesc *Culturi și limbi moștenite* s-a realizat o analiză pe un grup restrâns de studenți, pe baza discursurilor monologale și a interacțiunilor spontane pentru a evidenția fenomenele de interferență, instabilitate gramaticală și particularități fonetice și, mai schematic, nivelul competenței plurilingve și de mediere. Studiul pornește de la premisele teoretice ale plurilingvismului în contextul mobilității academice internaționale, punând accent pe modul în care limba moștenită este întreținută, folosită și valorificată în contexte educaționale și sociale reale. Rezultatele indică variații în competența lingvistică, corelate cu gradul de expunere formală la limba română, mediul de utilizare a limbii, nivelul de competență atins și implicarea familială în conservarea limbii moștenite. Datele obținute evidențiază o relație directă între gradul de expunere formală la limba română, achiziția limbii și nivelul de competență atins. Cercetarea confirmă direcțiile teoretice din studiile europene privind limbile moștenite și subliniază specificitatea limbii române moștenite în contextul revenirii temporare în spațiul lingvistic de origine.

Cuvinte-cheie: *bilingvism, limbă moștenită, competență lingvistică, interferență, sociolingvistică, româna ca limbă moștenită*

Introduction

La dynamique transformationnelle de la société contemporaine favorise les échanges culturels et intensifie les contacts linguistiques tant au niveau communautaire qu'individuel. Ce phénomène contribue à la reconfiguration des rapports linguistiques et encourage l'émergence du bilinguisme et du plurilinguisme. Toutefois, les situations linguistiques sont marquées par une grande hétérogénéité, à l'image des cadres communautaires et familiaux qui les engendrent. Dès lors, le processus d'acquisition des langues englobe une diversité de paramètres : le moment et le mode d'acquisition, le système de référence, la valeur identitaire liée à l'appartenance linguistique, ainsi que le niveau de compétence atteint. Un exemple pertinent de cette complexité est celui des familles d'émigrés de première génération (G1) ; pour leurs enfants, représentant la « deuxième génération » (G2), le roumain ne constitue plus une langue maternelle au sens strict, mais une langue patrimoniale, héritée.

Le terme de « langue héritée » (LH) a connu, au fil du temps, un processus de reconceptualisation et d'élargissement épistémologique, évoluant d'une définition administrative et éducative vers une approche multidimensionnelle, intégrant des valences psycholinguistiques, sociolinguistiques et identitaires. Conceptualisée pour la première fois dans l'espace canadien des années 1970, dans le contexte d'une initiative gouvernementale visant l'intégration des langues d'héritage au sein du curriculum scolaire (à travers le programme *Ontario Heritage Languages*), la notion de LH a été définie à partir de deux axes de référence : la langue maternelle et la langue autochtone (Aalberse *et al.* 2019, 2-3). Elle a en outre introduit, dans le champ du débat scientifique, la problématique du statut d'une langue, à savoir son caractère officiel ou non officiel dans l'espace social et géographique où elle est employée (Cummins 2005, 585-592). Les connotations étendues du terme « langue maternelle » (LM), ainsi que son interprétation restrictive, ont conduit à l'élaboration de solutions terminologiques alternatives (Platon 2021, 192-197³ ; Andreica *et al.* 2024, 13-43) qui ont complété et nuancé le champ sémantique du concept de LH. Dans les années 1990, le terme de LH s'est progressivement étendu dans le milieu académique nord-américain (Cummins 2005, 585) ainsi qu'en Europe (Valdés 2000 ; Polinsky & Kagan 2007 ; Polinsky 2018, etc.), ce qui a entraîné une reconfiguration de sa définition par l'intégration de paramètres tels que : le statut linguistique, le degré de

³ Il s'agit d'alternatives conceptuelles et terminologiques ayant pour objectif de décrire diverses catégories de référence, telles que : le contexte et la source de l'apprentissage (*langue natale*), l'ordre de l'acquisition et la succession temporelle (*première langue* – L1), le mode d'appropriation (*langue naturelle*), le système de référence (*langue de référence*), la fonction identitaire (*langue d'appartenance*), ainsi que la dynamique du processus d'apprentissage (*langue source* ou *langue de départ*).

normalisation et de standardisation, la relation des locuteurs avec la communauté d'origine, l'âge et les modalités d'acquisition, entre autres. (Fishman 1999 ; Valdés 2000 ; Polinsky & Kagan 2007 ; Rothman 2009 ; Benmamoun *et al.* 2010 ; Polinsky 2018 ; Aalberse *et al.* 2019 ; Kupisch 2021, etc.)⁴

Étant donné la diversité familiale et communautaire des locuteurs de LH, nous adopterons, à des fins opérationnelles, la définition proposée par les linguistes Valdés (Valdés 2000) Rothman (Rothman 2009) et Polinsky (Polinsky 2018), selon laquelle une langue d'héritage constitue la langue minoritaire d'un locuteur bilingue, dans un contexte où la langue de la société d'accueil détient un statut dominant. Il s'agit d'une langue acquise principalement dans le cadre familial (Avram *et al.*, 2024, 330), qui se caractérise par un ensemble distinct de particularités.

Les locuteurs, qualifiés de « bilingues précoces » (Avram *et al.* 2024, 330), présentent un profil linguistique caractérisé par des compétences inégales, les soi-disant « bilingues asymétriques » ("unbalanced bilinguals", Polinsky 2018, 4) : la langue du pays d'accueil détient un statut dominant, tandis que la LH, soumise aux interférences de la langue officielle, tend à se réduire sur le plan grammatical et lexical et à perdre progressivement son poids, sa qualité de référence (Neșu 2020, 12-13). En tant que telle, elle n'est plus la langue de scolarisation et elle reste mieux maîtrisée à l'oral qu'à l'écrit (Platon 2021, 196).

Par ailleurs, l'exposition linguistique inégale conduit à une spécialisation différentielle et situationnelle : la LH est souvent utilisée dans des contextes familiaux, professionnels ou sociaux (Myers-Scotton 2006, 3), à des fins affectives ou comme code linguistique distinct, lorsque les locuteurs ne veulent pas se faire comprendre par les autres. Les locuteurs du roumain langue d'héritage (RLH) apprennent ainsi la langue transmise par leurs parents immigrants, laquelle peut être déjà marquée par des erreurs, des inflexions régionales ou dialectales (Gogâță *et al.* 2022, 14). Toutefois, la LH peut également être revalorisée dans une double dynamique d'insertion, académique et sociale, par un retour à sa source.

La présente étude s'inscrit dans une perspective sociolinguistique et vise à analyser les formes d'expression orale en roumain de jeunes étudiants bilingues francophones, inscrits à l'Université de Médecine et de Pharmacie « Iuliu-Hațieganu » de Cluj-Napoca (UMFIH) et qui ont acquis le RLH dans un cadre familial restreint ou élargi, sans bénéficier pour leur grande majorité d'un processus formel de standardisation, ni d'un renforcement structuré de leurs

⁴ Pour une analyse détaillée de la définition du concept dans le milieu académique, de son évolution historique, des variations d'interprétation déterminées par les différences géographiques et culturelles, ainsi que pour une classification des locuteurs de la LH, voir Aalberse *et al.* 2019, 2-17 ; Kupisch 2021, 1-12.

compétences, ni d'un encadrement métalinguistique. Nous nous sommes donc fixé pour but de déterminer les variables qui ont mené au niveau actuel de compétence linguistique. Il s'agit de faire ressortir les spécificités linguistiques dans des parcours individuels de (re)construction des compétences linguistiques, à l'intersection du cadre académique et du contexte socioculturel roumain, afin de révéler les mécanismes qui sous-tendent leur formation.

Partant du postulat selon lequel le locuteur agit en tant qu'acteur social en mobilisant le langage comme ressource essentielle dans l'interaction, notre étude se penche sur l'ensemble d'interférences récurrentes entre le français et le roumain dans les discours oraux. L'analyse porte, d'une part, sur les structures lexicales marquées par la perméabilité linguistique, notamment les écarts lexicaux, sémantiques et morphosyntaxiques, et, d'autre part, sur la présence des inflexions régionales, des particularités dialectales héritées, ainsi que sur l'alternance des codes. Une attention ponctuelle est portée à la relation entre compétences communicatives, médiation et mobilisation du répertoire plurilingue des locuteurs.

Enfin, la pertinence de cette étude réside dans son approche qualitative, fondée sur l'observation directe des usages langagiers dans un cadre académique non formel. Alors que la plupart des recherches sur les LH examinent l'usage du RLH par la G2 en contexte migratoire francophone, notre démarche se distingue par le fait qu'elle analyse cette même génération, des étudiants francophones issus de parents immigrés, dans le pays d'origine de la langue héritée, la Roumanie, où ils vivent temporairement.

Le roumain langue d'héritage dans l'Union Européenne

Dans la littérature de spécialité portant sur le contact linguistique entre le RLH et les langues romanes, des travaux récents portent sur les phénomènes d'influence et d'interférence linguistique. Consécutivement à l'intégration de la Roumanie dans l'Union Européenne et à l'intensification de la migration, ces recherches privilégient les pays accueillant d'importantes communautés roumaines, notamment l'Italie et l'Espagne. Plusieurs contributions méritent d'être signalées : pour le cas de l'Italie, les études de Cohal (2014), Neșu (2020) et Gogăț (2022), et pour le contexte espagnol, les travaux de Jeanu (2012) et Buzilă (2016). Parmi les travaux consacrés au RLH en contexte francophone, nous devons citer les recherches de Bârlea et Bârlea (2000), Popa (2022), ainsi que celles d'Avram *et al.* (2024). En raison de la parenté entre les langues romanes, Bârlea et Bârlea soulèvent la difficulté à distinguer avec précision les éléments relevant du transfert linguistique de ceux qui relèvent d'un héritage

commun dû à une genèse commune. Ils avancent également l'idée que l'interférence constitue une forme particulière de transfert entre la langue-fournisseur et la langue-récepteur. Leur analyse sociolinguistique prend appui sur la notion de prestige culturel, qui façonne les interactions entre les deux langues et encadre la dynamique de l'échange linguistique entre la langue donatrice (le français) et la langue réceptrice (le roumain). Dans une perspective également sociolinguistique, Popa (2022) étudie les modes d'expression des Roumains établis en France et tente d'identifier un sociolecte, une variété du roumain propre au contexte hexagonal, qu'elle nomme *romceză*, à l'image des formes documentées dans les communautés roumaines d'Espagne (*rumañol*) et d'Italie (*rotaliană*). Son étude s'appuie sur un corpus dichotomique composé de transcriptions de conversations semi-dirigées menées auprès de 109 sujets (G1 et G2) et de messages échangés sur les réseaux sociaux au sein de la même communauté, présentant des traits linguistiques similaires. Son analyse du roumain parlé par la G2 met l'accent sur les mécanismes d'acquisition de la langue. Les conclusions de l'étude font ressortir le fait que *romceza* peut être considérée comme code minimal et uniforme seulement dans certaines situations de communication et par rapport à la norme objective de la langue.

Une équipe de spécialistes en linguistique, issus de différents centres universitaires et de recherche européens⁵, a élaboré un projet ayant pour principal objectif l'analyse du RLH parlé par la G2 en France. Ce projet, inscrit dans un cadre de recherche plus large portant sur le contact entre deux langues romanes⁶, a été coordonné par Elena Soare sous le titre *Les langues d'héritage dans le contexte du français comme langue dominante (2020–2022)*. Le groupe cible étant composé de soixante enfants bilingues âgés de 5 à 13 ans, les résultats de l'étude offrent un tableau complexe du RLH : phénomènes contrastifs entre le roumain et le français (au niveau morphologique, syntaxique, lexical, etc.) et identification des zones vulnérables, tant dans le discours oral que dans les productions écrites. Cependant, dans le contexte transnational européen, où les langues et les cultures sont des phénomènes dynamiques aussi bien sur le plan collectif qu'individuel, le RLH reste un champ encore insuffisamment exploré, ouvrant ainsi la voie à de nombreuses perspectives de recherche.

L'environnement multiculturel et plurilingue de l'UMFIH, associé aux besoins d'expression culturelle et identitaire des étudiants internationaux, a conduit à la création, à partir de l'année universitaire 2024-2025, d'un cercle

⁵ *Les langues d'héritage dans le contexte du français comme langue dominante (2020-2022)*, coopération entre l'Université Paris 8 Vincennes-Saint Denis, l'Université de Strasbourg, l'Université de Bucarest et l'INALCO. Site web consulté le 27 mai 2025 : <https://sites.google.com/site/soarelena/langues-d-heritage>

⁶ International Research Network « Heritage Languages in Europe » sur six ans (2020-2026).

étudiant de recherche visant à promouvoir les échanges culturels et à favoriser une meilleure compréhension de l'altérité. Intitulé *Cultures et langues héritées* (CLH), ce cercle encourage l'expression de soi à travers la culture et la langue d'origine. Dans ce cadre, les étudiants locuteurs du RLH, issus de familles linguistiquement homogènes ou mixtes, ont été invités à réfléchir à l'influence de leur langue et / ou culture héritée sur divers aspects de leur vie.

Méthodologie de la recherche

L'enquête est fondée sur l'observation directe des productions orales réalisées dans le cadre du cercle *Cultures et langues héritées*, organisé durant l'année universitaire 2024–2025. Les participants ont exposé, à l'aide d'un support PowerPoint, un contenu structuré sous forme de monologue (dix minutes), suivi d'échanges spontanés avec leurs collègues (entre cinq et dix minutes). Ils ont été encouragés à présenter des éléments liés à leur culture d'origine et à leur LH. Le CLH s'est proposé de faire ressortir des manifestations de la compétence linguistique, de la compétence plurilingue, de la médiation culturelle et linguistique, ainsi que de l'adéquation sociolinguistique, telles que définies dans le *Cadre européen commun de référence pour les langues* (CECR-VC) (CECR-VC 2018). La participation aux sessions a été volontaire et le nombre d'interventions limité, malgré le grand nombre d'étudiants intéressés, afin d'assurer un équilibre linguistique et culturel. Cette contrainte explique la composition restreinte du groupe étudié⁷. Sur le total de plus de quarante présentations du CLH, il y a eu treize présentations des étudiants de la G2, dont sept francophones. Le groupe cible n'a pas bénéficié de séances exclusives, mais leurs présentations ont été incluses dans les séances normales du CLH, et le groupe a suivi les mêmes règles que le reste des participants. En tant qu'espace pour la communication d'une production personnelle, chaque séance du CLH a duré entre une heure et demie et deux heures, selon le nombre d'étudiants inscrits et a disposé d'un public d'au moins vingt personnes par séance. Les étudiants internationaux suivent les cours de langue roumaine de l'année en cours, donc le niveau de compétence linguistique en roumain du public varie de A1.1 à A2+. Les étudiants de la G2 ne sont pas tenus de participer au cours, vu leur niveau de compétence linguistique, et reçoivent d'autres tâches à accomplir. Tous les étudiants ont disposé d'une liste thématique d'orientation et de repères théoriques, mais vu le statut de cercle étudiant de cette activité, ils ont dû faire leur propre recherche sur le sujet et choisir des thématiques qui leur conviennent,

⁷ Le nombre d'étudiants en médecine générale ayant un ou deux parents roumains ne dépasse pas sept par promotion.

sans aucune intervention de la part des cadres didactiques, qui n'ont que le rôle d'observateurs. En échange, les étudiants de la G2 ont dû inclure dans leur présentations la langue héritée (type d'acquisition et utilisation), accompagnée ou non d'un autre sujet, dans le but spécifique de notre recherche, mais les discussions après les présentations ont été menées exclusivement par le public et les étudiants modérateurs⁸. Les cadres didactiques présents lors des séances ont pris des notes détaillées sur la présentation orale de chaque étudiant de la G2.

Le groupe étudié est composé de sept individus (quatre hommes et trois femmes), inscrits en deuxième et troisième année à la Faculté de Médecine, section francophone. Sur le plan géographique, cinq sont originaires de France et un étudiant vient de Belgique. La répartition selon la structure familiale est équilibrée : trois étudiants proviennent de familles mixtes et quatre d'entre eux sont issus de familles où les deux parents sont d'origine roumaine. Dans les familles mixtes, la mère est roumaine dans les trois cas, tandis que le père est français pour deux étudiants et belge pour le troisième.

Deux types de productions ont été pris en compte : les monologues préparés à l'avance et les interactions spontanées. L'analyse s'est centrée sur les niveaux phonétique, lexical et morphosyntaxique et a mis en évidence les interférences linguistiques récurrentes, les phénomènes individuels et les influences du français, les variétés régionales du roumain et les registres populaires ou colloquiaux. Ces observations ont permis d'identifier certaines tendances générales, en lien avec le niveau de compétence linguistique et les trajectoires personnelles des locuteurs⁹.

Le profil du locuteur francophone de RLH

L'UMFIH propose un parcours académique en langue étrangère, français ou anglais, à un grand nombre d'étudiants internationaux, sur une durée de cinq à six ans, selon la spécialité choisie. D'après les données officielles de l'université, en janvier 2025¹⁰, parmi les 3 312 étudiants étrangers inscrits aux trois facultés de l'UMFIH, 1 777 poursuivent leur formation en français (1 050 à la Faculté de

⁸ Pour les années à venir et dans l'intérêt de nos recherches ultérieures, nous prenons en considération l'application d'un questionnaire sur d'autres aspects liés à la langue héritée, qui ne peuvent pas constituer l'objet du CLH tel qu'il a été défini pour l'année 2024-2025 (progrès de la langue, migration de retour, les connexions et la dynamique relationnelle, etc.)

⁹ Cette approche méthodologique est le point de départ d'une étude de plus grande envergure, à visée comparative et quantitative, que nous envisageons de mener au sein de la communauté universitaire de l'UMFIH.

¹⁰ Conformément au *Rapport sur l'état de l'université*, janvier 2024 - janvier 2025 : 57, 61, 65, 72, 73, *Raport-privind-starea-universitatii-pentru-anul-2024.pdf*, consulté : le 27 mai 2025.

Médecine, 384 en Médecine dentaire, 343 en Pharmacie). La majorité d'entre eux provient de zones francophones, principalement la France (1 484 étudiants) et la Tunisie (178 étudiants) (*Rapport sur l'état de l'université*, janvier 2024, janvier 2025 : 57, 61, 65, 72-73). Du point de vue de la répartition par années d'études, pendant les années précliniques (les trois premières années), la filière francophone compte environ 355 étudiants par an (200 à la Faculté de Médecine, 75 en Médecine dentaire, 80 à la Faculté de Pharmacie)¹¹. Pendant cette période, des cours de roumain (du niveau débutant au niveau B1 en roumain pour des objectifs spécifiques) sont prévus pour faciliter la communication avec les patients et favoriser l'intégration socioculturelle et académique en Roumanie.

Un bref aperçu des typologies des étudiants internationaux (Neamt 2025) peut apporter de nouveaux éléments dans l'analyse du RLH, dans le contexte particulier de son usage au sein du pays d'origine de la langue. En général, la mobilité étudiante est influencée par la position sociale et économique, qui détermine à son tour le type de formation recherchée. Les étudiants qui se trouvent dans des positions moins favorables poursuivent généralement des études locales, tandis que ceux disposant de plus de ressources ont accès à des institutions prestigieuses et à des programmes internationaux (Garneau 2007, 14). La mobilité de longue durée, en particulier dans le cadre de programmes académiques complets, est souvent définie statistiquement comme une migration internationale (King et Raghuram 2013, 127-137) ; pour notre groupe cible, la migration est généralement de longue durée, volontaire, et motivée par des accords intergouvernementaux et institutionnels (Ioniță et Vlad 2022), tout en aboutissant à l'obtention d'un diplôme internationalement reconnu. Aux déterminants de la migration identifiés par Harfi et Mathieu (Harfi et Mathieu 2006, 36), respectivement la qualité de l'éducation dans le pays d'origine, l'accès aux institutions du pays d'accueil, le coût de la vie, la reconnaissance des diplômes, les opportunités d'apprentissage des langues, la présence de réseaux d'accompagnement, nous rajoutons la LH comme facteur clé dans le choix du pays d'études.

Si le choix de la Roumanie par les étudiants francophones peut, en général, être motivé par des facteurs économiques, des liens familiaux, une proximité linguistique (Ioniță et Vlad 2022) ou l'accès aux études dans le pays d'origine, le RLH constitue, dans le cas particulier des locuteurs de cette langue, un facteur déterminant, qu'ils proviennent de pays francophones ou non. Même si les étudiants suivent leur formation en français¹², le RLH représente un atout incontestable durant les cours précliniques et cliniques, fondamentaux dans le parcours académique médical.

¹¹ *Offre éducationnelle - UMF Cluj. Oferta educațională - UMF Cluj*, consulté : le 25 mai 2025.

¹² Sauf cas de double nationalité, les étudiants francophones doivent suivre une filière dédiée aux étudiants internationaux.

Théoriquement, une fois à l'étranger, les étudiants internationaux tendent à s'orienter vers des communautés partageant des affinités culturelles, ce qui leur procure un sentiment d'appartenance dans un environnement perçu comme étranger. En ce sens, « l'étudiant peut se sentir chez lui – ou en tout cas dans un milieu plus familier – et entre pairs avec le fait d'être étranger pour trait commun » (Pleyers et Guillaume 2009, 75). Les étudiants internationaux s'ancrent ainsi dans leurs communautés d'origine, soutenus par les technologies de communication et des bulles communautaires relativement stables : « bien qu'ils soient une minorité à Cluj-Napoca, ils ne se sentent pas marginalisés, tant que la relation ontologique et comportementale se fait avec l'espace d'origine français ou avec le groupe de pairs partageant leur identité et leurs valeurs comportementales » (Ursa 2017, 343) (notre traduction)¹³. En théorie, ce schéma ne s'appliquerait pas aux étudiants francophones locuteurs du RLH. Leur bilinguisme leur permettrait, dans de nombreux cas, de s'intégrer rapidement, voire immédiatement, aussi bien au sein de la communauté académique que dans les groupes d'étudiants internationaux et dans la société d'accueil. Ils disposent d'un capital linguistique, social et culturel dont leurs collègues francophones sont souvent dépourvus (Neamț 2025) et qu'ils partagent, dans une certaine mesure, avec leurs pairs locaux.

Nous mettons en évidence ici les variables du public cible : la structure familiale, le degré de standardisation de la langue, reflétée dans le parcours de scolarisation antérieure en Roumanie, l'acquisition de la langue (Tableau 1), ainsi que le lieu de naissance et l'âge d'émigration¹⁴.

Tableau 1 . Variables des sujets de l'étude.

Sujet ¹⁵	Niveau		Études en Roumanie	Acquisition du roumain	Famille (G1)	Observations
	C1, C2	B2, B1				
CLH1	C1			en famille, films obligatoires, vacances en Roumanie	Parents roumains	préoccupation constante de la famille pour l'apprentissage du roumain, désir de la famille de revenir en Roumanie

¹³ „Deși sunt o minoritate în Cluj, nu se simt marginali, atâta timp cât raportul ontologic și comportamental se face cu spațiul francez de origine sau la grupul de colegi care le împărtășesc valorile identitare și comportamentale.” (Ursa 2017, 343).

¹⁴ A part les deux étudiants scolarisés en Roumanie, les cinq autres sont nées en France / Belgique.

¹⁵ Pour souci d'anonymat, les noms des participants sont notés à l'aide de l'acronyme du cercle et de numéros.

LE ROUMAIN LANGUE D'HÉRITAGE : LE CAS DES ÉTUDIANTS FRANCOPHONES EN MÉDECINE.
ORIENTATIONS ET APPROCHES

Sujet ¹⁵	Niveau		Études en Roumanie	Acquisition du roumain	Famille (G1)	Observations
	C1, C2	B2, B1				
CLH2		B1/B2.1		en famille, vacances en Roumanie	mère roumaine père français	manque d'intérêt pour la langue en famille
CLH3	C2		école primaire	à l'école, ensuite en famille vacances en Roumanie	mère roumaine, père français	famille bien intégrée en France
CLH4		B2		en famille, vacances en Roumanie	parents roumains	famille bien intégrée en France
CLH5	C2		école primaire	à l'école, ensuite en famille, vacances en Roumanie	parents roumains	intérêt pour la communication avec la famille élargie
CLH6	C1			en famille à l'église dans la communauté études universitaires, vacances en Roumanie	mère roumaine, père belge	études de Théologie en cours en Roumanie, pas d'accent, lexique colloquial occasionnel
CLH7		B2		en famille, vacances en Roumanie	parents roumains	accent régional, lexique familier

Résultats de l'analyse linguistique

En tant que langue de communication utilisée principalement dans le cadre familial et investie d'une forte charge affective, le RLH se manifeste dans les discours des étudiants par un ensemble de particularités linguistiques régionales. Dans le cas des locuteurs vivant et étudiant à l'étranger, donc hors de Roumanie, cette langue se distinguerait par « une composition lexicale restreinte, centrée autour d'un lexique minimal (parfois régional [...]), ainsi qu'une grammaire – tant syntaxique que morphologique – qui ne répond pas systématiquement aux

normes prescriptives » (Neșu 2020, 12-13) (notre traduction)¹⁶. Ce constat ne s'applique cependant pas aux étudiants de notre groupe, qui vivent et étudient en Roumanie : leur lexique et leur grammaire sont nettement plus développés, et leur niveau de compétence linguistique ne se situe pas en dessous du niveau B2/B2.1 Cette compétence semble déjà être consolidée pour la plupart, et un seul étudiant (CLH6) déclare avoir progressé depuis son arrivée en Roumanie.

La dynamique des relations intergénérationnelles, ainsi que la fréquence d'utilisation de la langue dans le cadre familial font ressortir des particularités phonétiques propres au sous-dialecte et des emprunts phonétiques au français, en l'absence d'un encadrement institutionnel. Ces traits font du RLH une composante active du « répertoire linguistique dynamique et évolutif d'un apprenant » (CECR-VC 2018, 28), autrement dit, du plurilinguisme. L'usage d'idiomes populaires, combiné aux interférences linguistiques entre le roumain et le français, même dans un contexte académique non formel, illustre la perméabilité linguistique propre au contexte familial transnational et social. Ce comportement linguistique, qui dévie inévitablement de la norme, permet à ces locuteurs plurilingues de passer d'une variété à une autre, d'un sous-dialecte à un autre, et d'utiliser différentes formes ou registres d'une même langue (Popa 2022, 15-21, 266-275). L'alternance ou la coexistence des registres (littéraire et populaire / vernaculaire) confère originalité et dynamisme au discours, sur fond de mobilisation des ressources linguistiques. Dans le contexte du CLH, les présentations étudiantes, au-delà de l'échange interculturel intrinsèque, pourraient contribuer à limiter la sédimentation progressive des idiomes populaires dans la langue.

Avant d'aborder l'analyse linguistique proprement dite, deux précisions s'imposent. Premièrement, les étudiants ont bénéficié d'un temps suffisant pour concevoir leur présentation PowerPoint ainsi que le discours oral. Parfois, celui-ci a été transcrit intégralement ou de manière schématisée, ce qui le rapproche d'une production écrite, soumise à révision. Peu d'écarts à la norme ont été relevés, donc c'est surtout au niveau des interactions spontanées que nous avons analysé les compétences linguistiques. Deuxièmement, pour les phénomènes linguistiques observés, nous avons donné soit un exemple qui peut être généralisé au niveau de l'utilisation des termes ou structures de la même catégorie, soit des exemples d'utilisation individuelle. Nous avons séparé le niveau avancé (NvA) et intermédiaire (NvI). Nous présentons ci-après les phénomènes identifiés dans les interactions orales et nos observations préliminaires, classés selon trois niveaux d'analyse linguistique.

¹⁶ „o compoziție lexicală redusă, organizată în jurul unui vocabular minimal (uneori regional [...]), cu o gramatică – sintaxă și morfologie – nu întotdeauna perfecte sub aspect normativ [...]” (Neșu 2020, 12-13).

1. Niveau phonétique : particularités phonétiques et phonologiques récurrentes, notamment au niveau de la prononciation et de l'accentuation.

- la consonne occlusive nasale palatale intervocalique [ń] : [vińe] = *vine*, [puńem] = *punem* ; utilisation individuelle (CLH2, NvI), influence régionale et sous-dialectale.
- la prononciation de /o/ initiale comme [uă] : [uăraş] = *oraş* ; utilisation individuelle (CLH4, NvI), influence régionale et sous-dialectale.
- accentuation de certaines voyelles sous l'influence du français : *díplómă* au lieu de *díplomă* ; utilisation généralisée entre tous les sujets, même si occasionnelle pour le niveau avancé. Accent tonique du français, utilisation généralisée, sauf CLH6 (NvA).
- renforcement de /r/ consonne fricative uvulaire voisée : *am crezut* ; phénomène généralisé entre tous les sujets de l'étude, plus ou moins perceptible, sauf CLH6 (NvA).

2. Niveau lexico-sémantique : fréquence élevée d'éléments lexicaux familiers et dialectaux, liés au processus d'acquisition du roumain dans un cadre familial, fortement corrélés à l'origine géographique et au positionnement socio-économique des membres de la G1.

- l'utilisation de l'adjectif *fain* (en roumain populaire et colloquial de Transylvanie, signifiant *chouette*, *sympa* ou *top*) ; utilisation individuelle (CLH4, NvI).
- le calque, ou la traduction littérale de certaines expressions idiomatiques colloquiales, telles que *a freca menta* (expression en roumain colloquial, signifiant *ne rien faire*, *traîner* en français standard et *avoir la flemme* en français colloquial) ; utilisation individuelle (CLH3, NvA) :

Il m'arrive souvent de penser et de formuler mes phrases dans ma tête en roumain [...]. Cela a provoqué [...] des moments amusants, surtout [...] quand j'apprenais encore à parler français. Par exemple, [...] quand j'ai dit, sûr de moi : « J'ai frotté la menthe ». [...] Mon camarade m'a regardé avec confusion et m'a demandé : « Pourquoi ? ». Et moi, content qu'il m'ait compris, je lui ai répondu avec assurance : « Parce que j'avais la flemme ». Au moment où j'ai vu son regard encore plus confus qu'avant, je me suis rendu compte qu'il ne m'avait absolument pas compris. »¹⁷ (CLH3) (notre traduction).

¹⁷ „Mi se întâmplă des, să gândesc și să formulez frazele în cap în limba română [...]. Acest lucru a provocat [...] momente amuzante, mai ales [...] când încă învățam să vorbesc franceza. De exemplu, [...] când am zis, convins de mine : « J'ai frotté la menthe. » [...] Colegul meu s-a uitat la mine confuz și m-a întrebat : « Pourquoi ? » Și eu mulțumit că m-a înțeles, i-am răspuns sigur pe mine : « Parce que j'avais la flemme. » În momentul în care i-am văzut privirea și mai confuză decât înainte, mi-am dat seama că nu m-a înțeles absolut deloc". (CLH3)

Cet exemple concret met en évidence l'influence de la langue de socialisation précoce sur la structuration de la pensée et la construction identitaire, montrant que la langue héritée ne constitue pas seulement un moyen de communication, mais aussi un repère dans la formation de soi et dans la manière d'appréhender le monde : « À travers ces moments, j'ai compris que certaines expressions ne se traduisent pas littéralement et que la langue dans laquelle j'ai grandi influence profondément ma manière de parler et de penser. »¹⁸ (CLH3) (notre traduction).

- l'usage familier du substantif *chestie*, -ii et son adaptation sémantique selon le contexte ; transfert positif, utilisation individuelle (CLH2, NvI).

Si dans *Le Dictionnaire orthographique, orthoépique et morphologique de la langue roumaine* (DOOM) (DOOM² 2005, 141) le substantif *chestie*, -ii est noté comme une variante familière du lexème indépendant *chestiune*, -i, dans l'édition la plus récente on le retrouve comme lexème indépendant, nom féminin, avec le sens de *chose*, équivalent colloquial de *truc*, *machin*, *bidule* et le sens d'événement, équivalent d'*affaire*, *histoire* (DOOM³ 2021, 362). *Chestie* suivi d'un adjectif est utilisé dans un sens générique, à connotation positive ou négative : *chestie bună*, *chestie normală*, *chestie urâtă*. Les équivalents français ont le même sens de *chose*, *objet* ou *histoire*, *affaire* : *truc bien* / *truc sympa*, *truc normal* / *histoire normale*, *mauvais truc* / *mauvaise histoire* / *mauvaise affaire*. Contrairement aux contextes où les jeunes en Roumanie utilisent le substantif *chestie* dans un registre familier-argotique en lui appliquant diverses resémantisations (Chircu 2013, Szathmary 2024), les locuteurs du RLH l'emploient dans des situations d'hésitation, par manque de vocabulaire ou par crainte d'incapacité à fournir des explications précises. Ainsi est-il utilisé comme un substitut linguistique : *Chestia aia mi s-a părut că nu ... / Mă gândeam mereu la chestia aia ...* (*Ce truc-là m'a semblé ne pas ... / Je pensais toujours à ce truc-là*). Dans d'autres contextes, *chestie* est employé pour renforcer une idée précédemment exprimée, remplissant une fonction anaphorique : „Când eram mic, mama m-a învățat să fac colțunași. Când am venit la Cluj, am continuat cu chestia aia, dar nu-mi ies așa bine ca ei.” (« Quand j'étais petit, ma mère m'a appris à faire des raviolis sucrés. Quand je suis arrivé à Cluj, j'ai continué avec ce truc, mais je ne les réussis pas aussi bien qu'elle. ») (CLH2). Nous y rajoutons l'influence de la fréquence d'emploi de l'équivalent français, *truc*, en français familier.

- l'emprunt : *un barbecue* au lieu de *grătar* (CHL6) ; quoique présent dans le vocabulaire roumain, est ici considéré comme un emprunt. Le mot *barbecue* a des occurrences très limitées en roumain et reste soumis à la

¹⁸ „Prin momentele astea, am realizat că unele expresii nu se traduc și că limba în care am crescut îmi influențează felul în care vorbesc și gândesc”. (CLH3)

prononciation anglaise, tandis qu'il s'agit d'un terme standard, largement intégré dans l'usage courant par souci de cohérence avec les autres francophones. Utilisation individuelle (CLH6, NvA).

3. Niveau morphosyntaxique : prédominance marquée d'occurrences reflétant l'influence conjointe du français et du RLH.

- la confusion du pronom réfléchi *se* avec la marque du conjonctif *să* : *să face* ; utilisation individuelle (CLH7, NvI).
- l'utilisation de la voyelle *o* au lieu de *a* dans la formation du passé composé de 3^{ème} personne singulier, particularité régionale : *o venit* ; utilisation généralisée.
- le non-respect des alternances *-ă > -e, -ăște > -ască* dans la structure conjonctive à la 3^{ème} personne : *să zice, să concepe, să hotărâște* ; utilisation généralisée entre les locuteurs de niveau intermédiaire (CLH2, CLH4, CLH7).
- l'utilisation en variation libre de *acum* / (*maintenant*) et de *aici* / (*ici*) et *acuma* et *aicea*, spécifiques au roumain familial et à l'oralité du discours : *vedeți aicea, am venit aicea, acuma vă prezint* ; utilisation généralisée entre les sujets de l'étude, doublée par l'utilisation généralisée des formes populaires et colloquiales du pronom ou de l'adjectif pronominal démonstratif (de proximité ou d'éloignement) : *ăsta, asta / ăla, aia*, etc.
- *se poate* au lieu de *se pot* : *glume proaste se mai poate întâlni* (CLH6) ; utilisation individuelle (CLH6, NvA).
- forme incorrecte du déterminant adjectival du nom au génitif : *poza bisericii ortodoxă română* (CLH6) ; utilisations généralisées.
- la réalisation du génitif / datif par des prépositions : *în fața la Casa de Cultură, datorită la bunica mea* ; réalisation génitivale spécifique au langage familial : *verișorul de încredere a lui mama* ; les formes de génitif réalisées à l'aide de l'article génitival *al* (forme erronée *a* dans l'exemple antérieur) suivi du morphème génitival *lui*. Utilisations généralisées, mais rarement entre les sujets scolarisés en Roumanie (CLH3, CLH5, NvA).
- manque d'accord des adjectifs : *discriminare prezent* ; utilisation individuelle (CLH6, NvA).
- la substitution du pronom relatif *care* par *că* : *am un unchi că este inginer* ; le pronom auquel nous nous serions attendues était *cine*, traduit du français *qui*. Le remplacement répétitif par *că* signale un niveau en dessous du A2 de compétence linguistique pour les éléments de relation. Utilisation individuelle (CLH2, NvI).
- confusions des prépositions : *cărți pe limba română* ; utilisations généralisées, sauf scolarisation antérieure en Roumanie (CLH3, CLH5, NvA).

- l'absence du morphème de l'accusatif *pe* : *care* au lieu de *pe care* ; (COD, transfert négatif) ; utilisation généralisée.
- confusion du pronom COI avec la forme du pronom COD, due à l'identité des formes des pronoms COD et COI en français, première et deuxième personne du singulier : *mă citește, m-a explicat* (CHL7) (transfert négatif) ; utilisation généralisée entre les sujets de niveau intermédiaire.
- l'absence de l'article défini : *evenimente culturale cum sunt concert* (absence de l'article défini en français, transfert négatif) ; utilisation généralisée entre les sujets de l'étude, sauf scolarisation antérieure en Roumanie (CLH3, CLH5, NvA).
- transfert négatif du marqueur comparatif d'identité *comme* : *evenimente cum sunt concert* ; utilisation individuelle (CLH5, NvA).
- confusion des genres : *doi mii* (CLH6) (transfert négatif) ; utilisation généralisée entre les sujets, sauf scolarisation antérieure en Roumanie (CLH3, CLH5, NvA).
- la phrase parataxique : [...] *și-apoi se ia și se pune în tavă [...], și-atunci când e gata se scoate din cuptor [...]* ([...] puis, on le prend et on le met dans le moule [...], et quand c'est prêt, on le sort du four [...]) ; utilisation individualisée (CLH2, NvI et CLH1, NvA).

L'influence du registre populaire se manifeste également dans l'organisation du discours. Parmi les caractéristiques de la langue populaire figure la phrase parataxique : il s'agit de propositions principales juxtaposées, reliées par *et* ou par d'autres conjonctions de coordination (*mais, or, cependant*), ainsi que par certains adverbes temporels (*ensuite, alors*) ayant perdu leur valeur adverbiale pour fonctionner comme de véritables conjonctions (Farcaș 2011, 315-324). Cette préférence des locuteurs pour la juxtaposition au détriment de la subordination confère de la fluidité et de la continuité au discours et renforce l'élan narratif.

De nombreux exemples apparaissent dans les exemples de préparation des produits traditionnels roumains, la nourriture constituant l'un des éléments constants des présentations. Ainsi, dans l'explication de la recette des *sarmale* (feuilles de chou farcies à la viande) ou des *cozonaci* (gâteaux briochés) : [...] *și-apoi se ia și se pune în tavă [...], și-atunci când e gata se scoate din cuptor* (« puis, on le prend et on le met dans le moule [...], et quand c'est prêt, on le sort du four »). La conjonction de coordination ne fonctionne pas seule comme connecteur discursif, mais s'associe à l'adverbe marquant la succession temporelle pour former un connecteur complexe, signalant le passage d'une étape à l'autre. La conjonction *și (et)* perd ici sa fonction de marque de coordination, ce qui explique la tendance des locuteurs à entamer une phrase par *și*. Dans la langue populaire, cette position initiale correspond à une stratégie de gestion du discours, un « temps de réflexion que le locuteur s'accorde avant de présenter la suite des

événements » (Teiuș 1980, 99) (notre traduction)¹⁹ et traduit une planification en temps réel et une tendance à la simplification syntaxique. Cependant, chez les locuteurs du RLH, cette construction peut refléter, sur le plan morphosyntaxique, soit un répertoire linguistique encore limité, soit une compétence narrative en cours d'acquisition, tant au niveau sémantique que pragmatique, en lien avec les descripteurs du CECR-VC relatifs à la gestion du discours et à la cohérence textuelle (CECR-VC 2018, 148-149). Une organisation discursive plus nuancée, marquée par des successions temporelles du type *după aceea, după ce, după*, est située déjà au niveau A2 du référentiel pour la compétence linguistique en roumain (Platon *et al.* 2023, 47). Or, ces marqueurs ne se retrouvent que dans les productions orales des locuteurs avancés scolarisés en Roumanie et ils fonctionnent comme des automatismes discursifs.

Nous avons regroupé les phénomènes linguistiques ci-dessus selon plusieurs catégories de fréquence (influence du registre familier, colloquial ou du sous-dialecte, lacunes dans la grammaire du roumain, transferts négatifs du français vers le roumain, emprunts lexicaux, calques vers le français, erreurs d'accent tonique, recours à la parataxe) et nous les avons rapportés au profil du groupe cible (Tableau 1). Nous avons pu ainsi établir une corrélation directe entre le niveau linguistique et les variables des sujets, comme suit :

1. Niveau phonétique

A. Utilisation généralisée

- NvI : accents régionaux ;
- NvI et NvA : accent tonique du français, renforcement de /r/ fricative uvulaire voisée (ex. : CLH6) ;
- NvA avec scolarisation en Roumanie : pas de manifestations significatives.

B. Utilisation individualisée

- NvI : prononciation régionale ;
- NvA : aucune occurrence ;
- NvA avec scolarisation en Roumanie : aucune occurrence.

2. Niveau lexico-sémantique

A. Utilisation généralisée : aucune généralisation.

B. Utilisation individualisée

- NvA : emprunt lexical et lexique familier occasionnel (CLH6), calque (CLH5) (niveau débutant de la compétence linguistique en français) ;
- NvI : lexique familier ;
- NvA avec scolarisation en Roumanie : aucune occurrence.

¹⁹ „Un răgaz de gândire pe care și-l ia locutorul înainte de a expune urmarea evenimentelor.” (Teiuș 1980, 99)

3. Niveau morphosyntaxique

A. Utilisation généralisée

- NvI : formes incorrectes du conjonctif présent de troisième personne et la confusion des pronoms COI et COD ;
- NvI et NvA : utilisation des formes populaires du passé composé à la troisième personne du singulier et de la forme de singulier féminin du déterminant adjectival du nom en génitif ;
- NvI et NvA sans scolarisation en Roumanie : confusion des prépositions avec le génitif / datif, régime prépositionnel de verbes erroné, l'absence de l'article défini et la confusion des genres ;
- NvA avec scolarisation en Roumanie : l'attrition de la langue roumaine acquise en primaire n'a pas eu lieu.

B. Utilisation individualisée

- NvI : confusions des pronoms ;
- NvA : CLH6 ;
- NvA avec scolarisation en Roumanie : aucune occurrence.

Le cas spécial de CLH6 peut être expliqué par l'implication constante dans la communauté diasporique roumaine, la scolarisation en roumain au niveau universitaire (des études en parallèle) et le besoin d'utiliser uniquement le roumain dans au moins l'un de ses deux futurs métiers. Son niveau C2 est encore en cours de consolidation : occasionnellement, l'utilisation des emprunts, des fautes d'accord de l'adjectif et du verbe ou le transfert dans la structure comparative d'identité émergent.

4. Organisation discursive : la parataxe

On ne remarque pas d'utilisation généralisée, mais une utilisation individualisée au NvI. L'utilisation au NvA, sous-tendue par l'usage intensif du RLH en famille, est assumée comme stratégie discursive (CLH1) d'adaptation de la langue au niveau du public.

Résultats de l'analyse

Dans le cadre des compétences communicatives langagières, l'adéquation sociolinguistique observée se situe entre les niveaux B2 et C1 (CECR-CV 2018, 144). Sur le plan pragmatique, la cohérence et la cohésion thématique se positionnent majoritairement aux niveaux B1 et B2 (CECR-CV 2018, 149, 148), tandis que l'aisance à l'oral varie de B2 à C2 (CECR-CV 2018, 151). La médiation, notamment la capacité à établir un espace pluriculturel, se situe aux niveaux B2 et C1. La stratégie visant à adapter son langage (CECR-CV 2018, 169-170) est observée surtout au niveau B2 :

les présentations sont bien adaptées au type de public et au cadre non formel, mais ne considèrent pas le niveau linguistique, souvent A2 pour les locuteurs du RLS.

Lors des échanges, les explications fournies restent à un niveau de complexité linguistique élevé, éloigné de la microlangue requise pour ce type de situation. Or, la microlangue suppose une simplification du discours en fonction du destinataire, tout en conservant une forme correcte (Platon 2019, 68-69). Si les enseignants recourent à ce procédé à tous les niveaux de la compétence communicative, les locuteurs du RLH adaptent surtout le lexique, en privilégiant le vocabulaire panroman (Mărcean *et al.* 2022, 144-146), en s'appuyant sur une intercompréhension intuitive fondée sur la transparence lexicale.

Les régionalismes restent néanmoins présents dans leurs discours et la sédimentation semble être définitive, sauf pour CLH6. Sur le plan grammatical, les simplifications n'affectent ni les temps, ni les modes verbaux ou les formes du génitif / datif. Cela paraît cohérent vu que les locuteurs du RLH ne sont pas des spécialistes du référentiel des niveaux pour le roumain et ne parviennent pas à ajuster spontanément leur discours. Leur stratégie discursive récurrente consiste alors à parler plus fort, plus lentement, et à accentuer plus nettement les voyelles. Cependant, c'est précisément à ce niveau que les compétences C1 et C2 liées à l'exploitation du répertoire plurilingue se manifestent à travers l'alternance des codes comme stratégie d'explication et de clarification. Cette alternance intervient dès que le locuteur du RLH perçoit une incompréhension de la part du public : des mots en français insérés dans des phrases en roumain, des phrases en roumain au sein d'un discours en français.

Concernant leur difficulté à simplifier leur discours, il est important de souligner qu'en décalage avec nos attentes relatives à l'usage fréquent du roumain en dehors du cadre universitaire, les locuteurs du RLH déclarent ne l'utiliser que de manière limitée, préférant s'exprimer en français avec leurs collègues et amis. Le roumain reste une langue fonctionnelle (avec les propriétaires, dans les magasins, etc.) ou familiale et les locuteurs du RLH restent dans les bulles communautaires. Dans la majorité des cas (à l'exception de CLH5 et CLH7), l'intonation française des locuteurs du RLH reste perceptible, conduisant les locuteurs natifs à les identifier comme étrangers et à leur répondre en anglais, selon les témoignages recueillis. Après deux, même trois années d'immersion linguistique, seul CLH6 déclare avoir constaté une amélioration notable de ses compétences, ce qui peut s'expliquer par son parcours universitaire complémentaire et un intérêt manifeste pour la communication avec ses pairs dans le pays d'origine.

Conclusions

L'analyse des productions orales des étudiants francophones locuteurs du roumain langue d'héritage révèle une large gamme de phénomènes linguistiques caractéristiques du bilinguisme précoce : interférences, emprunts, alternance

codique, régionalismes et structures issues de la langue populaire et colloquiale. Ces manifestations sont dues moins à l'influence du français comme langue dominante, mais à la transmission familiale inégale du roumain, souvent non standardisé.

Les compétences linguistiques observées sont directement proportionnelles aux variables des sujets, notamment le contexte d'acquisition (familial ou institutionnel), la fréquence d'usage dans la vie quotidienne et la scolarisation précoce en Roumanie. Ce dernier facteur semble assurer une meilleure stabilité morphosyntaxique, la planification de l'organisation discursive et l'absence des régionalismes au niveau lexical. À l'inverse, l'acquisition du roumain en milieu familial corrélée à l'utilisation domestique ou fonctionnelle mène à des compétences linguistiques moins stables, plus influencées par la langue dominante, avec des erreurs récurrentes au niveau morphosyntaxique et une prononciation régionale, mais dépassant néanmoins le niveau seuil.

Le cercle étudiant *Cultures et langues héritées* joue ici un rôle essentiel : il constitue un espace institutionnel où le RLH peut être valorisé et exploré à travers des pratiques discursives encadrées. En encourageant l'alternance des codes, la médiation interculturelle et l'ajustement du discours au public, ce dispositif pédagogique favorise l'affirmation identitaire, la confiance linguistique et le développement progressif de compétences orales plus structurées.

Ces résultats confirment les directions des recherches européennes sur les langues d'héritage, en soulignant la spécificité du RLH dans un contexte de retour temporaire et partiel dans le pays d'origine de la langue : une dynamique de revalorisation, d'évolution et consolidation linguistiques, dans un milieu plurilingue.

BIBLIOGRAPHIE

- Andreica, Alina, *et al.* 2024. „Experiența de lectură a vorbitorilor de limbă română ca limbă-moștenire. Studiu de caz”. In *Predarea limbilor moderne în domeniul medical. Perspective teoretice și abordări practice*, édité par Ovidiu Ursa, Oana Mureșan, 13-43, Cluj-Napoca : Éditions UMFIH.
- Aalberse, Suzanne, Backus, Ad, Muysken, Pieter (eds.). 2019. *Heritage Languages. A language contact approach*. John Benjamins Publishing Company.
<https://doi.org/10.1075/sibil.58?locatt=mode:legacy>
- Avram, Larisa, Mardale, Alexandru, Soare, Elena. 2024. « Langues d'héritage – traits généraux et tendances de développement. Le cas du roumain d'héritage parle en France », *Romanica Cracoviensia*, volume 24, no. 3, Institut de Philologie Romane – Université de Cracovie : 330-340.
- Bârlea, Gheorghe, Bârlea, Roxana-Magdalena. 2000. *Lexicul românesc de origine franceză*. Tirgoviste : Éditions Bibliotheca.
- Benmamoun, Ellabas, Montrul, Silvina, Polinsky, Maria. 2010. *Prolegomena to Heritage Linguistics*. White paper. Cambridge. MA : Harvard University Press.
- Buzilă, Paul. 2016. *El rumano hablo en España*. Bucarest : Éditions de l'Université de Bucarest.

- Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer*. 2018. Volume complémentaire avec de nouveaux descripteurs. Strasbourg : Conseil de l'Europe.
- Chircu, Adrian. 2013. „Observații asupra unui substantiv abstract devenit concret în limba română actuală : rom. chestie”. In *Études romanes. Hommages offerts à Florica Dimitrescu et Alexandru Niculescu*, édité par Dan Octavian Cepraga, Lupu Coman, Lorenzo Renzi, 188-98. Bucarest : Éditions de l'Université de Bucarest.
- Cohal, Alexandru-Laurențiu. 2014. *Mutamenti nel romeno di immigrati in Italia*. Milan : Francoangeli.
- Cummins, Jim. 2005. "A proposal for action : strategies for recognizing heritage language competence as a learning resource within the mainstream classroom", *The Modern Language Journal*, 89 : 585-592.
- Farcaș, Ioan-Mircea. 2011. „Aspecte privind interferența limbă literară – grai popular”. *Annales Universitatis Apulensis. Series Philologica*, no. 1 : 315-324.
- Fishman, Joshua. 1999. *Handbook of Language and Ethnic Identity*. Oxford : Oxford University Press.
- Garneau, Stéphanie. 2007. « Les expériences migratoires différenciées d'étudiants français. De l'institutionnalisation des mobilités étudiantes à la circulation des élites professionnelles ? ». *Revue européenne des migrations internationales*, 2007/1 (vol. 23) : 139-161.
- Gogăț, Cristina-Elena, et al. 2022. „Limba română ca limbă etnică – analiză contrastivă a contextelor de utilizare”. In *Actualități în predarea limbilor moderne în context medical*, édité par Ovidiu Ursa, Oana Mureșan, 13-27. Cluj-Napoca : Éditions UMFIH.
- Harfi, Mohamed et Mathieu, Claude. 2006. « Mobilité internationale et attractivité des étudiants et des chercheurs ». *Horizons stratégiques. Mouvements des hommes et des activités*, 2006/1 no. 1 (*La Documentation française*) : 28-42.
<https://doi.org/10.3917/hori.001.0028>.
- Ioniță, Anamaria, and Vlad, Monica. 2022. « Les (pluri)mobilités étudiantes : intégration, adaptation, réciprocité ». *Glottopol*, no. 37
<http://journals.openedition.org/glottopol/2090>.
- Jeanu, Ioana. 2012. *Interferențe lingvistice româno-spaniole*. Iasi : Lumen.
- King, Russell, and Raghuram, Parvati. 2013. „International student migration : mapping the field and new research agendas”. *Population, Space and Place*, 19(2) : 127-137.
<http://dx.doi.org/doi:10.1002/psp.1746>
- Kupinsck, Tania, Polinsky, Maria. 2021. „Language history on fast forward : Innovations in heritage languages and diachronic change.” *Bilingualism : Language and Cognition*. 25 :1. 1-12.
- Mărcean, Nora-Sabina, et al. 2022. « Intercompréhension en langues romanes : enseigner le roumain langue étrangère aux étudiants francophones ». *Actualități în predarea limbilor moderne în context medical*. édité par Ovidiu Ursa, Oana Mureșan, 117-56. Cluj-Napoca : Éditions UMFIH.
- Meyers-Scotton, Carol. 2006. *Multiple Voices. An Introduction to Bilingualism*. Cambridge University Press.
- Neamț, Nora-Sabina. 2025. "Exploring perceptions of academic and social integration through intercultural competence development among French-speaking medical students".

- Studia Universitatis Babeș-Bolyai Philologia. Romanian as a Second Language (RL2). Portraits of the RL2 User. 70/2025, no. 2 : 145-166.*
<https://doi.org/10.24193/subbphilo.2025.2.07>
- Neșu, Nicoleta. 2020. „Aspecte ale predării limbii române ca limbă străină (ls) și limbă etnică (le) în străinătate”. În *Predarea, receptarea și evaluarea limbii române ca limbă străină. Actualitate și perspectivă*, edité par Ioana Jieanu, Loredana Netedu, Paul Nanu, 11-18. Finlande : Editions de l'Université de Turku.
- Platon, Elena. 2019. "The „micro-language” – an assault on natural languages?”. *Lingua. Language and Culture*, XVIII, no. 2 : 67-83.
- Platon, Elena. 2021. *Româna ca limbă străină (RLS). Elemente de metadidactică*. Cluj-Napoca : Presa Universitară Clujeană.
- Platon, Elena, Sonea, Ioana, Vasii, Lavinia, Vîlcu, Diana. 2023. *Descrierea minimală a limbii române. A1, A2, B1, B2*. Cluj-Napoca : Presses Universitaires de Cluj.
- Pleyers, Geoffrey, and Guillaume, Jean-Francois. 2009. « Expériences de mobilité étudiante et construction de soi ». *Agora débats/jeunesses*, no. 50(4) : 68-78.
<https://shs.cairn.info/revue-agora-debats-jeunesses-2009-4-page-68>
- Polinsky, Maria & Kagan, Olga. 2007. „Heritage languages: in the “wild” and in the classroom”. In *Language and linguistics compass*. 5/1. Blackwell Publishing Ltd. 368-395.
- Polinsky, Maria. 2018. *Heritage Languages and Their Speakers*. Cambridge : Cambridge University Press.
- Popa, Delia-Florina. 2022. *Romceza. O perspectivă sociolingvistică asupra limbii române vorbite în Franța*. Cluj-Napoca : Casa Cărții de Știință.
- Rothman, Jason. 2009. „Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages”. *International Journal of Bilingualism*, 13(2) : 155-163. DOI :10.1177/1367006909339814
- Szathmary, Melitta. 2024. „Pledoarie” în această chestiune... sau (/și) chestie”. *Analele Universității București. Limba și Literatură română*, 71(1) : 91-105.
- Teiuș, Sabina. 1980. *Coordonarea în vorbirea populară românească*. Bucarest : Éditions Scientifiques et Encyclopédiques.
- Wiley, Terrence, et al. 2014. *Handbook of Heritage, Community, and Native American Languages in the United States*. première édition, New York : Routledge.
<https://doi.org/10.4324/9780203122419>
- Ursa, Anca. 2017. „Metode, strategii și reprezentări identitare în învățarea limbii române de către studenții francezi din România”. *Qvaestiones Romanicae. Travaux du Colloque International Comunicare și Cultură în Romania Europeană*. no. 5 : 335-345. Timisoara : „Jozsef Attila” Tudományi Egyetem, Kiado Szeged & Éditions de l'Université de l'Ouest de Timisoara.
- Valdés, Guadalupe. 2000. „The Teaching of Heritage Languages: An Introduction for Slavic-Teaching Professionals.” In *The learning and teaching of Slavic languages and cultures/Bloomington*.
- Vintilă-Rădulescu, Ioana (coord.). 2005. *Dicționar ortografic, ortoepic și morfologic al limbii române*. deuxième édition (DOOM2). Bucarest : L'Académie Roumaine. L'Institut de linguistique „Iorgu Iordan-Alexandru Rosetti”. Univers Enciclopedique.
- Vintilă-Rădulescu, Ioana (coord.). 2021. *Dicționar ortografic, ortoepic și morfologic al limbii române*. troisième édition (DOOM3). Bucarest : L'Académie Roumaine. L'Institut de linguistique „Iorgu Iordan-Alexandru Rosetti”. Univers Enciclopedique Gold.

PÉRDIDA Y SUSTITUCIÓN DE ARABISMOS EN ESPAÑOL: RECIPIENTES Y MUEBLES DE LA COCINA

Mihai ENĂCHESCU¹ 

*Article history: Received 8 June 2025; Revised 20 September 2025; Accepted 21 September 2025;
Available online 12 December 2025; Available print 30 December 2025*

©2025 Studia UBB Philologia. Published by Babeş-Bolyai University.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ABSTRACT. *Loss and Substitution of Arabisms in Spanish: Kitchen Containers and Furniture.* With the loss of Arab political and cultural influence, many Arabisms fell out of use and were replaced by words that largely derive from Latin, through learned borrowing. An inventory of 16 terms of Arabic origin has been selected, designating containers and pieces of furniture used in medieval kitchens. The study examines the linguistic evolution of these terms over time, how they disappeared, were preserved, or were replaced by other words, based on evidence from historical Spanish dictionaries and diachronic corpora.

Keywords: *Arabisms, kitchen containers, furniture, lexical substitution, lexical loss*

REZUMAT, *Pierderea și înlocuirea arabismelor în spaniolă: recipiente și mobilier din bucătărie.* Odată cu pierderea influenței politice și culturale arabe, multe arabisme ies din uz și sunt înlocuite de cuvinte ce provin în mare parte din latină, pe cale cultă. S-a selecționat un inventar de 16 termeni de origine arabă ce desemnează recipiente și piese de mobilier folosite în bucătăria medievală.

¹ **Mihai ENĂCHESCU** es profesor titular de Lingüística Románica y Lingüística Hispánica en el Departamento de Lenguas y Literaturas Románicas, Clásicas y Neogriego de la Universidad de Bucarest, y desde 2011 es doctor en filología. Sus intereses de investigación abarcan la semántica románica diacrónica y comparada, la historia del léxico español y la fraseología y paremiología comparadas dentro del ámbito románico. Estos temas se reflejan en numerosos artículos publicados tanto en Rumanía como en otros países. Su producción científica se articula en torno a tres líneas principales: la primera se centra en la semántica sincrónica y diacrónica desde una perspectiva comparativa entre dos o más lenguas románicas; la segunda aborda la historia del léxico español, especialmente los procesos de sustitución de arabismos por términos de otros orígenes; y la tercera analiza la fraseología y la paremiología comparadas en las lenguas románicas. Universitatea din București. E-mail: mihail.enachescu@lls.unibuc.ro.

Lucrarea urmărește istoria lingvistică a acestor cuvinte de-a lungul timpului, dispariția, păstrarea sau înlocuirea lor cu alți termeni, pe baza datelor din dicționarele istorice și corpusurile diacronice ale limbii spaniole.

Cuvinte-cheie: arabisme, recipiente, mobilier de bucătărie, substituție lexicală, pierdere lexicală

1. Introducción

El componente árabe fue, después del latino, la influencia más significativa en el léxico castellano hasta comienzos del siglo XVI (Lapesa 1986, 133). Durante toda la Baja Edad Media, los arabismos comenzaron a combatir con los cultismos latinos y con los extranjerismos europeos. Posteriormente, se inició un proceso de retroceso que, según Lapesa (1986, 155-156), se debió al declive de la influencia de la cultura árabe y al auge del Renacimiento europeo.

Con la pérdida del peso político y cultural del mundo árabe, muchos términos de origen árabe dejaron de utilizarse. Este fenómeno se explica sobre todo por factores extralingüísticos, como los cambios culturales, el avance técnico y la depreciación de la civilización árabe, lo que llevó a la pérdida de muchas de sus palabras. En otros casos, los arabismos fueron desplazados por cultismos latinos o extranjerismos europeos. Para reemplazarlos, el español ha recurrido a diversas estrategias: uso de cultismos de origen latino, vocablos heredados, creación de derivados y compuestos, o incorporación de préstamos de otras lenguas. Curiosamente, en ocasiones un arabismo ha sustituido otro arabismo.

No se puede dejar de mencionar, además, la actitud de desprecio y rechazo hacia los vocablos de origen árabe, que también contribuyó a su desaparición progresiva del español. Como señala Corriente Córdoba (2013, 203), “es innegable que desde la Edad Media a nuestros días ha disminuido considerablemente el número de arabismos del iberorromance en uso”.

La creciente actitud negativa para con el mundo islámico que caracteriza el tardío Medioevo español acaba por alcanzar el vocabulario de origen árabe. [...] Con los nuevos gustos y modos de la época prerrenacentista muchos aspectos de la cultura material heredada de los vecinos musulmanes quedaron desfavorecidos. En el plano lingüístico se ve el inicio del lento proceso de la eliminación de muchas palabras de origen árabe [...]. (Dworkin 2013, 648-649)

La lengua española incluye en su sistema léxico unas áreas terminológicas constituidas por arabismos (arquitectura, astronomía, botánica, medicina,

mineralogía, etc.) al lado de terminologías populares (gastronomía, indumentaria, instituciones, profesiones, etc.), muchos todavía vigentes, otros fuera de uso o reemplazados por otras palabras (Fasla 1998-1999, 245-247).

La pérdida y sustitución de arabismos es un tema tratado en algunos estudios previos (Enăchescu 2015, 2016, 2017a, 2017b, 2017c, 2018, 2019a, 2019b, 2020, 2021, 2022, 2023, 2024), donde se ha hecho el análisis de un importante número de arabismos agrupados en campos léxicos, campos terminológicos y terminologías populares².

Entre los arabismos del castellano medieval que designaban distintos tipos de recipientes y muebles de cocina, algunos han caído en desuso. En muchos casos, esto se debe al cambio de preferencias de la sociedad, que desde el Renacimiento abandona —e incluso rechaza— ciertos términos asociados, de algún modo, con la cultura islámica y sus prácticas culinarias. Otros arabismos han sobrevivido y se usan bastante actualmente, mientras que otros han sido sustituidos por otros vocablos.

Para la selección de los términos se ha recurrido a unos estudios fundamentales sobre los arabismos, como Corriente Córdoba (1992), Maíllo Salgado (1998) o Neuvonen (1941), pero también a los apartados dedicados a los arabismos de varios manuales de historia del español: Corriente Córdoba (2013), Dworkin (2012), Lapesa (1986), Penny (2006) o Torrens Álvarez (2007).

En primer lugar, se hará una presentación del inventario de los términos de origen árabe que serán el objeto de nuestro estudio. Se indicará el étimo, el significado según el DLE, su primer testimonio y un ejemplo recogido del corpus CDH. Asimismo, comprobaremos su uso actual mediante una búsqueda en el corpus CORPES XXI.

Luego se pasará al análisis de los términos, que se agruparán en dos apartados, los que aún están en uso y los que se encuentran hoy en día desusados. Se realizará un análisis de corpus apoyado en el banco de datos del CDH con el fin de reconstruir la historia lingüística de estas palabras y observar cómo ha evolucionado su uso a lo largo del tiempo, corroborado con los diccionarios históricos del español, tanto los académicos como otros diccionarios bilingües o glosarios de los siglos XV-XVIII.

2. Recipientes y muebles de la cocina

Acetre viene del ár. and. **assītl*, este del ár. clás. *saṭl* ‘vasija con asa, cubo’, a su vez del sir. *syṭla*, y este del lat. *sitŭla* (DAAL, s.v. *acéter*). Según el DLE, es un “caldero pequeño con asa que sirve para sacar agua de las tinajas” o bien,

² Véanse igualmente los estudios de Giménez Eguibar (2010, 2015, 2016a, 2016b) para otras aportaciones valiosas sobre el estudio del arabismo español.

en el ámbito religioso, un “caldero pequeño con asa en que se lleva el agua bendita para las aspersiones litúrgicas”, un significado derivado del primero. Se documenta por primera vez en el año 1042, en la forma *azeptre* (DCECH), en un texto en latín³. Reproducimos aquí la primera mención encontrada en el corpus CDH, de una fecha más tardía. Hay pocos ejemplos en el CORPES XXI, (incluye textos actuales, del siglo XXI), con tan solo 3 ejemplos en 3 textos.

E dezimos assí: que quando les acaeçier' aquesto, que tomen del orín del cini (e es el cobre de que fazen los **açetres**), e del áloen, e del açúcar blanco, tanto de lo uno como de lo ál, e muélanlo e denles d'ello con carne de oueja tres días, un pos otro, tanto quanto entendieren que avrán mester. (1250, Abraham de Toledo, *Moamín. Libro de los animales que cazan*)

Albarrada procede del ár. hisp. *albarráda*, y este del ár. clás. *barrādah*, ‘refrescadora’ (DAAL, *albarrada*2). Es lo mismo que *alcarraza*, según el DLE. No aparece en documentos, solamente en el *Glosario etimológico de las palabras españolas* de Eguilaz (“búcaro o jarro para beber”) y después en los diccionarios académicos a partir de 1914. Es homónima de *albarrada* ‘pared de piedra seca’. En el corpus del español del siglo XXI hay 36 ocurrencias en 17 textos, pero ningún ejemplo con este significado.

Alcarraza viene del ár. and. *alkarráza*, ár. clás. *kur[r]āz*, que a su vez viene persa *korāz* ‘buche’, por alusión a su forma (DAAL, s.v. *alcarraza*). Es una “vasija de arcilla porosa y poco cocida, que tiene la propiedad de dejar rezumarse cierta porción de agua, cuya evaporación enfría la mayor cantidad del mismo líquido que queda dentro” (DLE). Su primera documentación es del 1331, en la forma *carraço* (DCECH). Reproducimos el primer ejemplo encontrado en el corpus CDH, de 1569. En el CORPES XXI se encuentran 3 ejemplos en 2 textos, los dos de Colombia.

De muchas maneras nos enseñan los antiguos a enfriar el agua: o puesta en lugar alto al ayre frío para que le dé el sereno toda la noche; o en barriles bien atapados puestos en el profundo de algún pozo; o quando las **alcarrazas** o botijas se ponen sobre nieve o se ponen soterradas en la nieve, puestos los barriles d'el agua dentro en los arroyos o acequias por donde passa agua frigidíssima. (1569, Francisco Franco, *Tractado de la nieve y del uso de ella*)

Alcolla viene del ár. hisp. *alqúlla*, y este del ár. clás. *qullah* (DAAL, s.v. *alcolla*). Es una ‘ampolla grande de vidrio’ (DLE), y su primer testimonio data de 1256-1263 (DCECH). No se encuentra ningún ejemplo en el CORPES XXI.

³ En castellano se documenta en un documento notarial de c1181, fechado en 1056, pero se trata de una falsificación (DHLE).

Acuestas o por si mismo o en alguna su bestia o en carreta o en naue prometiendo de leuar algund onbre o vino o olio o otra cosa semeiante en odrez o en **alcollas** o en toneles. o pilares de marmol o redomas o otra cosa semeiante destas. (1256-1263, Anónimo, *Siete Partidas de Alfonso X*)

Alcuza tiene su origen en el ár. hisp. *alkúza*, ár. clás. *kūzah*, y este del persa *kuze* a través del arameo *kūz[ā]*, (DAAL, s.v. *alcuza*). Según el DLE, es una “vasija de hojalata o de otros materiales, generalmente de forma cónica, en que se guarda el aceite para diversos usos”; en unos países de América (Arg., Bol., Chile, Ec., Nic., Perú y R. Dom.) se puede referir también a una *vinagrera*, “pieza con dos frascos para aceite y vinagre” (DLE). Su primer testimonio se remonta a 1253, según el DCECH. Parece ser que todavía hay cierto uso actualmente, pues se han encontrado 31 ejemplos en 25 textos, 22 de estos en España.

Et que descenda sobresta piedra la uertud de figura de omne grand embuelto en un alquice. & antel una mugier que tiene en su mano una **alcuza** pora tener azeyt. (c1250, Alfonso X, *Lapidario*)

Aljáfana o **aljébana** vienen del ár. hisp. *aljáfna*, y este del ár. clás. *jafnah*. (DAAL, s.v. *aljafana*). El DLE remite en ambos casos a su sinónimo más usual, *jofaina*; *aljáfana* aparece con la mención desusado. *Aljáfana* aparece por primera vez en 1525, y *aljébana* en 1599. Para menciones de *aljébana* hemos recurrido al DH (1960-1996). No hay ocurrencias en el corpus del español actual.

[...] y después escaldar o remojar estas rebanadas de pan tostadas, con buen caldo de carnero en una **aljafana** o plato hondo, y después sacarlas y ponerlas en un gran plato alrededor. (1529, Anónimo, *Libro de guisados de Ruperto de Nola*)

Después hazer vnas rebanadas de pan, y tostarlas, y raer la quemadura, y después échalas en vna **aljebana** o plato ondo, y echalles por encima buen caldo de carne bien caliente. (1599, Diego Granado, *Libro del arte de cozina*)

Almirez o **almofariz** viene del ár. hisp. *almihrás*, y este del ár. clás. *mihrás* (DAAL, s.v. *almidet/z*). Es un “mortero de metal, pequeño y portátil, que sirve para machacar o moler en él” (DLE). Se documenta en un texto de 1475, según el DCECH, pero el corpus CDH propone un documento más temprano. Es una palabra en uso, según lo demuestran los 46 ejemplos en 31 textos, 29 de estos en España.

[...] & despues verted aquel agua & tornaldo a mojar con otra desta agua mesma & tornalda a verter & echad el antimono desque este bien enxuto en vn **almirez** bien linpio & molerlo gran rato. (a1453, Gómez de Salamanca, *Compendio de medicina*)

Almofía procede del ár. hisp. *almuhfiyya*, ‘oculta’ (DAAL, s.v. *almofía*). Su primera documentación es de 1490, en el *Universal vocabulario* de Alfonso de Palencia (DCECH). Es otro sinónimo, desusado, de *jofaina* (DLE); no se encuentran ocurrencias en el CORPES XXI.

Catinum. dizen vaso torneado o **almofia** bernegal. (1490, Alfonso de Palencia, *Universal vocabulario en latín y en romance*)

Altamía viene del ár. hisp. **ħaltamíyya* ‘vasija de barro’, adjetivo de *ħaltam* ‘sellado’, ya que al principio esta denominación se aplicaba a contenedores de vino (DAAL, s.v. *altamía*). El DLE indica que se trata de una “especie de taza”, una forma desusada, documentada por primera vez en el siglo XIII (DCECH). No se encuentran testimonios en el CORPES XXI.

[...] & bolverlo en uno & fervirlo en una sarten & quano fuere fervido echarlo en una **haltamia** & dexarlo elar & untarle mucho la corona con ello, dos vezes o tres al dia [...]. (c1275, Anónimo, *Libro de los caballos*)

Anafe procede del ár. hisp. *annáfih*, y este del ár. clás. *nāfih* ‘soplador’ (DAAL, s.v. *anafe*). Es un “hornillo, generalmente portátil” (DLE), documentado por primera vez en 1475, en la forma *alnafe* (DCECH). En el corpus CDH se encuentra otro testimonio muy cercano como fecha, que reproducimos abajo. En el CORPES XXI se encuentran 51 ejemplos en 37 textos, la mayoría en América, con tan solo dos ejemplos en España.

Costo vn **alnafre** de cobre, para haser perfumes, 754. (1477-1491, Anónimo, *Cuentas de Gonzalo de Baeza, tesorero de Isabel la Católica*)

Ataífor tiene su origen en el ár. hisp. *aṭṭayfūr*, y este del ár. clás. *ṭayfūr* ‘pájaro saltante’ (DAAL, s.v. *ataífor*). Su primera documentación, *atayfer*, que puede ser una errata, data de 1431-50 (DCECH). En el corpus CDH hemos encontrado la primera documentación en el *Vocabulario* de Nebrija. Se puede referir o bien a una “mesa redonda y pequeña usada por los musulmanes” o a un “plato hondo para servir viandas que se usaba antiguamente” (DLE). Parece estar casi desusado actualmente, pues se encuentran dos casos en dos documentos.

Ataifor morisco. abacus i. abax .acis. (1495, Antonio de Nebrija, *Vocabulario español-latino*)

Jarra procede del ár. hisp. *aljárra*, y este del ár. clás. *jarrah* (DAAL, s.v. *(al)gerra*), documentado en 1251 (DCECH). Es una “vasija de barro, porcelana, loza, cristal, etc., con cuello y boca anchos y una o dos asas” (DLE). Es una de las palabras más utilizadas hoy en día de este campo. En el CORPES XXI encontramos 2.182 ejemplos en 1.190 documentos, 893 de estos en España.

Dizen que un religioso avía cada día limosna de casa de un mercador rico: pan et miel et manteca et otras cosas de comer. Et comía el pan et los otros comeres, et guardava la miel et la manteca en una **jarra**. Et colgóla a la cabeçera de su cama, tanto que se finchó la **jarra**. Et acaesció que encaresció la miel et la manteca. (1251, Anónimo, *Calila e Dimna*)

Jofaina, aljofaina o ajofaina (la primera es la que se usa actualmente), vienen del ár. hisp. *aljufáyna*, y este del ár. clás. *jufaynah*, diminutivo de *jafnah*, étimo de *aljafana* / *aljébana* (DAAL, s.v. *alcofaina*). Es una “vasija en forma de taza, de gran diámetro y poca profundidad, que sirve principalmente para lavarse la cara y las manos” (DLE). Se documenta por primera vez en 1615-1617, según el DCECH, pero vamos a indicar aquí una mención anterior, que aparece en el DH (1960-1996). Se encuentran 106 ocurrencias en 71 documentos, 52 en España en el corpus del español actual; a estos se puede añadir una ocurrencia de *aljofaina*.

Axofayna llaman en algunas Partes d'España i en el Reyno de Toledo a vn vasso o conquilla de barro a que (en otras Partes d'España) llaman almofía y barreña. (1593, Diego de Guadix, *Recopilación de algunos nombres arábigos que los árabes pusieron a algunas ciudades y otras muchas cosas*)

Redoma viene del ár. hisp. **ra/uṭúma*, y este del ár. clás. *raṭúm* 'estrecha de vulva', influenciado por el prefijo latino *re-* (DAAL, s.v. *redoma*). El DCECH defiende en este caso que es una voz patrimonial, documentada en la Península desde el siglo X, en textos en latín, de origen desconocido. Es una “vasija de vidrio ancha en su fondo que va estrechándose hacia la boca” (DLE). Parece ser palabra actual, pues la encontramos 126 veces en 98 documentos, de los cuales 21 en España y 80 en Venezuela.

Los pueblos de la villa, páuperes et potentes, fazién grand alegría todos con instrumentos; adobavan convivios, davan ad non habentes sus carnes, sos pescados salpresos e recientes. Andavan las **redomas** con el

vino piment, conduchos adobados maravillosament; qui prenderlo quisiese non avrié falliment, non trayén en su pleito ningún escarniment. (1246-1252, Gonzalo de Berceo, *Los Milagros de Nuestra Señora*)

Taza se remonta al ár. hisp. *ṭássa*, este del ár. clás. *ṭassah*, y este del pahlavi *tašt* 'cuenco' (DAAL, s.v. *taça*). Es una “vasija pequeña, por lo común de loza o de metal y con asa, empleada generalmente para tomar líquidos” (DLE), documentada por primera vez en 1272, según el DCECH, fecha corregida por los datos del CDH con unos cincuenta años. Es la palabra más usual de nuestro inventario, hecho reflejado por las 17 377 ocurrencias en 3887 textos, 3088 en España.

[...] en aquesta ysla fueron primera mente fallados vasos e **taças** de tierra contra tierra de ytalía escontra ysla que dizen silicia. (c1223 Anónimo, *Semejanza del mundo*)

Zaque tiene su origen en el ár. hisp. *záqq*, y este del ár. clás. *ziqq* (DAAL, s.v. *zaque*). Es, según el DLE, un “odre pequeño”, cuyo primer testimonio se remonta a 1475 (DCECH), fecha corregida por los datos del CDH. Parece estar fuera de uso en el español actual, pues se ha encontrado un solo ejemplo en el corpus.

Está más una gran tina con todas sus pertenencias, los embudos y avenencias todos puestos en hazina, cinco **zaques** y una odrina, que se dize ser notorio ser el vuestro oratorio; no os hallan otra cortina. (a1435, Juan Agraz, *fragmento* (a1435, Juan Alfonso de Baena, «Poesías») (Suplemento al *Cancionero de Baena*))

3. Arabismos vigentes

Acetre ha tenido poco uso a lo largo del tiempo. En el corpus CDH se encuentran 56 ocurrencias en 37 textos, la mayoría, 53, en España. Hay 12 ejemplos medievales (siglos XIII-XV), 24 de los siglos XVII, luego su uso decrece: un ejemplo del XVIII, 9 del XIX y 10 del siglo XX. Pese a esto, parece tener un uso constante, sus pocas apariciones se deben a lo mejor a su especificidad.

Lo registran todos los diccionarios antiguos⁴: el *Vocabulario español-latino* (*urna aerea. situla aerea*) de Nebrija (1495), todos los repertorios bilingües de los siglos XVI-XVII⁵, así como el *Tesoro de la lengua castellana o*

⁴ Toda la información sobre los diccionarios antiguos se ha sacado del NTLLE.

⁵ Cristóbal de Las Casas, *Vocabulario de las dos lenguas toscana y castellana* (1570); Richard Percival, *Bibliothecae Hispanicae pars altera. Containing a Dictionarie in Spanish, English and*

española de Sebastián de Covarrubias (1611) y más tarde *Autoridades* (1726). Con el significado religioso ‘recipiente para llevar agua bendita’ se registra ya desde 1278, según el DHLE. Es el significado con el que sobrevive actualmente. Ejemplos más recientes (CORPES XXI) solo apuntan hacia este significado.

[...] I cruç gran de Limojas, & otra chica & otra fe fust, I scrino, & 1 II anpollas, & 1 I **açétel**, & 1 I frontal de quarayeyesma, & I vestiment blanco, & I casula de alcoton, & I vaçin, & II candeleros grandes de fierro. (1278, Anónimo, *Prior S. Pedro Viejo Huesca nombra procurador*)

A su lado, un diácono ataviado con dalmática de seda blanca le sostenía el **acetre**. El obispo hisopó la barquilla y el globo con agua bendita, y bendijo a Matías Pérez, que se hincó de rodillas dentro de la canasta y se santiguó. (2008, Berta Serra Manzanares, *Los ojos del huracán*)

Alcarraza aparece en el corpus diacrónico 81 veces en 49 documentos, 67 en España. Hay 13 ejemplos de los siglos XVI-XVII y 44 del siglo XX, lo que puede sugerir cierto uso actual. No aparece en Nebrija, pero sí en los repertorios de Palet (1604), Oudin (1607) o Vittori (1609). La recoge Covarrubias (1611), así como *Autoridades*. Un libro de alfarería la describe como una jarra de cuatro picos.

La jarra de cuatro picos está muy relacionada con las tallas de Triana (Sevilla) y las jarras de La Rambla (Córdoba), así como las que se hacían en Vera (Almería). Todas ellas son piezas muy antiguas cuyo primitivo nombre es **alcarraza**, de evidente raíz musulmana y no sólo en el nombre. Estas **alcarrazas** han encontrado lugar de honor en cuadros famosos como El aguador de Velázquez, el Bodegón de Zurbarán y el Bodegón de la **alcarraza** de Meléndez. (1997, Natacha Seseña, *Cacharrería popular. La alfarería de basto en España*)

Alcuza está registrada en 162 casos de 108 documentos, la gran mayoría, 152, en España. Hay 10 ejemplos medievales (siglos XIII-XV), 49 de los Siglos de Oro (XVI-XVII), y en el siglo XX se registra un uso no muy alto, pero contante, con 67 ocurrencias. Aparece en Nebrija y en todos los diccionarios antiguos, incluso el *Compendio* de Tamarid (1585). Covarrubias no la incluye, pero sí *Autoridades*. Conserva su significado tradicional hasta la actualidad.

Latine (1591); Juan Palet, *Diccionario muy copioso de la lengua española y francesa* (1604); César Oudin, *Tesoro de las dos lenguas francesa y española. Thresor des deux langues françoise et espagnolle* (1607), Girolamo Vittori, *Tesoro de las tres lenguas francesa, italiana y española. Thresor des trois langues françoise, italienne et espagnolle* (1609) y otros.

Según los datos del DLE, actualmente (en algunos países de América) se puede referir igualmente a un recipiente con dos contenedores, uno para aceite y otro para vinagre. El ejemplo que sigue podría referirse a este tipo de recipiente.

—Estamos pensando en un cuantoai. En un buen cajón de sastre, con un poco de cada cosa. O en una ensalada y una **alcuza**. Vivencias con unos granos de sal, unas gotas de aceite y un chorrito de vinagre. Lo que un sociólogo llamaría «una exploración de la realidad». (1988, Volodia Teitelboim, *En el país prohibido. Sin el permiso de Pinochet*, Chile)

Almirez o **almofariz** parece haber sido palabra usual a lo largo del tiempo. En el corpus se registran 521 ejemplos en 255 textos, 484 de estos en España. Hay 10 ejemplos medievales, 246 de los siglos XVI-XVII, y 140 del siglo XX. Aparece en todos los diccionarios antiguos: Nebrija (1495), Casas (1570), López Tamarid (1585), Percival (1591), Palet (1604), Oudin (1607), Vittori (1609), Rosal (1611), Covarrubias (1611), *Autoridades* (1726).

Es una palabra competida por la voz patrimonial *mortero*, del lat. *mortarium* (DCECH). En un texto del siglo XV podemos ver que las dos se usaban como sinónimos.

Toma seuo de venado o de carnero castrado o de cabron & derritelo en vna caçuela & pon ende vna poquita de argila & tanta resina como la quarta parte del seuo & vn poco de azeyte de oliuas & picalo todo muy bien en vn **almirez o mortero**. Unguento fosco para las llagas que hieden & en ellas no quiere crescer la carne. (1494, Anónimo, *Traducción del Compendio de la humana salud de Johannes de Ketham*)

Sin embargo, Maíllo Salgado (1998: 275-276) considera que no hubo nunca tal competición, puesto que pronto se hizo una diferencia semántica: el mortero es la denominación genérica, y puede ser de madera, piedra o metal, mientras que el almirez es de metal. Podemos comprobarlo en el ejemplo que sigue, del siglo XV, donde se habla de un almirez de cobre, y añadimos un ejemplo más reciente donde se explica asimismo que el almirez es un mortero de metal.

Lo que llevauan e se demandava por las gentes de aquellas partes eran ropas viejas traydas, que no toviesen pelo, e **almireces de cobre**, e candeleros de latón, & manillas de latón; y en especial llevauan de aquellas conchas, que eran allá mucho demandadas. (p1480-1484, Hernando Del Pulgar, *Crónica de los Reyes Católicos*)

Otro grupo importante lo constituyen las canciones religiosas particularmente las de Semana Santa y los villancicos que acompañados de zambombas, panderetas, castañuelas y **almireces (mortero de metal)**, se ejecutan en las iglesias durante la misa del gallo y también junto a los belenes. (1997, Tony Évora, *Orígenes de la música cubana. Los amores de las cuerdas y el tambor*, Cuba)

Almofariz aparece en los diccionarios a partir de *Autoridades* (1726), donde podemos leer: “Parece ser lo mismo que almirez”. *Autoridades* (1770) indica ya que se trata de una palabra anticuada. Los ejemplos escasean en el corpus, con tan solo una ocurrencia, a la que podíamos añadir 5 casos de *almofrez*, una variante anterior del siglo XV.

[...] y que soys más ordinaria / que en botica **almofariz**, (1580-a1627, Luis de Góngora y Argote, *Romances*)

Anafe tiene unas apariciones contadas, pero contantes, y parece ser hasta nuestros días palabra conocida por los hispanohablantes. En el corpus hemos encontrado 110 ocurrencias en 83 textos, 75 de estos en España. Hay un ejemplo medieval, 56 de los siglos XVI-XVII, y 36 del siglo XX. *Academia* (1770) es el primer diccionario que incluye este vocablo en su repertorio. Actualmente parece usarse más en América; de los ejemplos del siglo XX, 29 provienen de textos de América y 7 de España, el último ejemplo de España es de 1940; indicamos este y otro, reciente, de García Márquez. Los datos del CORDES XXI nos confirman la hipótesis de que es palabra viva en América, pero no en España. Así, de los 51 ejemplos encontrados, solamente dos son de España, mientras que el resto se reparte entre todas las zonas del español de América.

Es curioso el fenómeno emigratorio de las palabras. «**Anafe**» y «albóndiga» emprenden el camino por la costa africana del Mediterráneo hacia occidente, se asientan en España, son incorporadas al idioma castellano y, ya cristianizadas, embarcan para América, cruzan el charco y se desparrraman por los valles y cuencas de este continente. (1940, José Moreno Villa, *Cornucopia de México*)

Frente a esos dos aposentos, en el mismo corredor, estaba la cocina grande, con **anafes** primitivos de piedras calcinadas, y el gran horno de obra de la abuela, panadera y repostera de oficio, cuyos animalitos de caramelo saturaban el amanecer con su aroma succulento. (2002, Gabriel García Márquez, *Vivir para contarla*, Colombia)

Jarra ha tenido un uso constante y consistente a lo largo de las épocas, a partir de su introducción. Así, en el corpus CDH están presentes 1.821 casos en

741 documentos, 1.307 en textos de España. Hay testimonios copiosos desde los tiempos remotos, con 111 ejemplos medievales (a partir del siglo XIII), 212 Siglos de Oro, para llegar a 1.317 ejemplos del siglo XX. Pese a su uso constante, no aparece en Nebrija, ni en Covarrubias. Viene recogida solo en un diccionario del siglo XVII, Franciosini⁶ (1620), después sí en otros del XVIII y en *Autoridades* (1734).

Jofaina, aljofaina o ajofaina se encuentra en 243 ejemplos de 148 documentos, 164 de los cuales en España. Aparece en fecha muy tardía en el corpus, tenemos un ejemplo del siglo XVIII, 48 del siglo XIX y 194 del siglo XX. *Autoridades* (1726) es el primer diccionario que recoge esta palabra; en la entrada *aljofaina* podemos leer: “otros dicen jofaina y jufaina”. En *Autoridades* (1734), en la entrada *jofaina* se nos remite a *aljofaina*: “lo mismo que aljofaina”. En Terreros y Pando (1786)⁷ podemos leer: “aljfana, o ajufaina, jufaina, jofaina, almoffa, vaso que sirve comúnmente para lavarse. El nombre más común en Madrid es jofaina.”.

Redoma se encuentra en 1.010 casos en 413 documentos, 983 de estos en España. Hay 149 ejemplos medievales, 508 de los Siglos de Oro, y 177 del siglo XX. Se encuentra en todos los diccionarios antiguos: Nebrija (1495), Alcalá (1505)⁸, Nebrija (1516), Casas (1570), Percival (1591), Palet (1604), Oudin (1607), Vittori (1609), Rosal (1611), Covarrubias (1611) y *Autoridades* (1737).

Taza es, sin duda, la palabra más común en el corpus. Hay 9.156 ocurrencias en 1.727, 3.902 de estos en España, lo que la hace una palabra usual en todo el ámbito hispánico. Aparece en los textos a partir del siglo XIII, y hasta finales del siglo XV registramos 240 ocurrencias, luego su uso aumenta durante los siglos XVI-XVII con 968 ejemplos y se dispara en la época moderna (siglo XX) con 7.184 ejemplos.

El primer diccionario que recoge esta voz es Rosal (1611), luego aparece definida en Covarrubias (1611) y más tarde en *Autoridades* (1739).

4. Arabismos desaparecidos o sustituidos por otras palabras

Albarrada con el significado que nos interesa solo aparecerá en el *Glosario etimológico de las palabras españolas* de Eguilaz⁹, después en DRAE 1914 y a partir de ahí será incluida en todos los diccionarios académicos. Sin

⁶ Lorenzo Franciosini Florentín, *Vocabolario español-italiano, ahora nuevamente sacado a luz [...] Segunda parte.*

⁷ Esteban de Terreros y Pando, *Diccionario castellano con las voces de ciencias y artes y sus correspondientes en las tres lenguas francesa, latina e italiana.*

⁸ Fray Pedro de Alcalá, *Vocabulista arábigo en letra castellana. En Arte para ligeramente saber la lengua arábigo.*

⁹ Francisco López Tamarid, *Compendio de algunos vocablos arábigos introducidos en lengua castellana en alguna manera corruptos, de que comúnmente usamos.*

embargo, no hay rastros en los documentos, lo que nos permite considerar esta palabra como completamente fuera de uso, si lo ha tenido alguna vez.

Alcolla será recogida en *Autoridades* (1726), donde lleva ya la mención *anticuada*. En el corpus no se ha encontrado otra ocurrencia salvo la indicada arriba. Del DH (1960-1996) se ha sacado una segunda ocurrencia, donde se explica el significado de *alcolla* (“son ampollas grandes”), lo que sugiere su uso escaso.

Y embiaron al Rey vn presente para cenar: cient panes de trigo bueno y dos **alcollas** de vino (son ampollas grandes), vuas passas y higos, con diez gallinas. (1550, Pere Antoni Beuter, *Corónica General. 2a parte*)

Aljáfana no tiene mucha presencia en el corpus, con tan solo 5 ejemplos en un único documento del siglo XVI, un libro de recetas de cocina. Por lo que se refiere a su presencia en los diccionarios, la encontramos en Vittori (1609) o Minsheu (1617); Covarrubias no la recoge, pero sí *Autoridades* (1770) donde se menciona que es anticuada y se remite al sinónimo *aljofayna*. Tanto *aljáfana* como su variante *aljébana* no resisten a la competencia de *jofaina* / *aljofaina*. Indicamos abajo un ejemplo del libro de recetas de Ruperto de Nola, donde podemos ver que *aljáfana* es un tipo de plato hondo.

Muy buenos higos pasados tomarás, bien melados; y allanarlos bien uno a uno; y quitarles lo duro de los pezones; y tomar una **aljafana** o plato hondo que sean nuevos y muy limpios, y pon al suelo de la **aljafana** un lecho de rosas coloradas, quitando el blanco de ellas con unas tijeras, y sobre las rosas un poco de azúcar y después un lecho de los higos y desta manera haciendo un lecho de las rosas y azúcar, y otro de los higos. (1529, Anónimo, *Libro de guisados de Ruperto de Nola*)

Aljébana por su parte aparece solo en diccionarios más modernos, como el *Academia Histórico* (1933), donde se remite a *jofaina*. No se han encontrado ejemplos del corpus, el único encontrado es aquel sacado del DH (1960-1996), indicado en el apartado anterior. Todos los diccionarios que recogen esta voz la sitúan en Murcia.

Almofía es otra voz de poco uso, con 27 casos en 16 textos, todos de España. 5 son ejemplos medievales, 21 de los siglos XVI-XVII, y uno del siglo XX. Aparece recogido en varios diccionarios de los siglos XVI y XVII: Percival (1591), Palet (1604), Oudin (1607), Vittori (1609), Rosal (1611), Covarrubias (1611), *Autoridades* (1726). Reproducimos aquí el único ejemplo del siglo XX, que proviene de un libro de alfarería, donde se habla de las piezas típicas de la alfarería tradicional.

Piezas típicas eran la pañiella (medida para la sidra) y la **almofía** (puchero para cocer castañas). El vidriado lo preparaban las mujeres

mezclando plomo y estaño con arena en pequeños hornos. (1997, Natacha Seseña, *Cacharrería popular. La alfarería de basto en España*)

Altamía aparece en casos muy contados, 14 ocurrencias en 8 textos, todos en España. Es una palabra que se vuelve anticuada rápidamente, pues encontramos 13 ejemplos medievales y uno del Renacimiento. Rosal (1611), un diccionario etimológico, solo indica que se trata de un arabismo; luego *Autoridades* (1726) lo marca como anticuado. Es una voz que no sobrevive a la Edad Media. Seis de los ejemplos del corpus proceden de un libro de medicina, donde podemos ver algunos usos mágico-religiosos de la altamía.

Otrosi puede sanar la mordedura del can ravioso con estas santas palabras escritas en una **altamía** blanca. Dagelas a beber con agua bendita IX dias, uno en pos de otro; e mandale que diga a onor de las IX ordenes del cielo LX vezes el Pater Noster e XL vezes el Ave Maria; e por cierto sepas que sanara. (a1500, Fernando de Córdoba, *Suma de la flor de cirugía*)

Ataífor ha tenido poquísimo uso al parecer, si fuera a juzgar por los 5 casos en 5 textos, todos de España. Hay un ejemplo medieval, 2 de los Siglos de Oro, uno del XVII, y uno del siglo XX. Pese a esto, parece ser palabra conocida por los lexicógrafos, ya que aparece recogida en todos los diccionarios de la época: Nebrija (1495), Alcalá (1505), López Tamarid (1585), Percival (1591), Palet (1604), Oudin (1607), Vittori (1609), Covarrubias (1611), *Autoridades* (1726). Podemos ver abajo un ejemplo donde se nota el significado ‘mesa pequeña’.

Una caldera, pequeña y una sartén Tres orças viejas. Una mesilla como **ataífor**. Una tinaja de agua con su atapador de madera. (1568, Anónimo, «Secuestro de bienes» (*Inventarios de bienes moriscos del Reino de Granada*))

El significado ‘plato hondo’ aparece en una poesía del siglo XVI. Los poquísimos ejemplos del corpus no nos permiten saber cuál de los significados era más frecuente. Es de suponer que su asociación con la cocina árabe ha contribuido a su desaparición, después del ocaso de la influencia cultural árabe.

Aquesa vuestra gorrilla liçençado muy novel, me parece quesadilla, cobertera o escudilla, o suelo de algún pastel. Maridillo o caçoleta, rodajuela o **ataífor**, platelejo o salsereta, moldezillo de bragueta o conchilla de color. (c1540-1579, Sebastián de Horozco, *Cancionero*)

Gregorio Mayans, el autor de una historia del español del siglo XVIII, indica *ataífor* entre las palabras desusadas, y propone como sustituto *aparador*, lo que no coincide con los significados indicados por los diccionarios.

Hai total diferencia entre las palabras antiquadas i las no frecüentadas. Las antiquadas son aquellas que se dejaron de usar después que en su lugar se substituyeron otras, usando de éstas i no de aquéllas. Assí decimos por afruenta, afrenta; por afuciar, esperarzar; por al, otra cosa; por **ataifor**, **aparador**; por ayuso, abajo; [...]. (1737, Gregorio Mayans y Siscar, *Orígenes de la lengua española*)

Zaque aparece en 101 casos en 64 documentos, 95 de estos en textos españoles. Hay 4 ejemplos de la Edad Media, 38 de los siglos XVI-XVII y 32 del siglo XX, de los cuales podemos eliminar 5 nombres propios. Era palabra usual durante la Edad Media y los Siglos de Oro, pues aparece en todos los repertorios antiguos: Nebrija (1495), Alcalá (1505), López Tamarid (1585), Palet (1604), Oudin (1607), Vittori (1609), Rosal (1611), Covarrubias (1611), *Autoridades* (1739). Sin embargo, la palabra no parece resistir más allá del siglo XVII, dado que los ejemplos del siglo XX proceden de novela con ambiente histórico. Juan de Valdés la considera poco usada a mediados del siglo XVI.

Zaque lo mismo es que odre o cuero de vino, y a uno que sta borracho dezimos que sta «hecho un **zaque**». También he oído en la Mancha de Aragón llamar zaques a unos cueros hechos en cierta manera, con que sacan agua de los pozos; vocablo es que se usa poco; yo no lo uso jamás. (1535-1536, Juan de Valdés, *Diálogo de la lengua*)

Como sustituto podemos pensar en *odre*, palabra patrimonial muy presente en el corpus a partir el siglo XIII; la diferencia consta en el tamaño, al ser el zaque más pequeño.

5. Conclusiones

De todos los arabismos del español que se refieren a recipientes o mobiliario usados en cocina medieval, algunos han desaparecido. Con el declive de la influencia cultural árabe, salen del uso común muchos objetos asociados con los musulmanes. Es el caso de *albarrada*, *alcolla*, *aljáfana* (*aljébana*), *almofía*, *altamía*, *ataifor* y *zaque*.

Otros vocablos resultan desconocidos en el español europeo, como por ejemplo *anafe*, de uso casi exclusivo en el español de América.

Otras nueve palabras han sobrevivido y están en uso actualmente, unas más que otras; se trata de *acetre*, *alcarraza*, *alcuza*, *almirez*, *anafe*, *jarra*, *jofaina*, *redoma* y *taza*. Sin embargo, *acetre* se especializa como recipiente utilizado en la iglesia, mientras que *alcuza* puede designar hoy en día en algunos países de América un recipiente que puede llevar aceite y vinagre, no solo aceite.

Hay otros casos más especiales, cuando más de un arabismo cubre la misma esfera conceptual. *Almofía*, *aljáfana*, *aljébana*, *jofaina* se refieren a un tipo parecido de recipiente, y solamente ha sobrevivido *jofaina*. *Altamía* es un tipo de taza, y no consigue diferenciarse notablemente con respecto a *taza* y pierde la competencia, así como ocurrirá con *albarrada*, sustituida por *alcarraza*.

Por otra parte, el arabismo *zaque* será sustituido por *odre*, aunque hay una pequeña diferencia de significado.

Conviene destacar, como punto final, la dinámica diacrónica de este campo, que ha visto pérdidas, totales o parciales, conservaciones, pero también alguna sustitución.

REFERENCIAS BIBLIOGRÁFICAS

Corpus

CDH = Instituto de Investigación Rafael Lapesa de la Real Academia Española (2013): *Corpus del Nuevo diccionario histórico (CDH)* [on line].

<http://web.frl.es/CNDHE>. [Consulta: 10.05.2025].

CORPES XXI = REAL ACADEMIA ESPAÑOLA: Banco de datos (CORPESXXI) [on line]. *Corpus del Español del Siglo XXI (CORPES)*. <http://www.rae.es> [Consulta: 10.05.2025].

Diccionarios

DAAL = Corriente, Federico. 2008. *Dictionary of Arabic and Allied Loanwords*, Leiden: Koninklijke Brill NV.

DCECH = Corominas, Joan y Juan Antonio Pascual. 1980-1991. *Diccionario crítico etimológico castellano e hispánico*, 6 tomos. Madrid: Gredos.

DH = *Diccionario histórico de la lengua española* (1960-1996); disponible en línea <http://web.frl.es/DH.html> [Consulta: 10.05.2025].

DHLE = Real Academia Española (2013-): *Diccionario histórico de la lengua española (DHLE)* [en línea]. [Consulta: 10.05.2025].

DLE = *Diccionario de la Real Academia Española*, 23ª ed., (2014), Madrid: Gredos, [on line] <http://www.rae.es/recursos/diccionarios/drae> - [Consulta: 10.01.2021].

NTLLE = *Nuevo Tesoro Lexicográfico de la Lengua Española*. [online] <http://www.rae.es> - [Consulta: 10.05.2025].

Estudios

Corriente Córdoba, Federico. 1992. *Árabe andalusí y lenguas romances*. Madrid: Mapfre.

Corriente Córdoba, Federico. 2013. "El elemento árabe en la historia lingüística peninsular: actuación directa e indirecta. Los arabismos en los romances peninsulares (en especial, en castellano)". En *Historia de la lengua española*, Rafael Cano (coord.), 185-206. Barcelona: Ariel Letras.

- Dworkin, Steven N. 2012. *A history of the Spanish lexicon. A linguistic perspective*. Oxford: Oxford University Press.
- Dworkin, Steven N. 2013. "La transición léxica en el español bajomedieval". En *Historia de la lengua española*, Rafael Cano (coord.), 643-656. Barcelona: Ariel Letras.
- Enăchescu, Mihai. 2015. "El reemplazo de arabismos por latinismos: el caso de *albéitar*". En *ReCHERches. Culture et histoire dans l'espace roman*, no. 14, *Langue, grammaire et didactique en diachronie: domaine roman*, 59-68. Strasbourg: Presses Universitaires de Strasbourg.
- Enăchescu, Mihai. 2016. "Pérdida y reemplazo de arabismos en español: cargos militares y cargos públicos". En *Școala coșeriană clujeană. Contribuții*, vol. I, editado por Cornel Vilcu, BOJOGA, Eugenia Bojoga, y Oana Boc, 87-96. Cluj-Napoca: Presa Universitară Clujeană, pp..
- Enăchescu, Mihai. 2017a. "Pérdida y reemplazo de arabismos en español: metales y productos químicos". En *Probleme de filologie spaniolă și italiană*. ROȘCA, Angela Roșca y Radu Melniciuc (coord.), 61-67. Chișinău: CEP USM.
- Enăchescu, Mihai. 2017b. "Pérdida y reemplazo de arabismos en español: los nombres de minerales". *Studia romanística. Caminos del hispanismo. Lingüística*. 17, nº.1: 41-53.
- Enăchescu, Mihai. 2017c. "Pérdida y reemplazo de arabismos en español. Los nombres de oficios". *Studia Iberystyczne* 16. *Entre la lingüística y la didáctica*. 37-52.
- Enăchescu, Mihai. 2018. "Pérdida de arabismos en español: productos de belleza medievales". En *Pasión por el hispanismo III*, editado por Liana Hotařová, 55-70. Liberec: Universidad Técnica de Liberec/Liberec: Technická univerzita v Liberci.
- Enăchescu, Mihai. 2019a. "*Alcohol* – historia de una palabra viajera". *Quaestiones romanicae* VII/2: 105-113.
- Enăchescu, Mihai. 2019b. "Pérdida y reemplazo de arabismos en español: enfermedades y afecciones de la medicina medieval". *Colindancias* 10: 187-200.
- Enăchescu, Mihai. 2020. *Pérdida y sustitución de arabismos en español*, Szeged: Jate Press.
- Enăchescu, Mihai. 2021. "Continuidad y discontinuidad en la transmisión de las palabras patrimoniales competidas por arabismos: *oliva* y *aceituna*, *olio* y *aceite*, *olivo* y *aceituno*". *Studia UBB Philologia* LXVI, nº 4: 69-80.
- Enăchescu, Mihai. 2022. "Las palabras olvidadas: las denominaciones de la *alcachofa* en el castellano medieval". *Quaestiones Romanicae. Memorie – Uitare* IX/2: 461-469.
- Enăchescu, Mihai. 2023. "Pérdida y sustitución de arabismos en español: objetos de adorno". En *Perspectives de recherche en linguistique et philologie romanes*, editado por Dolores Corbella, Josefa Dorta, Rafael Padrón, 475-488. Strasbourg: Editions de Linguistique et de Philologie (EliPhi).
- Enăchescu, Mihai. 2024. "Pérdida y sustitución de arabismos en español: comidas, dulces e ingredientes de la cocina medieval". *Analele Universității din Craiova – Seria Științe filologice. Lingvistică* XLVI, nº 1-2: 74-89.
- Fasla, Dalila. 1998-1999. "Aportaciones terminológicas de la lengua árabe al español medieval, moderno y contemporáneo (notas para un análisis etnolingüístico y reflexiones críticas)". *RESLA* 13: 243-259.

- Giménez Eguibar, Patricia. 2010. "Algunas cuestiones sobre la pérdida de arabismos". *Romance Philology* 64: 185-196.
- Giménez Eguibar, Patricia. 2015. "Dos casos de sustituciones léxicas: los arabismos *alfayate* y *alfajeme*". En *Actas del IX Congreso Internacional de Historia de la Lengua Española (Cádiz, 2012)*, t. II, 1409-1423. Madrid: Iberoamericana Vervuert.
- Giménez Eguibar, Patricia. 2016a. "Ni contigo ni sin ti: el arabismo 'albéitar' en el léxico peninsular". En *Etimología e historia en el léxico del español: estudios ofrecidos a José Antonio Pascual (Magister bonus et sapiens)*, Mariano Quirós García, José Ramón Carriazo Ruiz, Emma Falque Rey, Marta Sánchez Orense, (coords.), 303-318. Madrid: Iberoamericana Vervuert.
- Giménez Eguibar, Patricia. 2016b. "Arabisms in the Spanish Lexicon of Trades: A Diachronic Perspective". En *Diachronic Applications in Hispanic Linguistics*, editado por Eva Núñez Méndez, 35-88. Cambridge: Cambridge Scholars Publishing.
- Lapesa, Rafael. 1986. *Historia de la lengua española*, 9ª edición. Madrid: Gredos.
- Maillo Salgado, Felipe. 1998. *Los arabismos del castellano en la Baja Edad Media*, 3ª edición. Salamanca: Ediciones Universidad de Salamanca.
- Neuvonen, Eero Kalervo. 1941. *Los arabismos del español en el siglo XIII*. Helsinki: Imprenta de la Sociedad de literatura finesa.
- Penny, Ralph 2006. *Gramática histórica del español*, 2ª edición actualizada. Barcelona: Ariel.
- Torrens Álvarez, Mª Jesús. 2007. *Evolución e historia de la lengua española*. Madrid: Arco Libros.

A CORPUS-BASED ANALYSIS OF THE EVENT-PERCEPTION CONSTRUCTIONS IN ROMANIAN

Mihaela GHEORGHE¹ 

Article history: Received 31 August 2025; Revised 2 November 2025; Accepted 7 November 2025;
Available online 12 December 2025; Available print 30 December 2025

©2025 Studia UBB Philologia. Published by Babeş-Bolyai University.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ABSTRACT. *A Corpus-Based Analysis of the Event-Perception Constructions in Romanian.* Consistent with cross-linguistic patterns, perception verbs in Romanian typically take two arguments: an Experiencer/Agent, and a Theme. The Theme argument can be expressed in three main syntactic forms: (i) as a determiner phrase (DP), (ii) as a finite or non-finite clause (CP), or (iii) as a restructured configuration. These options show different syntactic behaviours and constraints related to subject licensing, aspect, and the perception of events versus states. The paper presents a corpus-based analysis of Event-Perception Constructions (EPCs) in Romanian, focusing on the interaction between the type of perception verb (*a vedea* 'see', *a auzi* 'hear'), the semantic-aspectual class of the embedded predicate, and the choice of clausal complement (*că*-, *cum*-, or gerundive structures). Using data from the RoTenTen2021 corpus, the study examines EPCs with verbs that denote non-volitional or semi-volitional physiological and emotional events. The analysis shows that the distribution of complement types is influenced by both sensory salience (visual, auditory, or dual-modality events) and aspectual properties (atelic activities, semelfactives, punctual achievements). Gerunds are favored for processual construals, *cum*-clauses for manner or dynamic unfolding, and *că*-clauses for factual perception. Verbs with balanced perceptual profiles (e.g., *a plânge* 'cry', *a râde* 'laugh') occur freely with all complement types across modalities, while those with a strongly marked sensory channel (e.g., *a sforăi* 'snore', *a roşi* 'blush') display more restricted patterns.

Keywords: *Romanian language, perception verbs, event-perception construction, pseudo-relative clause, corpus linguistics*

¹ **Mihaela GHEORGHE** is professor of linguistics at Transilvania University in Braşov, at the Department of Theoretical and Applied Linguistics, and a senior researcher at the Institute of Linguistics „Iorgu Iordan – Alexandru Rosetti” of the Romanian Academy. Romanian syntax, mainly the syntax of relative clauses, represents the area of her research interest. Email: m.gheorghe@unitbv.ro

REZUMAT. *O analiză de corpus a construcțiilor de percepție evenimentială în limba română.* În concordanță cu modelele din alte limbi, și în limba română verbele de percepție au de obicei două argumente: Experimentator/Agent și Temă. Argumentul tematic poate fi exprimat în trei forme sintactice primare: (i) ca grup al determinantului (DP), (ii) ca propoziție subordonată (CP) sau (iii) ca structură sintactică reorganizată. Aceste opțiuni au comportamente sintactice și constrângeri diferite legate de autorizarea subiectului, de aspect și de tipul de percepție a obiectelor, a evenimentelor sau a stărilor. Articolul de față prezintă o analiză bazată pe corpus a construcțiilor de percepție a evenimentelor (EPC) în limba română, concentrându-se pe interacțiunea dintre tipul verbului de percepție (*a vedea, a auzi*), clasa semantico-aspectuală a predicatului din subordonată și alegerea complementului propozițional (structuri introduse prin complementizatorii *că-*, *cum-* sau construcții gerunziale). Folosind date din corpusul RoTenTen2021, studiul examinează EPC-urile cu verbe care denotă evenimente fiziologice și emoționale nonvolitive sau semivolitive. Analiza arată că distribuția tipurilor de compliniri este influențată atât de proeminența senzorială (evenimente vizuale, auditive sau cu modalitate duală), cât și de proprietățile aspectuale (activități atelice, semelfactive, realizări punctuale). Gerunziile sunt favorizate pentru construcțiile procesuale, subordonatele introduse prin *cum* pentru desfășurarea dinamică, iar subordonatele cu *că* pentru percepția factuală. Verbe cu profiluri perceptive echilibrate (de exemplu, *a plânge, a râde*) apar fără restricții cu toate tipurile de complementare, pe când cele cu un canal senzorial puternic marcat (de exemplu, *a sforăi, a roși*) prezintă limitări contextuale.

Cuvinte-cheie: *limba română, verb de percepție, construcție de percepție a evenimentelor, pseudo-relativă, lingvistică de corpus*

1. Preliminaries

1.1. *Complementation of perception verbs.* In Romanian, as in many other languages, perception predicates select two arguments. The argument receiving the theta-role of the Experiencer or of the Agent² is a [+animate] determiner phrase (DP) that occupies the subject position. The second argument, which takes the syntactic role of the direct object (DO), can be realized either as a DP (1a) or as a clausal complement (CP) (1b-e). When realized as a clause, the complement is introduced by the prototypical complementizers *că* ('that'), *să* ('to' = *SĂ_{SUBJ}*), and *dacă* ('if'), or, in specific constructions, by the wh-word *cum* ('how') (GR 2013, 88; 226–227; Nicula 2013, 316–323).

² Perception verbs can be either non-agentive: *vedea* 'see', *auzi* 'listen', *simți* 'feel', or agentive: *asculta* 'listen', *privi* 'look', *zări* 'see', *a se uita* 'gaze', etc. For a complete list, see Nicula 2012, 36.

1. a. *Ana a văzut* [_{DP} *filmul*] *ieri*.
'Ana saw the movie yesterday.'
- b. *Ana a văzut* [_{CP} *că Ion dansa cu Maria*].
'Ana saw that Ion was dancing with Maria.'
- c. *Ana n-a văzut* [_{CP} *să fi dansat Ion cu Maria*]
'Ana didn't see Ion dancing with Maria.'
- d. *Ana n-a văzut* [_{CP} *dacă Ion a dansat cu Maria*].
'Ana didn't see if Ion danced with Maria.'
- e. *Ana l-a văzut pe Ion* [_{CP} *cum dansa cu Maria*].
'Ana saw Ion dancing with Maria.'

In terms of non-finite clause structures, Romanian is distinct from other Romance languages because it only employs a gerund³ form after perception verbs (2a). Spanish (Aldama García 2016; Casalicchio 2016), Catalan (Rafel 2000) and Portuguese (Mateus et al. 2003) also accept the gerund structure (2b-d), but they also use an *infinitive* non-finite clause (3a-c), while other Romance languages⁴ like French and Italian rely solely on the infinitive (3d,e)⁵.

- | | | | | | | | |
|----|----|-------|--------|-------|-------|-------------------|------------------|
| 2. | a. | Rom. | O | văd | pe | Maria | <i>cântând</i> . |
| | | | cl.ACC | I.see | DOM | Maria | sing.GER |
| | b. | Sp. | Veo | a | María | <i>cantando</i> . | |
| | | | I.see | DOM | Maria | sing.GER | |
| | c. | Cat. | Veig | la | Maria | <i>cantant</i> . | |
| | | | I.see | DOM | Maria | sing.GER | |
| | d. | Port. | Eu | vejo | Maria | <i>cantando</i> . | |
| | | | I | see | Maria | sing.GER | |

³ The use of the infinitive within gerund contexts is also attested in Old and Pre-modern Romanian, but it is merely accidental:

(i) *Văzut-am flăcăii scuturându-și pletele (...) și fruntea lor a se încreți fără vreme (...), florile de pe capul copiilor a se vesteji și poporul căutând în beție uitarea necazurilor* (Russo, S, 228 *apud* Diaconescu 1977, 135)

'I have seen the lads shaking their hair (...) and their foreheads wrinkling without time (...), the flowers on the children's heads fading, and the people seeking in drunkenness to forget their troubles'.

(ii) *S-au auzit în tabăra vrăjmașilor multe sunete de trâmbițe a face gălceavă* (VF)

'In the enemy's camp, many sounds of trumpets were heard making noise' (*apud* Nicula 2013, 323)

⁴ The choice between infinitives and gerunds is more intricate in Romance languages. For an in-depth analysis from a generative grammar point of view, refer to Casalicchio 2016; 2024. He examines the Rhaeto-Romance dialects of northern Italy, particularly Gardenese, demonstrating that the interaction between the parameters [+/- progressive] _{v_{perc}} and [+/-Agr] gerunds accounts for the variations related to the non-finite complements of perceptive verbs.

⁵ Similarly, in Germanic languages, perception verbs select an infinitive structure (English, German, Dutch), either exclusively (German, Dutch) or alongside a structure with the present gerund or participle (English) (Niculescu 2013, 68).

3. a. Sp. Veo a María *cantar*.
 I.see DOM Maria sing.INF
 b. Cat. Veig *cantar* la Maria.
 I.see sing.INF DOM Maria
 c. Port. Eu vejo Maria (a)⁶ *cantar*.
 I see Maria (A) sing.INF
 d. Fr. Je vois Marie *chanter*.
 I see Marie sing.INF
 e. It. Vedo Maria *cantare*.
 I.see Maria sing.INF
 'I see Maria sing'

Regarding perception verbs, linguistic literature distinguishes between three types of perception: (i) non-epistemic/epistemically neutral/direct perception (4a,b), (ii) epistemic/epistemically positive/indirect perception (4c,d), and (iii) cognitive perception (4e). The specific type of perception is usually linked to particular patterns of complementation⁷, as many studies confirm (Alboiu & Hill 2013; Aldama García 2016; Avram 2003; Barwise 1981; Dik & Hengeveld 1991; Felser 1999; Guasti 1993; Nicula 2012; 2013; Usonienė 2001, among others).

4. a. O văd pe Maria *plângând*.
 cl.ACC I.see DOM Maria cry.GER
 b. O văd pe Maria *cum plânge*.
 cl.ACC I.see DOM Maria how cry.3SG
 c. O văd pe Maria *că plânge*.
 cl.ACC I.see DOM Maria that cry.3SG
 'I see Maria crying'
 d. *Văd că Maria a plâns*.
 'I see that Maria has been crying'
 e. Nu o văd pe Maria *angajându-se curând*.
 'I don't see Maria being hired anytime soon'

⁶ Alongside the bare infinitive employed as complement of a perception verb, European Portuguese also has a *Prepositional Infinitival Construction*, which consists of a DP, a P, and an infinitival verb that can be inflected for person and number (Rafel 2000, 78).

⁷ Using the semantic concept of evidentiality, Alboiu & Hill 2013, 275 argue that: "(i) a direct evidential reading arises when the perception verb selects a reduced CP complement; whereas (ii) an indirect evidential reading (inferential, evaluative) arises when the verb selects a full-fledged CP, the subject of which moves to the matrix clause (Raising-to-Object – RtoO)." The distinction largely aligns with the types of perception discussed here, in the sense that a broad semantic distinction is made between *first-hand evidentiality* (which indicates perceptual access to information) and *non-first-hand evidentiality*, encompassing both external sources (input from another person) and internal sources (inferential processes), see DTG 2019, s.v. *evidențialitate*.

In terms of their grammatical features, most perception verbs in Romanian are transitive, such as *a vedea* 'see', *a privi* 'look', *a zări* 'spot', *a auzi* 'hear', *a asculta* 'listen', and *a simți* 'feel'. The exceptions are *a se uita* 'look' (5a) and *a se holba* 'stare' (5b), which select a prepositional object, and exhibit restrictions in their complementation (5c-d).

5. a. Ion se uită la Maria *cum plânge*.
'Ion is watching Maria crying'
- b. Ion se holbează la Maria *cum plânge*.
'Ion stares at Maria as she cries'
- c. Ion_i se uită la Maria_j
Ion cl.REFL.3SG watch.3SG at Maria
pro_{i/*j} *plângând*.
cry.GER
'Ion is watching Maria as he cries'
- d. ?Ion_i se holbează la Maria_j
Ion cl.REFL.3SG stare.3SG at Maria
pro_{*i/*j} *plângând*.
cry.GER
'Ion stares at Maria as she cries'

1.2. Pseudo-relative clauses in Western Romance (PRs). There is extensive linguistic literature on a clausal argument of perception verbs that, in some Romance languages and dialects, exhibits the surface features of a relative clause: Italian (6a), French (6b), Spanish (6c), Catalan (6d), Galician (6e), and Calabrian (6f).

6. a. Ho visto Gianni *che correva*.
b. J'ai vu Jean *qui courait*.
c. He visto a Juan *que corria*.
d. He vist en Joan *que corria*.
e. Eu vin a Xoán *que corría*.
f. Vitti a Gianni *chi scappava*
have.I seen DOM John that ran.3SG.IMPERF
'I saw John running' (*apud* Rafel 2000: 68⁸)

⁸ The list of examples in Rafel 2000, 68 also includes Romanian (example (4g)): *Am văzut pe Ion că fugea*, which was glossed in the same way. I will comment on the inaccuracy of that reference in section 1.3. below.

Research in the field (Burzio 1986; Guasti 1993; Rafel 2000; Casalicchio 2016; Aldama García 2016; Graffi 2017; Cinque 2020, 244-275) has demonstrated that, despite their apparent similarity to relative constructions – mainly because of the overlapping of the complementizer *che/qui/que* with the homonymous *wh*-words – they are structurally different. These complements of perception verbs serve the discourse function of introducing a new entity within a given discourse and providing new information about that entity. The challenging aspect of understanding these constructions is that, although they are usually used to introduce an entity argument into discourse, they actually introduce an event argument. As a result, events are “disguised” as entities.

The syntactic features⁹ of PRs can be summarized as follows: (i) they have a CP structure; (ii) the initial DP is always coindexed with the subject in the subordinate clause; (iii) the initial DP can be a proper noun (a feature excluded for the antecedent of a restrictive relative clause); (iv) the verb in the matrix clause is a perception verb such as *to see*, *to hear*, *to feel*, or similar verbs that indicate a type of discovery: *to find*, *to meet*, *to discover*, *to surprise*, *to observe*, *to notice*, etc., through which the interlocutor is prompted to become aware of the presence of an entity or event; also, an interjection with a presentative meaning may appear as the predicate of the matrix clause; (v) the embedded clause is an event-type predicate; (vi) since it concerns immediate perception of an event, the temporal, modal, and aspectual features of the sentence complement depend on the corresponding features in the matrix.

For languages featuring *che/qui/que*, multiple interpretations have been proposed (see a review for (a) and (b) in Barron 2000, 12-14; Rafel 2000, 71-74; Alboiu & Hill 2013, 279): (a) the PR functions as a CP-type element with a DP raised to Spec, CP, creating a complex small clause; (b) the PR is a genuine relative clause that modifies the noun serving as the direct object of the perception verb; (c) the embedded clause is a relative clause without movement, involving a *wh*-element functioning as a secondary predicate or presentative relative clause. (Lambrecht 2000).

⁹ The theoretical framework used in the works from which I drew this synthesis is the current generative model, which divides sentence structure into three domains: CP (the functional domain of the complementizer, including complementizers, *wh*-elements, and other dislocated elements hosted by its left periphery), IP (the functional domain involving verbal inflection and pronominal clitics, responsible for, among other things, assigning the Nominative case), and vP (the lexical domain, responsible for generating the verb and arguments). Internal arguments are generated and marked casually within the lexical domain. Conversely, the external argument (the subject), although generated in the lexical domain, depends on the tense projection (TP), part of the IP domain, to be marked with the Nominative case (Chomsky 1981; see Cornilescu 2000; Stan 2005; Dragomirescu & Nicolae 2016 for Romanian). The prototypical PR can be schematically represented as in (i), adopted from Rafel 2000, 71:

(i) [_{CP} Jeani [_C qu(e) [_{IP} i [_{I'} [_{VP} pro i courait]]]]]

1.3. Are there PRs in Romanian? The presence of PRs in Romanian has received limited attention because there is no homonymy between a complementizer and a *wh*-word, thus lacking a syntactic parallel to Western Romance. However, since PRs in Romance languages correspond to gerund or infinitive structures and constructions with gerund complements of perception verbs have Romanian equivalents introduced by *cum/că*, this has led to discussions on the topic (Gheorghe 2004; 2015; SOR 2016, 468). I disagreed with Rafel (2000, 68), who grouped Romanian *că*-based constructions with PRs in Romance languages. I found the comparison forced because, in Romanian, the relative pronoun stems from Lat. *QUALEM*, not from Lat. *QUOD*, so the concept of homonymy in Romance languages doesn't apply. Nonetheless, I believe that the complement clauses of perception verbs that follow the PR model in Romance are more likely embedded clauses introduced by *cum*, especially since, in Old Romanian, the term was multifunctional, serving both as the complementizer *că* 'that' and the *wh*-word *cum* 'how' (SOR 2016, 468; GR 2013, 88). Following Alboiu & Hill 2013, I concluded that Romanian lacks PRs¹⁰ and that the only valid interpretation for the Romanian constructions headed by *cum* is that they are small clauses. The main difference between PRs in Romance languages and their Romanian equivalents is how the entity expressed by the DP is identified: in Romance PRs, the entity is cognitively inactive in discourse, introduced into the perception field of a perceiving subject through its action; in the Romanian equivalents, the entity is necessarily brought into the discourse universe and then connected with the event through a syntactic reorganization mechanism (i.e., small clauses).

2. A corpus-based analysis of EPCs in Romanian.

2.1. Methodology. In analyzing the data on EPCs in Romanian, I examined the collocations of the transitive perception verbs *a vedea* 'see' and *a auzi* 'hear' with complements that denote events: full CPs (headed by the complementizer *că*), defective CPs (headed by the complementizer *cum*), and non-finite clauses (gerunds).

I used the Romanian Web 2021 (RoTenTen2021¹¹) corpus, which I interrogated with Corpus Query Language (CQL) formulas that enable searching for complex grammatical or lexical patterns. The CQL formulas I employed are the following:

¹⁰ See Casalicchio 2016 for the same interpretation.

¹¹ The corpus is available through SketchEngine: <https://app.sketchengine.eu>.

- (i) [tag="P.*"] + [lemma="vedea/auzi/privi"] + [word="pe"] + [] + [word="că"];
- (ii) [tag="P.*"] + [lemma="vedea/auzi/privi"] + [word="pe"] + [] + [word="cum"];
- (iii) [tag="P.*"] + [lemma="holba/uita"] + [word="la"] + [] + [word="cum"];
- (iv) [tag="P.*"] + [lemma="vedea/auzi/privi"] + [word="pe"] + [] + [tag="V.*"].

The results I obtained through CQL (iii) were filtered manually to retain only instances with gerunds.

The corpus analysis did not focus on quantitative comparisons but aimed to identify as many specific usages of these parallel constructions as possible, highlighting restrictions and incompatibilities that might be overlooked when only working with constructed examples.

For some patterns, the query generated tens of thousands of examples, and I randomly sampled 200 examples for manual labeling based on the following criteria: (i) the category of the verb in the main clause (auditory perception vs. visual perception); (ii) the semantic-syntactic features of the verbs in the embedded clause according to aspect (telicity vs. atelicity) and type of event (process vs. state vs. result); (iii) the syntactic pattern (constructions with *că/cum*/gerund).

Due to time and space constraints, I have utilized only a portion of the EPC collection for the data analysis in section 2.2 below, limiting the discussion to a category of verbs that occur in the embedded clause of the perception verb or as a gerundial complement: verbs denoting non-volitional or semi-volitional physiological or emotional events. The corpus was filtered for verbs in the embedded clause, and the inventory that resulted includes: (i) psychological events: *căscă* 'yawn', *leșina* 'faint', *roși* 'blush', *sforăi* 'snore', *strănuta* 'sneeze', *tuși* 'cough', *transpira* 'sweat', *vomita* 'vomit'; (ii) emotional events: *ofta* 'sigh', *plânge* 'cry', *râde* 'laugh', *suspina* 'sob'.

2.2. The data analysis. In Romanian EPCs, the choice of the complement (*că-/cum-/gerund*) depends on the interaction between the type of perception verb and the aspectual class of the embedded predicate. The construction with *că* 'that' is the most flexible, accepting activities, achievements, and even states. *Cum* 'as' is more restrictive and is linked to process-oriented activities, especially in visual perception. The gerund works with atelic activities and is not used with stative verbs or punctual achievements.

In this section, I will present data from the corpus, and I will comment on the limits of ECP distribution in Romanian. The analysis of actual examples will confirm that the theoretical generalizations hold, but will also reveal that the semantics of the embedded verb can affect the distribution of the patterns.

2.2.1. Constructions with verbs denoting physiological events. This group includes verbs that describe bodily processes and reflexes with a clear physiological basis, such as *a căsca* ‘yawn,’ *a leșina* ‘faint,’ *a roși* ‘blush,’ *a sforăi* ‘snore,’ *a strănuta* ‘sneeze,’ *a tuși* ‘cough,’ *a transpira* ‘sweat,’ and *a vomita* ‘vomit.’ From an aspectual perspective, most of these verbs are atelic and durative, indicating ongoing activities or states (*a căsca*, *a sforăi*, *a transpira*), while *a tuși* ‘cough’ and *a strănuta* ‘sneeze’ are semelfactive in their basic reading but can also denote continuous atelic events when interpreted iteratively. The perceptual profile varies considerably across the set: some predicates are accessible to both *a vedea* ‘see’ and *a auzi* ‘hear’ (e.g., *a strănuta*, *a tuși*), while others are primarily auditory (*a sforăi*) or visual (*a roși*, *a transpira*), and a few have a strongly limited sensory channel (*a leșina*, *a vomita* - mostly visual attestations in the corpus). Regarding complementation, these verbs show the full range of options available in EPCs – gerund, *cum*-clauses, and *că*-clauses – though distribution patterns depend on both the sensory salience of the event and its aspectual profile. Gerunds usually encode a process-oriented view, *cum*-clauses emphasize manner or the unfolding of the event, and *că*-clauses present it as a factual perception. The following analyses demonstrate how these factors interact for each verb, based on corpus examples and constructed instances where relevant.

For *a căsca* ‘yawn’, only the gerund appears in the corpus, both with *a vedea* ‘see’ (7a) and *a auzi* ‘hear’ (7b), but constructions with complementizers *cum* (7c) and *că* (7d) are also perfectly possible. The variant with *că* could be omitted by speakers for reasons of euphony, but semantically and syntactically, there are no restrictions.

7.
 - a. Este contagios, respectiv caști atunci când *vezi* pe altcineva *căscând*. (libertateapentrufemei.ro)
‘It’s contagious, meaning you yawn when you see someone else yawning.’
 - b. Căscatul contagios se poate declanșa atunci când *auzim* pe cineva *căscând*. (carti-bune.ro)
‘Contagious yawning can be triggered when we hear someone yawning.’
 - c. L-au văzut *cum căsca*...
‘They saw him yawning...’
 - d. L-au văzut *că căsca*...
‘They saw him yawning...’

In case of the verbs *a leșina* ‘faint’ and *a roși* ‘blush’, the corpus rendered only examples with *a vedea* as a perception verb (8a,b) and (9a,b). The absence of contexts with *a auzi* ‘hear’ is understandable because this perception verb requires a detectable auditory input, and events like fainting or blushing do not have a clear “auditory profile”.

In terms of complementation, there is a contrast between the two verbs: *a leșina* appears only with gerunds (8a) and *cum*-clauses (8b), while *a roși* appears with gerunds (9a) and *că*-clauses (9b).

For *a leșina*, the absence of contexts with *că*-clauses is not surprising. After verbs of perception, they mainly occur when the complement introduces a finite clause representing a situation regarded as a confirmed fact, not necessarily perceived as happening in real-time. *A leșina* ‘faint’ is a punctual event, an achievement-type event, which is more naturally expressed through direct, event-based descriptions, hence the preference for *cum* (which indicates how the event unfolds from the observer’s perspective) and for the gerund (which emphasizes the ongoing process), and the oddness of an example like (8c). On the other hand, for the verb *a roși* ‘blush’, there are attested constructions with *că*, profiling a change of state outcome observable in its result (skin is red), but not with *cum*. Still, examples with *cum* can be acceptable (9c).

8.
 - a. Oare fiindcă îl văzuse pe Harry leșinând în tren și își spuse că nu era mare lucru de capul lui?
(literaturapetocuri.ro)
‘Was it because he had seen Harry fainting on the train and thought he was not very bright?’
 - b. Când ies din incinta tunsătoriei le văd pe dudu cum leșină prin stația de autobuz. (misterx.ro)
‘When I leave the barber shop, I see the ladies fainting at the bus stop.’
 - c. ??Le-am văzut că leșinau...
‘I saw them fainting’
9.
 - a. Ce să spun era înduioșător s-o vezi pe regină roșind și zâmbind. (docplayer.org)
‘It was touching to see the queen blushing and smiling.’
 - b. Dacă îi văd pe unii că roșesc, la ei se reped mai aprig. (historia.ro)
‘If they see some people blushing, they rush at them more fiercely.’
 - c. Îi văd pe unii cum roșesc...
‘I see some people blushing...’

The verb *sforăi* ‘snore’ naturally pairs with the verb *auzi* ‘hear’ because it is atelic and durative, making it suitable for imperfective views, mainly as an auditory event. The corpus supports this with both *că* (10a) and gerundive (10c) complements. *Că* matches the factual auditory perception reading (*aud că sforăie*), while the gerund naturally fits with processual auditory perception. *A vedea* ‘see’ is less central here but still possible in contexts where visual cues of snoring - such as an open mouth or body movement - are noticeable. The corpus only shows the gerund (10b), which describes the event as an ongoing visual process. *Cum* is completely absent as a complementizer here, probably because snoring usually doesn't require manner-focused description, as it is mainly identified by the sound rather than a visually dynamic unfolding that needs “manner” encoding. However, examples with *cum* are not excluded (see the constructed example under 10d).

10. a. O aude pe Ana sforăind ca un motor, în fundul
camerei. (natgeo.ro)
‘He hears Ana snoring like a motor at the back of the room.’
- b. Aruncând o privire o văzu pe Maşa sforăind în
continuare. (beautyboutiquesalon.ro)
‘Glancing over, he saw Masha still snoring.’
- c. (...) îl aud pe unul că sforăie, pe altul că tuşeşte (...) (taifasuri.ro)
‘(...) I hear one snoring, another coughing (...)’
- d. O aude cum sforăie.
‘He hears her snoring.’

Sneezing is a brief, periodic physiological event that is perceivable through both modalities. It is punctual (completed), but can occur in sequences, allowing for both perfective and imperfective interpretations. The verb *vedea* ‘see’ naturally applies to the visible bodily movement associated with sneezing; *auzi* ‘hear’ applies to the distinctive sound. This dual perceptibility explains why all three complement types are attested with the two perception verbs. The corpus data shows a balanced visual-auditory profile, due to the event’s strong cues in both modalities. *A vedea + a strănuta* are the options for factual *că*, while *cum* is used for a dynamic manner; There is no gerund with *vedea* because the process-oriented visual interpretation of the event is secondary to the auditory one. *A auzi + a strănuta* have the same interpretation, i.e., *că* for factual events, and *cum* for ongoing or repeated events, but this time with sounds. Example (10d) also has an exclamative reading, due to the syntactic configuration of the matrix (a surrogate imperative). In this context, *cum* has a stronger modal value than in other similar EPCs.

11.
 - a. Când *auzea* pe cineva *strănutând*, se întorcea, zâmbind (confluente.org)
 - b. Tot de politețe ține să spui "sănătate" după ce *auzi* pe cineva *că strănută*. (radiounirea.ro)
'It is also polite to say "bless you" after hearing someone sneeze.'
 - c. Când *vedeți* pe cineva *că strănută* și nu are mască, puteți suna la 112. (mediafax.ro)
'When you see someone sneezing and not wearing a mask, you can call 112.'
 - d. Câteva săptămâni parașutele le inspirăm noi, da' să-l *vezi* pe Ricci *cum strănută*! (digitalreviews.ro)
'We'll inspire the parachutes for a few weeks, but just wait until you see Ricci sneeze!'

A tuși 'cough' is a short, often repeated physiological event that can both be heard (primary) and seen (secondary, through bodily movements). Regarding aspect and (im)perfectivity, *a tuși* 'cough' is an atelic verb when repeated and punctual when isolated. *A vedea* 'see' naturally fits in contexts where the visual component is prominent (such as shoulder movement, hand to mouth, facial expression), while *a auzi* 'hear' emphasizes the distinctive sound. Concerning the complement type distribution, both *a vedea* 'see' and *a auzi* 'hear' select *că* (for factual construal of ongoing or habitual actions), as well as gerunds and *cum* for process-oriented auditory or visual depiction, emphasizing repeated or sustained coughing. There is no gerund attested for the auditory domain, but it is perfectly possible (12 e). The absence of *cum* in the visual domain suggests it is more natural with auditory than with visual perception (12f).

12.
 - a. Tot la televizor, îl *vezi* pe Arafat *tușind* în pumn. (kmkz.ro)
'Also on television, you see Arafat coughing into his fist.'
 - b. A două zi, călătorind împreună cu mașină, *l-am auzit* pe MB *cum tușea*. (dcmedical.ro)
'The next day, traveling together by car, I heard MB coughing.'
 - c. (...) îl *aud* pe unul *că sforăie*, pe altul *că tușește* (...) (taifasuri.ro)
'(...) I hear one snoring, another coughing (...)'
 - d. Să nu intrăm în magazine dacă *vedem* pe cineva *că tușește*, este febril sau transpiră. (libertatea.ro)

- 'Let's not go into stores if we see someone coughing, feverish, or sweating.'
- e. *Îl auzeam tușind* dincolo de perete.
'I could hear him coughing on the other side of the wall.'
- f. *?Îl văd cum tușește.*
'I see him coughing.'

A transpira 'sweat' is a non-auditory, gradual physiological process, making *a vedea* 'see' the only natural perception verb here. Auditory perception is irrelevant since there is no inherent sound associated with sweating. The complement type distribution shows that with *a vedea* 'see', all three complements are possible, even though the corpus only presented an example with the gerund (13a). *Că-* complements can be chosen for factual visual construal, related to an observed result (wet clothes, glistening skin) (13b), while gerund and *cum*-complements are used for process-oriented depiction, framing the observation in real time. *A transpira* 'sweat' is durative and atelic, so all three complements match the imperfective nature of the event.

13. a. *Când vedem pe cineva transpirând* putem fi siguri că îi este cald sau este tulburat (trotusaeauto.ro)
'When we see someone sweating, we can be sure that they are hot or upset.'
- b. *Îl văd că transpiră* instantaneu.
'I see him sweating instantly.'
- c. *Îl vedeam cum transpiră* din ce în ce mai mult.
'I could see him sweating more and more.'

A vomita 'vomit' has both visual and potential auditory components, but the corpus shows only visual attestations (14a). *A vomita* 'vomit' is telic, often punctual, but can have iterative phases, so both *că* and *cum* complements are possible in Romanian (see 14b,c). Gerund supports a processual construal even for telic events by stretching the moment of observation.

- 14 a. *M-ați văzut pe mine vomitând* când vorbesc românește? (ebihoreanul.ro)
'Have you seen me vomiting when I speak Romanian?'
- b. *L-a văzut că vomită* și s-a speriat.
'He saw him vomiting and got scared.'
- c. *Când l-a văzut cum vomită*, s-a speriat.
'When he saw him vomiting, he got scared.'

2.2.2. Constructions with verbs denoting emotional events. This group includes verbs whose lexical semantics denote emotional manifestations with a strong physiological component, such as *a plânge* ‘cry’, *a râde* ‘laugh’, *a suspina* ‘sob’, and *a ofta* ‘sigh’. These predicates are perceptually salient in both the visual and auditory domains to varying degrees, depending on the nature of the event: crying and laughing typically have a balanced dual profile, while sighing (*a suspina*, *a ofta*) tends to have a primary auditory component, with visual cues playing a secondary role. In terms of aspect, these verbs are atelic and readily compatible with imperfective construals, which allows them to occur with all three types of complement – gerund, *cum*-, and *că*-clauses – without strong grammatical restrictions. The corpus data show that the distribution across complement types often correlates with the sensory modality that is most salient for the event: gerunds are frequent for processual depictions, *cum* highlights manner or dynamic unfolding, and *că* encodes a factual perception. The analyses below illustrate how these tendencies play out for each verb, with examples covering both *a vedea* ‘see’ and *a auzi* ‘hear’.

A plânge ‘cry’ is a dual-modality physiological act, with clear auditory cues (sobbing, weeping) and strong visual cues (tears, facial expressions, posture). The corpus confirms this dual profile with examples of both *a vedea* ‘see’ (15a,d,f) and *a auzi* ‘hear’ (15b,c,e). All three complement types are found in both modalities: the gerund (*plângând*) describes the crying as an ongoing process; *cum*-clauses depict the manner or dynamic unfolding of the act; and *că*-clauses are used for factual descriptions, showing the crying as a confirmed state or event. Since *a plânge* ‘cry’ is atelic and compatible with imperfective construal, there are no aspectual restrictions on complement choice, and the full range of patterns is available to speakers.

15. a. O singură dată l-am văzut *plângând* (istorielocala.ro)
‘I saw him crying only once.’
- b. Ți se rupe sufletul când îl *auzi plângând*. (desprecopii.ro)
‘It breaks your heart when you hear him crying.’
- c. Mută-te în altă cameră pentru a nu-l mai *auzi* pe bebeluș
că plânge (qbebe.ro)
‘Move to another room so you can't hear the baby crying.’
- d. Nu-i nici o bucurie să-l *vezi* pe bebe *că plânge*.
(femeia.ro)
‘It's no fun to see your baby crying.’
- e. Îl *auzeam* pe tata *cum plânge*, pe mama pe fundal care
vorbea cu altcineva la telefon. (wordpress.ro)
‘I could hear my father crying, my mother in the background talking to someone else on the phone.’

- f. E teribil de dureros *să vezi* pe altul *cum plânge*, când tu nu mai ai nicio lacrimă. (wordpress.ro)
'It is terribly painful to see someone else crying when you have no tears left.'

A râde 'laugh' is also a dual-modality event, with prominent auditory cues (the sound of laughter) and equally salient visual cues (facial expressions, body movement). The corpus includes attestations for both *a vedea* 'see' (16b,d) and *a auzi* 'hear' (16a,c,e), covering all three complement types in each modality, except (16f), which is constructed, but acceptable. The gerund (*râzând*) provides a processual reading; *cum*-clauses highlight manner (e.g., *cum râde încetișor* 'how he laughs softly'); and *că*-clauses represent laughter as a factual perception. Since laughter is atelic and durative in repeated bouts, all complements are natural and semantically flexible.

16. a. Îl *aud* pe Viorel *râzând* în hohote. (ning.com)
'I hear Viorel laughing loudly.'
- b. Scriu comedie fiindcă îmi place *să-i văd* pe oameni *râzând*. (teatral.ro)
'I write comedy because I like to see people laughing.'
- c. Îmi crește inima când *aud* pe cineva *că râde* din cauza și în preajma mea. (nemira.ro)
'My heart swells when I hear someone laughing because of and around me.'
- d. Dacă te *vede* pe tine *că râzi*, va începe să râdă și el. (suntmamica.ro)
'If he sees you laughing, he will start laughing too.'
- e. Îl *auzi* pe străin *cum râde* încetișor. (editura.ro)
'You hear the stranger laughing softly.'
- f. Îl *vezi* *cum râde* încetișor.
'You see him laughing softly.'

A suspina 'sigh/soft sob' has mainly an auditory profile but is also visually noticeable through bodily cues such as chest movement and facial tension. The corpus includes both *a vedea* 'see' (17b) and *a auzi* 'hear' (17a,c) example, covering all complement types except the factive *că* and *cum* in the visual domain. The lack of *cum* in the context of *vedea* probably reflects the low visual salience of sighing unless sound is involved, though the construction is still possible. The *că*-complement characterizes the act as a factual perception, and probably contexts with *suspina* 'sob' are less frequent.

17.
 - a. Mi s-a părut că *aud* pe cineva *suspinând* sus, lângă marginea mormântului. (wordpress.ro)
'I thought I heard someone sobbing up by the edge of the grave.'
 - b. Vei sta tu nepăsător *să-i vezi* pe păcătoși *suspinând* în durerile morții. (rcrwebsite.com)
'Will you stand by indifferently and watch sinners groaning in the pains of death?'
 - c. În același timp o *auzeam* pe Monica *cum suspina* la telefon. (doxologia.ro)
'At the same time, I could hear Monica sobbing on the phone.'
 - d. O *aud* că *suspină*.
'I hear her sobbing.'
 - e. O *văd* că *suspină*.
'I hear her sobbing.'
 - f. O *văd* *cum suspină*.
'I hear her sobbing.'

A ofta 'sigh (long exhale)' shows a balanced auditory - visual profile, as the action produces both a perceivable sound and visible bodily movement. The corpus attests all three complement types, but not for both *a vedea* (18a) and *a auzi* (18b,c). However, the missing configurations are still possible; see the constructed examples (18d-f). Due to its atelic and imperfective nature, *a ofta* allows all complements without semantic restriction.

18.
 - a. Nu suport *să-i văd* pe bărbați *oftând*. (asa-si-asa.ro)
'I can't stand seeing men sighing.'
 - b. Îi *auzi* pe unii *că oftează*. (jurnalul.ro)
'You hear some people sighing.'
 - c. Zâmbetul îmi dispare atunci când îl *aud* pe Itachi *cum oftează*. (forumz.ro)
'My smile disappears when I hear Itachi sigh.'
 - d. Îi *aud* *oftând*.
'I hear them sighing.'
 - e. Îl *văd* *cum oftează*.
'I see him sighing.'
 - f. Îl *văd* *că oftează*.
'I see him sighing.'

3. Conclusions

This study provides a corpus-based analysis of Event-Perception Constructions (EPCs) in Romanian, examining how perception verb types (*a vedea* ‘see’ vs. *a auzi* ‘hear’), the semantic and aspectual features of the embedded predicate, and the choice of clausal complement (*că*-, *cum*-, or gerundive structures) interact. By analyzing a set of verbs representing non-volitional or semi-volitional physiological and emotional events, several conclusions emerge. First, the distribution of complements is strongly influenced by sensory salience. Events with a balanced dual profile (e.g., *a plânge*, *a râde*) easily combine with all complement types across modalities. Verbs with a primary auditory profile (*a sforăi*, *a ofta*) or a primary visual profile (*a roși*, *a transpira*, *a leșina*, *a vomita*) tend to have more restricted patterns, often lacking certain complements in the less salient modality. Second, aspectual class interacts with the choice of complement. Atelic, durative verbs favor gerunds for processual interpretations, whereas punctual achievements (*a leșina*) are more naturally expressed with *cum*- clauses or gerunds, which depict the unfolding of the event. Semelfactives (*a strănuta*, *a tuși*) can resemble activities when interpreted iteratively, broadening their compatibility with all complement types. Third, the different complementizers serve distinct functions: (i) *că*-clauses present the embedded event as a factual perception, with fewer restrictions on aspect but some sensitivity to modality; (ii) *cum*-clauses encode manner or dynamic unfolding, being more common in visual EPCs but also appearing in the auditory domain when the manner of sound production is salient; (iii) gerundives consistently indicate processual readings and are the most flexible across modalities. Finally, the analysis supports the idea that Romanian *cum*-clauses in EPCs should be viewed as evidential CPs with an eventive interpretation rather than as pseudo-relatives. Corpus evidence shows they pattern closely with gerundives in their event-focused sense, with an added emphasis on the manner and unfolding of the perceived event. This differs from Western Romance PRs, which have different syntactic triggers and discourse functions. Future research could expand the dataset to include quantitative frequency comparisons, explore EPCs with volitional action verbs for contrast, and study diachronic changes in the use of *că*-, *cum*-, and gerundive complements in Romanian.

WORKS CITED

- Alboiu, Gabriela, and Virginia Hill. 2013. "On Romanian Perception Verbs and Evidential Syntax." *Revue roumaine de linguistique* LVIII (3): 275–298.
- Aldama García, Nuria. 2016. *Pseudo-relatives Complement of Perception Predicates*. M.A. Diss., Universidad del País Vasco. <https://addi.ehu.es/handle/10810/21499>.
- Avram, Larisa. 2003. "An Aspectual Analysis of Gerunds." *Revue roumaine de linguistique* 48 (1–4): 203–219.
- Barron, J. 2000. "The Morphosyntactic Correlates of Finiteness." In *Proceedings of the LFG00 Conference*, edited by M. Butt and T. Holloway King, 1–19. CSLI Publications, Berkeley. <http://csli-publications.stanford.edu> (accessed 06.05.2011).
- Barwise, Jon. 1981. "Scenes and Other Situations." *The Journal of Philosophy* 78: 369–397.
- Burzio, Luigi. 1986. *Italian Syntax: A Government-Binding Approach*. Dordrecht: D. Reidel Publishing Company.
- Casalicchio, Jan, and Peter Herbeck. 2024. "Pseudorelative Clauses, Infinitives, and Gerunds with Spanish Perception Verbs: A Comparative View." *Syntax*, 1–34. <https://doi.org/10.1111/synt.12271>.
- Casalicchio, Jan. 2016. "The Use of Gerunds and Infinitives in Perceptive Constructions: The Effects of a Threefold Parametric Variation in Some Romance Varieties." In *Theoretical Approaches to Linguistic Variation*, edited by Ermenegildo Bidese, Federica Cognola, and Manuela Caterina Moroni, 53–88. Amsterdam: John Benjamins. <https://doi.org/10.1075/la.234.03cas>.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Cinque, Guglielmo. 2020. *The Syntax of Relative Clauses: A Unified Approach*. Cambridge: Cambridge University Press.
- Cornilescu, Alexandra. 2000. "The Double Subject Construction in Romanian." In *Comparative Studies in Romanian Syntax*, edited by Virginia Motapanyane, 83–133. Amsterdam / Lausanne / New York / Oxford / Shannon / Singapore / Tokyo: Elsevier.
- Diaconescu, Ion. 1977. *Infinitivul în limba română*. București: Editura Științifică și Enciclopedică.
- Dik, Simon C., and Kees Hengeveld. 1991. "The Hierarchical Structure of the Clause and the Typology of Perception-Verb Complements." *Linguistics* 29: 231–259.
- Dragomirescu, Adina, and Alexandru Nicolae. 2016. "Originea formelor verbale «nonfinite» cu flexiune: infinitiv vs. supin." *Diacronia* 4 (1 August): art. A54, 1–16. <https://doi.org/10.17684/i4A54ro>.
- DTG. 2019. *Dicționar de termeni gramaticali*. Gabriela Pană Dindelegan (coord.), Adnana Boioc Apintei, Adina Dragomirescu, Alexandru Nicolae, and Rodica Zafiu. București: Editura Univers Enciclopedic Gold.
- Felser, Claudia. 1999. *Verbal Complement Clauses: A Minimalist Study of Direct Perception Constructions*. Amsterdam: John Benjamins.
- Gheorghe, Mihaela. 2004. *Propoziția relativă*. Pitești: Paralela 45.
- Gheorghe, Mihaela. 2015. "Notă sintactică despre construcțiile pseudo-relative în limba română." In Ionuț Pomian (coord.) and Nicolae Mocanu (ed.), *Înspre și dinspre Cluj. Contribuții lingvistice. Omagiu profesorului G.G. Neamțu la 70 de ani*, Cluj-Napoca: Scriptor Argonaut.

- GR. 2013. *The Grammar of Romanian*. Edited by Gabriela Pană-Dindelegan. Oxford: Oxford University Press.
- Graffi, Giorgio. 2017. "What Are 'Pseudo-Relatives'?" In Roberta D'Alessandro, Gabriele Iannàccaro, Diana Passino, and Anna M. Thornton (eds.), *Di tutti i colori. Studi linguistici per Maria Grossmann*, 115–132. Utrecht: Utrecht University Repository.
- Guasti, Maria Teresa. 1993. *Causative and Perception Verbs: A Comparative Study*. Torino: Rosenberg & Sellier.
- Lambrecht, Knud. 2000. "Prédication seconde et structure informationnelle: La relative de perception comme construction présentative." *Langue française* 127: 49–66.
- Mateus, Maria Helena, Ana Maria Brito, Inês Duarte, Isabel Hub Faria, et al. 2003. *Gramática da Língua Portuguesa*. Lisboa: Editorial Caminho.
- Nicula, Irina. 2012. *Modalități de exprimare a percepțiilor fizice. Verbele de percepție în limba română*. București: Editura Universității din București.
- Nicula, Irina. 2013. "The Romanian Verbs of Perception *vedea* and *auzi*: Between Direct and Indirect Perception." *Revue roumaine de linguistique* LVIII (3): 313–327.
- Niculescu, Dana. 2013. *Particularități sintactice ale limbii române din perspectivă tipologică. Gerunziul*. București: Editura Muzeului Național al Literaturii Române.
- Rafel, Joan. 2000. "From Complementizer to Preposition: Evidence from Romance." *Probus* 12 (1): 67–91.
- Stan, Camelia. 2005. *Categoria cazului*. București: Editura Universității din București.
- SOR. 2016. *Syntax of Old Romanian*. Edited by Gabriela Pană Dindelegan. Oxford: Oxford University Press.
- Usonienė, Aurelija. 2001. "On Direct/Indirect Perception with Verbs of Seeing and Seeming in English and Lithuanian." *Working Papers* 48. Lund: Lund University, Department of Linguistics: 163–181.

TRABAJO Y TRAVAIL: EVOLUCIÓN SEMÁNTICO-DISCURSIVA EN ESPAÑOL Y FRANCÉS

Timea TOCALACHIS¹ 

Article history: Received 16 June 2025; Revised 30 September 2025; Accepted 1 October 2025;
Available online 12 December 2025; Available print 30 December 2025

©2025 Studia UBB Philologia. Published by Babeş-Bolyai University.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ABSTRACT. *Trabajo and Travail: A Semantic-Discursive Evolution in Spanish and French.* This article aims to analyze the semantic and etymological evolution of the word *trabajo* in the Spanish language, from its origins in Vulgar Latin to its contemporary usage, by establishing a comparison with the French term *travail*. The goal is to understand how both terms, initially associated with suffering, transformed their meanings to encompass the concept of productive activity and human effort in various facets of society, particularly in economic and social contexts. The methodology combines a diachronic and comparative approach. Through the analysis of representative historical texts obtained from sources such as the *Corpus Diacrónico del Español (CORDE)*, the *Diccionario Histórico de la Lengua Española (DHLE)*, the French *FRANTEXT* corpus and the *Corpus de Français Médiéval (CFM)*, the first occurrences of *trabajo* and *travail* are traced. The semantic transformations of both words are examined, focusing on their evolution from the Vulgar Latin *tripalium*, an instrument of torture, to their resemanticization across different historical periods. The methodology also includes an analysis of the metaphorical load of both terms in various historical and cultural contexts, considering how their usage has varied according to socioeconomic realities and the prevailing values of each period. In sum, this article seeks to offer a deeper understanding of how language reflects social,

¹ **Timea TOCALACHIS** es profesora titular en la Universidad Babeş-Bolyai de Cluj-Napoca, la Facultad de Ciencias Económicas y Gestión de los Negocios, Departamento de Lenguas Modernas y Comunicación en los Negocios. Enseña el español aplicado a los negocios, lenguaje especializado aplicado, comunicación en los negocios. Sus áreas de interés incluyen la enseñanza del español para los negocios, comunicación intercultural en los negocios, análisis del discurso, lexicografía, lexicología. Es coautora para la parte española del volumen *Mic dicționar englez-german-francez-italian-spaniol-român de termeni economici, juridici și politici* (*Breve diccionario inglés-alemán-francés-italiano-español-rumano de términos económicos, jurídicos y políticos*), por la editorial Casa Cărții de Știință de Cluj-Napoca. Correo electrónico: timea.tocalachis@econ.ubbcluj.ro

economic, and cultural dynamics over time. The evolution of *trabajo* and *travail* not only illustrates linguistic change, but also a transformation in how societies value human effort, offering a comprehensive perspective on the interaction between language, culture, and history in the Hispanic and Francophone worlds.

Keywords: *semantic change, trabajo/travail, prototypical semantics, diachronic corpus, lexical conceptualization*

REZUMAT. *Trabajo și travail: evoluție semantico-discursivă în spaniolă și franceză.* Acest articol își propune să investigheze evoluția semantică și etimologică a termenului *trabajo* din limba spaniolă, de la originile sale în latina vulgară până la accepțiunile sale contemporane, prin intermediul unei comparații sistematice cu termenul francez *travail*. Obiectivul principal constă în elucidarea modului în care ambii termeni, inițial ancorați în sfera suferinței și constrângerii, au cunoscut o transformare semantică profundă, ajungând să desemneze activitatea productivă și efortul uman în diverse domenii ale vieții sociale, în special în contexte economice. Metodologia adoptată îmbină o abordare diacronică cu una comparativă. Analiza se bazează pe corpusuri istorice relevante, precum *Corpus Diacrónico del Español (CORDE)*, *Diccionario Histórico de la Lengua Española (DHLE)*, *Frantext* și *Corpus de Français Médiéval (CFM)*, urmărind atestările timpurii ale termenilor *trabajo* și *travail*. Transformările semantice sunt interpretate în raport cu originea lor comună, *tripalium* — instrument de tortură din latina vulgară — și cu procesele succesive de resemantizare. De asemenea, este analizată încărcătura metaforică a acestor termeni în contexte istorice și culturale variate, evidențiindu-se influența determinantă a factorilor socioeconomiici și a valorilor dominante. În concluzie, studiul oferă o perspectivă comprehensivă asupra modului în care limbajul reflectă și încorporează transformările sociale, economice și culturale. Evoluția lexicală a termenilor *trabajo* și *travail* nu doar ilustrează mecanismele schimbării lingvistice, ci și metamorfozarea concepției colective privind valoarea și sensul muncii în spațiile hispanofon și francophon.

Cuvinte-cheie: *schimbare semantică, trabajo/travail, semantică prototipică, corpus diacronic, conceptualizare lexicală*

1. Introducción

El léxico de una lengua no es solo un inventario funcional de palabras para nombrar el mundo; es también un espejo dinámico de las concepciones culturales, ideológicas y sociales de una comunidad. En este sentido, el estudio de la evolución semántica de términos aparentemente comunes, como *trabajo* en español y *travail* en francés, ofrece valiosas pistas sobre cómo las sociedades han concebido y valorado la actividad humana a lo largo de la historia.

Ambos términos derivan del latín vulgar *tripalium*, un instrumento de tortura conformado por tres palos que servía para inmovilizar y castigar. Esta conexión etimológica con el sufrimiento físico no es trivial: revela una concepción antigua del trabajo como una forma de imposición, castigo o dolor. Aunque muchas investigaciones etimológicas tradicionales se han enfocado en los orígenes formales o morfológicos de los términos *trabajo* y *travail*, estas suelen privilegiar una perspectiva diacrónica lineal sin considerar su evolución semántica como fenómeno discursivo y cognitivo y a menudo dejan en segundo plano su dimensión ideológica. Entre estas obras destacan el *Diccionario crítico etimológico castellano e hispánico* de Corominas (1980–1991, vol. V, 152 s.), que vincula *trabajo* con el latín *tripalium*, y el *Diccionario Etimológico Romance* (DÉRom) de Buchi y Schweickard (2008, s.v. *TRIPALIUM*), que profundiza en su evolución dentro del latín romance. En el ámbito francófono, los léxicos canónicos como los de Littré o Bréal también respaldan la derivación desde *tripalium*, aunque críticos como Franck Lebas (2023, 22–23) han sugerido raíces alternativas en *trabs* o *trabacula*, conectadas con estructuras madereras en obras arquitectónicas medievales; y Delport (1984, 122–124) propuso una etimología basada en el prefijo latino *trans-* (‘atravesar’), lo que evidencia la complejidad y polisemia de una historia léxica que supera la mera forma.

La lingüística histórica tradicional, basada en el enfoque diacrónico clásico—como el de Ullmann (1962) o Aitchison (1981)—tiende a clasificar los tipos de cambio semántico en categorías como ampliación, restricción o peyoración, pero sin siempre indagar en las implicaciones sociales o cognitivas de estos procesos. Estudios enciclopédicos, como el *Oxford Research Encyclopedia of Linguistics*, reconocen estos cambios en el significado como fenómenos nucleares de la disciplina, pero sugieren que el modelo tradicional ha sido renovado por enfoques cognitivos y funcionales.

La presente investigación propone estudiar la evolución semántica y discursiva de *trabajo* y *travail* desde la Edad Media hasta el siglo XXI, aplicando como marco teórico el modelo de semántica prototípica de Dirk Geeraerts (1997, 17–22). Este modelo permite entender la estructura del significado léxico como una red radial, con un núcleo más representativo y periferias donde habitan significados extendidos, metafóricos o socialmente marcados. La hipótesis de partida es que estos términos funcionan como “nudos semánticos” (65), en los que se condensan tanto significados heredados como nuevas actualizaciones discursivas.

Complementariamente, recurrimos al enfoque funcional-discursivo de Blank (1997), que reconoce al discurso y a los sistemas ideológicos como fuerzas activas en la reconfiguración del significado. Esta combinación teórica permite

comprender no solo qué cambios semánticos ocurrieron, sino también las causas de estos y la modalidad en la que respondieron a transformaciones sociales, culturales y económicas concretas en cada período histórico. Según Blank (1997, 28–30), los cambios semánticos obedecen tanto a necesidades comunicativas internas como a presiones ideológicas externas, a menudo invisibilizadas en las fuentes léxicas tradicionales. El objetivo no es solo describir el cambio de significado, sino también valorar el papel del léxico como archivo de las ideologías del trabajo, en línea con lo propuesto por Crespo (2007, 86–88) y Carpenter (2021, 23–25). Como destaca Jerome Baghana (2016, 95), el discurso sobre el trabajo en textos históricos revela estructuras de dominación simbólica, desde las obligaciones feudales hasta el ethos neoliberal. De ahí que el análisis semántico se complemente con una lectura crítica de los usos discursivos, atendiendo al contexto social, político y económico en el que cada acepción se activa. No obstante, aunque los enfoques clásicos (Corominas, Ullmann, Aitchison) han proporcionado un marco sólido para la descripción etimológica y semántica, y los estudios más recientes (Crespo, Carpenter, Baghana) han innovado en metodologías y contextos específicos, ambos conjuntos de investigaciones han abordado estas dimensiones de forma parcial y sin una comparación sistemática. El presente trabajo se diferencia al articular de manera explícita las perspectivas etimológica, semántica y discursiva en un marco común, lo que permite avanzar en la comprensión global del fenómeno y superar las limitaciones de los estudios previos.

Metodológicamente, se opta por un enfoque basado en la lingüística de corpus, que combina análisis semántico cualitativo y cuantitativo. Para el español, se empleará el Corpus Diacrónico del Español (CORDE), que permite rastrear usos reales del término *trabajo* desde el siglo XIII. Para el francés, se recurrirá tanto al corpus FRANTEXT como al Corpus de Français Médiéval (CFM), cubriendo un amplio espectro cronológico. Esta investigación se justifica por su doble aportación: por un lado, contribuye al campo de la semántica cognitiva diacrónica al aplicar un modelo radial a dos términos centrales del imaginario occidental. Por otro lado, ofrece una lectura crítica sobre cómo el lenguaje del trabajo construye y refleja ideologías sociales. Como se mostrará, el término *trabajo*, lejos de ser neutro, condensa una memoria discursiva donde confluyen violencia, virtud, sacrificio y productividad. En este marco, la tesis central que se defiende es que el análisis de *trabajo/travail* como nudo semántico revela no solo los mecanismos de cambio lingüístico, sino también la manera en que el léxico conserva y proyecta las representaciones ideológicas del trabajo en distintas épocas históricas.

2. Marco teórico

El marco teórico de este estudio se articula a partir de tres perspectivas interrelacionadas: la semántica cognitiva, el modelo de prototipos aplicado al cambio léxico, y la dimensión discursiva del significado. Estas aproximaciones permiten comprender cómo conceptos como *trabajo* y *travail* no solo sufren transformaciones léxicas, sino que también reflejan y reproducen estructuras ideológicas y culturales a lo largo del tiempo.

2.1. Semántica cognitiva y significado encarnado

La semántica cognitiva se desarrolla como una reacción a las limitaciones del estructuralismo y del formalismo generativo, defendiendo que el significado está profundamente enraizado en la experiencia humana. Para Lakoff (1987, 5–6), el conocimiento lingüístico se construye mediante categorías que no responden a límites estrictos, sino a núcleos prototípicos. Estas categorías se forman a través de mecanismos como la metáfora conceptual, la metonimia y la extensión radial del significado.

Langacker (1987, 45–47) coincide en que la gramática y el léxico forman parte de una misma red conceptual, donde las unidades lingüísticas reflejan esquemas de atención, enfoque, y control cognitivo. Este marco permite considerar cómo términos como *trabajo* pueden encarnar distintas conceptualizaciones, desde lo físico hasta lo emocional, dependiendo de los marcos socioculturales.

2.2. El modelo prototípico y la semántica diacrónica

El modelo prototípico fue adaptado al estudio del cambio léxico por Geeraerts (1997), quien propuso una concepción radial del significado léxico, donde los cambios semánticos no necesariamente sustituyen significados anteriores, sino que los reorganizan jerárquicamente. En su enfoque, el significado léxico se estructura en torno a un “núcleo” que agrupa los usos más representativos y “periferias” que contienen extensiones, metáforas y usos socialmente marcados (Geeraerts, 1997, 21–24). Este modelo es particularmente útil para analizar conceptos cargados de valor ideológico como *trabajo* y *travail*, en los que coexisten significados heredados de sufrimiento y significados modernos ligados a productividad, ciudadanía o autorrealización. Geeraerts (85) señala que la categorización prototípica puede reflejar estructuras de poder al naturalizar ciertas acepciones como “centrales” frente a otras consideradas “marginales”. El presente estudio privilegia el modelo prototípico de Geeraerts por su idoneidad para integrar las dimensiones etimológica, semántica y discursiva en un análisis diacrónico. Si bien la teoría de la metáfora conceptual de Lakoff y Johnson y la gramática cognitiva de Langacker han aportado perspectivas

fundamentales, ambas tienden a centrarse en mecanismos específicos del cambio léxico. En cambio, la semántica prototípica permite concebir el significado léxico como una red radial, capaz de articular acepciones nucleares y periféricas. Esta amplitud explicativa la convierte en el marco más adecuado para los objetivos de esta investigación.

2.3. Semántica discursiva y cambio ideológico

Complementariamente, se adopta la perspectiva de la semántica discursiva propuesta por Blank (1997), quien considera que los cambios léxicos están condicionados no solo por mecanismos internos de la lengua, sino también por la reconfiguración de los campos discursivos. Para Blank (1997, 41–43), cada término posee un “campo semántico” que se activa en contextos específicos y se reconfigura según las necesidades comunicativas y simbólicas de una época. Desde esta perspectiva, la evolución de *trabajo* y *travail* debe entenderse como el resultado de negociaciones discursivas: desde discursos religiosos sobre la penitencia, hasta discursos liberales sobre el mérito, o neoliberales sobre la empleabilidad. Esta dimensión será clave para interpretar los usos que aparecen en los corpus.

2.4. Contribuciones desde la lingüística histórica y sociolingüística

El análisis se enriquece con estudios que integran la semántica con la variación histórica y social. Nevalainen (2006, 34–36) subraya la importancia de correlacionar cambios léxicos con variables extralingüísticas, como el género discursivo, el estatus social y el cambio de registros. McMahon (1994, 102–105) destaca la necesidad de combinar evidencias cualitativas con datos cuantitativos en el análisis diacrónico. Asimismo, Fischer (1992, 15–22) argumenta que los corpus históricos permiten rastrear con precisión los momentos de bifurcación semántica y los puntos de inflexión ideológica, elementos cruciales para un estudio como el presente.

2.5. Estudios aplicados al léxico del trabajo

Estudios como los de Begoña Crespo (2007, 86–88) muestran cómo el léxico económico refleja cambios ideológicos, pasando de una visión feudal del trabajo a un discurso burgués de la productividad. Carpenter (2021, 23–25) examina cómo el lenguaje laboral en textos legales y periodísticos reconfigura la identidad del trabajador moderno. Finalmente, Baghana (2016, 95) propone que términos como *travail* son lugares de sedimentación ideológica, donde se inscriben tensiones entre disciplina, emancipación y autoexplotación.

3. Metodología

Este estudio se inscribe en una metodología mixta, de carácter cualitativo y basada en la lingüística de corpus, anclada en la semántica cognitiva diacrónica. El objetivo es identificar y analizar los cambios semánticos que han afectado a los términos *trabajo* (español) y *travail* (francés) desde una perspectiva comparativa, centrada en la relación entre significado léxico, discurso e ideología.

3.1. Enfoque diacrónico y comparativo

La dimensión diacrónica permite rastrear la evolución del significado de ambos términos desde la Edad Media hasta el siglo XXI, lo cual es esencial para identificar patrones de continuidad y ruptura en su estructuración semántica. Como señalan McMahon (1994, 103–105) y Nevalainen (2006, 34–36), el estudio del cambio léxico no debe limitarse a relaciones formales entre significados, sino que debe correlacionarse con los contextos discursivos e históricos en los que estos términos circulan. El enfoque comparativo, por su parte, permite establecer similitudes y divergencias entre las trayectorias de *trabajo* y *travail*, examinando no solo sus orígenes comunes en el latín vulgar *tripalium*, sino también sus resignificaciones particulares en el contexto hispánico y francófono. Esta comparación se articula mediante el modelo prototípico de Geeraerts (1997, 21–24), que permite observar cómo las redes semánticas se reorganizan a lo largo del tiempo en torno a diferentes núcleos prototípicos.

3.2. Corpus y criterios de selección

Se han empleado tres corpus principales para la recolección de datos:

- **CORDE (Corpus Diacrónico del Español)**: proporciona ejemplos textuales del término *trabajo* desde el siglo XIII hasta el XXI, abarcando distintos géneros (literatura, crónicas, textos jurídicos).

- **FRANTEXT (ATILF/CNRS)**: corpus de textos franceses desde el siglo XVI hasta el XX, útil para analizar la evolución moderna de *travail*.

- **Corpus de Français Médiéval (CFM)**: corpus lingüísticamente anotado que contiene textos en francés antiguo y medio desde los siglos IX al XV, cubriendo el período inicial de uso de *travail*.

Los criterios de selección de ejemplos incluyen:

- Presencia explícita de *trabajo* o *travail* (en sus variantes gráficas);
- Representatividad del género discursivo (textos religiosos, jurídicos, económicos, literarios);

- Claridad del contexto semántico (identificación del sentido predominante en el uso);

- Distribución cronológica equilibrada por siglos.

Para cada siglo se han recogido entre 4 y 6 ejemplos representativos por término, lo que permite una triangulación entre el modelo teórico, los datos empíricos y las implicaciones discursivas.

3.3. Herramientas y técnicas de análisis

Los datos se han tratado mediante un análisis cualitativo interpretativo basado en la semántica prototípica. Se han clasificado las acepciones de *trabajo* y *travail* en:

- **Núcleo prototípico** (esfuerzo físico, sufrimiento corporal);
- **Extensiones metafóricas** (trabajo espiritual, trabajo emocional);
- **Resignificaciones ideológicas** (trabajo como virtud, derecho, sacrificio o carga).

Asimismo, se han realizado gráficos de frecuencia léxica por siglo, siguiendo la metodología de Fischer (1992, 18–21), para observar patrones de aparición y expansión semántica.

3.4. Limitaciones

Se reconocen algunas limitaciones:

- El corpus de francés medieval no incluye una cobertura cuantitativa exhaustiva;

- Las marcas discursivas pueden ser sutiles o ambiguas, lo que dificulta una clasificación automática;

- No se ha incorporado una dimensión oral (por falta de fuentes), limitando el análisis a textos escritos.

Sin embargo, la triangulación entre corpus, teoría y contexto histórico permite una interpretación robusta de los cambios semánticos y discursivos.

4. Análisis Semántico Comparado

El análisis semántico comparado de los términos *trabajo* (español) y *travail* (francés) se realiza mediante una lectura cronológica de sus usos en los corpus CORDE, FRANTEXT y el Corpus de Français Médiéval (CFM), aplicando el modelo prototípico de Geeraerts (1997, 21–24). Este enfoque permite observar cómo, a lo largo de los siglos, cada término articula sentidos distintos a partir de un núcleo común de esfuerzo físico y sufrimiento.

Siglos XIII–XV: El núcleo prototípico del sufrimiento físico

En ambos contextos lingüísticos, los términos *trabajo* y *travail* presentan en sus primeras documentaciones una fuerte carga de sufrimiento físico y esfuerzo extremo. En español, expresiones como “gran trabajo me dio libramiento de esta pena” (CORDE, XIII) o “sus trabajos fueron muchos e peligrosos” (CORDE, XV) aluden al padecimiento corporal y los peligros enfrentados. En francés medieval, encontramos ejemplos como “en grant travail de cors estoit” (CFM, XIII) y “ce travail estoit douloureux et long” (CFM, XV), donde el término aún conserva la impronta de tortura y aflicción. Este núcleo semántico primitivo, heredado del latín *tripalium*, se mantiene como prototipo en ambos idiomas durante la Edad Media, con escasa diferenciación ideológica.

Siglos XVI–XVII: Incorporación de valores religiosos y morales

A partir del siglo XVI, ambos términos comienzan a resignificarse en contextos cristianos. En español, el “trabajo humano” se asocia con la virtud y el mérito: “por el trabajo humano se alcanza honra” (CORDE, XVI). En francés, aparece el “travail honnête” como elevación espiritual: “le travail honnête élève l’âme” (FRANTEXT, XVI). Esta evolución refleja el discurso cristiano que dignifica el sufrimiento como vía de redención, en línea con el análisis discursivo de Blank (1997, 43–45).

Siglos XVIII–XIX: Nacionalización del trabajo

Durante los siglos XVIII y XIX, los discursos nacionalistas y republicanos transforman el campo semántico del trabajo en ambos idiomas. En español, se idealiza como motor de la nación (“el trabajo como motor de la nación”, CORDE, XVIII), mientras que en francés se convierte en emblema republicano (“travail, honneur de la République”, FRANTEXT, XVIII). Esta resignificación coincide con el auge de la burguesía y el liberalismo económico, que convierten el trabajo en un valor cívico y moral (Crespo, 2007, 86; Carpenter, 2021, 23–25).

Siglos XX–XXI: Fragmentación y nuevas periferias semánticas

En el siglo XX, ambos términos muestran una creciente fragmentación semántica. En español, aparece el trabajo asalariado como categoría legal y social (“trabajo asalariado en tiempos modernos”, CORDE, XX). En francés, emergen connotaciones de malestar: “le stress lié au travail moderne” (FRANTEXT, XX).

Ya en el siglo XXI, se incorporan nuevas periferias semánticas vinculadas al trabajo digital, emocional y afectivo. Ejemplos como “el teletrabajo impone nuevas fatigas” (CORDE, XXI) y “le télétravail et la fatigue numérique” (FRANTEXT, XXI) reflejan una reconfiguración simbólica del esfuerzo laboral en el contexto tecnológico contemporáneo.

Convergencias y divergencias

El análisis revela una evolución parcialmente paralela: ambos términos comparten un núcleo original y desarrollan extensiones ideológicas similares (virtud, nación, estrés), pero con matices culturales. Mientras que en español se observa una progresiva revalorización moral del trabajo, en francés persiste una ambivalencia discursiva, que oscila entre el deber cívico y la alienación emocional (Baghana, 2016, 95).

5. Resultados del Análisis

Los resultados del análisis semántico y corpus-based permiten trazar una evolución estructurada del campo léxico de *trabajo* y *travail*, donde se identifican fases comunes, momentos de bifurcación discursiva y una reorganización prototípica con impacto ideológico.

5.1. Persistencia del núcleo prototípico

Ambos términos mantienen hasta el siglo XV un núcleo prototípico asociado al esfuerzo físico intenso y al sufrimiento, que remite directamente a su origen etimológico común en *tripalium*. Este sentido se encuentra ampliamente documentado en textos narrativos y religiosos, donde *trabajo/travail* designan una carga, una pena o una tortura. Esta coincidencia sugiere que el campo semántico inicial estuvo ampliamente condicionado por las representaciones medievales del cuerpo, el dolor y la expiación.

5.2. Formación de extensiones ideológicas

A partir del Renacimiento, los términos se resignifican en función de valores religiosos y morales. El trabajo se transforma en una práctica dignificadora, lo que da lugar a extensiones semánticas que incluyen “trabajo honesto”, “trabajo redentor”, “deber laboral”. Estas extensiones consolidan en ambos idiomas una visión moralizada del esfuerzo, en línea con los valores cristianos y humanistas. Se detecta aquí la emergencia de una semántica periférica que, sin desplazar el núcleo de sufrimiento, lo recontextualiza.

5.3. Nacionalización del campo léxico

Durante los siglos XVIII y XIX, el trabajo se convierte en categoría cívica. Esta fase implica una resemantización más intensa: en español, el trabajo aparece vinculado a la identidad nacional (“el trabajo dignifica al ciudadano”);

en francés, al republicanismo (“travail, égalité, liberté”). Esta reorganización discursiva afecta la estructura prototípica, desplazando el núcleo de sufrimiento hacia la periferia y promoviendo un nuevo núcleo ideológico: el trabajo como valor central del sujeto moderno.

5.4. Fragmentación semántica contemporánea

En el siglo XX, la estructura prototípica se complejiza. Surgen nuevos subcampos semánticos: trabajo asalariado, trabajo alienado, trabajo informal, trabajo emocional. El trabajo ya no es unívoco, sino que se disgrega en múltiples dimensiones contextuales. Esta fragmentación se acentúa en el siglo XXI, con la digitalización del trabajo (*teletrabajo*, *fatiga digital*), lo que genera nuevas cargas simbólicas asociadas al estrés, la vigilancia y la autoexplotación (Crespo, 2007, 87; Baghana, 2016, 95). En los discursos contemporáneos sobre los medios digitales, la noción de *trabajo* se resignifica a través de formas emergentes como *teletrabajo*, *home office* o *empleo remoto*. Estos términos no solo remiten a la desmaterialización del espacio laboral, sino que también se cargan de connotaciones ideológicas y prácticas vinculadas a la flexibilidad y la adaptabilidad (Hoyos, Cetina Canto y Centeno Ley 2024, 51–52). El estudio semántico realizado por estos autores muestra, por ejemplo, que *teletrabajo* se asocia directamente con *computadora*, *distancia*, *comodidad*, *tiempo* e *internet*, mientras que en un segundo plano se vincula a valores como *organización*, *productividad* y *responsabilidad* (54). Por su parte, el término *empleo remoto* privilegia las asociaciones con *tiempo* y *comodidad*, seguido de *flexibilidad*, *productividad* y *organización* (55). Así, el vocabulario que designa estas nuevas formas de actividad laboral digital no solo refleja condiciones técnicas, sino que también configura representaciones sociales del trabajo, donde las nociones de *tiempo* y *responsabilidad* emergen como núcleos comunes de sentido (56).

5.5. Diferencias culturales en la reorganización prototípica

Aunque las trayectorias generales son convergentes, se observan diferencias relevantes: en español, el trabajo tiende a conservar una carga positiva y moralizante más marcada, especialmente desde el discurso nacional-católico del siglo XX. En cambio, el término *travail* en francés conserva una ambivalencia discursiva, con una base negativa más persistente, incluso en contextos contemporáneos. Esta divergencia sugiere una reconfiguración prototípica asimétrica: mientras el español desplaza el sufrimiento a la periferia, el francés lo mantiene más cerca del núcleo.

6. Discusión Crítica

Los resultados obtenidos a partir del análisis corpus-based y prototípico no solo ilustran un proceso de cambio semántico, sino también un fenómeno de construcción discursiva e ideológica del significado. A través de las trayectorias de *trabajo* y *travail*, se evidencia cómo el léxico funciona como un archivo simbólico donde se sedimentan valores, normas y estructuras de poder (Blank, 1997, 41–45).

6.1. El léxico como archivo de ideologías

Tal como sostiene Begoña Crespo (2007, 86–88), los términos económicos y laborales no son neutrales, sino que actúan como vehículos ideológicos que naturalizan formas históricas de organización social. En este sentido, *trabajo* y *travail* constituyen nodos léxicos donde se articulan narrativas de redención (medieval), virtud (moderna) y productividad (contemporánea). La semántica de estos términos está profundamente enraizada en sistemas morales y políticos, lo que explica su resistencia al vaciamiento conceptual.

6.2. Reconfiguración prototípica y hegemonía cultural

El modelo de Geeraerts (1997, 21–24) permite observar cómo el campo semántico del trabajo se reorganiza en función de hegemonías ideológicas. En el español moderno, el discurso nacionalista y católico impulsa la centralización de acepciones positivas del trabajo como virtud, esfuerzo redentor o motor económico. En cambio, en el francés contemporáneo se mantiene una prototipicidad ambivalente, donde *travail* oscila entre el deber y la carga emocional, reflejando un ethos más escéptico o crítico respecto a la cultura del trabajo (Baghana, 2016, 95). Esta reorganización se corresponde con estructuras culturales más amplias. En el contexto hispánico, el trabajo ha sido tradicionalmente exaltado como vía de ascenso moral y social, lo que se refleja en su semantización positiva. En el mundo francófono, sin embargo, las representaciones colectivas conservan huellas de alienación, sufrimiento y resistencia, visibles en usos literarios, jurídicos y mediáticos.

6.3. Discurso, subjetividad y agencia

Otro aspecto relevante es la manera en que estos términos median la construcción de subjetividades. Como señala Carpenter (2021, 23–25), el lenguaje laboral define no solo tareas, sino identidades: el trabajador moderno se concibe a través de marcos lingüísticos que le atribuyen mérito, responsabilidad,

o vulnerabilidad. Las acepciones periféricas más recientes, como *teletrabajo*, *autoexplotación* o *fatiga digital*, muestran cómo el campo semántico incorpora nuevas formas de subjetividad neoliberal, donde el control externo se internaliza bajo la forma de rendimiento personal.

6.4. Valor interpretativo del análisis semántico

Finalmente, este estudio confirma que el análisis semántico, lejos de limitarse a la historia del significado, constituye una herramienta crítica para el estudio de las ideologías. La evolución de *trabajo* y *travail* no es solo un fenómeno lingüístico, sino también cultural, económico y político. La perspectiva cognitiva, al centrarse en categorías, prototipos y metáforas, permite acceder a una cartografía simbólica donde se inscriben las tensiones de la modernidad: entre sufrimiento y realización, alienación y autorrealización, disciplina y creatividad.

7. Conclusiones

Este estudio ha demostrado que la evolución semántica de los términos *trabajo* y *travail* no puede entenderse al margen de los contextos históricos, ideológicos y discursivos que los han modelado. Partiendo de un núcleo prototípico compartido —el sufrimiento físico vinculado al instrumento de tortura *tripalium*— ambos términos han atravesado procesos de resignificación cultural que han redefinido su papel en el imaginario colectivo de las sociedades hispánica y francófona. La aplicación del modelo prototípico de Geeraerts (1997) ha permitido observar cómo los significados no se reemplazan linealmente, sino que se organizan en redes donde coexisten sentidos antiguos y nuevos, centrales y periféricos, según su activación contextual. Así, el trabajo se transforma en virtud, derecho, deber, o incluso carga emocional, dependiendo del momento histórico y del discurso dominante.

Los resultados extraídos del corpus confirman una evolución convergente pero asimétrica: en español, *trabajo* se resignifica progresivamente con connotaciones positivas, influido por el discurso religioso y nacionalista; en francés, *travail* conserva una ambivalencia estructural, donde la noción de esfuerzo obligado sigue siendo central, incluso en su modernidad digital.

Desde una perspectiva cognitiva y crítica, esta investigación ha puesto de relieve que el léxico no solo nombra la realidad, sino que también la construye y la carga de valores. Estudiar *trabajo* y *travail* como unidades prototípicas permite entender cómo el lenguaje configura nuestra relación con la actividad humana, el cuerpo, el tiempo y el poder. El léxico del trabajo, lejos de ser transparente, constituye un archivo semántico de las ideologías que han moldeado

la historia social y económica del Occidente. Si bien este estudio se ha centrado en un análisis semántico-discursivo, futuras investigaciones podrían explorar más sistemáticamente las conexiones con la sociología y la filosofía del trabajo, campos donde conceptos como alienación, precariedad o realización personal podrían enriquecer la interpretación de los resultados.

El presente estudio abre además posibilidades de continuidad. Una primera línea consistiría en extender el análisis comparativo a otras lenguas romances, con el fin de identificar regularidades y especificidades culturales en la evolución del léxico del trabajo. Una segunda línea, más ligada al presente, implicaría examinar cómo el discurso laboral se resignifica en la era digital, donde fenómenos como la economía de plataformas, la cultura del emprendimiento y el teletrabajo generan nuevas metáforas, categorías y narrativas ideológicas en torno al trabajo.

BIBLIOGRAFÍA

- Aitchison, Jean. 1981. *Language Change: Progress or Decay?* London: Fontana Press.
- Baghana, Jerome. 2016. *Travail et subjectivités: Discours, représentations, mutations*. Paris: Éditions L'Harmattan.
- Blank, Andreas. 1997. *Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen*. Tübingen: Max Niemeyer Verlag.
- Bréal, Michel. 1897. *Essai de sémantique: Science des significations*. Paris: Hachette.
- Buchi, Éva, y Wolfgang Schweickard, dirs. 2008-. *Dictionnaire Étymologique Roman (DÉRom)*. Nancy: ATILF. Consultado el [30.09.2025].
<https://apps.atilf.fr/derom/>.
- Carpenter, Marisa. 2021. *Discourses of Labour and Identity in Late Modern Spanish*. Amsterdam: John Benjamins.
- Corpus de Français Médiéval (CFM). Université de Lausanne. <https://cfm.helsinki.fi/>
- Corominas, Joan. 1980–1991. *Diccionario crítico etimológico castellano e hispánico*. Vol. V. Madrid: Gredos.
- Crespo, Begoña. 2007. "The Semantic Field of Labour in Late Medieval English: Ideological Representations." *Journal of Historical Pragmatics* 8(1): 85–91.
- Delport, Marie-France. 1984. "Trabajo-trabajar(se): étude léxico-syntaxique". *Cahiers de linguistique hispanique médiévale*, no 9: 99-162.
- Fischer, Olga. 1992. "Syntax and Semantics in Diachronic Perspective." En *The Syntax of Early English*, edited by O. Fischer, A. van Kemenade, W. Koopman, and W. van der Wurff, 15–21. Cambridge: Cambridge University Press.
- FRANTEXT. ATILF – CNRS & Université de Lorraine. <https://www.frantext.fr/>
- Geeraerts, Dirk. 1997. *Diachronic Prototype Semantics: A Contribution to Historical Lexicology*. Oxford: Clarendon Press.

- Hoyos, Yanko Mézquita, Thelma Cetina Canto, y Guadalupe Centeno Ley. 2024. "Hacia una exploración semántica del trabajo a distancia." *Revista de Estudios Clínicos e Investigación Psicológica* 14 (27): 50–57.
<https://doi.org/10.56342/recv.vol14.n27.2024.32>.
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar, Vol. I: Theoretical Prerequisites*. Stanford: Stanford University Press.
- Lebas, Franck. 2023. "L'étymologie de "travail". En *Idées reçues sur le travail. Emploi, activité et organisation*, editado por Marie-Anne Dujarier, 22-23. Paris. Éditions du Cavalier Bleu.
- Littre, Émile. 1872. *Dictionnaire de la langue française*, t. 4. Paris: Hachette, s.v. travail.
- McMahon, April M. S. 1994. *Understanding Language Change*. Cambridge: Cambridge University Press.
- Nevalainen, Terttu. 2006. *An Introduction to Early Modern English*. Edinburgh: Edinburgh University Press.
- Real Academia Española. CORDE (Corpus Diacrónico del Español).
<https://www.rae.es/recursos/banco-de-datos/corde>.
- Traugott, Elizabeth Closs. 2017. "Semantic change." In *Oxford Research Encyclopedia of Linguistics*, edited by B. Arnold. Oxford: Oxford University Press.
- Ullmann, Stephen. 1962. *Semantics: An Introduction to the Science of Meaning*. Oxford – New York: Basil Blackwell.

CULTURAL SENSITIVITY IN AI TRANSLATION OF CHINESE LANGUAGE: EVALUATING AI'S CAPABILITY TO GRASP CULTURALLY RICH TEXTS

Iulia Elena CÎNDEA¹, Jiong WANG²

Article history: Received 13 June 2025; Revised 12 November 2025; Accepted 20 November 2025; Available online 12 December 2025; Available print 30 December 2025

©2025 Studia UBB Philologia. Published by Babeş-Bolyai University.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ABSTRACT. *Cultural Sensitivity in AI Translation of Chinese Language: Evaluating AI's Capability to Grasp Culturally Rich Texts.* The widespread adoption of Artificial Intelligence (AI) in translation has fundamentally reshaped global communication by enhancing efficiency, cost-effectiveness, and accessibility. Despite these advancements, critical questions persist about AI's capacity to navigate the inherent complexities of language, particularly regarding cultural sensitivity and contextual accuracy. These challenges are especially pronounced in Chinese, a language abundant in idiomatic phrases, literary allusions, and culturally embedded references, where mistranslations risk significant misunderstanding.

-
- ¹ **Iulia Elena CÎNDEA** is a lecturer of Chinese Language, Translation, and Cultural Studies at Lucian Blaga University of Sibiu, Faculty of Letters and Arts, Chinese Chair. She obtained her Ph.D. from the University of Bucharest, Doctoral School of Literary and Cultural Studies. Her research interests span cultural studies, translation studies, cultural sociology, and cross-cultural communication, with a particular focus on Sino-Romanian intellectual and literary exchanges. She has published on topics including the sociology of translation, China's soft power strategies, and the role of translation in knowledge production. Her current work critically examines the intersections of translation, sinology, and geopolitical influence in post-1989 Romania, investigating how knowledge production on China is shaped by historical legacies, political alliances, and economic interests. She also explores decolonial perspectives in China studies, analyzing how Romanian intellectual traditions engage with and challenge Western-centric narratives while negotiating China's expanding cultural diplomacy. E-mail: iulia.cindea@ulbsibiu.ro
- ² **Jiong WANG** is a lecturer at Beijing Language and Culture University (BLCU), where she teaches university students majoring in English and train in-service and pre-service teachers of English. Wang also teaches short-term international students learning Chinese as a second language and served as Director of the Training Department and the E-learning Department at BLCU. E-mail: wangjiong@blcu.edu.cn

This study examines the capabilities of leading AI translation tools in rendering culturally rich Chinese texts into Romanian and English. Drawing on a curated corpus of classical literature, contemporary literature and news articles, the research evaluates each tool's performance in terms of accuracy, cultural sensitivity, contextual understanding, emotional tone, and overall fluency. Both qualitative and quantitative analyses reveal considerable variation among popular platforms. While high fluency and coherence are attainable, some persistent gaps remain in the handling of idiomatic expressions and nuanced cultural elements. These findings reaffirm the need for expert post-editing and a judicious, human-centered integration of AI into translation workflows.

Keywords: *AI translation tool, cultural sensitivity, hermeneutics, Chinese culture, post-editing, hybrid translation*

REZUMAT. Sensibilitate culturală în traducerea asistată de IA a limbii chineze: evaluarea capacității de redare a textelor cu densitate culturală.

Adoptarea pe scară largă a inteligenței artificiale (IA) în traducere a remodelat fundamental comunicarea globală, sporind eficiența, costurile și accesibilitatea. În pofida acestor progrese, persistă întrebări critice privind capacitatea IA de a gestiona complexitățile inerente ale limbajului, în special sensibilitatea culturală și acuratețea contextuală. Provocările sunt deosebit de pronunțate în chineză, limbă bogată în expresii idiomatiche, aluzii literare și referințe puternic ancorate cultural, unde erorile de traducere pot genera neînțelegeri majore. Studiul de față examinează capacitățile principalelor instrumente de traducere bazate pe IA în redarea în română și engleză a textelor chineze bogate cultural. Pe baza unui corpus atent selectat, cuprinzând literatură clasică, literatură contemporană și articole de presă, cercetarea evaluează performanța fiecărui instrument în termeni de acuratețe, sensibilitate culturală, înțelegere contextuală, ton emoțional și fluență generală. Analizele calitative și cantitative evidențiază variații considerabile între platformele populare. Deși fluența și coerența ridicate sunt posibile, persistă anumite lacune în tratarea expresiilor idiomatiche și a elementelor culturale nuanțate. Aceste constatări reafirmă necesitatea post-editării de către traducători profesioniști și a unei integrări cumpătate, centrate pe om, a IA în fluxurile de traducere.

Cuvinte-cheie: *instrumente de traducere bazate pe IA; sensibilitate culturală; hermeneutică; cultureme chineze; post-editare; traducere hibridă*

Introduction

Translation studies have long rejected a purely linguistic definition of its object, positioning translation as a socially situated, historically contingent practice (Gentzler 2014; Marais 2022). Recent work argues that the arrival of artificial-intelligence systems intensifies this interdisciplinarity, demanding theoretical input from anthropology, ethics, history, and epistemology (Nergaard &

Arduini 2022). Accordingly, research on machine translation (MT) must be anchored in a framework that acknowledges both its algorithmic foundations and its cultural embeddedness (Rundle 2014; Zheng, Tyulenev & Marais 2023).

Empirical studies confirm that state-of-the-art NMT narrows the gap with human translators on surface metrics, yet still underperforms on culturally embedded texts such as classical Chinese literature or dialectal media (Venuti 1995; Castilho et al. 2018).

Among languages that pose such challenges, Chinese stands out for its dense idiomatic legacy, wide array of literary allusions, and culturemes. Converting Confucian moral injunctions, Daoist paradoxes, or even everyday colloquialisms into equally nuanced target languages demands more than grammatical fidelity; it entails preserving the cultural and historical resonance that underpins meaning. Subtle oversights, such as mistranslating a classical phrase or ignoring the emotional register of a literary passage, may distort or occlude the source text's core significance. This problem becomes even more acute for specialized genres like philosophical treatises, classical poetry, or the colloquial dialogues embedded in modern Chinese media.

Responding to these concerns, the present study focuses on five key dimensions of translation quality—idiomatic accuracy, cultural references, contextual understanding, emotional tone, and overall fluency—in order to capture both linguistic precision and deeper interpretive demands (Castilho et al. 2018). By systematically evaluating AI outputs for classic Chinese literature (e.g., Confucian texts, Daoist scriptures), contemporary colloquial conversations, mass-media news, and technical manuals, the research aims to ascertain how well current tools address the layered nature of Chinese language across diverse registers and domains.

Central to this investigation is an appreciation of hermeneutical considerations. Recent philosophical debates, particularly those revolving around John Searle's (1980) critique of "strong AI" and Schleiermacher's emphasis on the translator's empathetic engagement, highlight the gap between advanced pattern recognition and genuine semantic or cultural "understanding." Although recent NMT systems show strong gains in surface fluency, they still struggle with historically anchored, affective, and symbolic dimensions, especially outside training domains. This gap underscores the continuing relevance of human expertise—both as post-editors who correct errors and as co-creators who guide AI toward culturally faithful renderings.

Hence, this study explores how effectively current leading AI platforms handle cultural nuance and interpretive subtlety in Chinese-to-English or Chinese-to-Romanian translation. The findings elucidate each system's capabilities and limitations, illustrating why a hybrid approach—in which human translators validate or refine machine output—remains indispensable for culturally saturated

or context-sensitive material. The ultimate objective is to illuminate both the practical potential and the philological constraints of AI translation tools, thereby contributing to ongoing discussions in computational linguistics, translation studies, and cross-cultural communication.

1. Theoretical Background

Recent developments in machine translation have spurred renewed reflection on what precisely constitutes “translation” and how scholarly inquiry might deal with its increasingly transdisciplinary character. While discussions of AI translation often focus on computational or linguistic aspects, emerging perspectives suggest that translation cannot be circumscribed by such discrete disciplinary boundaries (Gentzler 2014; Marais 2022). Instead, it entails broader social, cultural, and historical processes that intersect with ethics, technology, and epistemology (Nergaard & Arduini 2022). Consequently, the study of AI-driven translation must engage with a wider theoretical foundation that includes not only computational linguistics but also translation history, anthropology, and hermeneutics.

This recognition of translation’s scope also demands a rethinking of AI-based workflows in terms of historical contextualization and cultural immersion (Rundle 2014). For instance, classical Chinese texts, with their deep intertextual links and symbolic references, cannot be approached solely as computational challenges but must be viewed as historically and culturally embedded phenomena. The conceptual frameworks offered by Zheng, Tyulenev, and Marais (2023) propose “(re-)conceptualizing translation” to underscore the fluid boundaries between translation studies and technology, noting that the interplay of algorithmic processes and translator agency calls for an integrated theoretical lens.

The transition from SMT to NMT represents a fundamental shift in how translation systems process language. Whereas SMT approaches relied on phrase-based probabilistic models often prone to syntactic dislocations and constrained context windows, NMT leverages deep neural networks capable of learning context-aware representations across entire sentences (Bahdanau et al. 2014). A key inflection point in NMT design arrived with the Transformer (Vaswani et al. 2017), whose self-attention mechanism obviates the need for sequential recurrence. By modeling global dependencies, Transformers accommodate more coherent and fluent outputs, thereby addressing some of the limitations of recurrent neural networks (RNNs) and long short-term memory (LSTM) models. Building on the foundations of Transformer architectures, LLMs scale both training corpora and model parameters in ways that substantially enhance zero-shot and few-shot translation capabilities (Brown et al. 2020;

Chowdhery et al. 2022). Empirical evidence indicates that these models not only approximate human-level performance for specific high-resource language pairs but also exhibit emergent behaviors, managing nuanced tasks such as style transfer and context-based disambiguation more effectively than earlier NMT systems. Nonetheless, their escalating size raises computational and ethical challenges, including the significant carbon footprint associated with large-scale model training (Bapna et al. 2022, 266-280; Strubell et al. 2019, 3645–3650) and the propagation of biases embedded within uncensored data (Bender et al. 2021, 5185–5198). Further concerns center on “hallucination,” where LLMs produce convincingly phrased yet inaccurate results. These issues underscore the importance of domain- or language-specific tuning, particularly in under-resourced settings, and highlight the ongoing need for human oversight to mitigate errors and ensure responsible deployment (Son & Kim 2023, 574; Guerriero et al. 2023, 1500–1517).

Cultural Sensitivity and Contextual Understanding

Despite the leaps in syntactic fluency, cultural adaptation persists as a formidable challenge. Scholars note that translation tools, even those fine-tuned with domain-specific corpora, can flatten linguistic variety by ignoring deeper cultural references—particularly in the context of literary Chinese or idiomatic expressions. Chinese rhetorical forms and historical allusions often require interpretive insight, not merely lexical equivalences (Venuti 1995). When dealing with Confucian classics or Daoist texts, mismatched conceptual metaphors can cause significant distortions in the target output. Similarly, the multiplicity of dialects in modern Chinese media complicates text standardization, indicating that purely corpus-driven AI systems risk producing literal but context-poor translations.

NMT excels at sentence-level coherence but often struggles to incorporate cultural cues that might be peripheral in a purely statistical sense (Castilho et al. 2018, 241–262). The model may capture grammatical well-formedness while overlooking connotations tied to unique historical events, region-specific humor, or social hierarchies. Thus, context must be understood not only as contextual references within a particular passage but also as a broader cultural knowledge base. Models that fail to incorporate these deeper layers risk generating translations that appear superficially correct yet diminish the text’s cultural and aesthetic depth.

Human-based evaluations remain indispensable in MT research, incorporating expert reviewers or professional translators who can detect omissions, distortions, or culturally insensitive phrasing (Zaretskaya 2018, 139-160). Studies

on post-editing processes reveal that advanced NMT outputs, while proficient, still require interventions to restore nuance or reintroduce connotations lost in the algorithmic leap from source to target. Furthermore, domain specialization (e.g., legal or medical) and culturally laden material exacerbate the mismatch between raw AI translations and the contextual adequacy demanded by stakeholders.

A central question in hermeneutical inquiry is whether neural architectures can truly “understand” a text’s cultural substrata or whether they merely produce outputs that appear syntactically correct (Searle 1980, 417-424; Bender & Koller 2020, 5185-5198). Searle’s Chinese Room Argument, in which Searle imagines himself in a room with a set of rules in English for manipulating Chinese characters and although he does not understand Chinese, he can follow the rules to produce appropriate Chinese responses to Chinese inputs. To an outside observer, it appears as if Searle understands Chinese, but in reality, he is merely following syntactic rules without any understanding of the semantics (Searle 1980). The experiment demonstrates how a system proficient in symbol manipulation may lack genuine intentionality, while Lonergan’s (1958) account of intentional consciousness suggests that meaningful comprehension involves self-reflection, moral agency, and experiential depth. Despite major strides in pattern recognition, today’s large language models have no intrinsic access to these human dimensions of understanding. Consequently, even advanced tools can overlook the deeper cultural or philosophical resonances central to classical Chinese works and other context-rich texts.

From a hermeneutic standpoint, scholars such as Schleiermacher (1838) and Gadamer (1960) emphasize that interpretation is inextricably tied to the translator’s existential horizons and empathetic engagement with the author’s voice. Language models like GPT-4 excel at mapping linguistic patterns (Bender & Koller 2020), yet they cannot internalize the lived, historical standpoint that undergoes human interpretation. This theoretical gap manifests practically when AI systems generate translations that, while grammatically sound, fail to capture emotional subtleties, ethical nuances, or symbolic layers. Researchers therefore continue to advocate a “human-in-the-loop” paradigm, ensuring that professional translators and domain experts refine AI-generated text in order to preserve moral, cultural, and existential dimensions of meaning (Zaretskaya 2018).

The longstanding philosophical question of whether AI truly “understands” language (Searle 1980) gains renewed significance in light of large language models that convincingly emulate interpretive depth. Bender and Koller (2020) present a sharp critique of conflating form with meaning, emphasizing that pattern recognition alone does not equate to genuine semantic awareness. Such arguments resonate with hermeneutic theories put forth by Gadamer (1960) and Schleiermacher (1838), who prioritize the translator’s existential horizon and empathetic immersion—dimensions that purely data-driven models cannot replicate.

From an ethical perspective, Floridi (2013) accentuates how accountability and transparency become central concerns when machines process culturally or historically sensitive material. If the technology mechanically flattens nuanced references or perpetuates biases, the cost may be eroded cultural integrity and ethical oversight. The balancing act between leveraging AI's efficiency and preserving the moral and existential dimensions of translation emerges as a major concern for practitioners and ethicists alike.

Cognizant of these limitations, Green et al. (2014) demonstrate that post-editing by professional translators improves both the efficiency and accuracy of machine outputs, especially in content where figurative language or domain-specific knowledge is non-negotiable. This collaborative approach, explored in multiple studies, underscores the practical synthesis of high-speed automation and interpretive fidelity, the latter being a hallmark of human expertise in reading cultural cues and rhetorical nuance.

2. Methodology

In order to examine the nuanced challenges that culturally rich Chinese texts pose to AI translation tools; this study adopted a multifaceted methodological approach. The ensuing sections describe the selection of the corpus, the choice of translation systems, and the criteria employed for both quantitative and qualitative evaluation. Although these procedures are structured around established best practices in machine translation research, they are equally informed by insights drawn from philological scholarship, which underscores the importance of historical context, idiomatic richness, and interpretive depth in textual analysis.

2.1. Corpus Selection

To evaluate whether AI systems preserve *cultural meaning* we require texts that systematically elicit (a) idioms and culturemes, (b) intertextual and historical allusions, (c) register and politeness management, (d) elliptical/zero-pronoun structures, and (e) emotionally marked narrative voice. Chinese provides all of these in concentrated form, but no single genre covers them evenly. We therefore constructed a two-part corpus that intentionally spans high-context literary discourse and lower-context expository prose, enabling a contrast between culture-dense and culture-light conditions while holding Chinese as the source language constant.

1. Literary texts:

- Classical literature subcorpus (sourced from the *Chinese Text Project*): *The Analects* (论语, 2.1; 2.4), *Dao Dejing* (道德经, ch. 1), *The Art of War* (孙子兵法, opening aphorisms), *Dream of the Red Chamber* (红楼梦, Baoyu passages). These provide aphoristic density, parataxis, archaic lexicon, and canonical allusions that are well known to challenge literal systems.
- Contemporary literature subcorpus (five excerpts): Mo Yan (*Life and Death Are Wearing Me Out*), Yu Hua (*To Live*), A Cheng (*The Chess King*). These authors contribute colloquial idiom, rural/dialectal flavor, euphemism, irony, and culture-bound lexis, i.e., phenomena where tone and social meaning are easy to flatten.

2. Non-literary articles (sourced from *The Chairman's Bao*): five short texts on education policy, public health, social issues, and technology. These balance the corpus with formal register, domain terminology, and explicit propositional content—useful as a *control* condition in which we expect systems to perform comparatively well.

Selection followed three practical principles. First, phenomenon coverage: each passage was chosen because it exemplifies at least one of the target stressors (idioms/allusions, politeness and particles, ellipsis/zero pronouns, emotionally salient voice). Second, comparability: passages were kept short and self-contained to minimize length effects and facilitate consistent rating across systems and target languages. Evaluation required stable anchors. Each criterion was scored on a five-point scale (1–5) from poor to exceptional. For the literary set, published translations by experienced sinologists in English and Romanian served as primary references. For the news items, editor-style references ensured consistent terminology. These references ground judgments of idiomatic accuracy, cultural faithfulness, contextual understanding, emotional tone, and overall fluency, so that scoring reflects interpretive adequacy rather than surface overlap.

Finally, the corpus is deliberately bilingual on the target side—English as a high-resource language and Romanian as a comparatively lower-resource one—to expose how resource asymmetries affect idiom mapping, register calibration, and error profiles. In sum, the literary–news pairing offers a diagnostic “challenge-plus-control” design. If modern LLMs truly surpass earlier baselines, the gains should be visible not only in the expository set but, crucially, in the culture-dense classical and contemporary fiction where meaning is inseparable from context, intertext, and social pragmatics. This corpus therefore fits the study’s aim: it makes cultural sensitivity measurable, comparable across systems, and attributable to more than general improvements in fluency.

2.2. Translation Tools

We evaluated a mixed set of neural machine translation (NMT) and large-language-model (LLM) systems selected for global prominence, distinct architectures, and widespread use among both non-specialists and professional translators. The pool comprised Google Translate, Baidu Translate, DeepL Translator, and Wordvice AI (NMT/MT services), alongside the newer strong-LLM: GPT-4, Vertex AI API, and DeepSeek.

To keep comparisons meaningful across genres, we adopted a strand-specific inclusion policy. All seven systems were screened on the classical Chinese strand. Because Google, Baidu, and DeepL exhibited floor effects there, subsequent strands (contemporary novels and non-literary texts) were restricted to the strong-LLM trio, which are closer peers in capability.

2.3. Evaluation Framework

In line with recognized practices for translation quality assessment, this study employed a multi-dimensional framework, adapting elements from Castilho et al. (2018) and Son & Kim (2023). The criteria were chosen specifically to illuminate whether an AI system could handle more than mere surface-level accuracy—namely, how it performs in domains of idiomatic usage, cultural context, and expressive fidelity.

1. Idiomatic Accuracy

Particular attention was given to the manner in which AI systems interpreted set phrases (成语), proverbs, or culturally contingent metaphors. An accurate rendering of idiomatic language serves as a key indicator that the system grasps not only lexical correspondences but also the underlying cultural or rhetorical function of such expressions.

2. Cultural References

As both classical and modern texts frequently employ historical or social allusions, translations were scrutinized for evidence that the cultural essence was preserved or rendered intelligible in the target language. Cases of mistranslation often arose from literal or decontextualized output, in which the immediate cultural or historical connotation was lost.

3. Contextual Understanding

Effective translations require recognizing contextual cues that inform word sense disambiguation, register shifts, and pragmatic subtleties. These range from understanding the domain (technical vs. literary) to detecting implicit references within a dialogue (e.g., forms of address or shifts in politeness).

4. *Emotional Tone*

Especially relevant for literary and journalistic texts, emotional tone encapsulates the tenor or emotive force intended by the original author. Whether conveying reverence in a Confucian passage or the urgency of a public health notice, capturing tone ensures the reader perceives the text's affective dimension authentically (Venuti 1995).

5. *Overall Fluency*

Finally, fluency describes the extent to which translations read as coherent, grammatical, and stylistically appropriate in the target language. While fundamental to any translation evaluation, fluency alone does not guarantee cultural or contextual fidelity (Castilho et al., 2018), hence its integration alongside other criteria.

3. Results

3.1. *Classical literature*

A careful appraisal of five classical literature excerpts reveals significant divergences in how current AI translation tools handle cultural nuance, idiomatic patterns, and contextual accuracy. These results echo earlier observations indicating that deeper semantic modeling is indispensable when facing philosophically and culturally rich material.

Comparative readings of *The Analects* (verses 2.1 and 2.4) highlight the inherent challenge of conveying Confucian moral teachings in translation. Expressions such as the “North Star” metaphor exemplify pitfalls for literal or decontextualized machine processing. For instance, Google Translate handles English syntax with relative ease but sometimes falters in Romanian by presenting “Beichen” untransformed, diluting Confucius’s intended cultural resonance. Baidu Translate likewise produces syntactically coherent yet semantically imprecise renderings, occasionally misrepresenting key life stages in Confucius’s self-reflection.

The cosmological image “居其所而众星共之” means the North Star stays fixed and other stars align/revolve by it. Literal renderings flatten cosmology into *location* (“lives there”) and misconstrue “共之” as *share it*, losing the idea of moral centripetal order. Google (RO) leaves “Beichen” unassimilated and literalizes structure („Beichen rămâne la locul său, iar stelele îl împart.”) which sounds like property division, not alignment. Baidu (EN/RO) similarly yields “*resides in its place and the stars share it.*” By contrast, GPT-4, Vertex AI, DeepSeek more effectively retain the metaphorical sense of governing “like the

North Star,” integrating context awareness in line with hermeneutical demands for interpretive depth (Gadamer 1960). These models anchor to the spatial frame evoked by “居” (“reside”) and to the high-frequency English verb “share,” rather than the Classical Chinese astronomical-moral frame (“orient / revolve / pay court”).

A parallel challenge appears in rendering the *Dao Dejing*, whose aphorisms are elliptical and paradox-driven. Google Translate and Baidu Translate sometimes collapse paradox into tautology in both Romanian and English, thereby diluting the Daoist tension between “the nameless” and “the named.” The line 「道可道，非常道；名可名，非常名。」 is a case in point: we repeatedly obtain outputs like “Calea care poate fi spusă este Calea fără nume,” i.e., “The Dao that can be named is the nameless Dao,” which erases the contrast “not the *constant* Way/Name.” The error stems from mis-scoping negation: in 非常道 / 非常名, 非 modifies 常 (“constant”), yielding “The Way that can be spoken is not the constant Way; the Name that can be named is not the constant Name,” but the systems attach 非 to 道 / 名 and default to the high-frequency collocation “nameless Dao,” collapsing polarity. This is emblematic of what Searle (1980) calls “syntactic manipulation”: fluent strings without semantic penetration. By contrast, Vertex AI and GPT preserve the paradox intact —*the Way that can be spoken is not the constant Way; the Name that can be named is not the constant Name*—suggesting that broader training corpora and stronger context modeling yield higher fidelity on culturally dense material.

The Art of War introduces yet another layer of complexity by combining specialized military terminology with succinct, declarative phrasing. While Google Translate performs reasonably in English, it slips into awkward literalism in Romanian (e.g., “Soldații sunt chestiuni importante ale țării”). Baidu Translate struggles to parse essential strategic vocabulary, distorting references such as “five constant factors.” Tools like DeepL reach moderate clarity but sometimes soften Sun Zi’s didactic tone. Wordvice AI, featuring domain-specific adaptability, produces concise and coherent outputs, yet strong LLMs again maintain the highest consistency by preserving both rhetorical gravity and lexical precision.

Turning to *Dream of the Red Chamber*, the emotive, colloquial style of passages such as “girls are made of water, boys are made of mud” demands sensitivity to figurative language. Google Translate occasionally yields disjointed Romanian phrasing, reducing the text’s lyrical quality. Baidu Translate, for example, unwittingly introduces extraneous characters or familial references that do not exist in the original, skewing the novel’s depiction of gender roles and social critique. Strong LLMs more successfully capture Jia Baoyu’s nuanced perspective, conveying both freshness and repulsion in a manner that aligns with the text’s heightened emotional register.

3.2. Contemporary novels

Across five passages drawn from Yu Hua, Mo Yan, and Ah Cheng, we treated the paired human translations (Romanian and English) as the benchmark and examined three systems (GPT-4, Vertex AI, DeepSeek) in both languages against five criteria: idiomatic accuracy, treatment of cultural references, contextual understanding, emotional tone, and overall fluency.

In the opening passage from *To Live*, the benchmark preserves both the blunt register and the rural material culture embedded in the Chinese. Key items such as “先生” (polite address), “拉屎” (vulgar), “马桶” (chamber pot), and “粪缸” (the manure vat for night soil) are carried over without euphemism, while the pace of “慢吞吞地” is rendered as a leisurely amble and the comic simile “打着饱嗝...像青蛙叫唤” becomes a frog-like belch. GPT-4 and Vertex match this well in English and in Romanian, keeping the low register and the specific rural container (“hârdăul cu fecale / manure vat”). By contrast, DeepSeek’s Romanian dilutes the scene twice: “mergea la necazuri” is a euphemism that does not mean defecate, and “găleata de gunoi” generalizes the specialized “粪缸”; a small agreement slip (“rângâind ca un broască”) further cools the voice. These micro-choices matter because the passage’s effect depends on the friction between social prestige (“domnule / sir”) and a bodily routine described in plain farm language.

The second Yu Hua excerpt magnifies the role of culture-specific references and sonic images. The source’s unit of land “亩 (mu)” —the social titles “阔老爷 / 阔少爷,” and the onomatopoeic line “钱在钱上面哗哗地流” together stage memory as material plenitude: copper coins clink (“铜钱碰来撞去”), money rushes over money, and the child’s call-and-response—“‘爷爷，你为什么动呀？’ ‘是风吹的。’” —anchors the tenderness of the scene. The benchmark keeps *mu* in English and Romanian, lets the titles breathe as “rich old master / stăpânul bogat,” and reproduces the money-soundscape rather than replacing it with abstraction. GPT-4 aligns most closely: it preserves *mu* transmits the dialogue verbatim and maintains the narrator’s deadpan warmth. Vertex is equally strong in Romanian, but its English replaces *mu* with “acres” and over-converts the quantity; this looks helpful but is materially inaccurate and erases a visible Chinese cultureme. DeepSeek is broadly correct in both languages, though it tends to add explanatory phrasing that slightly formalizes the rustic narrator. Here the criteria pull together, as illustrated also in table 1: idiomatic accuracy is high across models, yet the best handling of cultural reference is the one that resists domestication, keeps *mu*, and trusts a brief gloss to do the intercultural work.

Table 1. Comparison of key source concepts in a contemporary literary excerpt (Yu Hua, *To Live*), contrasting human EN/RO baselines with GPT-4, Vertex, and DeepSeek.

Source concepts	Human EN baseline	Human RO baseline	GPT-4	Vertex	DeepSeek
打着饱嗝...像青蛙叫唤	"belch... like a frog's croak"	„râgăit... ca orăcăitul de broască"	Accurate, neutral cadence	Accurate, natural rhythm	Accurate; sometimes "echoing like" - stylistic
阔老爷/阔少爷	"rich old master / rich young master"	„stăpânul... tânărul stăpân"	EN/RO accurate	EN/RO accurate	EN/RO accurate
钱在钱上面哗哗地流	"money... rushing over money"	„banii curg... zornăie"	EN accurate, slightly prosy	RO very idiomatic („foşnesc curgând")	EN concise; RO idiomatic
亩	Noted in footnote or gloss	Often implicit	EN sometimes adds a short gloss in note	EN/RO keep „mu," no over-explanation	EN/RO keep „mu"; may add brief gloss

The first Mo Yan passage from *Life and Death are Wearing Me Out*—crisp skin in hot oil, “肌肉爆裂的噼啪声,” the “贪官污吏,” and the vow to be ground to powder under a “石磨” or pounded to paste in an “铁臼”—tests whether systems can balance grotesque physicality with folk-Buddhist cosmology (“阎王”). The human benchmark does so with unflinching concreteness (“moara de piatră / millstone,” “piuă/mojar / mortar”) and a sardonic, stoic voice (“To hell with them!"). GPT-4 and Vertex reproduce the tool pair accurately in both languages and keep “King of Hell / Regele Iadului,” avoiding secular dilution. DeepSeek’s Romanian slips on the implement—rendering “铁臼” as “piuliță” (a nut), not *piuă/mojar*—a precise example of how world-knowledge about material culture feeds idiomatic accuracy. The consequence is more than terminological: the image of bodily reduction into paste depends on the mortar and pestle pairing, not on a hardware fastener.

In the companion Mo Yan courtroom passage from the same novel (see table 2), the evaluation turns on three details: cadence, the courtroom object, and the semantics of injustice. The benchmark hears “连珠炮般的话语” as a machine-gun flow, keeps the magistrate’s “惊堂木” as a specifically Chinese sounding block (the Romanian explains it succinctly as “plăcuța de lemn cu care se cere ordine în sală”), and renders “冤枉” as *wronged* rather than merely *innocent*. GPT-4 handles all three with poise; when it uses “gavel,” it often

appends “(sounding block),” which is an acceptable compromise. Vertex matches the rhythm but occasionally paraphrases “冤枉” as “wrongful death,” shifting the meaning from moral injustice to legal causation and thereby narrowing the text’s karmic dimension. DeepSeek remains fluent but defaults to the generic “gavel,” losing the cultural grain of the Chinese courtroom.

Table 2. Comparison of key source concepts in a contemporary literary excerpt (MO Yan, *Life and Death are Wearing Me Out*), contrasting human EN/RO baselines with GPT-4, Vertex, and DeepSeek.

Source concepts	Human EN baseline	Human RO baseline	GPT-4	Vertex	DeepSeek
焦干...油汪里...肌肉爆裂的噼啪声	“scorched... slick of oil... muscles bursting (popping)”	„pârjolit... baltă de ulei... mușchi care plesnesc”	EN vivid, tight cadence	EN/RO vivid; RO keeps “unsuros” later	EN sharp; RO “pocnetele mușchilor”
阎王 / 判官	“King of Hell / judges”	„Regele/Împăratul Iadului; judecătorii”	Terminology consistent	Terminology consistent	Terminology consistent
连珠炮般的话语	“rapid-fire barrage”	„ca un foc neîntrerupt / șuvoi de cuvinte”	EN “machine-gun-like torrent”	EN/RO good	EN good; RO good
惊堂木	“gavel / sounding block”	„bloc/plăcuță de lemn / masa judecătorească”	EN “gavel” (acceptable domestication)	EN “gavel,” RO “blocul de lemn”	EN “gavel,” RO “ciocănelul”
冤枉	“wronged / falsely accused”	„nedreptățit / nevinovat”	Accurate	Accurate	Accurate

Ah Cheng’s *The Chess Master* paragraph tests institutional literacy (for reference, table 3). The source is dense with Cultural Revolution nomenclature: “插队” (being sent to production brigades in the countryside), “运动” (the Movement), “留城政策” (policy allowing only children to remain in the city), and “机关的铝牌编号” (work-unit inventory tags). The benchmark names the period explicitly in Romanian, retains the administrative texture in English, and keeps the lupine self-image of “野狼似的转悠.” All three systems perform competently here in both languages. GPT-4 and Vertex keep the policy nuance and the tag detail; DeepSeek’s English usefully mirrors the agentive “我送走” (“sent off by me”), but its Romanian ends one sentence with the slightly meta “într-un mod numit corect,” where the benchmark’s “cel puțin în acte, pe deplin justificat” is more idiomatic and bureaucratically faithful.

Table 3. Comparison of key source concepts in a contemporary literary excerpt (Ah Cheng The Chess Master), contrasting human EN/RO baselines with GPT-4, Vertex, and DeepSeek.

Source concepts	Human EN baseline	Human RO baseline	GPT-4	Vertex	DeepSeek
插队 (sent-down to countryside)	“sent off to the countryside to join production brigades”	„trimiși la brigăzile de producție”	EN/RO correct	EN/RO correct	EN/RO correct
污点 (political blemishes)	“black marks / stains”	„pete politice”	Accurate	Accurate	Accurate
运动 (the Movement)	“the movement” with CR implicature	„Mișcarea” (implicit CR)	EN “Movement (运动)” clarifies	RO keeps implicit; context clear	EN/RO clear
机关的铝牌编号	“work-unit aluminium inventory tags”	„plăcuțe de inventar ale instituției”	EN preserves state ownership nuance	RO preserves nuance	EN/RO preserve nuance
留城政策	“policy allowing... to remain in the city”	„politica de a rămâne în oraș”	Accurate	Accurate	Accurate

Taken together, these passages show consistent strengths and bounded weaknesses. Idiomatic accuracy is now high for all three systems in English and, with occasional dips, in Romanian. The most frequent failures are small nouns that carry a great deal of world knowledge. Cultural references are best served when systems resist over-domestication. Contextual understanding is generally robust—the prestige/earthiness irony in Yu Hua and the karmic justice frame in Mo Yan are correctly read—though the “wrongful death” paraphrase shows how semantic neighborhoods can mislead. Emotional tone is where small register choices have outsized effects: a single euphemism (“mergea la necazuri”) or a lyrical embellishment that the source does not license can shave the rustic deadpan or the grotesque stoicism that define these narrators. Overall fluency remains strongest in English; Romanian is solid but more vulnerable to agreement errors and to the loss of rural or craft vocabulary.

For the purposes of this study, these examples do more than illustrate score differences; they explain them. When a system names the right object,

keeps the right unit, and holds its nerve on register, it reads as culturally literate and emotionally faithful—and its scores rise across all five criteria. Where it converts away a cultureme, generalizes a tool, or softens the register, the loss propagates: idiomatic accuracy drops, cultural reference weakens, tone shifts, and fluency begins to feel generic. The upshot is practical. With light, targeted post-editing GPT-4 and Vertex routinely approach the benchmark translations in both languages, while DeepSeek benefits most from Romanian register control and concrete-noun verification Their reliance on extensive multilingual corpora and robust neural architecture seems to support greater flexibility with archaic forms and metaphorical references. Figure 1 illustrates the scores across the five criteria obtained by each of the LLM tools.

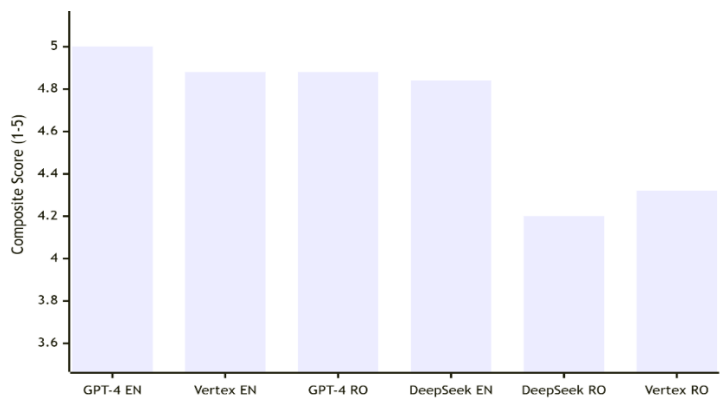


Figure 1. Performance of strong LLM Translation Tools throughout five criteria: Idiomatic Accuracy, Cultural Sensitivity, Contextual Understanding, Emotional Tone, Overall Fluency

3.3. Non-Literary texts

Against the same criteria, results converge tightly on the human benchmark: Vertex and DeepSeek are marginally stronger in Romanian, while GPT-4 remains highly competitive in English. Performance clusters around four pressure points: the semantics of policy metaphors, the precision of numerals and units, the handling of institutional terminology, and low-level negation.

The grading-reform notice (The Chairman’s Bao 2024b) typifies the metaphor constraint. The source signals a methodological shift to “decouple” (脱钩) evaluation from GPA arithmetic while adopting an A-F letter scheme. The best outputs anchor 脱钩 to GPA—“abandons GPA in favor of letter grades”—and avoid the misleading calque “decuplează studenții de la note,” which would imply abandoning grades altogether. Vertex and DeepSeek handle this anchoring

most consistently in Romanian; GPT-4 is solid in English, preserving policy intent and newsroom neutrality.

The bereavement/re-entry vignette (The Chairman's Bao 2024c) shows how a single noun can skew context. Whereas one human draft misread 旅游文化班 as a swimming class, all three systems correctly render a travel/tourism culture course, maintain kinship address (阿婆 → "Grandma"/„bunica"), and carry the closure metaphor 走出了人生的阴霾 into natural Romanian (e.g., „a ieși din umbra vieții") and fluent English. Here, the models matched a baseline human draft on a terminology point.

Data-dense reporting further narrows the gap. In the excerpt on AI and jobs (The Chairman's Bao 2023), all three models preserve sample size (5,300 workers), scope (seven member countries), the headline statistic (60%), and the conditional stance attributed to Mathias Cormann, with no numeric drift. Likewise, the Beijing vaccination notice (The Chairman's Bao, 2021) tests polarity and units; the LLMs retain the two-dose schedule, keep the fee exact (93.5 yuan ≈ \$14.2 per dose), and correctly scope eligibility around 医保 ("enrolled in Beijing medical insurance" / „înscriși în sistemul de asigurări medicale"), avoiding a polarity flip present in one human draft.

Finally, the cultural-heritage brief on the 200-year-old Chinese wallpaper (The Chairman's Bao 2024a) probes entity chains and chronology. The models consistently maintain Westport / County Mayo, the 1730 build date, the Browne/Hughes lineages, and the conservation quote. Occasional enrichments (e.g., "Westport House") are contextually plausible but should be flagged when strict source fidelity is required. Register remains appropriately restrained.

In short, when terminology is conventional and figures are explicit, GPT-4, Vertex AI, and DeepSeek meet—or slightly exceed—human baselines under light post-editing. Residual, correctable risks are (i) semantic drift in policy metaphors, (ii) context-skewing noun misreads, and (iii) polarity flips in conditional statements. A brief editorial protocol—lock numerals/units verbatim, verify negation, gloss institutional acronyms once, and prefer precise reformulations—consistently lifts all three systems to publication-ready quality across non-literary texts.

4. Discussions

As Fig. 2 illustrates the strong LLMs attain the highest aggregate scores for cultural sensitivity, idiomatic fidelity, contextual alignment, and tonal control, while Google, DeepL, and Baidu are serviceable on straightforward content yet more prone to figurative loss in archaic or philosophically dense texts. This mirrors prior work linking superior outcomes to larger, more diverse corpora and stronger architectures (Castilho et al. 2018; Son & Kim 2023).

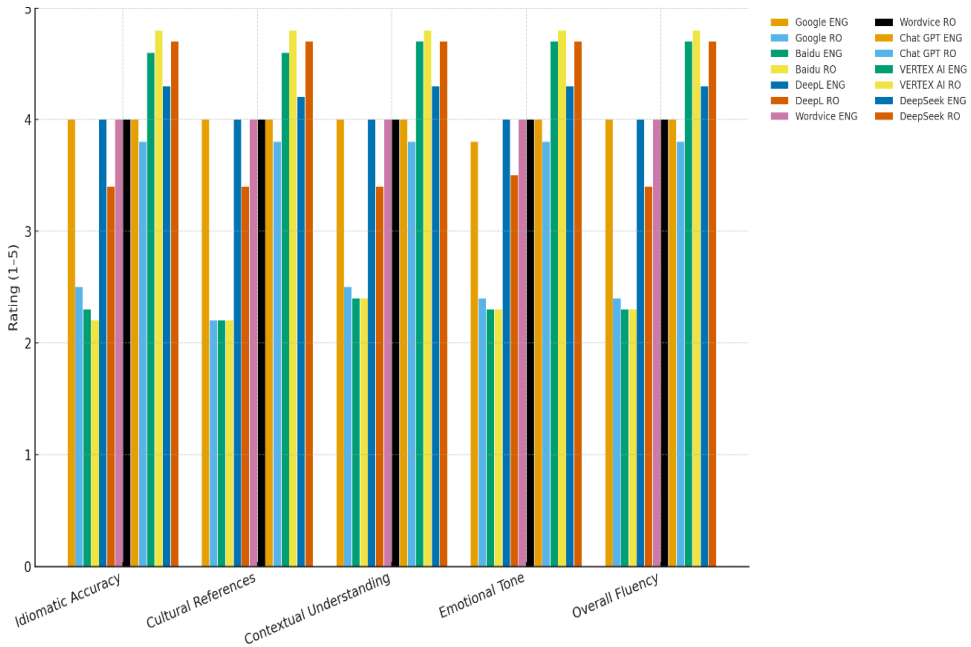


Figure 2. Comparison of Ratings of 6 Translation Tools across Different Criteria According to the Present Case Study

In the classical strand, top systems preserve metaphor scope and culture-specific references; weaker systems default to literal, monosemous renderings. In contemporary fiction, the triad separates under literary pressure along three axes: (i) culturemes/units, (ii) concrete-noun specificity, and (iii) register control. GPT-4 and Vertex more reliably retain culture-bound items and precise implement nouns, sustaining imagery and tone; DeepSeek's Romanian occasionally domesticates units or generalizes objects, softening stylistic effects. In non-literary news, outputs converge on the human benchmark when terminology is conventional and numerals/units explicit.

From a hermeneutic vantage point, these limitations reflect a larger philosophical tension: advanced NMT systems demonstrate remarkable facility at pattern recognition, yet they lack the historically and existentially grounded understanding that shapes deep literary or moral interpretation (Searle 1980; Schleiermacher 1838; Gadamer 1960). As Schleiermacher and Gadamer (1960) argued authentic comprehension transcends syntactic mapping, necessitating lived insight and a horizon of experience. AI, by contrast, operates within algorithmic constraints, capturing "surface equivalences" without fully embodying the cultural memory or ethical vision that animate classical texts and moral teachings.

Even with advanced context modeling, the existential dimension—encompassing self-reflection, intentionality, and moral agency—remains beyond neural mimicry. Accordingly, human oversight is focal: light, targeted post-editing—locking numerals/units, verifying negation, glossing culture-specific artefacts once, and resisting over-domestication—consistently restores rhetorical and cultural fidelity.

Taken together, these three strands suggest a stable division of labor. On classical passages, top systems lead but still require expert post-editing for metaphor scope, diachronic lexicon, and culturemes. On non-literary news, the strong-LLM triad reaches near-parity with humans under a light checklist. On contemporary novels, fine-grained cultural literacy separates GPT-4/Vertex from DeepSeek in Romanian, with GPT-4 generally most stable in English and Vertex strongest in Romanian cadence.

From a broader vantage, domain adaptation—whether via specialized subcorpora, glossaries of archaic Chinese, or bilingual commentary—could mitigate many of these diachronic mismatches. Yet the consistent need for philological discernment confirms that bridging centuries of linguistic evolution demands more than raw computational might. While scaling up corpora or refining contextual embeddings can somewhat lessen anachronisms, the interpretive subtlety inherent in classical Chinese—encompassing allusions, metaphorical frameworks, and shifting script reforms—ultimately requires the informed human-AI partnership described throughout this research. By weaving philological insight into large-scale machine translation pipelines, the field may progress toward a more sensitive, historically resonant handling of premodern Chinese texts, fully honoring the language's diachronic richness. In short, large models narrow—but do not erase—the gap between structural equivalence and cultural resonance; targeted human intervention remains the decisive step from fluency to fidelity.

Conclusion

Advances in neural architectures and large language models have propelled AI translation to new heights of fluency and efficiency, yet this study affirms that fully capturing the cultural, idiomatic, and contextual richness of Chinese texts remains elusive. In our design, six systems were evaluated on classical texts; after observing floor effects for Google Translate, DeepL, and Baidu Translate, we restricted the contemporary-novel and non-literary strands to the strong-LLM triad (GPT-4, Vertex AI, DeepSeek). While GPT-4, Vertex AI, and DeepSeek demonstrate notably higher proficiency, they still fall short of the interpretive depth that human translators provide. News articles or straightforward expository content typically benefit most from automated output, requiring minimal post-editing. By contrast, specialized or historically layered genres underscore the need for domain adaptation and expert

intervention, highlighting the persistent gap between surface accuracy and the deeper hermeneutic engagement that brings texts vividly to life.

Limitations include a finite, unidirectional corpus and the absence of targeted fine-tuning experiments. Future work should expand genre coverage, add bidirectional evaluation, quantify inter-annotator agreement on error typologies, and test domain-adapted models for archaic lexicon and cultureme retention.

In sum, large models narrow—but do not erase—the gap between structural equivalence and cultural resonance. The decisive step from fluency to fidelity remains human: translators ensure that both classical aphorisms and contemporary journalism are rendered not only with linguistic precision, but with the cultural memory and ethical vision that give these texts their enduring force.

We advocate a stance of critical openness toward AI in translation. Welcoming the efficiencies of LLM-assisted workflows does not diminish the translator’s centrality; rather, it intensifies professional obligations. The human translator retains ultimate ethical and moral responsibility for the published text and must demonstrate heightened linguistic, cultural, and methodological competence—capable of auditing LLM outputs for hallucinations, conceptual drift, inappropriate domestication, register shifts, and other category errors. The question, therefore, is not whether to oppose LLMs, but how to integrate them responsibly. This entails principled protocols for their use—transparent prompts, documented post-editing, verifiable decision trails, and clear lines of accountability—so that AI functions as an assistive companion under informed human supervision, and fidelity to meaning, culture, and context remains securely human-guaranteed.

WORKS CITED

- Ah Cheng. 2005. 《棋王》 [*The Chess Master: Chinese-English Bilingual Edition*]. The Chinese University of Hong Kong Press.
- Bahdanau, D., K. Cho, and Y. Bengio. 2014. “Neural Machine Translation by Jointly Learning to Align and Translate.” *arXiv preprint arXiv:1409.0473*. <https://arxiv.org/abs/1409.0473>.
- Bapna, A., et al. 2022. “Multilingual Mix: Example Interpolation Improves Multilingual Neural Machine Translation.” In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*: 266–280.
- Bender, E. M., and A. Koller. 2020. “Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data.” In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*: 5185–5198.

- Bender, E. M., et al. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*: 610–623.
- Brown, T., et al. 2020. "Language Models Are Few-Shot Learners." In *Proceedings of the 34th Conference on Neural Information Processing Systems*.
- Cao Xueqin. 2025. "Hongloumeng 紅樓夢 (Dream of the Red Chamber)." *Gushiwen*. Accessed March 6, 2025.
https://m.gushiwen.cn/guwen/bookv_ca3ee0e7ebd7.aspx.
- Castilho, S., et al. 2018. "Evaluating MT for Professional Use." In *Translation Quality Assessment: From Principles to Practice*, edited by J. Moorkens, S. Castilho, F. Gaspari, and S. Doherty, Springer: 241–262.
- Chowdhery, A., et al. 2022. "PaLM: Scaling Language Modeling with Pathways." *arXiv preprint arXiv:2204.02311*. <https://arxiv.org/abs/2204.02311>.
- Confucius. n.d. "Lunyu 論語 (The Analects), Book 2 'Wei zheng 為政,' 2.1 and 2.4." *Chinese Text Project*. Edited by Donald Sturgeon. Accessed March 2, 2025.
<https://ctext.org/lunyu-zhushu/wei-zheng/ens>.
- Floridi, L. 2013. *The Ethics of Information*. Oxford University Press.
- Gadamer, H.-G. 1960. *Truth and Method*. Seabury Press. (Original work published 1960.)
- Gentzler, E. 2014. "Translation Studies: Pre-Discipline, Discipline, Interdiscipline, and Post-Discipline." *International Journal of Society, Culture & Language* 2 (2): 13–24.
- Green, S., J. Heer, and C. D. Manning. 2014. "The Efficacy of Human Post-Editing for Language Translation." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*: 439–448.
- Guerreiro, Nuno M., et al. 2023. "Hallucinations in Large Multilingual Translation Models." *Transactions of the Association for Computational Linguistics* 11: 1500–1517.
- Laozi. n.d. "Daodejing 道德經, chap. 1." *Chinese Text Project*. Edited by Donald Sturgeon. Accessed March 4, 2025.
<https://ctext.org/wiki.pl?if=gb&chapter=492483&remap=gb>.
- Lonergan, B. 1958. *Insight: A Study of Human Understanding*. Philosophical Library.
- Marais, K., ed. 2022. *Translation beyond Translation Studies*. Bloomsbury.
- Mo Yan. 2006. *《生死疲劳》 [Life and Death are Wearing Me Out. Chinese Edition]*. Zhejiang Publishing House.
- Nergaard, S., and S. Arduini. 2022. "Translation: A New Paradigm." *Translation. A Transdisciplinary Journal* 1 (1): 8–17.
- Rundle, C. 2014. "Theories and Methodologies of Translation History: The Value of an Interdisciplinary Approach." *The Translator* 20 (1): 2–8.
- Schleiermacher, F. 1998. *Hermeneutics and Criticism and Other Writings*. Edited and translated by A. Bowie. Cambridge University Press.
- Searle, J. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3 (3): 417–424.
- Son, J., and B. Kim. 2023. "Translation Performance from the User's Perspective of Large Language Models and Neural Machine Translation Systems." *Information* 14 (10): 574.

- Strubell, E., A. Ganesh, and A. McCallum. 2019. "Energy and Policy Considerations for Deep Learning in NLP." In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*: 3645–3650.
- Sunzi. n.d. "Sunzi bingfa 孫子兵法 (The Art of War), chap. 1 'Ji 計 (Laying Plans).'" *Chinese Text Project*. Edited by Donald Sturgeon. Accessed March 2, 2025. <https://ctext.org/wiki.pl?if=gb&chapter=934017&remap=gb>.
- The Chairman's Bao. 2021. "Beijing Opens Vaccination Applications to Foreigners." *The Chairman's Bao*. Accessed October 24, 2025. <https://www.thechairmansbao.com/beijing-opens-vaccination-applications-to-foreigners>.
- The Chairman's Bao. 2023. "Study Finds Up to 27% of Jobs Could Be Lost to AI." *The Chairman's Bao*, July 13. Accessed October 24, 2025. <https://www.thechairmansbao.com/study-finds-up-to-27-of-jobs-could-be-lost-to-ai>.
- The Chairman's Bao. 2024a. "200-Year-Old Hand-Painted Chinese Wallpaper Undergoes Restoration at Irish Home." *The Chairman's Bao*, May 16. Accessed October 24, 2025. <https://www.thechairmansbao.com/200-year-old-hand-painted-chinese-wallpaper-undergoes-restoration-at-irish-home>.
- The Chairman's Bao. 2024b. "Chinese Universities Discuss Dropping GPA Grading for Applicants." *The Chairman's Bao*. Accessed October 24, 2025. <https://www.thechairmansbao.com/chinese-universities-discuss-dropping-gpa-grading-for-applicants>.
- The Chairman's Bao. 2024c. "Elderly Woman Defies Expectations with Active Lifestyle." *The Chairman's Bao*, May 23. Accessed October 24, 2025. <https://www.thechairmansbao.com/elderly-woman-defies-expectations-with-active-lifestyle>.
- Vaswani, A., et al. 2017. "Attention Is All You Need." In *Proceedings of the 31st Conference on Neural Information Processing Systems*: 5998–6008.
- Venuti, L. 1995. *The Translator's Invisibility: A History of Translation*. Routledge.
- Yu Hua. 2017. *《活着》 [To Live. Chinese Edition]*. Beijing: Beijing October Literature and Art Publishing House.
- Zaretskaya, A. 2018. "Machine Translation Post Editing at TransPerfect." In *Translation Quality Assessment: From Principles to Practice*, edited by J. Moorkens, S. Castilho, F. Gaspari, and S. Doherty, Springer: 139–160.
- Zheng, B., S. Tyulenev, and K. Marais. 2023. "Introduction: (Re-)Conceptualizing Translation in Translation Studies." *Translation Studies* 16 (2): 167–177.

POLARITY IN THE VERBAL DOMAIN. PRELIMINARY RESULTS AND CONSIDERATIONS¹

Hajnalka DIMÉNY² 

Article history: Received 7 May 2025; Revised 20 October 2025; Accepted 20 November 2025; Available online 12 December 2025; Available print 30 December 2025

©2025 Studia UBB Philologia. Published by Babeş-Bolyai University.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ABSTRACT. *Polarity in the Verbal Domain. Preliminary Results and Considerations.* The paper presents preliminary results of an extensive investigation of polarity in the verbal domain of Hungarian. Polarity has manifold and wide-ranging connections to different aspects and subdomains of the language system. The present study is concerned with the distribution of polarity in the lexicon and its possible motivations.

In language studies, it has long been hypothesized that negative polarity is richer in its representation than positive polarity on a lexical, constructional and idiomatic level as well. The preliminary results of the investigation show that, in the verbal domain, this is indeed the case. This is not surprising, but the nature of the phenomenon and its ramifications deserve further linguistic exploration.

Other important considerations presented in the study regard the definition of polarity. From a quantitative approach, it seems obvious that affective and evaluative meaning cannot be separated from socio-cultural knowledge and subjective expectations. This last category does not differentiate itself distinctly

¹ I would like to thank the anonymous reviewers for their valuable comments and suggestions, which have greatly helped in improving the quality and clarity of this paper.

The work was funded by the Domus program of the Hungarian Academy of Sciences.

² **Hajnalka DIMÉNY** is a university assistant at Babeş-Bolyai University Faculty of Letters, Cluj-Napoca. She has defended her PhD in 2021 with the title *Main features of verbal meanings and lexicalization patterns of Hungarian verbs. Criteria for a semantic typology of Hungarian verbs*. She is interested in semantic phenomena and works in the paradigm of cognitive linguistics, following her supervisor, Sándor SZILÁGYI N. She is also a proponent of science popularization, and works a lot with students interested in linguistic problems, as the head of the Márton Gyula Linguistic College. Contact: hajnalka.dimeny@ubbcluj.ro.

from other, more robust cases of polarity. The connection is also strengthened by structural properties of verbs and by the negativity bias. Because of this, I believe that, besides affective content and evaluative meaning, although strongly connected to these, negative polarity verbs also encode non-preferred alterations to events, while positive polarity verbs encode preferred alterations or outstanding forms of events. For an event to be outstanding in some respect, is relevant in fewer contexts, than excelling in non-preferred ways. This explains the higher elaboration of the negative domain, but makes polarity as a verbal phenomenon even harder to grasp and define.

Keywords: *polarity, verbal meaning, distribution across the lexicon, socio-cultural entrenchment*

REZUMAT. Polaritatea în domeniul verbal. Rezultate și considerații preliminare. Lucrarea prezintă o cercetare preliminară în cadrul unei investigații extinse asupra polarității în domeniul verbal al limbii maghiare. Polaritatea are conexiuni multiple și de amploare cu diverse aspecte și subdomenii ale sistemului lingvistic. Studiul de față se concentrează asupra distribuției polarității și a posibilelor sale motivații.

În studiile lingvistice, există de multă vreme ipoteza că polaritatea negativă are o reprezentare mai bogată decât polaritatea pozitivă, atât la nivel lexical, cât și la nivel construcțional și idiomatic. Rezultatele preliminare ale cercetării arată că, în domeniul verbal, această ipoteză se confirmă. Acest lucru nu este surprinzător, însă natura fenomenului și implicațiile sale merită o explorare lingvistică mai profundă.

Alte considerații importante prezentate în lucrare privesc definirea conceptului de polaritate. Dintr-o perspectivă cantitativă pare evident că sensurile afective și evaluative nu pot fi separate de factorii și așteptările socio-culturale. Această ultimă categorie nu se diferențiază clar de alte cazuri mai robuste de polaritate. Legătura este consolidată și de proprietățile structurale ale verbelor și de către tendința spre negativitate. Din acest motiv, consider că, pe lângă conținutul afectiv și semnificația evaluativă – deși strâns legate de acestea – verbele cu polaritate negativă codifică și modificări ale evenimentelor considerate ca fiind nedorite sau nepreferate, în timp ce verbele cu polaritate pozitivă codifică forme remarcabile ale evenimentelor. Un eveniment remarcabil într-un anumit sens este relevant în mai puține cazuri, decât manifestarea modurilor nedorite sau nepreferate. Acest lucru explică reprezentarea lexicală mai bogată al domeniului negativ, dar face ca polaritatea, ca fenomen verbal, să fie și mai dificil de înțeles și de definit.

Cuvinte-cheie: *polaritate, sens verbal, distribuție în lexicon, factori socio-culturale*

1. Introduction

In language studies, it has long been hypothesized that negative polarity has a higher level of lexical sophistication (or lexical diversity) than positive polarity on a lexical, constructional and idiomatic level as well. However, substantial research on polarity in lexical elements, especially in verbal lexemes, has not attracted significant attention to date. The paper presents preliminary work for an extensive investigation which aims to contribute to this subject of linguistics by offering a comprehensive analysis of polarity in verbal meanings, on the one hand, and testing the accuracy of automatic annotation tools, on the other hand.

The term *polarity* is used broadly in the paper, referring to positive or negative semantic content in terms of emotive states, social judgements or subjective/moral values expressed by the lexical elements. This procedure is motivated on the one hand by the observation that positive or negative semantic content expressed by verbs can be related to (1) emotions (e.g., *giggle, enjoy, rejoice; fear, fret, shock*), but also to (2) (dis)satisfaction with some features of experience (e.g., *beaver away, brighten, encourage; gorge, misbehave, ramble*), and (3) moral/social values (e.g., *appease, compliment, revive; demoralize, embarrass, reprimand*). On the other hand, extensive approaches to positive and negative content also embrace a wider stance (Martin and White 2005; see chapter 3 for details). Starting from these, and because the aim was to have a comprehensive grasp of polarity in the verbal system, a wider approach to the phenomenon was adopted.

Polarity phenomena in natural languages are the entailment of the basic evaluative and affective character of language. Consequently, they are pervasive, quite diverse, and most likely to some extent universally realized. Besides space and time, polarity is one of the main organizational patterns in the intrafield relational structure (Faber and Mairal Usón 1997b). Thus, an extensive analysis of polarity goes beyond formulating statements about the investigated linguistic system proper. Polarity phenomena can be correlated with meaning extension and analogical changes in the language system, with grammatical and constructional phenomena through semantic prosody, they are tied inseparably to usage-based properties and affective-cognitive foundations of language, to the linguistic expression and reinforcement of biases, and so on. Thus, the comprehensive study of polarity in an entire subsystem of a language (the verbal system in our case) can have far-reaching applications.

The preliminary results of the analysis of 29 905 Hungarian verb meanings processed by one annotator, lead to several important considerations and to the formulation of two hypotheses. First, they suggest that in the verbal

domain, indeed negative polarity is lexically more diverse and more richly represented than positive polarity. This is not surprising, but the nature of the phenomenon, its ramifications and motivations deserve further linguistic investigation. This is carried out in chapters 4.1. and 4.2. Second, the data makes it possible to investigate the correspondences of polarity and word length, testing whether negative polarity verbs are longer on average than positive polarity verbs. According to the Pollyanna hypothesis, evaluatively positive terms are used more often and more diversely than evaluatively negative words (Boucher and Osgood 1969). In the same time, Zipf's law states that words used more frequently are shorter in length and carry less information (Zipf 1936, 1949). From these, we can hypothesize that negative polarity words carry more information and hence are longer in average than evaluatively positive words. This hypothesis is tested in chapter 4.3.

Although further developments of the dataset are needed, the case study suggests that there are important and wide-ranging correspondences to be investigated. Moreover, the data lays the foundations for a future comparison of manual and automated polarity labelling based on dictionary entries, so as to be able to test the accuracy of automated labelling.

Polarity is often encoded in lexical elements of the language. At the same time, the lexicon of a language expresses different characteristics of the community using it, even other factors than emotions and evaluation. Thus, polarity phenomena in natural languages, originally most probably the expression of evaluative and emotive content (see the negativity bias as a cognitive-affective pattern, Jing-Schmidt 2007), are deeply intertwined with socio-cultural and pragmatic factors. To be able to grasp polarity as an element extensively present in the lexicon, the study uses a dataset of all verbal entries of an explicatory dictionary. Because these constitute an abstract, expansive representation of the language, the findings are representative of the entirety of the verbal system, with the possibility of extending it to the verbal system of other natural languages.

Linguistic analyses have so far focused mainly on negative polarity items (see Giannakidou 2011 for an overview), on the language of basic emotions (e.g., Wierzbicka 1999), on negative emotive intensifiers (Jing-Schmidt 2007; Kugler 2014; Szabó, Vincze and Bibok 2023), on the usage frequency of positive and negative polarity words (Garcia, Garas and Schweitzer 2012), and on sentiment analysis (e.g., Cambria and Hussain 2015; Hua et al. 2024). However, only a few studies have been carried out regarding the distribution of positive and negative lexical items in the entire lexicon of natural languages, and its ramifications to other subdomains of the language. The most notable is that of Boucher and Osgood (1969), they, however, examined only a small portion of the vocabulary, though on a large language set.

In response, the present study brings new insights regarding the distribution of polarity verbs in the lexicon of the Hungarian language, but also makes quantitative claims regarding the possible socio-pragmatic motivations of it. The findings constitute important data for studies of polarity, of the verbal semantic network, and also corroborate usage-based approaches to language which state that “language is an embodied and social human behavior”, and that “usage has an effect on linguistic structure” (Bybee and Beckner 2009, 827). Nonetheless, the results presented are only preliminary, and are to be treated as such. Because of the size of the dataset, automated labelling is planned in the future.

2. The data of the study

As mentioned earlier, the study is based on a dataset of 29 905 Hungarian verb meanings and their relevant semantic features. The backbone of the data was compiled by Sándor Szilágyi N., and it has been complemented and adjusted for the present study by the author.³ This database comprises the entries of the second largest Hungarian explicatory dictionary (Pusztai 2006), treating each separate meaning and its dictionary interpretation as separate data. A powerful aspect of the present study is (1) that it is based not on an arbitrarily chosen, small fragment of the lexicon, but on a large, extensive sample of the language comprised in an explicatory dictionary, and (2) that it deals with all word meanings separately. Its limitations are also twofold, and are related (1) to the use of one single annotator and (2) to the fact that dictionary data never represents the language use of any speaker. The first one is to be addressed in the future.

The coding for event type was done by the author and followed the lexical domains of the functional lexicological model introduced by Martín Mingorance (1990). The term lexical domain refers here to “the set of lexemes which together lexicalize all or part of a conceptual domain” (Faber and Mairal Usón 1999: 59). This model was also used by Faber and Mairal Usón in their work for a typology of predicate schemata of English (1997a, 1997b) and for constructing a lexicon of English verbs (1999). The main reason for this choice is the practical procedures used by the authors in determining the lexical domains, which make the model easily applicable to the present data. The premise of this method is that dictionary entries contain the keywords for identifying the different relations between word meanings (e.g., causation, part-whole relation,

³ Szilágyi N. and Dimény (2021). The data underlying this article are not yet available online because of copy rights, but will be gladly shared on reasonable request to the authors.

hyponymy, and hypernymy). Specifically, the use of one verb in the definition of another suggests that the two meanings are related, while the phrasing of the dictionary entry elaborates this relation. Therefore, starting from entirely specific events, in a few steps, I was able to categorize each verb meaning into a broad category of abstract event type and several subtypes.

The lexical domains in the Functional Lexicological Model are the following: EXISTENCE, POSITION, CHANGE, POSSESSION, PERCEPTION, EMOTION, COGNITION, SPEECH, GENERAL ACTION, MOVEMENT, and SENSE PERCEPTION (Faber and Mairal Usón 1999: 88, annotations follow the original). This event typology is not entirely adopted from Faber and Mairal Usón, but adapted to the study data and cross-checked with other verb ontologies, specifically WordNet (Princeton University 2010) and the Rich Event Ontology⁴ (Windisch Brown et al. 2017). The motivations for using a somewhat different approach in the verbal typology are manifold. (1) The nature of the semantic content of Hungarian verbs was taken into consideration. In these, the encoding of non-dynamic events seems to be more prevalent than among English verbs (for a few examples see *örül* 'be happy', *szomorkodik* 'be sad'; *zöldell* 'be green', *rozsdállik* 'be rusty'). This resulted in the differentiation of STATES (like being happy) and ATTRIBUTES (such as something being green). (2) Verbs expressing BEHAVIOR have a specific form in Hungarian, besides their specific meaning structure (many of them end in *-kodik/-kedik/-ködik*)⁵, and because if this they were considered to be a class of their own. (3) A greater focus has been placed on the nature of the primary figure (or subject argument) to the verb, because in the dictionary the different meanings of verbs are mainly separated based on collocational properties, with strong emphasis on the nature of the primary (and in many cases also the secondary) figure (human being, animals, inanimate things, or natural phenomena etc.). Related to this, (4) the intentionality of the primary figure also received a role. This is also a main feature in separating different verbal meanings in explicatory dictionaries (e.g., verbs referring to human activity in their primary meaning and to inanimate or complex things in their figurative meanings, like *megy* 'go' and *fordul* 'turn', or the English verbs *work* and *run*). This resulted in the more numerous class of verbs encoding ACTION/ACTIVITY, including all intentional activity, even speech and sense perception, beside the classes of HAPPENING, EXISTENCE and PERCEPTION, which encode non-intentional events. Finally, (5) taking into consideration previous literature on verb ontologies

⁴ The Rich Verb Ontology comprises 161 verb classes and a multiple nodes deep hierarchy, with, for example, the following classes on the mid-level: COGNITIVE EVENT, LIFE EVENT, INTENTIONALLY ACT, MOTION (annotations follow the the authors). This was too detailed for the present purposes, was provided important insights for the ontology.

⁵ Although, this suffix is not exclusive to this class of verbs, it can also refer to other activity types (e.g., *fésülködik* 'to brush one's hair', *tanárkodik* 'to be working as a teacher').

(specifically WordNet, the Rich Event Ontology and Leonard Talmy's semantic model [2000]), MOVEMENT has been distinguished from the more general class of ACTIVITY. Not incidentally, verbs encoding movement are considered to be prototypical verbs by several authors (e.g., Langacker 2008, Tolcsvai Nagy 2016).

These considerations resulted in the following basic event types: HAPPENING, ACTION/ACTIVITY, MOVEMENT, BEHAVIOR, EXISTENCE, PERCEPTION, STATE, and ATTRIBUTE/CHARACTERISTIC. These are expressed most generally in the dictionary entries by verbs such as *történik* 'happen'; *tesz, csinál* 'do'; *megy* 'go', *mozog* 'move'; *viselkedik* 'behave'; *létezik* 'exist', *él* 'live, be alive'; *érez* 'perceive'; *van* 'be', *mutatkozik* 'appear as', and the adjective *képes* 'be able'. Within some of these basic event types, we can speak of subtypes, most prominently about subtypes of activity verbs. At this moment, these were identified only partially based on the dictionary entries and on the semantic literature available. For example, we can speak of verbs of speaking, verbs of eating and drinking (or digestion), verbs of sound, verbs of transfer, verbs of possession and change of possession, etc. These are only used as examples in this study. Naturally, it could be argued that a different event typology would lead to different results. Remarks regarding this are made in the Discussion part.

As it can be surmised, a typology of few classes was aimed for, with a high emphasis on the nature and characteristics of the primary figure of the event structure. The main reasons for this were (1) the nature of the dataset and (2) the aims of the study at the present moment. For more detailed analyses a more granular typology will be needed in the future.

The coding for polarity is also based on the phrasing of the dictionary entries, because these can be considered a representation of the shared knowledge of linguistic encoding. In the next chapter, I elaborate on the exact method I have used and on the challenges of coding for polarity on this type of data and in general. Here, I will only mention that 456 meanings have multiple labelling, because the dictionary entries sometimes group similar usages which may have different polarities. Verbs referring to being in an emotional state are a typical case of this, because they can express both positive and negative polarity depending on the argument they take. See the entries in (1) for an example.

- (1) *elragad*
'carry away'

<Érzelem, indulat>	hatalm-á-ba	kerít	és	heves szav-ak-ra,
feeling emotion	power-GEN.3SG-ILL	get.hold.of.3SG	and	fierce word-PL-SUB
cselekedet-ek-re	készítet	vki-t.		
act-PL-SUB	lead.3SG	someone.3SG-ACC		

'<Feeling, emotion, temper> takes over someone, and leads him/her to some impulsive words or acts.'

In (1), the polarity of the verb (and consequently of the sentence, or vice versa) depends on the argument the verb takes. If someone is carried away by a positive feeling or by an event in a positive sense, the verb is associated with positive polarity. But in the case of an argument that expresses a negative feeling or state, negative polarity is associated. In this case, being carried away itself becomes negative.

Multiple labelling is used to suit the aims of the study; by using the database, to be able to account for the encoding of polarity in the language in general.

3. Theoretical prerequisites of polarity

In linguistics, the term *polarity* is most commonly used to refer to the distinction between positive and negative semantic content. This distinction seems quite intuitive, and thus, polarity is often left undefined. On the one hand, the term *polarity* can refer to a grammatical feature of an item associated with affirmation or negation. In this case, we speak about *polarity items*, i.e., linguistic items expressing affirmation or negation (see Giannakidou 2011 for an overview). On the other hand, on a lexical-semantic level, *polarity* refers to the affective evaluation of the speaker expressed by linguistic elements, used mainly in sentiment analysis. In this case, polarity can be defined as an evaluative stance, or as a positive or negative attitude, encoded in the meaning of a lexical item. However, this evaluative stance can be defined in several ways. Martin and White (2005, 39) offer a summary of different approaches to evaluative language analysis and its terminology.

Table 1. A summary of approaches to evaluative language analysis and its terminology as compiled by Martin and White (2005)

Approaches to evaluation	'entity focused'	'proposition focused'
Chafe & Nichols 1986		evidentiality
Ochs & Schiefflen 1989	affect specifiers	affect intensifiers
Biber & Finnegan 1989	affect	evidentiality
Wierzbicka 1990b	emotion	
Bybee & Fleischman 1995	evaluation	modality
Niemeier & Dirven 1997	emotion	
Conrad & Biber 2000	attitudinal stance	epistemic stance
Hunston & Thompson 2000	opinions about entities	opinions about propositions
Hunston 2000	'status' and 'value' on the 'autonomous plane'	'status' and 'value' on the 'interactive plane'

The authors themselves use the term *appraisal* to refer to the more general evaluative meaning in language. It is “one of the three major discourse semantic resources construing interpersonal meaning (alongside involvement and negotiation). Appraisal itself is regionalised as three interacting domains – ‘attitude’, ‘engagement’ and ‘graduation’” (Martin and White 2005, 34–35). *Affect* refers to the expression of emotional experiences, *judgement* is connected with behavior, and *appreciation* is concerned with the value of things, including natural phenomena and semiosis.

The more recent and more popular computational linguistic analyses include all these in their investigation of evaluative language, and use the term *sentiment* to refer to them. In computational linguistics, a distinction between entity-level, propositional-level and aspect-based sentiment analysis is being held. E.g., in their systematic review of 727 primary studies on sentiment analysis, Hua et al. (2024, 296) refer to the subject of their investigation as “views and feelings (i.e. sentiment”), but delimit their interest to how these are expressed in user reviews, social media posts, and open-ended survey question responses.

The present paper follows this line of wide approach to evaluative meaning, and uses the terms *polarity* to refer to the emotive states, social judgements and subjective/moral values pertaining to the meaning of words (cf. affect, judgement, and appreciation in Martin and White [2005] mentioned above). This decision is motivated on the one hand by the observation that positive or negative semantic content expressed by verbs can be related to (1) emotions (e.g., *giggle*, *enjoy*, *rejoice*; *fear*, *fret*, *shock*), but also to (2) (dis)satisfaction with some features of experience (e.g., *beaver away*, *brighten*, *encourage*; *gorge*, *misbehave*, *ramble*), and (3) moral/social values (e.g., *appease*, *compliment*, *revive*; *demoralize*, *embarrass*, *reprimand*).

Starting from these, and because the aim was to have a comprehensive grasp of polarity in the verbal system, a wider approach to the phenomenon was adopted. The study only investigates polarity expressed on a lexical level, in verbs, but in the process of coding for polarity a sentence level analysis was used, given the nature of the data.⁶ Thus, a sentiment polarity assigned to the sentence explaining a certain meaning, is assigned to the meaning in question. Sentence level sentiment analysis determines whether a positive, negative or neutral opinion is expressed in a sentence (Liu 2012).⁷ It takes content words appearing in a sentence, and assigns a value based on the polarity expressed by them.

⁶ Preverbal forms of verbs that appear as independent entries in the dictionary, are not treated separately, nor they are structurally analysed. Although there are cases in which the preverbal prefix is the morpheme carrying polarity meaning (cf. Evellei 2012), in this investigation they are not treated differently, and polarity is assigned to the whole of the lexical form.

⁷ Liu (2012: 7) uses the terms *sentiment analysis* and *opinion mining* interchangeably to refer to “the field of study that analyzes people’s opinions, sentiments, evaluations, appraisals, attitudes,

In the present study, the coding for polarity was performed by only one person at the moment, the author. However, some measure of accuracy is ensured by the fact that all labelling is based on the dictionary entries, which are the result of the collective work of lexicographers and are the representation of a shared knowledge of the language. Naturally, there are plans for expanding the team of annotators for testing the accuracy of the present annotation.

Three classes make up the polarity coding in the database. Entries containing negative words were coded as “negative”, and entries containing positive words were coded as “positive”. Those not containing neither negative, nor positive words were coded as “neutral”. By the terms *positive words* and *negative words*, I mean the following:

- 1) they refer to an event, state or experience that is in itself widely considered, subjectively, to be positive (good, acceptable) or negative (bad, not acceptable), e.g., *győz* ‘to win’, *öröm* ‘joy’, *jó, jól* ‘good, well’, *hálás* ‘grateful’, etc.; and *rossz* ‘bad’, *aggodalmaskodik* ‘to worry’, *szégyen* ‘shame’, *megaláztatás* ‘humiliation’, *erőszakosan* ‘aggressively’, etc.; here the motivation can vary from emotional experiences (like happiness, worrying) to evaluation of things, experiences or behavior.
- 2) words with positive or negative connotation (rarely also noted in the dictionary by information regarding style/register, like *vulgar*, *euphemistic*), e.g., *tisztel* ‘to respect’, *kedves* ‘nice’, *érték* ‘value’, etc.; and *megbán* ‘to regret’, *sajnál* ‘pity’, *le-* ‘down’ (in some cases), etc.

Although based on keywords expressing evaluation, the coding takes into account the whole of the dictionary entry (which is practically a sentence). This means that in case words with different polarities appear in an interpretation (which is sometimes the case), the coding is adjusted to the meaning in question. See example (2) for a demonstration. The [+] and [-] signs in subscript indicate the polarity of the lexical item in question.

- (2) félrevezet_[-]
‘mislead’

Vki-t	a	helyes _[+]	útírány-tól	eltérít.
someone.3SG-ACC	ART	right	direction-ABL	divert.3SG
‘Lead someone in the wrong direction.’				

and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes.” That is, although a difference can be made between all these relations to the texts produced by humans, the term *sentiment* is nowadays used very broadly. This way, practically sentence level sentiment analysis was performed when detecting the polarity of a dictionary entry, which was then assigned to a verb meaning.

In an automatic analysis, the presence of the word *helyes* 'right' would lead to a "positive" polarity label, because other words in the entry carry neutral polarity (*valaki* 'someone', *eltérít* 'divert', *útirány* 'way, direction'). However, the presence of the verb *eltérít* 'divert' changes the polarity suggested by the word *helyes* 'right', the verb expressing that the right way is bypassed. Not taking context into consideration would lead to an erroneous classification of the meaning, labelling it with positive polarity instead of negative polarity. The human annotator eliminates this problem, as well as that of idiomatic expressions, questions, conditionals and grammatical elements such as negation. These were also taken into account, and idiomatic expressions were treated as a whole. Idioms are often listed as separate meanings of a verb in the dictionary. In an automatic classification, these would most likely cause problems and inaccurate categorization. However, the relatively detailed dictionary entries to the meanings and the human annotator, deal with them easily and, most importantly, accurately. All ambiguous cases were cross-checked against the larger, though much less recent Hungarian explicatory dictionary (Bárczi and Országh eds. 1959–1962).

For another case of polarity assignment, see example (3) and (4) below. The lexical elements encoding polarity, which determine the polarity assignment of the verbal meaning in question, are marked with bold. The content words which are not marked this way, do not encode polarity.

- (3) *adományoz*_[+]
'donate'

<Nagyobb bigger	összeg-et, sum-ACC	érték-et _[+] > value-ACC	jótékony _[+] charitable	vagy or
közösségi community	cél-ra _[+] purpose-SUBL	ad. give.3SG		

'Donate a large sum of money or valuables to charity or for community purposes.'

- (4) *akadékoskodik*_[-]
'demur'

Vki-t	túlzott _[-] ,	kicsinyes _[-]	aggályoskod gátol. ás-sal _[-]	
someone.3SG-ACC	exaggerated	petty	worry-INS	hinder

'Someone hinders another person with excessive, petty concerns.'

- (5) változik
'change'

Folyamatosan	más,	lesz,	mint	amilyen	volt
continually	different	will.be	like	such.as	was

'It continually becomes different than it used to be.'

The database still needs some improvements and testing for accuracy. At the moment it only makes possible qualitative analyses and the formulation of hypotheses regarding quantitative aspects. The use of a dataset based on dictionary entries has some advantages that deserve to be mentioned. Most prominently, the data is not only an arbitrarily chosen fragment of the lexicon, but the representation of the entire verbal system of the standard variant of Hungarian. Another benefit is that the database is built not on verbs, but on verb meanings, and this makes it possible to account not only for the primary meanings, but for all meanings listed in the dictionary separately. Thus, the data reflects the entire verbal domain of the language.

An analysis performed on the data at this point, leads to several important lessons and observations which will help to improve the dataset. Overall, the database created specifically for the study of verbal semantics is a useful and effective tool for investigating polarity information encoded in verbs and for formulating domain-general hypotheses.

4. Results and discussion

All findings and correlations are to be treated carefully, as preliminary results. Here they are used to formulate hypotheses and to refine future investigations. Although some of the findings are in line with previous studies, because of the use of a single annotator, the correspondences found can only be regarded as guidelines in subsequent analyses, which are of great variety: regarding polarity more generally, motivations of its lexical encoding, correspondences with meaning extension mechanisms, possible role in analogical changes, and so on.

4.1. *Distribution of polarity in the study data*

A total of 21 061 verb meanings are labeled “neutral” in the database, and 456 meanings are labeled “positive” and/or “negative” and/or “neutral”, because some meanings encode more than one polarity. This gives us a total of 70.43% of neutral polarity in all verb meanings. While the language of the Internet (and most

likely everyday language, too) is “emotionally charged”, as shown by Garcia, Garas and Schweitzer (2012), in the overall lexicon of a language, neutral words predominate, which is not surprising. Only a total of 1 977 verb meanings are encoded as “positive” in the dataset, which shows a 6.61% positive polarity, along with 6 411 verb meanings encoding “negative” polarity, which constitutes 21.44%. The remaining 1,52% of the verbs (456 entries) show multiple coding for polarity.

Taking these multiple labelling cases into account, we have the following distribution of polarity in the database: 21 334 instances of “neutral” label (70.21%), 2 247 instances of “positive” polarity (7.40%), and 6 804 instances of “negative” polarity (22.39%). Comparing these ratios to the previous ones, we see no significant difference in the values of the two accounts, as shown in *Table 2* below. The cases of multiple coding do not eventuate significant differences.

Table 2. Summary of the distribution of polarity in the data

Category labelling			Multi- ple labell- ing	Total	Category labelling including cases of multiple labelling			
“neut- ral”	“posi- tive”	“nega- tive”			“neu- tral”	“posi- tive”	“nega- tive”	Total
21 061	1 977	6 411	456	29 905	21 334	2 247	6 804	30 385
70.43%	6.61%	21.44%	1,52%	100%	70.21%	7.40%	22.39%	100%

Overall, the data suggests that in the second largest Hungarian explicatory dictionary, approximately 70% of the verb meanings express neutral events, about 7% of the data encode positive polarity, and approximately 22% encode negative polarity. These ratios seem to support the old claim that negative polarity has a richer lexical representativity than positive polarity, at least in the verbal domain. A similar analysis on the entire lexicon of the language and on the adjectival domain would shed light on the correctness of the supposition of richer representativity of negative polarity, but the findings are in line with the cognitive-affective phenomenon of negativity bias which is “an automatic tendency to pay significantly more attention to unpleasant than pleasant information” (Jing-Schmidt 2012: 417). While positive polarity is more frequent (Boucher and Osgood 1969; Garcia, Garas and Schweitzer 2012, and others), negative polarity is more richly expressed in language.

Moving on to a more detailed look, we can see that approximately 30% of verb meanings encode polarity, positive or negative (or both, in a few cases). Interestingly, but not surprisingly, the vast majority of these verbs refer to some kind of activity performed by a person, while only 3 verbs express perception. *Table 3* summarizes the occurrence of the different (more general) event types among the verbs encoding polarity. Naturally, these proportions can only have full explanatory power if we compare them to the occurrence of event types in general. This is why we can see this in the last columns of *Table 3*.

Table 3. The proportion of event types among the verb meanings expressing polarity and among all data

Event type	Number of occurrences	Percentage	Number of occurrences	Percentage
	among the verb meanings expressing polarity		in the data	
ACTION/ACTIVITY	5 560	62.87%	20 751	69.39%
HAPPENING	1 710	19.34%	5 248	17.55%
MOVEMENT	578	6.54%	2 243	7.50%
STATE	520	5.88%	723	2.42%
BEHAVIOR	331	3.74%	367	1.23%
ATTRIBUTE/CHARACTERISTIC	111	1.26%	383	1.28%
EXISTENCE	31	0.35%	133	0.44%
PERCEPTION	3	0.03%	57	0.19%

Looking at the third and fifth columns of *Table 2*, we can see that state and behavior are amply represented among the meanings comprising polarity, while perception verbs are highly underrepresented. A total of 71.92% of state verbs and 90.19% of behavior verbs encode polarity. ACTION/ACTIVITY, HAPPENING, MOVEMENT, ATTRIBUTE/CHARACTERISTIC and EXISTENCE are encoded with a polarity of 23-32%. This distribution of polarity in the verbal semantic system reflects the affective foundations of language on the one hand, and its fundamental social functions on the other hand. STATES and BEHAVIORS are two semantic territories that are particularly highly loaded with affective and evaluative components.

It is even more interesting to see how positive and negative polarity are distributed among the event types. ATTRIBUTES/CHARACTERISTICS is the only event type that shows a higher proportion of positive polarity, 18.81% against 10.05%.⁸ In all other event types, we see a strong inclination toward the encoding of negative polarity. In behavior verbs, we can see a ratio of almost one to six, while with other event types (excepting perception, where we have very few cases of polarity⁹), we have a ratio of one to three or one to four. These ratios can be tied to how the lexicon is shaped to our affective, social and communicative needs.

⁸ The reason for this might lie in the specificity of the event typology. An attribute/characteristic is usually not considered a separate event type. However, Hungarian verbs can express features such as having a particular color, shape or other specific attribute in verbal form, and I considered these to be different from state verbs. They are also stative verbs, but the feature encoded is generally external and many of them cannot be gradual, contrary to most state verbs. If we were to not distinguish between the two event types, but take them as one, negative polarity would show a much higher encoding rate (68.26%) than positive polarity (32.79%) overall. This result suggests that the event typology needs to be refined on this point.

⁹ Furthermore, perception verbs are sometimes treated as state verbs, together with verbs referring to attributes/characteristics.

Table 4. The proportion of positive and negative polarity encoded in different event types

Event types	Proportion of positive polarity encoded	Proportion of negative polarity encoded
ACTION/ACTIVITY	5.53%	20.42%
HAPPENING	7.89%	24.89%
MOVEMENT	6.23%	19.66%
STATE	13.98%	58.21%
BEHAVIOR	13.49%	76.72%
ATTRIBUTE/CHARACTERISTIC	18.81%	10.05%
EXISTENCE	5.84%	18.98%
PERCEPTION	1.75%	3.51%

To illustrate the nature of the correspondences between polarity and event type and their possible motivations, in the following chapter I present examples of verbs encoding polarity and argue for their socio-pragmatic grounding.

4.2. Polarity and its socio-pragmatic grounding

There are several verbs in Hungarian expressing negative polarity and encoding eating in general, but in a particular way. A few examples would be: *csámcsog* ‘munch’, *csömöszöl* ‘eat greedily and improperly much, ram’, *dőzsöl* ‘eat immoderately at a feast’, *fal* ‘eat greedily’, *güzmölődik* ‘eat idly’, *habzsol* ‘eat hastily and greedily, often loudly, devour’, *malackodik* ‘eat improperly, staining one’s clothes’, *nyámmog* ‘eat slowly, unwillingly’, *pofázik* ‘eat greedily’, *bezabál* ‘eat greedily, hastily, and/or immoderately much’, *felzabál* ‘eat up something greedily, like a gobbler’. As seen from the examples, eating too much and greedily, too fast or too slowly, and loudly or unwillingly are characteristics of eating that (in Hungarian) need to be spoken about specifically and succinctly (though not very often as a quick search in the Hungarian Gigaword Corpus shows).

By contrast, there are only a few eating verbs that encode positive polarity: *eddegél* ‘eat slowly, leisurely’, *eszik-iszik* ‘eat and drink to one’s liking’, *falatozik* ‘eat calmly and in good taste’, *jóllakik* ‘eat until satiated’ and *lakmározik* ‘eat copiously and in good taste’. As shown in the explanations, these meanings refer to situations when eating is calm, copious and generally in good taste. We should note nonetheless that there is no verb to express only eating in good taste. “Positive” eating (in general) can only be expressed with adverbial phrases of the verb *eszik* ‘eat’. I believe this to be a kind of linguistic efficiency, where the expected form of the activity is expressed by the more general and the most frequently used verb (*eszik* ‘eat’), expressing the default form of eating, without added components, and all important/relevant alterations are encoded in separate lexemes to make possible their rapid, brief phrasing. Faber and Mairal Usón (1997b: 25) also observe that “in language we tend to

find words for what draws our attention, and what surprises us is generally what differs from our expectations or deviated from the socially-expected norm". This type of markedness (and unmarkedness) fits the more general linguistic view of semantic markedness. This is related to human perception and the conceptual processing of linguistic expressions, and the degree of markedness depends on the ease of their mental processing (Ladányi 2017: 520). Although not enumerated among the cases of semantic markedness, the default form of events as a case of semantic unmarkedness easily fits the list, and it might have both universal linguistic and typological motivations. Semantic markedness of events does not necessarily motivate the morphological structure of linguistic expressions (cf. Ladányi 2017: 521), but are basically expressed in entrenchment and lexical richness. Semantically unmarked forms of events are more entrenched and lexically less rich, while semantically marked forms of events are less entrenched and have a richer lexical representation. However, there are cases where semantic markedness of events in lexical elements is expressed in the morphological structure also. See the example of verbs with a negative suffix expressing scarcity below.

In this view of semantically marked forms of events, besides affective content and evaluative meaning, although strongly connected to these, negative polarity verbs encode non-preferred alterations to the event, while positive polarity verbs encode preferred alterations or outstanding forms of the event. For an event to be outstanding in some respect, is relevant in fewer contexts, than excelling in non-preferred ways. Therefore, the latter cases are more often lexicalized than the former cases, and the expected form is expressed by a verb that is basic category-like. Thus, the richer lexical representation of negative polarity in the verbal domain is motivated and, in some cases, can be explained not by affective or evaluative content per se, but by socio-cultural factors, especially expectations regarding the typical and preferred features of events. This is also the case with behavior verbs, but also speaking verbs, verbs expressing movement, mental activity and others.

In the view of the negativity bias, the richer lexical representation of the negative experiences might be motivated on the neural level of attention allocation (Jing-Schmidt 2012: 423). The negativity bias refers to "an automatic tendency to pay significantly more attention to unpleasant than pleasant information" (Jing-Schmidt 2012: 417). It has been observed on many areas of life that "negative events have a greater impact on people's behaviour than positive events" (Jing-Schmidt 2012: 418). This goes in line with our result regarding the richer lexical representation of the negative experiences, but we must notice that not all negative polarity words refer to experiences that would necessarily frighten us. Just think about eating improperly or greedily. It does not actually refer to an activity which one would avoid at all costs. I believe, this is where a view of negative polarity as alteration to the more common or

preferred way of events can help complete the picture. Not incidentally, this view can be easily reconciled with the negativity bias. If negative is more important in terms of behavior, because “bad signals the need for change” (Jing-Schmidt 2012: 419), it might also be useful to have a richer lexical representation for it, and thus to be able to briefly and succinctly talk about it if needed. Moreover, Jing-Schmidt (2012: 429) also notes that in the psychology of fear, novelty or strangeness is considered one of the natural clues of fear. In parallel to this, familiar and common experiences are not related to strong emotions, but in turn, they are “more entrenched and more expected in verbal communication” (Jing-Schmidt 2012: 434). Naturally, I do not think that an unusual way of eating is something to be feared, but in a broader view of language use, the example of eating verbs falls in line with emotion words both in terms of their relation to the familiarity or novelty/strangeness of experiences and in terms of their morphological features. This latter strengthens the fact that polarity should be regarded as a wider affective-cognitive-social category with some linguistic expressions showing stronger and others showing weaker polarity, and/or stronger or weaker intensity, with some units not even related to emotions, but to expected/not expected forms of events.

The morphological structure of some Hungarian verbs also makes a case for polarity as an alteration to the default form of events. These are derived verbs in the dataset, most of which express negative polarity with a positive or neutral polarity stem and a negative polarity suffix. These verbs (and many others) refer to some kind of inadequate or unfit social behavior, while having a negative suffix expressing scarcity: *-talan/-telen*, *-(a)tlan/- (e)tlen*. That is, they have the following structure: positive/neutral stem + negative suffix + verbal suffix (e.g., *bizalm-atlan-kodik* ‘distrust’, from the stem *bizalom* meaning ‘trust’, *enged-etlen-kedik* ‘disobey’, from the stem *enged* meaning ‘allow’, *ill-etlen-kedik* ‘behave inappropriately’, from the stem *illő* meaning ‘appropriate’, *szemérm-etlen-kedik* ‘be/act shameful’, from the stem *szemérem* meaning ‘modesty’, *udvari-atlan-kodik* ‘be/act impolite’, from the stem *udvarias* meaning ‘polite’).

Moreover, there are no verbs in Hungarian to express the positive equivalent of these actions/behaviors. Instead, besides expressing scarcity, verbs can also refer to the exaggeration of an action, again with a negative connotation, even without a negative suffix. This can be illustrated by *bizalm-as-kodik* ‘take liberties with someone’ (from the stem *bizalom* ‘trust’) and *szemérm-es-kedik* ‘being/acting too shy’ (from the stem *szemérem* ‘modesty’) among others. Here, neither the stem, nor the suffix encode negative polarity. What is more, both verbs also have a form with a positive stem + negative suffix + verbal suffix structure (see *bizalm-atlan-kodik* and *szemérm-etlen-kedik* above). So that there is a verb to express the lack of trust/shame, and there is a verb to express the

abundance of these, but there is no verb to express the adequate amount of trust/shame in someone's behavior.¹⁰

Some other examples to support the above claim, would be verbs which have polarity words in their dictionary entry, but would not be generally considered encoding affective or evaluative content: *normalizálódik* 'the situation settles down, gets back to **normal**_[+]' (positive polarity), *stagnál* 'to stagnate, **not to evolve**_[-]' (negative polarity), *különködik* 'to behave **eccentrically**_[-]' (negative polarity), *támogat* '**to support**_[+] financially, morally' (positive polarity), *szövegel* 'to talk a lot or **unnecessarily**_[-]' (negative polarity), *beskatulyáz* 'to **hastily**_[-] or **forcefully**_[-] label someone/something' (negative polarity), etc. In the view of lexicalizing alterations to the default form of experiences, the polarity of these verbs becomes transparent and its motivations clear. Naturally, polarity is a scale, and these verbs do not show strong polarity or high intensity. Nonetheless, they do fit the case of polarity as an alteration to the default form of events. Behavior that adheres to societal norms, which does not surprise, is not something that needs to be remarked upon. While uncommon, unexpected manifestations, demeanors, manners and attitudes are experiences that attract attention, and at times need to be spoken about, preferably with precision, but succinctly.

We can see a similar case with sound symbolic-verbs. They can be considered marked expressions in terms of structural iconicity (Ladányi 2017: 524–526), and they refer to unfamiliar, uncommon, even odd forms of activities (see for example *boggle*, *bumble*, *lumber*, *totter*, *trudge*).

Some adjectives also underpin the claim that negative polarity is related to alterations to the default (familiar/expected/preferred) form of events. These do not refer to negative affect or attitude, but their structure and meaning are analogical to that of negative emotive words. Adjectives such as *unusual*, *unchanging*, *unmarried*, *unexpected* are the case here. We see them in Boucher and Osgood (1969), who analyze positive and negative evaluative words in their study on semantic differential. However, these adjectives have an unclear evaluation in their work (1969, 7). This is not surprising, since *unusual*, *unchanging* and *unmarried* are most likely considered to be neutral in terms of affective content, which is the subject of interest of their study. Nonetheless, the structure of the adjectives shows the effect of the Pollyanna hypothesis, in that negative polarity affixes are added to positive or neutral polarity stems, this being characteristic of negative polarity elements when derived from positive or neutral polarity items. I believe that the results of Boucher and Osgood

¹⁰ Naturally, there are also verbs with a negative stem and a negative suffix expressing the presence of a negative emotion or situation. E.g., *erőszakoskodik* 'be violent' with the stem *erőszak* 'violence', *finnyáskodik* 'be fastidious' with the stem *finnyás* 'fastidious', *garázdálkodik* 'be rowdy' with the stem *garázda* 'rowdy' etc. The detailed and extensive analysis of all verbs with the *-kodik/-kedik/-ködik* suffix is not carried out here.

(Osgood 1964; Boucher and Osgood 1969) also point to the fact that socio-cultural expectations are also reflected in the structure and distribution of words encoding polarity.

Boucher and Osgood present evidence based on data from 13 languages and a large scale of positive–negative oppositions that “there is a universal human tendency to use evaluatively positive (E+) words more frequently, more diversely and facily, than evaluatively negative (E–) words” (Boucher and Osgood 1969, 1). This is in line with the findings of the more recent study of Garcia, Garas and Schweitzer (2012). Their comprehensive analysis of 1 034 English, 2 902 German and 1 034 Spanish affective words as used in internet texts, “together with their emotional scores obtained from extensive human ratings” (2012, 3), show that words with a positive emotional content are used more often in the languages studied, and they carry less information in their semantic content in contrast to negative emotion words. Although somewhat different considerations are used in defining the terms *positive* and *negative polarity*, analyses on different subsets of linguistic elements show similar results and a higher frequency of usage of positive polarity words (see also Jing-Schmidt [2007] regarding the positivity bias).

Communicative functions of language leave their “imprint” (Diessel 2017) not only on linguistic structure in the sense of constructional patterns, but also on the overall semantic network of the lexicon. This is because experience not only determines how we talk, but also what is there to be talked about and what is worth talking about. It is not only linguistic structure that makes use of statistical information for speaking, but the occurrence of specific experiences or experience types too. The impact of these can be grasped, among others, by looking at the distribution of lexical items encoding polarity.

Perhaps our human trait that we feel comfortable in familiar situations, also pertains to the problem, in that unusual, unchanging and so on, are among the conspicuous, unexpected, perhaps unwanted situations, and so, among the experiences relevant to be talked about. In terms of the human psyche, what is frequent is good, because is known, and thus one knows what to expect. In this way, words expressing experiences known to us (most likely because they are frequent, common in our everyday experiences), fall in the same category as words expressing positive experiences. (Unless, of course, if they are really dangerous or frightening. Those could not be considered positive, even after becoming familiar with them.)

In the light of the above enumerated examples, I believe we can conclude that alterations to the default mode of experiences are an important organizational principle in the lexicon in general, and it should be considered a part of linguistic polarity phenomena along with affective-emotive content. This is substantiated by structural and semantic features of the linguistic units analysed in the paper, which all fit with the theory of semantic markedness (cf.

Ladányi 2017) and the negativity bias pattern (cf. Jing-Schmidt 2007, Kugler 2014). In this light, the definition of negative polarity should be refined by adding non-preferred or not expected alterations to the default forms of events, and positive polarity by adding preferred or outstanding alterations to its definition. By *default*, here we mean that it is familiar, considered appropriate, suitable, or fitting the socio-cultural norms and subjective expectations of the speaker. This is mainly expressed by the most frequent basic level-like verb for a certain event.¹¹

Thus, in verbal meaning (but most certainly in other categories too) 'positive' and 'negative' should be understood not only in terms of positive and negative emotive content, approved or disapproved behavior, positive or negative moral or subjective evaluation of things and experiences (cf. Martin and White 2005), but more broadly as compliance to the preferred and non-preferred form of experiences.

After having a wide-spread look at polarity in verbal meanings, in the next chapter I touch upon the question whether the semantic load of polarity verbs shows correlations with word length, as surmised by the Pollyanna hypothesis and Zipf's law.

4.3. *Polarity and word length*

As a sequel to the finding that in the verbal system negative polarity is lexically more richly represented in the lexicon than positive polarity, I also used the dataset to test whether negative polarity verbs are longer on average than positive polarity verbs. This is a follow-up to the Pollyanna hypothesis and Zipf's law, the first stating that evaluatively positive terms are used more often and more diversely than evaluatively negative words, and the second stating that words used more frequently are shorter in length and carry less information (Boucher and Osgood 1969; Zipf 1936, 1949). We can infer from these two that evaluatively negative words carry more information and hence are longer in average than evaluatively positive words.

I have used the dataset to make a preliminary assessment of the possible correlations by counting word length in two ways: based on the constituent sounds and based on the constituent syllables.¹² However, no significant difference showed up between the two methods. I only present them both below for the sake of accuracy.

¹¹ Cf. Faber and Mairal Usón: "Although semantic parameters codifying sociocultural norms are different..., it is relatively easy to deduce what such norms are, because they are precisely what is not lexicalized" (1997b: 25, highlight in the original).

¹² Consonants denoted with letter combinations in Hungarian (*cs, dz, gy, ly, ny, sz, ty, zs* etc) were treated as one sound.

The results show no substantial difference in word length between positive and negative polarity verbs, but the ANOVA tests¹³ suggest that the correspondences of word length and polarity is not random in the data ($F(2, 29902) = 102,2402$, $p = 0,0000$ for the correspondences of polarity and word length counted with constituent sounds, and $F(2, 29902) = 83,7886$, $p = 0,0000$ for the correspondences of polarity and word length counted with constituent syllables, respectively).

There is a slight variation with regard to polarity, with evaluatively negative verbs being on average longer than evaluatively neutral verbs, but they are also shorter than evaluatively positive verbs. The table below presents the exact numbers, and the figures show the distribution of the datapoints in a means plot.

Table 5. Average word length in the data

	Average word length counted with constituent sounds	Average word length counted with constituent syllables
Evaluatively negative verbs	7.76	3.07
Evaluatively positive verbs	7.83	3.12
Evaluatively neutral verbs	7.36	2.92

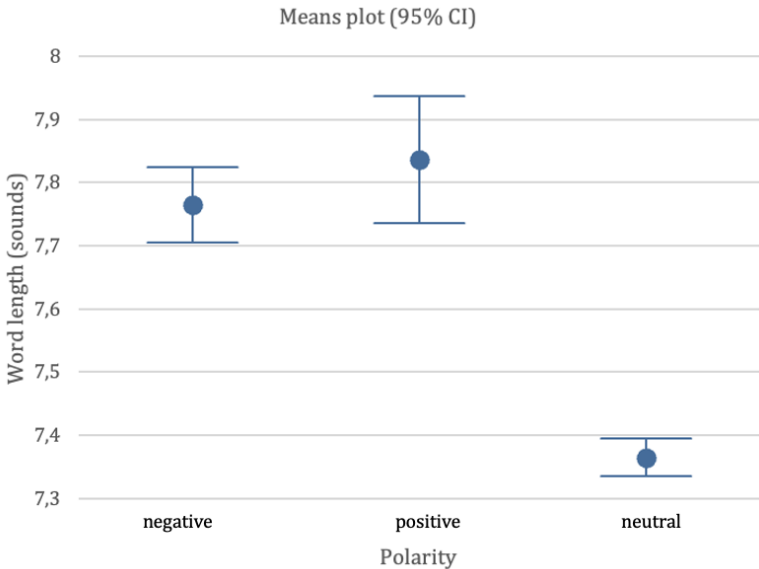


Figure 1. Distribution of polarity and word length (constituent sounds)

¹³ One-way ANOVA using Microsoft Excel's StatPlus add-in. I have tested for the correlation between polarity and word length using constituent sounds and syllables too, with similar results.

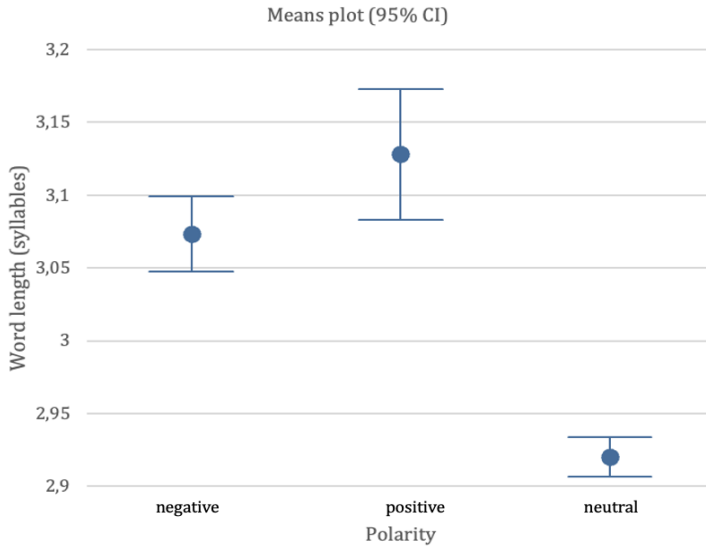


Figure 2. Distribution of polarity and word length (constituent syllables)

We must note first of all, that positive polarity verbs are somewhat longer on average than negative polarity verbs, which challenges the hypothesis that negative polarity verbs are longer. Nonetheless, neutral verbs are shorter than verbs with more semantic load (positive and negative polarity). Thus, we can conclude that in the view of the verbal lexicon of Hungarian, shorter words possibly also carry less information. This statement parallels Zipf's law, according to which words used more frequently are shorter in length and carry less information, although on the basis of the present dataset we cannot make inferences regarding frequency of use, and that is not the aim here.

Naturally, it might also be the case that when looking at the lexicon of a language, there is no difference in word length between positive and negative polarity verbs, because in both cases there is a semantic overload compared to non-polarity verbs. This way, we should see a difference only between polarity verbs and non-polarity verbs – which the present analysis suggests. Taking into consideration that the omega squares are 0,0067 and 0,0055 respectively (suggesting that only 0,06 and 0,05 percent of the variation among the length of the verbs is explained by polarity), it seems obvious that there is more to the relation of word length and polarity, given that polarity is not the only linguistic feature that might be correlated with word length, and so the correspondence between the two factors needs further investigation. This claim is corroborated by post hoc tests, which almost all show correlation between positive and

neutral, and negative and neutral categories. These correlations are not discussed here, because of the nature of the study, but descriptive statistics can be found in Appendix 1 for information. The correspondences of word length and polarity will be studied in more depth in the future.

Naturally, we must also take into consideration that perhaps the data is not representative, and with the use of more annotators the distribution of verbs encoding polarity would be somewhat different. This could lead to different correlations. In all cases further analyses are required to test the correlations between word length and polarity. The preliminary study shows most importantly that correspondences regarding word length and polarity in the view of the lexicon (as opposed to frequency studies) must also be tested on a somewhat different dataset: polarity verbs versus non-polarity verbs. Also, other factors must be considered in the investigation of this problem, which could be correlated with word length.

5. Conclusions

In the paper I presented the preliminary results and the important considerations of a larger study regarding the distribution of polarity in the verbal domain of Hungarian, with possibilities of generalizing the results to other domains of the language and other languages. The study of the problem is motivated by the long-standing hypothesis that negative polarity has a richer lexical representativity than positive polarity on a lexical, constructional and idiomatic level as well. To be able to test this, I have created a database of 29 905 Hungarian verb meanings and coded them for polarity, using one annotator (up to this moment) and a three-category typology: “positive”, “negative” and “neutral”.

The preliminary results of the investigation show first of all that, in the verbal domain, indeed negative polarity is lexically more diverse than positive polarity. Naturally, this correspondence should also be tested on data from more annotators, and this is planned in the future, but we should note that other scholars have also observed this distribution among verbs in other languages. Faber and Mairal Usón (1997b: 25) write that “in language we tend to find words for what draws our attention, and what surprises us is generally what differs from our expectations or deviated from the socially-accepted norm”; and that “the opposition of *good* and *bad* consistently appears as a structuring device within the semantic domains studied” (1997b: 30, highlights in the original). Moreover, these preliminary results fall in line with other linguistic and affective-cognitive patterns, which are worth mentioning again.

Based on the distribution of polarity among the event types and on the example of Hungarian verbs derived with a negative suffix expressing scarcity

(*-talan/-telen*, *-(a)tlan/-(-e)tlen*) and on adjectives derived with a negative suffix, but usually not considered to have negative polarity (*unusual*, *unchanging*, *unmarried*, *unexpected*), I have argued that besides affective content and evaluative meaning, the definition of polarity should also include the concept of default events, where negative polarity verbs encode non-preferred alterations to the event, while positive polarity verbs encode preferred alterations or outstanding forms of the event. For an event to be outstanding in some respect, is relevant in fewer contexts, than excelling in non-preferred ways. Therefore, the latter cases are more often lexicalized than the former cases, and the expected form is expressed by a verb that is basic category-like. Thus, the richer lexical representation of negative polarity in the verbal domain is motivated and explained not only by affective or evaluative content, but also by socio-cultural factors, especially expectations regarding the typical and preferred features of events. Naturally, the negativity bias has its imprint on the expression of negative polarity in language. But besides the fact that negative experiences have a greater impact on us than positive events, novelty or strangeness are also considered natural clues of fear (Jing-Schmidt 2012: 418, 434). Obviously, from an evolutionary and a neural perspective the negativity bias is the first and strongest motivation for the richer lexical representation of negative polarity in the affective field, but the factor of novelty/strangeness as an important trigger can easily be transferred to social experiences, and this is what might have happened.

I believe that this view is corroborated by the example of eating verbs and adjectives presented in the paper, which fall in line with emotion words both in terms of their relation to the familiarity/novelty of experiences and in terms of their morphological features. Moreover, this approach fits with the view of semantic markedness (Ladányi 2017), in the sense that semantically unmarked forms of events are more entrenched and lexically less rich, while semantically marked forms of events are less entrenched and have a richer lexical representation. Alterations to the default mode of experiences might be an important organizational principle in the lexicon, its ties with polarity in structural and semantic features being only one example of its manifestation.

Another intriguing problem investigated in the paper is the correspondence of polarity and word length. If more frequent words are shorter in length (Zipf 1936, 1949) and positive polarity words are used more frequently (Boucher and Osgood 1969), than negative polarity verbs should be longer on average than positive polarity verbs. The analysis did not confirm this hypothesis, but made clear an important aspect in the investigation of polarity distribution in the entire verbal domain (as opposed to usage data), namely that the dataset should also be able to deal with polarity verbs as one group as opposed to non-polarity verbs. It might be the case that the difference in word

length cannot be found between positive and negative polarity verbs, but among polarity and non-polarity verbs when looking at the lexicon of a language. Naturally, it might also be that this distribution is not characteristic to the lexicon of a language, but can be captured only in usage data with regard to the frequency of positive polarity expressions.

Based on these results, I believe that polarity should be regarded as a wider affective-cognitive-social category with some linguistic expressions showing stronger and others showing weaker polarity, and/or stronger or weaker intensity, with some expressions not even related to emotions, but to expected/not expected forms of events.

The research will continue based on these considerations stemming from important realizations regarding diverse and far-reaching connections of polarity in the lexicon, with a special attention to the structuring property of polarity in the lexicon, of which the details still need to be revealed.

WORKS CITED

- Bárczi Géza, and Országh László (eds.). 1959–1962. *A magyar nyelv értelmező szótára [Hungarian Explicatory Dictionary]*. Budapest: Akadémiai Kiadó.
- Boucher, Jerry, and Charles E. Osgood. 1969. "The Pollyanna Hypothesis." *Journal of verbal learning and verbal behavior* 8: 1–8.
- Bybee, Joan L., and Clay Beckner. 2009. "Usage-based theory." In *The Oxford Handbook of Linguistic Analysis*, edited by Narrog, Heiko, and Bernd Heine, 827–55. Oxford: Oxford University Press.
<https://www.unm.edu/~jbybee/downloads/BybeeBeckner2010UsageBasedTheory.pdf>
- Cambria, Erik, and Amir Hussain. 2015. *Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis*. Heidelberg–New York–Dordrecht–London: Springer. <https://sentic.net/sentic-computing.pdf>
- Diessel, Holger. 2017. „Usage-based linguistics." In *Oxford Research Encyclopedia of Linguistics*, edited by Aronoff, Mark, 1–26. New York: Oxford University Press.
- Faber, Pamela B., and Ricardo Mairal Usón. 1997a. "The paradigmatic and syntagmatic structure of the semantic field of existence in the elaboration of a semantic macronet." *Studies in language* 21, no. 1: 129–67.
- Faber, Pamela B., and Ricardo Mairal Usón. 1997b. "Towards a typology of predicate schemata in a Functional-Lexematic Model." In *Towards a Functional Lexicology*, edited by Wotjak, Gerd, 11–37. Frankfurt: Peter Lang.
- Faber, Pamela B., and Ricardo Mairal Usón. 1999. *Constructing a Lexicon of English Verbs*. Berlin–New York: Mouton de Gruyter.
- Garcia, David, Antonios Garas, and Frank Schweitzer. 2012. "Positive words carry less information than negative words." *EPJ Data Science* 1, no. 3: 1–12.
<https://epjdatascience.springeropen.com/articles/10.1140/epjds3>

- Hua, Yan Cathy, Paul Denny, Jörg Wicker, and Katerina Taskova. 2024. "A systematic review of aspect-based sentiment analysis: Domains, methods, and trends." *Artificial Intelligence Review* 57: 296. <https://link.springer.com/article/10.1007/s10462-024-10906-z>
- Giannakidou, Anastasia. 2011. "Negative and positive polarity items: Variation, licensing, and compositionality." In *Semantics: An international handbook of natural language meaning*, edited by Maienborn, Claudia, Klaus von Heusinger, and Paul Portner, 1660–712. Berlin: Mouton de Gruyter.
- Jing-Schmidt, Zhuo 2007. "Negativity bias in language: A cognitive-affective model of emotive intensifiers." *Cognitive Linguistics* 18, no. 3: 417–443.
- Mingorance, Martín L. 1990. "Functional grammar and lexematics." In *Meaning and Lexicography*, edited by Tomaszczyk, Jerzy, and Barbara Lewandowska-Tomaszczyk, 227–86. Amsterdam/Philadelphia: John Benjamins.
- Martin, Jennifer R. and Peter Rupert Robert White. 2005. *The language of evaluation. Appraisal in English*. New York: Palgrave Macmillan.
- Kugler Nóra 2014. "A nyelvi polaritás kifejezésének egy mintázata, avagy milyen a *félelmetesen jó*?" *Magyar Nyelvőr* 138, no. 2: 129–139.
- Ladányi Mária. 2017. "Alaktan." In *Nyelvtan [Grammar]*, edited by Tolcsvai Nagy Gábor, 503–660, Budapest: Osiris.
- Langacker, Ronald W. 2008. *Cognitive grammar. A basic introduction*. New York: Oxford University Press.
- Oravecz Csaba, Váradi Tamás, and Sass Bálint. 2014. The Hungarian Gigaword Corpus. In: *Proceedings of LREC 2014*.
- Osgood, Charles E. 1964. "Semantic differential technique in the comparative study of cultures." *American Anthropology* 66: 171–200.
- Princeton University. 2010. *About WordNet*. WordNet. Princeton University. <https://wordnet.princeton.edu/>
- Pusztai Ferenc. 2006. *Magyar értelmező kéziszótár [Hungarian explicatory dictionary]*. 2nd, extended edn. Budapest: Akadémiai Kiadó.
- Szabó Martina, Vincze Veronika, and Bibok Károly 2023. "Thank you for the terrific party!" – An analysis of Hungarian negative emotive words. *Corpus Linguistics and Linguistic Theory* 19, no. 3: 451–485.
- Szilágyi N., Sándor, and Dimény Hajnalka. 2021. *Magyar igék szemantikai adatbázisa [Semantic database of Hungarian Verbs]*. Cluj-Napoca: Babeş-Bolyai University.
- Talmy, Leonard. 2000. *Toward a cognitive semantics I. Concept structuring systems*. London: MIT Press.
- Tolcsvai Nagy Gábor. 2016. "Van: változatok a létezés nyelvi konstruálására [Van 'there is': Variations of the linguistic construal of existence]." *Magyar Nyelvőr* 140, no. 1: 7–22.
- Wierzbicka, Anna. 1999. *Emotions across languages and cultures*. New York: Cambridge University Press.
- Windisch Brown, Susan, Claire Bonial, Leo Obrst, and Martha Palmer. 2017. "The Rich Event Ontology." In *Proceedings of the Events and Stories in the News Workshop*, edited by Caselli, Tommaso, Ben Miller, Marieke van Erp, Piek Vossen, Martha Palmer, Eduard Hovy, Teruko Mitamura, and David Caswell, 87–97. Vancouver: Association for Computational Linguistics. <https://aclanthology.org/W17-2712/>

Zipf, George Kingsley. 1936. *The Psychobiology of Language*. London: Routledge.

Zipf, George Kingsley. 1949. *Human Behavior and the Principle of Least Effort*. New York: Addison-Wesley Press.

Appendix 1.

Descriptive statistics of the ANOVA analyses.

Comparisons among groups (Factor 1 - Polarity)					
Polarity vs. word length counted with constituent sounds					
Scheffe					
<i>Group vs. Group (Contrast)</i>	<i>Difference</i>	<i>95% Confidence Interval</i>		<i>Test Statistic</i>	<i>p-value</i>
Neg vs Pos	-0,0717	-0,2113	0,0679	1,2568	0,4540
Neg vs Neu	0,4000	0,3200	0,4799	12,2451	0,0000
Pos vs Neu	0,4717	0,3448	0,5985	9,1029	0,0000
Tukey-Kramer					
<i>Groups</i>	<i>Difference</i>	<i>Test Statistic</i>	<i>p-value</i>	<i>Significant</i>	
Neg vs Pos	-0,0717	1,7774	0,8527	No	
Neg vs Neu	0,4000	17,3172	0,7460	No	
Pos vs Neu	0,4717	12,8734	0,7460	No	
Bonferroni					
<i>Alpha/N</i>	0,0167				
<i>Group vs. Group (Contrast)</i>	<i>Difference</i>	<i>95% Confidence Interval</i>		<i>Test Statistic</i>	<i>p-value</i>
Neg vs Pos	-0,0717	-0,2083	0,0649	1,2568	0,6265
Neg vs Neu	0,4000	0,3218	0,4782	12,2451	0,0000
Pos vs Neu	0,4717	0,3476	0,5957	9,1029	0,0000
Fisher LSD					
<i>Group vs. Group (Contrast)</i>	<i>Difference</i>	<i>Test Statistic</i>	<i>p-value</i>	<i>Significant</i>	
Neg vs Pos	-0,0717	1,2568	0,2088	No	
Neg vs Neu	0,4000	12,2451	0,0000	Yes	
Neg vs Neu	0,4717	9,1029	0,0000	Yes	
Dunnett's test (Control Group = 1)					
<i>Group vs. Group (Contrast)</i>	<i>Difference</i>	<i>Test Statistic</i>	<i>p-value</i>	<i>Significant</i>	
Pos vs Neg	0,0717	1,2568	0,3534	No	
Neu vs Neg	-0,4000	12,2451	-0,0016	Yes	

Comparisons among groups (Factor 1 - Polarity)					
Polarity vs. word length counted with constituent syllables					
Scheffe					
<i>Group vs. Group (Contrast)</i>	<i>Difference</i>	<i>95% Confidence Interval</i>		<i>Test Statistic</i>	<i>p-value</i>
Neg vs Pos	-0,0547	-0,1169	0,0075	2,1544	0,0982
Neg vs Neu	0,1530	0,1174	0,1887	10,5215	0,0000
Pos vs Neu	0,2078	0,1513	0,2643	9,0047	0,0000
Tukey-Kramer					
<i>Groups</i>	<i>Difference</i>	<i>Test Statistic</i>	<i>p-value</i>	<i>Significant</i>	
Neg vs Pos	-0,0547	3,0468	0,7661	No	
Neg vs Neu	0,1530	14,8797	0,7460	No	
Pos vs Neu	0,2078	12,7345	0,7460	No	
Bonferroni					
<i>Alpha/N</i>	0,0167				
<i>Group vs. Group (Contrast)</i>	<i>Difference</i>	<i>95% Confidence Interval</i>		<i>Test Statistic</i>	<i>p-value</i>
Neg vs Pos	-0,0547	-0,1156	0,0061	2,1544	0,0936
Neg vs Neu	0,1530	0,1182	0,1879	10,5215	0,0000
Pos vs Neu	0,2078	0,1525	0,2630	9,0047	0,0000
Fisher LSD					
<i>Group vs. Group (Contrast)</i>	<i>Difference</i>	<i>Test Statistic</i>	<i>p-value</i>	<i>Significant</i>	
Neg vs Pos	-0,0547	2,1544	0,0312	No	
Neg vs Neu	0,1530	10,5215	0,0000	Yes	
Neg vs Neu	0,2078	9,0047	0,0000	Yes	
Dunnett's test (Control Group = 1)					
<i>Group vs. Group (Contrast)</i>	<i>Difference</i>	<i>Test Statistic</i>	<i>p-value</i>	<i>Significant</i>	
Pos vs Neg	0,0547	2,1544	0,0585	No	
Neu vs Neg	-0,1530	10,5215	-0,0012	Yes	

LITERATURE'S RESPONSE TO INFORMATION OVERLOAD IN DON DELILLO'S *WHITE NOISE*

Cristina Mirela NİCOLAESCU¹ 

*Article history: Received 27 July 2025; Revised 16 November 2025; Accepted 25 November 2025;
Available online 12 December 2025; Available print 30 December 2025*

©2025 Studia UBB Philologia. Published by Babeş-Bolyai University.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ABSTRACT. *Literature's Response to Information Overload in Don DeLillo's White Noise.* The present article examines the novel *White Noise* through the intersecting theoretical frameworks of Paul Virilio's dromology, and Jean Baudrillard's simulacral theory, proposing that DeLillo constructs an epistemological model of modern life as mediated risk perception. Virilio's theories on the acceleration of information flow frame the novel's temporal anxiety, whilst Baudrillard's notion of hyperreality explains the characters' dependence on simulated knowledge. Together, these frameworks reveal how DeLillo's fiction translates abstract theoretical notions into lived experience, positioning narrative as both a mirror and critique of late modern consciousness. This paper argues that in *White Noise*, DeLillo transforms narrative fragmentation and sensory saturation—both formal strategies and moral critiques of late-capitalist modernity—into a distinctive literary aesthetics of excess. By examining how his style enacts a crisis of perception, the study demonstrates the impossibility of coherence in a hypermediated age, drawing on theoretical frameworks from Virilio and Baudrillard to illuminate the sociocultural stakes of perceptual excess.

Keywords: *postmodern fiction, consumerism, digital culture, media saturation*

¹ **Cristina Mirela NİCOLAESCU** has a PhD in English Literature on Postmodern Fiction. Her major interests are postmodern fiction, narrative theory, media theory, and cultural studies. She holds a professor position at the Department of Western Languages, Yozgat-Bozok University, Türkiye. She received a PhD in 2013 from the University of Bucharest, Romania. The research findings in the field of English Literature have been presented in international conferences and symposia, some of them published as academic articles in indexed journals. She authored a few academic books and over 50 research articles. The most representative books are: *Technologies of Perspective on Contemporary Literary Texts*, Cambridge Scholars Publishing UK, 2022, ISBN 978-1-5275-8666-6 and *Iconicity and Arbitrariness of Linguistic Expressions in Speech Acts*, Cambridge Scholars Publishing UK, *Female Characters in Regency and Victorian Literature*, Nobel Bilim Yayın Evi, Türkiye, 2021, ISBN 978-6257492-85-0, ISBN: 978-1-0364-4166-1, 2025; Email: chris2013x@yahoo.com

REZUMAT. Răspunsul literaturii la supraîncărcarea informațională în *White Noise* de Don DeLillo. Prezentul articol examinează romanul *White Noise* prin intermediul cadrelor teoretice intersectate ale dromologiei lui Paul Virilio și ale teoriei simulacului formulate de Jean Baudrillard, propunând că DeLillo construiește un model epistemologic al vieții moderne ca percepție mediatizată a riscului. Teoriile lui Virilio privind accelerarea fluxului informațional conturează anxietatea temporală a romanului, în timp ce noțiunea lui de hiperrealitate a lui Baudrillard explică dependența personajelor de cunoașterea simulată. Împreună, aceste cadre evidențiază modul în care ficțiunea lui DeLillo traduce noțiuni teoretice abstracte în experiență trăită, poziționând narațiunea ca oglindă și critică a conștiinței modernității târzii. Lucrarea susține că în *White Noise* DeLillo transformă fragmentarea narativă și saturația senzorială — atât strategii formale, cât și critici morale ale modernității târziu-capitaliste — într-o estetică literară distinctă a excesului. Prin examinarea modului în care stilul său înscenează o criză a percepției, studiul demonstrează imposibilitatea coerenței într-o epocă hipermediată, valorificând cadre teoretice din Virilio și Baudrillard pentru a ilumina mizele socioculturale ale excesului perceptiv.

Cuvinte-cheie: ficțiune postmodernă, consumism, cultură digitală, saturație mediatică

Introduction

According to recent scholarship on DeLillo, *White Noise* has been reexamined through the lenses of algorithmic governance, postdigital anxiety, and ecological precarity. Such readings suggest that DeLillo's insights into media saturation have grown even more resonant in the twenty-first century. The novel is situated within the emergence of what Boxall (2012) and Osteen (2023) describe as "the data sublime," a condition in which information itself becomes an agent of ontological uncertainty. This line of interpretation has been further developed in the context of digital surveillance by Kavadlo (2022), who argue that DeLillo's recurring motifs of the supermarket and television anticipate the algorithmic architectures of the cloud era.

White Noise has been extensively examined in relation to consumerism, mass media, and simulation (Toth 2021). However, few studies have investigated how DeLillo's stylistic techniques—his polyphonic narrative, tonal flattening, and recursive syntax—formally enact the sensory and interpretive saturation his prose depicts. By offering a detailed formal analysis of overload as both a theme and narrative strategy, this study addresses that gap. Moreover, it engages with recent developments in media ecology and cognitive literary studies (Zunshine 2021; McCann 2023; Melley 2024) to argue that DeLillo's representation of perceptual excess anticipates the attention crises and algorithmic mediation characteristic of twenty-first-century life.

In an age defined by algorithmic saturation, environmental crisis, and the fragmentation of public discourse, Don DeLillo's *White Noise* continues to serve as one of the most incisive explorations of informational excess in late capitalist culture. The novel's prescient depiction of media excess—television, radio, computer screens, and the constant hum of background data—has only gained relevance in the digital era. DeLillo's world, in which information circulates endlessly and meaning collapses under its own weight, anticipates what Shoshana Zuboff (2019) identifies as the “age of surveillance capitalism,” where knowledge and identity are commodified into data streams.

Recent DeLillo scholarship (Boxall 2012; Duvall 2008; Kavadlo 2022; Cowart 2022; LeClair 2020; Osteen 2023) has renewed attention to *White Noise* as a novel that anticipates twenty-first-century information economies. By synthesizing these critical insights with Virilio's, and Baudrillard's theories, this article situates *White Noise* at the intersection of media ecology, risk epistemology, and posthuman subjectivity.

Ultimately, this paper argues that *White Noise* transforms the experience of *white noise* from a technological symptom into an existential condition—a state in which identity, perception, and mortality are redefined through mediation. The novel enacts, rather than merely describes, the saturation of human consciousness by external data systems. Through its fragmented narrative, recursive dialogue, and ironic tone, *White Noise* invites readers to inhabit the dissonance of meaning in an era where information itself becomes the primary substance of reality. This argument both extends and revises existing DeLillo scholarship by proposing that *White Noise* models the informational subject as the defining figure of postmodern and postdigital life.

A common theme in many works that examine media saturation, and fragmented consciousness is the metaphor of *noise*, which can be generally interpreted as sensory, informational, linguistic, or existential clutter. Don DeLillo provides a prophetic literary description of a cultural phenomenon, capturing the chaotic and fractured texture of late twentieth-century life. The characters are surrounded by an unrelenting stream of information, which includes everything from emergency announcements to supermarket price tags, from academic jargon to consumer advertising, and from television broadcasts to store signs. The lines separating meaningful communication from meaningless noise are blurred in this noisy setting. The idea of information overload is not new, however, new forms of cognitive, affective, and cultural exhaustion have been brought about by the specific forms this situation has taken in the late 20th and 21st centuries, which are marked by the widespread use of digital communication, electronic media, and advertising. *White Noise* is remarkably prescient in capturing these settings.

The media ecology tradition, especially Marshall McLuhan's well-known statement that "the medium is the message" (McLuhan 2021, 7), is in line with DeLillo's work. In *White Noise*, media technologies influence not just how information is disseminated but also how people think and feel. The novel depicts the breakdown of the boundaries between representation and reality, which Jean Baudrillard defined as the emergence of the simulacrum in a hypermediated society:

Postmodernity is said to be a culture of fragmentary sensations, eclectic nostalgia, disposable simulacra, and promiscuous superficiality, in which the traditionally valued qualities of depth, coherence, meaning, originality, and authenticity are evacuated or dissolved amid the random swirl of empty signals (Baudrillard 1994, 19).

However, *White Noise* does more than just portray these events; it enacts them through its own formal methods, such as its recursive structures, paratactic narrative, and characters' fixation on pictures, codes, and signals. At the core of this novel is the belief that literature continues to be an essential platform for contemplating and navigating the intricacies of modern communication, either in spite of or as a result of its interaction with these forces. Literary narratives emphasize the qualitative experience of living in these conditions, while some critical views of excess concentrate on the quantitative proliferation of data. The impact of noisy information on relationships, memory, identity, and even mortality perceptions is masterfully dramatized. The well-known "airborne toxic event" literally depicts the subject's widespread but imperceptible contamination by invisible forces, whether they be ideological, chemical, or informational: "The flood of information and distraction constantly reshapes the way our minds work" (Carr 2010, 19). By foregrounding the relationship between narrative fragmentation and cultural saturation, this study expands the field's understanding of how DeLillo's prose mediates the psychological consequences of postmodern excess.

Methodology

While four decades of criticism have canonized *White Noise* as the quintessential postmodern novel, the originality of this paper lies in reframing its narrative form, rather than its content, as the primary site of cultural critique. Earlier scholars—such as Duvall, Cowart, Boxall, and LeClair—have interpreted DeLillo as a chronicler of systems and mediated experience, yet their analyses often remain at the level of thematic or sociological observation. In contrast, this

study argues that DeLillo's stylistic and rhythmic excess constitutes a literary enactment of cognitive and informational overload, transforming form itself into critique.

By integrating Virilio's theory of acceleration and Baudrillard's concept of simulation, this paper adopts an interdisciplinary methodology that bridges literary form, cultural theory, and the philosophy of technology. This framework situates *White Noise* simultaneously within postmodern aesthetics and within contemporary debates on digital culture, attention economies, and the phenomenology of media saturation.

The contribution of this study is therefore twofold: first, it extends the formalist interpretation of DeLillo's prose through the lens of late-modern theoretical paradigms; and second, it demonstrates the novel's continued relevance for understanding the ontological and perceptual conditions of twenty-first-century mediated life. Ultimately, this approach builds upon the foundational insights of DeLillo scholarship while introducing a new conceptual vocabulary—the literature of excess—to describe the persistence of overload aesthetics in the postdigital age.

The literature of excess

Literature has used a variety of artistic techniques to address the dilemma of information overload. In postmodernism, pastiche, intertextuality, and narrative recursion—forms that mirror the saturation of signs without attempting to resolve it—supplanted the fragmentation of modernism. The novel's serial format, rambling conversations, and use of media snippets serve as excellent examples of these tactics. Its plot resists conclusion, reflecting the continuous flow of information, and its characters communicate in soundbites, lists, and slogans. Last but not least, this article addresses the larger issue of literature's place in an era of excess as well as DeLillo's book. Literature continues to be an essential discipline for slowing down, attending, and reflecting in a time where data threatens to surpass meaning, attention is a limited resource, and noise spreads more quickly than signal. One such technique is provided by *White Noise*, a story that struggles with the impossibility of creating meaning in a world full of tools for doing so.

This idea appears to be an unavoidable aspect of contemporary living in the era of big data, continuous notifications, and digital surveillance. However, authors and theorists understood the cultural, cognitive, and existential ramifications of living in an information-rich environment long before smartphones and social media. One of the most insightful literary reactions to this phenomenon

is Don DeLillo's novel, which captures the confusion and concerns of a culture overrun by advertising, media signals, technical noise, and competing narratives. As a novel that reflects, embodies, and critiques the fragmentation of meaning, the erosion of subjectivity, and the crisis of epistemology brought about by the saturation of media and data in postmodern culture, this study argues that *White Noise* is a paradigmatic work in this sense. DeLillo's portrayal of media saturation is represented as the pervasiveness of radio and television creates a semiotic environment where the lines separating representation and reality become increasingly hazy.

The fear of too much information existed before the twentieth century. The speed, saturation, and ubiquity of information, which are made possible by electronic media, are what set the late 20th century—and by extension, *White Noise*—apart from other periods. This change is encapsulated in Jean Baudrillard's idea of the simulacrum, which describes a world in which images circulate as self-referential signs within a closed system of media and commodities rather than referring to any underlying reality. Through their never-ending streams of messages and signs, the supermarket, television, and radio are not only incidental elements; rather, they are active agents that shape perception and create reality: "We live in a world where there is more and more information, and less and less meaning" (Baudrillard 1994, 19).

From the disintegration of modernism to the parodic pastiche of postmodernism, DeLillo's novel is one of a long tradition of literary reactions to cultural burden. *White Noise*, however, is a prime example of the postmodern state that Fredric Jameson defined as the cultural logic of late capitalism—a world of shallowness, fractured subjectivities, and commodified experience in which mass media codes permeate narrative:

I believe that the emergence of postmodernism is closely related to the emergence of this new moment of late, consumer or multinational capitalism... its formal features in many ways express the deeper logic of that particular social system (Jameson 1991, 9).

This condition is addressed both formally and philosophically in DeLillo's novel. Jack Gladney, the main character, teaches Hitler Studies at a generic American college and leads a life that is influenced by television commercials and grocery store aisles just as much as by academic endeavors. Media intrusions, such as disembodied radio voices, ads that are chanted like incantations, and tabloid headlines that are taken as fact, interrupt his family's daily rituals. The novel's central "airborne toxic event" literalizes the unseen dangers that permeate contemporary life, whether they be data, poisons, or ideas. Information is not just too much; it is unavoidable, pervasive, and a part of reality.

This state of overburden is reflected in the narrative structure. Coherence is continuously postponed in the textual context created by fragmented episodes, paratactic patterns, and digressive dialogues. The abundance of product labels, brand names, and technobabble fills the text with excessive semiotics, mirroring the media environment it portrays. Despite being set in 1980s America, the observations on media, communication, and the decline of meaningful discourse are still incredibly timely. The novel seems to anticipate 21st-century algorithmic worlds, where DeLillo's diagnoses are expanded by data-driven personalization, targeted advertising, and tailored feeds. The dynamics of overload makes it harder to distinguish between distraction and information.

Furthermore, Shoshana Zuboff argues that data mining and behavioral prediction permeate daily life: "Surveillance capitalism unilaterally claims human experience as free raw material for translation into behavioral data" (Zuboff 2019, 7). The critic emphasizes how predictive analytics and data extraction mold behaviors, eroding autonomy and causing anxiety: "The problem of postmodernism ... this problem is at one and the same time an aesthetic and political one" (Zuboff 2019, 56).

Although DeLillo's characters aren't specifically subjected to digital surveillance, they nonetheless live in societies that gather, classify, and commercialize data, influencing their anxieties and wants in ways they can hardly understand. They go through similar experiences, despite the fact that it predates digital surveillance: they are haunted by the mediated depictions of death, disaster, and violence; they are terrified by threats they cannot fully perceive; and they are overtaken by data they cannot control. Our contemporary obsession with invisible data flows, algorithmic profiling, and invisible infrastructures of control is foreshadowed by the novel's concentration on the invisible toxic event—a threat whose existence is only acknowledged through technological mediation.

White noise refers to both the background hum of technological gadgets and the mental fog of too much information—an atmosphere in which it is impossible to discern what is significant from what is meaningless. The interconnectedness of technology, media, and subjectivity is highlighted by this background noise, which permeates both the storyline and the experiences of the characters. Furthermore, the novel reflects the breakdown of stable forms under the weight of informational abundance by engaging with genre conventions—academic satire, catastrophe narrative, and domestic fiction—only to destroy them. The repetitive patterns, ironic detachment, and humorous tone reflect what Jameson refers to as the decline of affect in postmodern culture, where authenticity and depth vanish in the midst of the deluge of signals and pictures.

Narrative distance and the illusion of objectivity

Rather than merely depicting the saturation of human consciousness by external data systems, the novel performs this saturation formally and stylistically. Through its fragmented narrative structure, recursive dialogue, and pervasive irony, *White Noise* immerses readers in the dissonance of meaning characteristic of an age in which information itself becomes the dominant substance of reality.

Information excess has significant psychological and existential ramifications that go beyond cultural and aesthetic aspects. The home is depicted as being heavily influenced by the media throughout the novel. In the Gladney household, radios and televisions are kept on not because of their programming but rather because of their ambient presence, or how they add to the background sounds of daily life. DeLillo already shows how media gadgets invade home spaces, altering the texture of perception and the rhythms of existence (Akers and Martin 2021).

This dynamic is best illustrated in one of the novel's passage where the Gladney children huddle in front of the TV and watch commercials, news stories, and trivia with equal disinterest. Jack's teenage son Heinrich exhibits the ability to engage with media in a detached manner, demonstrating a postmodern subjectivity influenced by ongoing exposure to conflicting streams of information that produce hyperreality. The absurdity of mediated life is captured by DeLillo in a sardonic tone, as institutional life as both meaningless and self-important: "I've got a German department that's small but proud. It's a status department, a money department. They're all out to get tenure before the place falls apart" (4). Jack's voice occasionally becomes unaffectionately bland and rhythmically sparse. His detached narrative style, which expresses emotional intimacy through ironic mimicry rather than sentiment, is typified by the exchange's recurring symmetry, as in:

Who will die first?' my son said.
'It's a toss-up,' I said.
'I hope it's me,' he said.
'I hope it's me,' I said (100).

As a result of this saturation, the lines separating the real from the simulated are blurred, and media events begin to resemble or even replace lived experience (Brown 2020; Goldberg 2021; Virilio 1998). To Baudrillard, the hyperreal arises when signs circulate in an unending play of images rather than referring to any reality outside of themselves.

The family's speech patterns are influenced by media language, which turns private conversations into a patchwork of catchphrases, commercials, and

pieces of popular culture. Also, techniques such as fragmentation and repetition are abundant: "‘Toyota Celica,’ he said. ‘Toyota Celica,’ I said. He said it again. I repeated it. We said it a third time, then a fourth, still without any hint of irony" (155).

The narrative voice veers between academic jargon and media-inflected clichés, while Babette's recitation of product names and media slogans highlights the decline of private language. The reduction of emotional depth to the monotonous repetition of cultural indicators often referred to as the waning of affect in postmodern society: "The supermarket shelves have been rearranged. I'd always believed there were psychic data in the supermarket—energy waves, information floods" (36).

The supermarket serves as a microcosm of cognitive fatigue and media saturation. It is a place where signs, symbols, and semiotic excess are consumed in addition to things. A dense field of stimuli, including the bright packaging, background music, PA announcements, and promotional signage, overwhelms the senses and confuses the subject:

The supermarket is a fairyland of sorts, an enchanted place of power and ritual. All around us the fluorescent lights hum in the way that the hum of the world's machinery is heard in the early morning, when the streets are empty and the silence is an absence of human sounds (DeLillo 1985, 3).

Noise is the ubiquitous, mechanical hum that pervades human experience from the very beginning of the book. The buzz of the fluorescent lights represents the continuous technical background—a faint yet persistent sound that echoes the "world's machinery". This hum is more than just sound; it is an indication of the infrastructure of modernity, implying that noise permeates daily areas even during quiet times. A microcosm of postmodern existence, where noise is both tangible and metaphorical, is the supermarket, a place of materialism and media saturation. It is the auditory representation of the deluge of goods, signals, and data that will overtake the characters during the course of the book. The "absence of human sounds" emphasizes how the cacophony of technology supplants or undermines the inherent human presence.

Jack Gladney acknowledges the supermarket's function as a theater of signs in his thoughts about it. Brightly colored packaging, recurring product names, and corporate slogans all contribute to what Baudrillard refers to as semiotic saturation in consumer culture, where processing signs is just as important to consumption as purchasing commodities. Choice becomes illusory in this setting; variety obscures the fundamental similarity of goods, and consuming turns into a way to participate in the flow of signals: "You never consume the object in itself (in its use-value); you are always manipulating objects (in the broadest sense) as signs" (Baudrillard 1998, 60).

The way the supermarket affects perception is similar to the wider cultural consequences of information overload, which include the loss of critical thinking skills, desensitization, and cognitive fragmentation. According to Jack, the supermarket is a contradictory place where having too many options gives one a sense of security while also dulling consciousness. It is both reassuring and unnerving. The concurrent attraction and repulsion that define postmodern reactions to media saturation is reflected in this ambivalence.

The depiction of cognitive stress is aided by its linguistic saturation. Throughout the story, DeLillo uses a variety of media-culture-derived lists, repeats, and interjections, such as product slogans, brand names, medical terms, and bureaucratic jargon. This overuse of language is a reflection of information saturation, a state in which language no longer functions as a tool for clarity but rather as a cause of alienation, confusion, and distraction. The novel also considers the loss of narrative coherence in an information-burdened society. The din of media constantly interrupts Jack's attempts to create a meaningful narrative, whether it be about his family, his academic field, or his own mortality. David Harvey's observations regarding time-space compression in postmodernity are reflected in this: conventional narratives of progress, coherence, and meaning are undermined by the acceleration and fragmentation of experience:

Time-space compression is a sign of the intensity of forces at work at this nexus of contradiction and it may well be that crises of overaccumulation as well as crises in cultural and political forms are powerfully connected to such forces (Harvey 1989, 258).

The psychological effects of living in a world full of frightening facts are highlighted by Babette's covert use of Dylar, a fictitious medication intended to reduce fear of death. The yearning to escape not only mortality but also the overwhelming sensations that keep it ever-present is reflected in Dylar's vow to eradicate the fear of death from consciousness. "The white noise of the television murmurs in the other room, a constant presence, a comfort, a signal that life goes on" (DeLillo 1985, 20).

The "white noise" of the television offers a steady background, a comforting buzz that life goes on in spite of uncertainty or dread. However, this sound is also intrusive and unavoidable —it permeates awareness and is constantly present. The metaphor also applies to the psychological domain, where the noise anesthetizes consciousness, fills in internal voids, and diverts attention from existential dread. It implies a coping and denial technique in which sound takes the place of quiet and introspection. This emphasizes the idea that the saturation of contemporary life makes it difficult to distinguish between internal and exterior media.

Through recognizable structures, the characters' reliance on routines—such as shopping, watching TV, and going to lectures—can be interpreted as defence mechanisms. In a world with more knowledge than can be comprehended, these rituals offer momentary steadiness. A loop of reliance and disempowerment is created, though, as these tactics also serve to strengthen the identical systems of consumption and media saturation that first cause a burden: “The television is a machine for thinking with” (DeLillo 1985, 48). This seemingly straightforward statement highlights the ambivalence of noise as it is mediated by technology. The continuous stream of sounds and visuals from the TV affects perception and thought processes. Noisy information crystallizes in the noise of television; it is not merely background but also formative, influencing the characters' mental processes.

The expression frames television, and consequently media noise, as an epistemic instrument that influences the acquisition of knowledge. However, this “thinking” is passive, disjointed, and spectacle-heavy, undermining the idea of logical comprehension and highlighting the damaging impacts of media cacophony.

White Noise provides a foresighted reflection on how excessive communication affects subjectivity, perception, and thought through its portrayal of media-rich settings. Media technologies alter experience by creating the frameworks that allow reality to be understood as well as by disseminating information. The characters inhabit a world where media noise permeates every aspect of daily life, influencing social interactions, emotional reactions, and intellectual processes (Ferguson 2021; Kaufman 2022).

These cultural circumstances are reflected in DeLillo's aesthetic techniques—his use of fragmentation, repetition, and parody—and in the ways that the locations and activities—from the supermarket to *the airborne toxic event*—act as metaphors for the ubiquitous, frequently imperceptible forces of information and media in forming human experience. Through a variety of formal techniques, DeLillo creates a literary aesthetics of excess that mirrors, replicates, and challenges the fragmented, recursive, and confusing aspects of existence in a media-rich society. By doing this, DeLillo contributes to a larger postmodern literary tradition that uses irony, satire, narrative fragmentation, and reflexivity to address societal excess.

The novel's episodic, fractured structure is one of its most remarkable elements; it reflects the interruptions and diversions of modern life. Each of the three sections of the book—“Waves and Radiation”, “The Airborne Toxic Event”, and “Dylarama”—offers a unique but marginally related take on the main topics of mortality, media, and communication. Unlike realist fiction, which follows a continuous, causal path, the story in these sections develops in discrete incidents.

The disarray represents what David Harvey refers to as the time-space compression of postmodernity: conventional narratives of continuity, progress, and coherence are broken by the speed at which cultural and technological change is occurring. Interruptions—advertising, news updates, and family discussions that veer off topic—intersperse Jack Gladney's experiences, echoing the rhythms of channel-surfing, magazine flipping, or scrolling through internet feeds. In line with Frederic Jameson's assessment that postmodern narratives frequently exhibit a "waning of affect" and a collapse of depth into surface, the end product is a text that defies totalizing meaning.

Furthermore, conventional narrative hierarchies are undermined by this structure. Moments that seem insignificant, like shopping excursions, pointless discussions, and product lists, take up just as much narrative space as pivotal or epiphanies. Jean Baudrillard's concept of the simulacrum, which holds that distinctions between the profound and the banal vanish in the never-ending flow of messages, is echoed by this flattening of significance.

One of this novel's primary narrative techniques is repetition, which captures the cyclical patterns of cognitive overload as well as the repetitions of media culture. Phrases, slogans, and images recur throughout the book: the kids mimic commercial jingles; Jack and Babette keep coming back to the subject of death; the signs and brands of the supermarket resurface like song refrains. The non-linear temporality of media environments, where news, entertainment, and advertising cycles produce content loops that supplant narrative progression, is reflected in this recurrence: "That's just noise," he said. "It's white noise. It's a comfort." "Noise," she said. "Just noise" (DeLillo 1985, 52). The paradox of noise as useless and significant is encapsulated in this brief interaction. As language repeats itself in loops, reinforcing and destabilizing meaning, the use of the word *noise* echoes the novel's stylistic motifs.

The term *white noise* refers to a pervasive, sometimes calming, but empty background hum. The characters' differing reactions—perceiving noise as either *comfort* or *just noise*—reflect opposing viewpoints on media saturation: noise as a sign of cultural emptiness or as a necessary buffer against anarchy. Echoing the larger issue of semiotic saturation, this verbal repetition creates a sort of cacophony inside language itself, demonstrating how communication gets saturated and fragmented.

From a formal standpoint, repetition is a thematic and structural technique. It simulates the stagnation of burdened consciousness, in which an abundance of information results in paralysis rather than knowledge. People are reduced to passive consumers of unending, repeating content because there is so much media messaging in modern culture that it overwhelms their ability to think critically. The discussions in the novel serve as an example of this circumstance.

Characters interrupt one another, questions remain unresolved, and conversations often repeat themselves. A feeling of cognitive exhaustion is created by this recursive pattern, which makes it seem as though the characters and the reader are trapped in an endless cycle of unresolved information.

In a similar vein, DeLillo blurs the lines between media reality and narrative fiction by incorporating media fragments such as news bulletins, weather reports, and television conversations. The reader is reminded of the manufactured character of story and the inability to isolate fiction from the media landscapes it reflects by these intrusions, which upend the fictitious universe. This is consistent with Brian McHale's description of postmodern fiction as ontologically unstable, focused more on highlighting the representational processes than on portraying reality:

Postmodernist writing seeks to foreground the ontological duality of metaphor, its participation in two frames of reference with different ontological statuses ... texts often prolong this hesitation ... as a means of foregrounding ontological structure (McHale 1987, 134–135).

Jack Gladney's first-person narrative sheds light on how awareness is shaped by media exposure. His narrative voice fluctuates between media-infused banalities, neurotic contemplation, and scholarly detachment. This hybrid voice represents a subjectivity made up of opposing discourses, including media noise, consumer culture, family life, and scholastic authority. A mediated consciousness, one that finds it difficult to access genuine emotion or unmediated reality, is revealed by Jack's propensity to filter experience through language, whether it be academic, scientific, or media-derived. In order to emulate the informational density of his surroundings, his narration frequently veers into lists, product descriptions, and commentary on television shows. Katherine Hayles noted that information technologies create fragmented, data-saturated identities, which is consistent with this artistic decision.

I understand human and posthuman to be historically specific constructions that emerge from different configurations of embodiment, technology, and culture. My reference point for the human is the tradition of liberal humanism; the posthuman appears when computation rather than possessive individualism is taken as the ground of being, a move that allows the posthuman to be seamlessly articulated with intelligent machines (Hayles 1999, 33).

DeLillo's narrative tactics implicate the reader in the feeling of overwhelm. Disorientation, deferral, and saturation are characteristics of the reading experience created by the novel's broken episodes, recursive dialogues, and profusion of

signs. Instead of leading the reader to conclusion or coherence, the novel encourages a reading style that reflects the characters' fragmented, distracted attention just in the way in which the novel thematizes cognitive and cultural disorientations and enacts them through fragmentation, repetition, intertextuality, irony, and mediated narrative. The broader postmodern understanding that conventional narrative forms are inadequate to convey the intricacies of life in a media-rich, information-rich environment is reflected in these tactics.

The aesthetics of overload and the crisis of meaning

The paradoxes of modern life—the simultaneous abundance and lack of meaning, the confusion of representation and reality, and the breakdown of cohesive subjectivity—are brought to light by DeLillo's aesthetic of overload. Thus, the novel serves as both a formal intervention in the literary response to the surplus of information and a representation of its cultural context. From the supermarket to the evacuation center, certain locations and settings are presented as miniature representations of media and information saturation, influencing the story and the characters' perceptions of reality.

The medical center is where DeLillo's spatial investigation of excessive communication comes to a head. It is a place where patients are turned into case numbers, bodies into data, and health into probabilistic results. Readouts, scans, and risk assessments—all types of informational abstraction that supplant embodied experience—are the main focus of medical consultations. These systems act as a mediator for his fear of dying, which is exacerbated by his incapacity to understand the consequences of the information that is being given to him. Informational medicine's dehumanizing consequences are reflected in the center's impersonal interactions, bureaucratic procedures, and antiseptic surroundings.

The contradiction of modern informational systems is best shown by the medical center: they promise control and clarity yet frequently result in uncertainty and powerlessness. A major tension in the novel's depiction of overburdened environments is the gap between existential understanding and informational knowledge, which Jack's experience highlights. The locations, such as shops, shelters, and medical facilities, are metaphors for the larger cultural state of suffocation. Subjectivity dissolves into roles and data points, signs multiply, and attention becomes fragmented in these settings. The conversion of lived experience into mediated, systematized, and commercialized forms is dramatized in these settings.

DeLillo's depiction of information-rich environments forces readers to acknowledge the ways in which systems, media, and architecture influence perception and experience in ways that are frequently imperceptible but significantly

impactful. In order to further explore the impacts of excessive communication on bodies, surroundings, and narratives, the themes of contamination and pollution stand out, both literally and figuratively. Contamination serves as both a physical and figurative motif in *White Noise*. The airborne toxic event, the most obvious example of pollution in the book, dramatizes environmental disaster, but DeLillo is more concerned with a more subtle kind of contamination: the inundation of informational pollutants into human consciousness. DeLillo's connection between environmental toxicity and cognitive and cultural burden shows how the logic of invisibility, ubiquity, and cumulative harm is shared by both types of contamination:

The alert tone pierced the night. It was followed by the voice: 'A hazardous airborne chemical has been released in the vicinity of Blacksmith Village. Residents are advised to shelter in place and await further instructions.' The voice was calm, but the tone was sharp, urgent. The words fell into a static murmur (DeLillo 1985, 109).

The emergency broadcast is a prime example of how, in times of crisis, loudness can both facilitate and obstruct understanding. The "static murmur" and "alert tone" create a sensory burden: the signal is urgent but partially muffled by background noise, representing the disjointed delivery of vital information. The uneasiness brought on by mediated communication is heightened by the contrast between the "calm voice" and the "sharp, urgent tone". Here, the noise represents the epistemic ambiguity that defines contemporary risk, simultaneously communicating danger and confusing comprehension. The characters' general inner struggle is reflected in their incapacity to completely comprehend the nature of the airborne hazardous event.

White Noise is a metaphor for pollution in its widest sense, which includes the overabundance of chemicals, information, pictures, and data that infiltrate human habitats, bodies, and minds in addition to the contamination of air and water. This perspective, which draws from the theories of Paul Virilio (1994, 1998) and Jean Baudrillard (1994), places DeLillo's writing in a tradition that links the technological advancements of modernity to new kinds of danger, uncertainty, and toxicity.

Despite being physically present, the toxic cloud turns into a media event just as much as an environmental one, demonstrating how pollution functions in the postmodern era through both material and semiotic means. The notion of contamination is literalized by Jack's exposure to Nyodene D. His body is marked with invisible particles that are said to shorten his life, creating a site of poisonous inscription. However, this contamination is never completely understandable; it exists as a collection of risk evaluations, medical records, and

possibilities. The situation created is a reflection of larger worries about the body's susceptibility in the information age. The body is no longer only biological; it is now better understood and controlled by informational systems, which turn flesh into data through scans, health records, and risk algorithms. Contamination is not just chemical but also informational, permeating future and sense of self through probabilities and numbers.

A feeling of exhaustion is brought on by the combined effects of informational and environmental pollutants. Jack feels the psychological toll of living inside systems that constantly barrage the body with information about its weaknesses in addition to the fear of contamination. This supports Virilio's claim that as technology advances, the body finds it difficult to keep up with the speed of information, leading to new types of trauma:

The age of paradoxical logic begins with video recording, holography and computer graphics... as though, at the close of the twentieth century, the end of modernity were itself marked by the end of a logic of public representation (Virilio 1994, 63).

In this situation, the fear of dying is inextricably linked to Jack's dread of informational collapse, which is the incapacity to understand, manage, or break free from the institutions that govern his life: "All plots tend to move deathward. This is the nature of plots" (26). Like his surroundings, his body is teeming with invisible forces—chemicals, information, and indicators—that conflate the concepts of inside and outside, self and system.

In *White Noise*, language itself turns into a vehicle for contamination. Jargon, slogans, brand names, and bureaucratic speech abound in the novel's linguistic landscape; these types of language mimic the logic of pollution in that they expand, saturate, and obfuscate rather than clarify. The protagonists' cognitive burden is exacerbated by this verbal excess, which reflects the saturation of their surroundings. Language, commercial culture, and the airborne toxic event are all expressions of a more profound existential state. They represent the ways that internal fears are externalized in contemporary life.

White Noise provides a foresighted analysis of how contemporary systems manage, create, and sustain excess, foreshadowing current worries about data pollution, attention economies, and ecological crises. In this novel, contamination shapes the novel's narrative techniques, spaces, and subjectivity representations. It is not just a thematic element, but it reflects, analyzes, and reproduces the conditions of excess and fragmentation that define modern life through its portrayal of media saturation, environmental disaster, consumer culture, and mediated subjectivity. However, the novel does more than simply depict information excess; it aestheticizes it, turning the state of cognitive saturation into a collection of

formal, spatial, and narrative techniques that force the reader to share the characters' bewilderment and confusion. *White Noise* is a literary reaction to the encumbering cultural situation as well as a diagnostic. It also considers the wider ramifications of DeLillo's writings for comprehending the function of literature at a time when the abundance of signs, information, and media environments defines our existence.

The novel employs the logic of excess both structurally and conceptually, as we have seen throughout this study. The feeling of living in a world where information is accumulating more quickly than it can be digested or comprehended is mirrored by fragmented episodes, recursive conversation, lists of signs and products, and the proliferation of narrative voices.

DeLillo's purposeful opposition to narrative closure is seen in his failure to provide closure or coherence, which represents the persistent, unresolved character of excess itself. The novel's ambiguity and open-endedness push readers to the boundaries of understanding and emphasize the psychological and epistemological ramifications of living in a world where meaning is constantly postponed.

White Noise serves as a key work for comprehending how literature deals with the aesthetics of overload. It demonstrates how literature may provide a space for critical reflection in the midst of these pressures and how narrative can express the intricacies of living in environments that are crowded with signs, data, and noise. In reaction to the deluge of information that permeates modern life, DeLillo's novel emphasizes the value of developing awareness, skepticism, and introspection. By doing this, it reaffirms literature's timeless value as a window into cultural circumstances and a tool for readers to regain agency in the excess-filled landscapes that characterize the contemporary world. DeLillo explores the metaphor of noise in these passages on a number of levels: as a symbol of fragmented communication and epistemological uncertainty; as media and linguistic clutter that disrupt and shape cognition; as existential hum that conceals underlying fears and uncertainties; as environmental and technological soundscapes that permeate everyday life and physical spaces. The metaphor is essential to comprehending the novel's indictment of modern culture because of the intimate relationship between real and figurative noise, which allows readers to feel the novel's thematic problems on a visceral level.

Conclusion: implications for literary studies and contemporary culture

Future research could expand on this approach to see how authors continue to balance the aesthetics and ethics of excess in the digital age in contemporary writing that faces comparable issues. In the twenty-first century, these issues

are still very much relevant. DeLillo's description of the dynamics of information surplus has been exacerbated by the emergence of digital technology, social media, and algorithmic societies. The novel's insights into the saturation of bodies and minds by invisible forces are echoed by concerns about data pollution, attention economies, and privacy degradation.

White Noise examines the impact of an abundance of information on narrative, subjectivity, and epistemology in addition to depicting a time of media saturation. It expresses the contradiction of a society where information is abundant yet meaning is elusive through its fragmented form, thematic preoccupations, and repetitive structures. By enacting the state of cognitive and cultural excess through its structure, narrative voice, and stylistic devices, the novel challenges readers to consider not only the issues of media saturation but also the boundaries of narrative coherence in a time of excess. The depiction of media spectacle as a force that reshapes reality itself is among its most remarkable features. The characters in the novel frequently internalize television, advertising, and news reporting clichés as they view their lives through the prism of media narratives. Jack and his family encounter the “airborne toxic event” through a variety of media frames, including news bulletins, emergency announcements, and scientific jargon, despite the very real dangers involved. Their perception of the event is influenced by the ways in which information about it is shared rather than by firsthand experience.

The novel transcends its historical moment, offering profound insights into how literature mediates information saturation and shapes evolving conceptions of subjectivity, mortality, and social cognition. By performing the effects of overload through its narrative, stylistic, and structural strategies, the novel functions simultaneously as a critique of mediated existence and an aesthetic experiment in representing cognitive excess.

Implications for literary studies include: recognition of literature's distinctive capacity to represent cognitive and cultural phenomena that elude traditional realist and linear narrative forms, demonstration of how postmodern techniques—fragmentation, repetition, and intertextuality serve as formal correlatives for the experiential conditions of information saturation, and engagement with current debates surrounding digital culture, algorithmic mediation, and attention economies, underscoring the novel's continuing relevance for theorizing mediated subjectivity in the twenty-first century.

By foregrounding these dynamics, this article advances a more nuanced understanding of DeLillo's achievement, positioning *White Noise* not merely as a cultural artifact of postmodernity but as a conceptual model for analyzing literary responses to the complex informational and technological ecologies that define contemporary experience.

WORKS CITED

- Akers, Alan and Helen Martin. 2021. "Narrative Fragmentation and Cognitive Overload in DeLillo's *White Noise*." *Contemporary Literature*, vol. 62, no. 1, 89–112.
- Baudrillard, Jean. 1994. *Simulacra and Simulation* (S. F. Glaser, Trans.). US: University of Michigan Press.
- . 1998. *The Consumer Society: Myths and Structures*. Los Angeles: Sage.
- Boxall, Peter. 2012. *Don DeLillo: The Possibility of Fiction*. Routledge.
- Carr, Nicholas. 2010. *The Shallows: What the Internet Is Doing to Our Brains*. New York: W. W. Norton & Company.
- Brown, Claire. 2020. "Media Saturation and the Hyperreal in Don DeLillo." *Modern Fiction Studies*, vol. 66, no. 2, 245–270.
- Cowart, David. 2020. *Don DeLillo: The Physics of Language*. University of Georgia Press.
- DeLillo, Don. 1985. *White Noise*. New York: Viking Press.
- Duvall, John N. 2021. *The Cambridge Companion to Don DeLillo*. Cambridge University Press.
- Ferguson, Margaret. 2021. "Noise, Information, and Postmodern Subjectivity." *American Literary History*, vol. 33, no. 4, 600–625.
- Goldberg, David. 2021. "The Supermarket as Cultural Microcosm in *White Noise*." *Studies in the Novel*, vol. 53, no. 3, 321–345.
- Harvey, David. 1989. *The condition of Postmodernity*. Malden, USA: Blackwell Publishing Inc.
- Hayles, Katherine N. 1999. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: University of Chicago Press.
- Jameson, Fredric. 1991. *Postmodernism, or, the Cultural Logic of Late Capitalism*. Durham, USA: Duke University Press.
- Kaufman, Sarah. 2022. "Repetition, Noise, and Narrative Structure in DeLillo." *Narrative*, vol. 30, no. 3, 310–335.
- Kavadlo, Jesse (ed.). 2022. *Don DeLillo in Context*. Cambridge University Press.
- LeClair, Thomas. 2020. *In the Loop: Don DeLillo and the Systems Novel*. University of Illinois Press.
- McHale, Brian. 1987. *Postmodernist Fiction*. UK: Methuen.
- McLuhan, Marshall. 2021. *Understanding Media: The Extensions of Man*. MIT Press.
- Osteen, Mark. 2023. "Consumerism and the Noise of Culture: Re-Reading *White Noise* in the 21st Century." *Contemporary Literature Review*, vol. 54, no. 2, 31–56.
- Toth, Emily. 2021. "Fragmented Narratives and the Epistemology of Overload." *American Literature*, vol. 93, no. 2, 237–261.
- Virilio, Paul. 1994. *The Vision Machine*. USA: Indiana University Press.
- . 1998. *The Information Bomb*. Verso.
- Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism*. New York: PublicAffairs.

METAPHORICAL FRAMING OF OBESITY IN ONLINE NEWS DISCOURSE

Anișoara POP¹ 

Article history: Received 21 July 2025; Revised 26 September 2025; Accepted 10 October 2025; Available online 12 December 2025; Available print 30 December 2025

©2025 Studia UBB Philologia. Published by Babeș-Bolyai University.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ABSTRACT. *Metaphorical Framing of Obesity in Online News Discourse.* For people living with obesity, public discourse is replete with messages concerning the risks and warnings that their overweight and size pose to society and themselves. Moreover, bias and stigma are frequently associated with obesity, people suffering from it being perceived as unattractive, repulsive, and even responsible for their condition.

This paper is based on a content analysis of obesity discourse in Digi24 news published online between 2017 and 2022. It aims to find out how obesity is framed through metaphors and what its subsequent conceptualisations and evaluations are at the linguistic level (individual versus collective responsibility, positive versus negative evaluation), as well as at the level of associated imagery. Results demonstrate that Romanian online news metaphorically frames obesity in ways that both inform and reinforce stigma toward the persons living with it. Overcoming the limited understanding, stereotypes, and stigma surrounding obesity is a slow but essential process, to which online news can contribute by indirectly shaping public perceptions and societal attitudes towards obesity.

Keywords: *obesity, metaphorical framing, linguistics, online news discourse, image anchoring, stigmatisation*

REZUMAT. *Reprezentări metaforice ale obezității în discursul știrilor online.* Pentru persoanele care suferă de obezitate, discursul public este invadat de avertismente ce accentuează riscurile pe care obezitatea și dimensiunea corporală

¹ **Anișoara POP** is Professor, PhD at *GE Palade* University of Medicine, Pharmacy, Science, and Technology of Targu Mures, *Petru Maior* Faculty of Sciences and Letters, Romania. Her main interests include Medical English, specialised medical translation, multimodal discourse analysis, pragmatics of advertising, and narrative medicine. E-mail: anisoara.pop@umfst.ro.

le reprezintă pentru sine și pentru societate. De asemenea, prejudecățile și stigmatizarea sunt frecvent asociate cu obezitatea, persoanele cu obezitate putând fi percepute ca neatrăgătoare, respingătoare chiar și responsabile pentru afecțiunea de care suferă.

Lucrarea de față este o analiză de conținut a discursului despre obezitate pe un eșantion de știri Digi24 publicate online între anii 2017 - 2022. Ea pune în lumină modul în care obezitatea este încadrată metaforic la nivel lingvistic și al imaginilor care ancorează discursul obezității dar și conceptualizările și evaluările acesteia pozitive sau negative: obezitatea ca responsabilitate individuală versus colectivă. Renunțarea la percepțiile limitate legate de obezitate este un proces lent dar esențial la care discursul știrilor online poate contribui prin modelarea acestora.

Cuvinte-cheie: *obezitate, cadre metaforice, lingvistică, discursul știrilor online, ancorare prin imagini, stigmatizare*

1. Introduction

Obesity is a chronic, multifactorial disease involving a combination of various comorbidities and worldwide spread, which can induce a physical, psycho-emotional or occupational imbalance and whose management takes time and requires specific individual and group interventions (https://connected4health.pixel-online.org/index_nogranated.php). Obesity stigma and negative stereotypes that are damaging to the health, dignity, and quality of life of people with obesity also abound and are perpetuated through the public discourse (Puhl and Heuer 2010; Foster et al. 2003).

According to the literature, “the obesity epidemic discourse has replaced that of AIDS as the most vilified disease of our century” (Harjunen 2016, 47). It is, therefore, evident that obesity has become a stringent public health concern, and online media and news can play a key role in shaping public perceptions and societal attitudes towards obesity.

The objective of this paper is to analyse how obesity is represented in Romanian Digi24 online newsfeeds and to identify predominant metaphorical framings and their implications for understanding and perceiving obesity. To the best of our knowledge, there is no analysis of the metaphorical representation of obesity in Romanian online news discourse. A content discourse analysis framework was chosen to study the linguistic devices, more specifically, metaphor types (mapping positive versus negative evaluations and individual versus collective responsibility), alongside image anchorage and other disparate linguistic choices associated with them (modality, specialised terminology).

2. Metaphors we suffer by

There is a vast literature on metaphor and communication. We frequently use metaphors in everyday conversations because we think metaphorically (Landau and Keefer 2015, 131). Metaphors are employed to express feelings, emotions, and attitudes, and to conceptualise thought. In turn, metaphors work to create meanings and shape our cognitions, and thus we can discuss metaphor-enriched cognition (Landau, Meier, and Keefer 2010, 1046). Metaphor is a staple component of communication because it allows us to understand and communicate abstract concepts through concrete mappings.

In medical communication, disease may be metaphorically represented as a dangerous animal attacking, lurking in our DNA, and running in families. It can be *treacherous, insidious, tricky, merciless, cruel, ruthless, evil, aggressive, malignant* and *benign*. Virginia Woolf in her essay “On Being Ill” has remarked how pain cannot be rendered in sensible language when she stated, with a poetic resonance, that “The merest schoolgirl, when she falls in love, has Shakespeare, Donne, Keats to speak her mind for her; but let a sufferer try to describe a pain in his head [. . .] and language at once runs dry” (Woolf 1926). Later on, Elaine Scarry in her book *The Body in Pain* claimed that “pain destroys language” (Scarry 1985, 4).

Besides illness, metaphor as a cognitive and linguistic device is also used in relation to subjective and sensitive experiences such as death and dying, employed euphemistically in order to discuss and relate to complex and risky situations in a non-threatening and indirect manner. Thus, war terminology with all that it entails in terms of danger, risk, damage, weaponry, war loss or war winning after fighting a valiant battle, expresses the patients’ affliction, feelings, and attitudes towards their illness. According to how it is used, metaphor can favour, but also impede, medical communication, since, as Semino (2020, 3) apodictically puts it, the use of violence metaphors for cancer is potentially harmful to cancer patients.

In public communication, metaphor is a form of rhetoric that tries to persuade or dissuade people, steering them towards a particular reasoning and solution. By perceiving a health problem as a common enemy, for instance, public discourse metaphors attempt to persuade us to fight against it and eventually destroy it. If the health problem is perceived metaphorically as burden on the health system or as contagion, responsibility falls on the individual, patients feel responsible and, thus, incriminated. The present analysis will focus on the online news discourse from the perspective of metaphor framing and its image anchoring of one of the most severe afflictions of our modern times: obesity.

2.1. *Analysing metaphorical framing of obesity in online news*

The current paper is part of a metaphor-based content analysis of online Digi24 newsfeeds, covering the period between October 2017 and March 2022 (<https://www.digi24.ro/cautare?q=obezitate>). The corpus compilation was based on the keyword search: “obezitate”. It retrieved a total of 539 articles that referred to news on health, healthy/unhealthy eating, food, cancer, adolescents, etc. The articles ranged in length from 160 to 600 words.

Qualitative data analysis was performed manually, applying inductive in vivo coding (labels) and thematisation (Manning 2017), thus preserving the original words and phrases encountered in the news rather than the researcher’s own conceptual terms. The following research questions were formulated to address the core issues of metaphorical framing of obesity in online news:

Q1: What metaphors are employed to frame obesity?

Q2: What evaluations do they convey? and

Q3: What images anchor the obesity discourse?

The working hypotheses are that, through its linguistic choices, Romanian news media shape understanding of obesity as: 1) a *collective endeavour* which is associated with a positive perception and determination to fight it, but, at the same time, 2) news communicates an *individual negative evaluation* associated with personal sin and blame that perpetuates obesity stigma.

Stigma communication theories (Smith, 2007; Barry et al, 2009) were employed to examine stigmatising metaphors associated with obesity in the selected newsfeeds corpus. The taxonomy was based on Barry et al. (2009), who distinguish the following types of obesity metaphors: a) sinful behaviours; b) disability; c) time crunch; d) obesogenic environment; e) obesity as a response to unrealistic messages of physical attractiveness. This categorisation distinguishes a continuum of blame, starting with individualistic (gluttony and food addiction) and ending with external, collective blame, the latter residing outside individual control (junk food advertising, the big industries). I have also employed Smith’s (2007) taxonomy, according to which metaphors fulfil the functions of stigma communication: they label, mark, assign personal responsibility, and link to peril the stigmatised person.

In terms of logistics, all translations from Romanian into English were performed by the author. Likewise, online news articles were coded, while those discussed herein were referenced as case studies in the final Primary resources section of the Works Cited.

Manual in vivo coding and thematisation (Manning 2017) was employed in order to preserve the authenticity of the newsfeed language and thus avoid researcher bias in developing initial concepts. The content analysis of metaphorical framings of the corpus resulted in the following major types: obesity as fight/war,

excess weight as a societal burden, obesity as epidemic, obesity as food addiction, obesogenic environment and food addiction, and obesity as economic construct. Figure 1 below illustrates the process of *in vivo* coding and inductive thematisation of the most frequent metaphorical framing of obesity, i.e. obesity as fight/war:

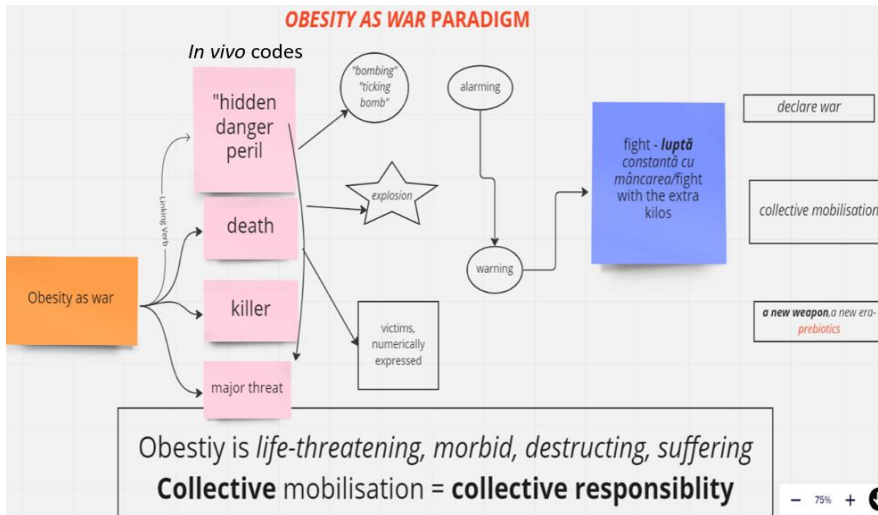


Figure 1. In vivo coding and thematisation of *Obesity as war/fight* (my compilation)

In the following sections, the metaphorical framings identified in the corpus will be detailed, together with image anchorage and disparate linguistic evidence such as terminology (semantic field) and modality, which support the current analysis and interpretation.

2.1.1. Obesity as fight

As in previous studies (Cotter, Samos, and Swinglehurst 2021), the most frequent thematic framing identified in the corpus was that of fight/war against obesity. Obesity rates have spiked “dramatically”, it is surrounded by “perils” and “danger”, and it represents one of the “most serious *threats* to people’s health” (*italics mine* for emphasis throughout the paper):

- Losing extra kilos as fight: Obesity is “a major threat” that evolves “rapidly”. People “*fight* with the extra kilos”; “I’m back to the feeling of constant *struggle* with food.”
- Obesity as hidden danger/dangerous: Obesity is “morbid”, and it is mapped as “danger” and a “ticking bomb”. “Obesity is *dangerous* at puberty”[1]; it is a “major *danger*” as there is “hidden *danger* behind the extra kilos”.

Even more, within the context of the COVID-19 pandemic, obesity is indirectly personified as “death” and “killer”:

- Death/mortality framework: “Obesity increases the risk of *death*” (headline) [2]; “Several diseases hide behind obesity,..... which increases the risk of *mortality* in case of infection with the coronavirus”[3]. “Researchers have found out why *obese* people are at higher risk of developing severe COVID-19 and *death*”[4]. From a pragma-linguistic perspective, through the use of endorsements from experts (i.e., *researchers*) and a dangling comparative (i.e., *at higher risk*), the online discourse intends the highest possible risk level to be read.
- Explosion framing of obesity is associated with its labelling as case: “The *cases* of obesity have not decreased [...] on the contrary, they have *exploded*”[5]. By considering people “cases” and “obese”, instead of employing people-first terms (i.e., people suffering from obesity), metaphors assign a label (Smith 2007), categorise, dehumanise, and reduce them to a situation.
- Weapon. The fight and weaponry metaphors are frequently employed but in conjunction with epistemic modality, which questions the success of such interventions: “A *new weapon* that *might open* a new era in the *fight against* obesity”[6]; “We *might* soon have a *new weapon to fight* against infantile obesity...”[7].
- Alarm. The online news frequently issues an “*alarm signal*” that the number of patients with obesity requiring emergency treatment has skyrocketed and that “obesity is life-threatening”. “Pandemic is putting children at *high risk* of obesity” as it has developed at “*alarming rates*” and has reached a “critical growth”: “*Alarm signal*. Increasingly more patients with morbid obesity arrive at the Emergency Department (News headline) [8]”.

Since obesity is a salient social concern, metaphorical framings of war/fight and all its associated mappings from the semantic field of hidden danger, weapon, alarm, explosion, and mortality are expected to influence the way in which people think about obesity, lowering its rates. Nonetheless, the epistemic modality inherent in the verb (i.e., *might* soon have a new weapon) acts as a downtoner of the verb force. By fighting against obesity, health authorities underline that collective responsibility and mobilisation are crucial. Fear of the hidden danger and death is an effective tactic for spurring people into action. However, even though the fight metaphor is intended to influence public opinion towards a collective response, fight is generally construed as

damage, destruction, aftermath and suffering. Such negative implicatures are to be blocked in the 'obesity as fight' metaphor, should the public discourse framing achieve its mobilising purpose. Thus, although fighting or declaring war on disease is not a novel metaphor, it is one intended to have a new positive contextual implication that transcends its negative power.

2.1.2. *Excess weight as societal burden*

Within the general fight against obesity, the situation when a medical term is employed, i.e., "excess weight" instead of "fat" or "obesity" (RO: *exces de greutate/greutate în exces*) deserves special consideration. On the one hand, "excess weight" or other professional terminology that makes patients vulnerable is intended to sound neutral versus the negatively loaded term "fat" or "obese": "Patients with *excess weight* are more predisposed to suffer from a severe form of COVID-19", "Young people are more *vulnerable*"[9]. On the other hand, "excess weight" is perceived as a burden on the system and the healthcare workers. COVID-19 patients with underlying obesity needed "extra equipment", "extra team members", and "extra effort" to be transported to hospital, placing considerable strain on healthcare staff, underlining that the experience was painful for both patients and physicians. This type of discourse blames the sufferer and induces a feeling of guilt, which is evident in the negative judgment:

.....*extra equipment and extra effort* were needed from everyone's part... it put the medical staff to a *heavy test* - most of them weighed over 200 kilos. For these patients, *additional crews* were needed to transport them to hospital. [10]

Patients who represent a major threat to the healthcare staff and are a different type of helpless and incapable victim are likely to induce disease avoidance and reluctance response in medical staff, patients themselves, and people in general:

Evaluating and treating morbidly obese patients requires *extra effort* from everyone.....The experience *marks* both the doctor and the patient They feel *helpless*. [10]

This negative evaluation of obesity as societal burden assigns blame and individual responsibility to patients with obesity, is weight-stigmatising and runs contrary to the positive *fight-as-collective-responsibility* in the metaphorical framing of obesity.

2.1.3. Obesity as economic construct

A further type of metaphorical framing of obesity is that of economic construct expressed through terminology belonging to the stock market and industrial semantic field (quota, cost, pressure on the health budgets, fuel, trend) that imply objective, collective measures and responsibility:

- quota, trend - obesity is “maintained at high *quotas*” and it represents “a worrying *trend*” [22],
- food as fuel is associated with business/marketing measures:

“....nutritious and healthy food should be available, quality *fuel* for the body.”

The shift of responsibility from the patient to the external economic and marketing environment is evident in the association of obesity with the obesogenic environment. At the linguistic level, this shift is achieved in conjunction with:

- passive agency: “the consumer *is sold* juices, and sweets with over 300 calories and large amounts of sugar, even though this is forbidden by law in educational units”,
- impersonal verbal constructions (passive reflexive in Romanian: *se cere*, subjunctive), without explicitly mentioning the agent: “advertising of ultraprocessed foods should be banned..... sweets should not be displayed at cash desks in supermarkets”. These constructions are often used in official language, scientific texts, or news reporting to express necessity, obligation, and to avoid responsibility.

The obesogenic environment is commonly associated with sub-themes belonging to the semantic categories of:

- war: alarm/ing, warning: “These are *alarming* figures, which in the near future will but increase and trigger one of the most severe challenges for our public health.” [23],
- “cost, major pressure on the health budget.” [23]

Although linked directly with ultraprocessed foods and therefore, external responsibility of the obesogenic environment (“increase in consumption of ultraprocessed food”), by associating obesity with personal unhealthy eating habits, as well as assigning blame for the pressure exerted on the health budget and society, obesity also falls inside the paradigm of individual responsibility:

“it implies a *burden for society*..... [for] the cost of medical therapies, [it] *costs governments loads of money*, causes *economic loss*.... and decrease of productivity.” [23]

2.1.4. Obesity as epidemic contagion

Obesity as an infectious disease and pandemic, or the obesity epidemic, represents another frequent metaphorical framing (Boero 2007; Hilton, Patterson, and Teyhan 2012). Epidemic grounds the obesity metaphor by transferring one property (quantity) from vehicle (pandemic) to tenor (obesity): “Obesity has to be regarded as *an infectious disease*”[11].

Obesity as pandemic and global complex health challenge adds to its numerical and spread interpretation (“*global complex health challenges* as obesity”[12]). The “plague” or “epidemic” framings of obesity underline the potential devastation of the threat. However, using contagion for something which is not can socially isolate people living with obesity from the general population (e.g., obesity versus normal weight). It is a potential segregation that is harmful and contributes to stigmatisation (Sontag 2002, 133-4, 149).

Saying that “Europe is confronted with an obesity epidemic”[13] may entail its contagious interpretation, involving societal causes, measures of control and isolation and hence the need to avoid obesity and people who live with it. The “contagion” reading in the obesity as epidemic metaphor, as in the excerpt from the online news discourse below, is negative, promoting fear of catching it and, therefore, avoidance of people with obesity:

Subheading: “If we are surrounded by people who are overweight, we are more likely to gain weight ourselves. // Body text: “*The idea that obesity can spread like the flu or chickenpox sounds strange*, but a study of military families in the United States shows it is true.”[14]

Lieberman, Tybur, and Latneret (2012) warned against reacting towards patients with obesity as if they were suffering from infectious diseases, whereas Park, Van Leeuwen, and Chochorelou (2013) claimed that participants in their study had increased discomfort levels at the prospect of having physical contact with a person with obesity. Such perceptions build on a negative bias that causes viewers to react unfavourably, with negative emotions such as disgust, fear or behavioural attitudes (rejection) and discomfort towards otherwise healthy individuals (Park, Van Leeuwen, and Chochorelou 2013). That this interpretation of stigma offers social power and induces a tendency to separate from the stigmatised is evident in an online article which, in order to distance itself from the possible negative perception of obesity as contagion, underlines that “obesity is *nontransmissible*” [15].

To conclude, framed as pandemic, obesity is a cause of alarm, but it also highlights the need for concerted efforts and measures of control. Nevertheless, infectious diseases are a distinct domain in medicine; therefore, the interpretation of obesity as contagion and avoidance is stigmatising and, consequently, should be reconsidered.

3. Metaphors we eat by – Obesity and food as drug, addiction, and sport

Alongside obesity metaphors, there are metaphors we eat by, i.e., related to food and diet. A specific metaphorical framing of food in the surveyed corpus is that of peace, comfort, self-medication, and drugs. Metaphors from the semantic sphere of food as drug, food as addiction, food as the new smoking, and food as sport and entertainment encourage the perception of obesity as sinful behaviour:

- Obesity as addiction – “Obesity is *the new smoking*. Foods are no longer our daily nourishment..... but *comfort* and *drug*, food is *self-medication*” [16, 17],
- Food as sport, entertainment: “Chubby and cute does not mean healthy”; “Stuffing with food is the parents’ and grandparents’ favourite *sport*.” [18]

The negative conceptualisation of food as sinful behaviour with its sub-categorisation as parents’ sinful behaviour towards their children – chubby *and cute*/(RO: *grăsuț și frumos*) – is supported by adjacent metaphors from the field of fight, i.e., threat, alarming: “obesity is one of the *biggest threats* to human health, it has been increasing *alarmingly* among children” [19].

In framing obesity as a sin or addiction to food, with people not trying harder to get healthier, it is implied that obesity is within personal control and responsibility. Likewise, parents and grandparents are to be blamed for their sinful behaviours towards their children and grandchildren.

Regarding the choices from the lexical field of obesity, these can range from positive or euphemistic to loaded. The use of medical terminology (medical jargon) has euphemistic connotations in the noun phrases: extra kilos, excess of fat, excess weight, putting on weight, corpulence (RO: kilograme în plus, exces de grăsime, greutate în exces, îngrășare, corpolență). Likewise, the corresponding adjectives – overweight/supraponderal, corpulent/corpolent, larger persons/persoane corpolente, do not sound pejorative and do not add stigmatising labels, as it happens especially in the context of pediatric obesity, where labels like fatty and big belly perpetuate the negative load and reinforce obesity stigmatisation (Flint and Reale 2014, 1925):

- derogatory terms: fatty/”graso”, fat and pretty/”gras și frumos” [20],
- antithetic ironic juxtaposition and inherent negative value judgment: big bellies, small brains/”burți mari, creier mic”. [21]

4. Obesity discourse and image anchorage – Dehumanised body parts

The images anchoring the obesity discourse in the analysed corpus (N=29 case studies) were categorised according to a grid that included the presence (+) or absence (-) of the following variables:

- a) person and age: adult male [M]/Adult female[F]/child [C] or teenager [T];
- b) cropped, headless body;
- c) body part with obesity: central [cent], extremities [ext], upper [up];
- d) eating junk food/engorging [Junk];
- e) obesity/fat;
- f) faceless individual, back towards the camera;
- g) inappropriately fitting clothes, disgraceful;
- h) sedentary versus active;
- i) neuter image/other (Table 1):

Table 1. Categorisation grid employed for interpretation of image anchorage in the corpus - cases 1-14 out of 29 (my compilation)

.	(M)/(F) /(C) (T) (a)	Cropped d (b)	Body part (c)	Junk (d)	Fat +/- (e)	Faceless (f)	Disgrace ful (g)	Sed/Act (h)	Other (i)
1	T	+	cent	-	-	+	-	sed	
2	M	+	centl	-	+	+	-	act	sport
3	M	+	cent	-	+	+	-	sed	
4	C	+	cent	+	+	+	-	sed	
5									Healthy food
6	M	+	ext	-	+	+	-	sed	
7	C	+	ext	-	+	+	-	sed	
8	M	-	up	-	+	-	-	sed	ICU
9	-								ICU
10	M	-	up	-	+	-	-	sed	ICU
11	M+F+C	-		-	+	+	+	sed	
12	M	+	cent	-	+	+	+	sed	
13	M	+	cent	-	+	+	+	sed	
14	M+F+C	-		-	+	+	+	sed	

In the entire corpus, images of adult males (44.8%) and children/teenagers (44.8%), rather than females with obesity, anchor the textual metaphors. Apart from an empirically evident small number of images related to sport, healthy food, and active people, in the surveyed corpus of online news, visual exposure to obesity was most commonly instantiated through objectified, dehumanised

imagery, neatly cropped, headless men and women (48.2%) with central obesity, visceral or belly fat and adipose tissue (51%), and a focus on other fat body parts.

This type of image anchorage of obesity for the textual content underlines a simplified and simplistic perception of obesity, demonstrating poor coverage of factors beyond personal responsibility. Such representations of the human body affected by obesity are a mark or visual cue likely to trigger aversive responses (Park, Schaller, and Crandall 2007; Smith 2012) and to lead to exclusionary and stigmatising attitudes towards people with obesity (Buckels and Trapnell 2013).

Presenting people with obesity as an object with the intention of rendering a symbol of a collective problem, albeit a stringent and severe health issue, can have damaging effects and a negative impact on individuals suffering from obesity. While technically this presentation is a legal way of representing real people, i.e. cropping out the head for anonymisation and privacy, it is not present in the case of slender bodies, where, on the contrary, the faces are the focus. A curtailed trunk can be interpreted, therefore, as a metaphoric representation of no mouth and, thus, no voice, no head, no opinions or ideas.

Alternatives to headless body images that anchor obesity messages are the faceless persons with their backs to the camera (65%), associated with an “ashamed” interpretation, somehow at par with the sinful behaviour category identified at the level of the textual metaphor. [24, 25, 26] (Pop 2022, 21)

Other positions in which people living with obesity are visually depicted in the studied corpus include:

- engorging junk food (20%), likely to even further increase repulsion and aversion (Puhl et al. 2013) [27];
- fatty thrombophlebitic legs [7];
- showing a bare abdomen, with body parts and junk food priming [2], [12], [18], [21];
- sedentary, lazy (79%) [22];
- helpless [28];
- dressed in inappropriately fitting clothing, disgraceful (31%) and, most commonly, headless [12], [13], [21], [22], [29].

Discussion

Roland Barthes (as cited in Acetoso 2017, 1) asserted that “language is never innocent”: it can seduce or it can wound. Starting from the premise that public perceptions can be shaped through metaphors, this article has analysed how obesity is framed in a corpus of Digi24 online news articles, in order to understand how it is locally communicated and evaluated.

Based on content analysis of metaphorical framing in the corpus, and in line with the hypotheses of news position towards obesity, either as collective or individual responsibility, positive or negative perception, the following most frequent categories have been identified, described and illustrated: obesity as fight (i.e., major threat, hidden danger, ticking bomb, explosion, weapon, fight, victims); obesity as societal burden; obesity as economic construct; obesity as epidemic: contagion; obesity as food addiction, and food as sport.

Through the decision to fight against obesity, it is evident that there is great determination in the public online Romanian news discourse to eliminate it. The predominance of the “fight” metaphor, with all its dangers, weaponry, and victims, demonstrates that obesity is seen as an outside, collective evil, and therefore, assigned a positive evaluation.

Furthermore, obesity is assigned individual responsibility and thus a negative, blaming and stigmatising evaluation through its mapping as societal burden and cost on the health system.

The “epidemic” metaphorical mappings of obesity as globally virulent, threatening, and uncontrollable are meant to warn against the obesity pandemic - if one is not careful, one can become obese. However, perceiving obesity as contagion, which it is not, assigns personal responsibility and stigma, and it can trigger avoidance and stereotypical attitudes such as isolation and removal from groups of persons living with it.

Images that anchor textual meanings of metaphorical framing overtly link interpretations to bias and stigmatisation. Although images of active and sports-involved persons with obesity are present, depictions of headless and idle people represent the pervasive imagery associated with obesity. Whether for reasons of privacy laws, consideration, and civility, and therefore, well-intended, when these images refer to bulging stomachs and individuals gulping junk food, they can have a dehumanising, damaging effect.

Conclusion

The analysis of Romanian online newsfeeds reveals that, while obesity is predominantly framed through collective risks and societal struggles, it is also associated with individual vulnerability, exposure to harm, negative portrayals, and stigma. At the visual level, online news discourse reinforces stigmatising representations of bodies affected by obesity, thereby amplifying social marginalisation. By examining both the messages about obesity and their linguistic realisation, discourse analysis can raise critical awareness of the stigma attached to obesity and contribute to reshaping public understanding of the condition. To conclude, journalists and policymakers in particular could refine their communication by adopting less stigmatising discourse strategies when reporting on issues related to obesity.

WORKS CITED

Primary resources – Online news case studies

- [1] <https://www.digi24.ro/magazin/stil-de-viata/viata-sanatoasa/obezitatea-periculoasa-la-pubertate-ce-trebuie-sa-faca-parintii-699167>.
- [2] <https://www.digi24.ro/stiri/actualitate/obezitatea-creste-riscul-de-deces-din-cauza-covid-19-cu-aproape-50-si-ar-putea-reduce-din-eficienta-unui-vaccin-1358596>.
- [3] <https://www.digi24.ro/stiri/actualitate/social/pandemia-a-dus-la-o-crestere-a-cererilor-pentru-operatii-de-micsorare-a-stomacului-1365137>.
- [4] <https://www.digi24.ro/stiri/sci-tech/descoperiri/cercetatorii-au-aflat-de-ce-persoanele-obeze-au-un-risc-mai-mare-de-a-se-imbolnavi-grav-si-a-muri-de-covid-1765139><https://www.digi24.ro/stiri/actualitate/sanatate?p=2>.
- [5] <https://www.digi24.ro/magazin/stil-de-viata/viata-sanatoasa/medic-dieta-nu-este-altceva-decat-temelia-kilogramelor-in-plus-de-ce-nu-e-bine-sa-incepi-dieta-de-dimineata-1540309>.
- [6] <https://www.digi24.ro/stiri/externe/mapamond/medicamentul-care-ar-putea-deschide-o-noua-era-in-lupta-cu-obezitatea-o-femeie-care-l-a-folosit-a-pierdut-o-cincime-din-greutate-1449955>.
- [7] <https://www.digi24.ro/stiri/sci-tech/descoperiri/descoperire-ce-ar-putea-ajuta-copiii-obezi-sa-slabeasca-740730>.
- [8] <https://www.digi24.ro/stiri/actualitate/sanatate/semnal-de-alarma-tot-mai-multi-pacienti-cu-obezitate-morbida-ajung-la-urgente-1435387>.
- [9] <https://www.digi24.ro/stiri/externe/mapamond/studiu-pacientii-cu-exces-de-greutate-sunt-mai-predispusi-sa-sufere-o-forma-grava-de-covid-19-tinerii-sunt-mai-vulnerabili-1513645>.
- [10] <https://www.digi24.ro/stiri/actualitate/sanatate/semnal-de-alarma-tot-mai-multi-pacienti-cu-obezitate-morbida-ajung-la-urgente-1435387>.
- [11] <https://www.digi24.ro/stiri/sci-tech/descoperiri/obezitatea-trebuie-privita-ca-o-boala-infectioasa-867733>.
- [12] <https://www.digi24.ro/stiri/sci-tech/descoperiri/descoperire-revolutionara-o-mutatie-genetica-rara-ii-fereste-pe-oameni-de-obezitate-1591697>.
- [13] <https://www.digi24.ro/stiri/actualitate/sanatate/europa-se-confrunta-cu-o-epidemie-de-obezitate-provocata-de-alimentele-ultraprocesate-1135134>.
- [14] <https://www.digi24.ro/stiri/sci-tech/descoperiri/obezitatea-trebuie-privita-ca-o-boala-infectioasa-867733>.
- [15] <https://www.digi24.ro/stiri/actualitate/sanatate/solutia-pentru-complicatiile-severe-ale-obezitatii-1860091>.
- [16] <https://www.digi24.ro/magazin/stil-de-viata/emotiile-care-ne-ingrasa-provocarile-societatii-moderne-pe-care-romanii-le-trateaza-cu-mancare-964722>.
- [17] <https://www.digi24.ro/magazin/stil-de-viata/emotiile-care-ne-ingrasa-provocarile-societatii-moderne-pe-care-romanii-le-trateaza-cu-mancare-964722>.

- [18] <https://www.digi24.ro/stiri/actualitate/sanatate/pacient-in-romania/grasut-si-frumos-nu-inseamna-grija-parentala-doua-greseli-pe-care-le-fac-parintii-duc-la-situatii-deseori-umilitoare-pentru-copii-1595049>.
- [19] <https://www.digi24.ro/stiri/actualitate/educatie/absolventii-de-gimnaziu-ar-putea-sustine-si-o-proba-de-sport-la-evaluarea-nationala-proiect-1842803>.
- [20] <https://www.digi24.ro/magazin/stil-de-viata/viata-sanatoasa/poti-fi-gras-si-sanatos-817283>.
- [21] <https://www.digi24.ro/stiri/sci-tech/descoperiri/concluziile-unui-studiu-burti-mari-creier-mai-mic-1063144>.
- [22] <https://www.digi24.ro/stiri/actualitate/sanatate/oms-numarul-copiiilor-obezi-a-crescut-de-10-ori-in-ultimii-40-de-ani-808713>.
- [23] <https://www.digi24.ro/stiri/actualitate/sanatate/europa-se-confrunta-cu-o-epidemie-de-obezi-tate-provocata-de-alimentele-ultraprocesate-1135134>.
- [24] <https://www.digi24.ro/stiri/actualitate/sanatate/romania-si-ue-topul-obezi-tatii-767411>.
- [25] <https://www.digi24.ro/stiri/actualitate/sanatate/britanicii-sunt-sfatuiti-sa-manance-mai-putin-pentru-a-si-reduce-riscul-de-a-muri-din-cauza-covid-19-1344057>.
- [26] <https://www.digi24.ro/stiri/externe/ue/fumatul-nu-va-mai-fi-principala-cauza-de-cancer-in-urmatorii-ani-amenintarea-imensa-la-adresa-sanatatii-publice-1001967>.
- [27] <https://www.digi24.ro/magazin/stil-de-viata/viata-sanatoasa/bolile-provocate-de-kilogramele-in-plus-la-copii-876596>.
- [28] <https://www.digi24.ro/stiri/actualitate/sanatate/drama-fetitei-care-la-10-ani-avea-63-de-kilograme-nu-esti-buna-eu-sunt-slab-eu-pot-tu-nu-poti-1161561>.
- [29] <https://www.digi24.ro/magazin/stil-de-viata/viata-sanatoasa/in-marea-britanie-persoanele-obeze-somere-ar-putea-primi-tratament-pentru-slabit-pentru-a-se-reintoarce-la-munca-2969195>.

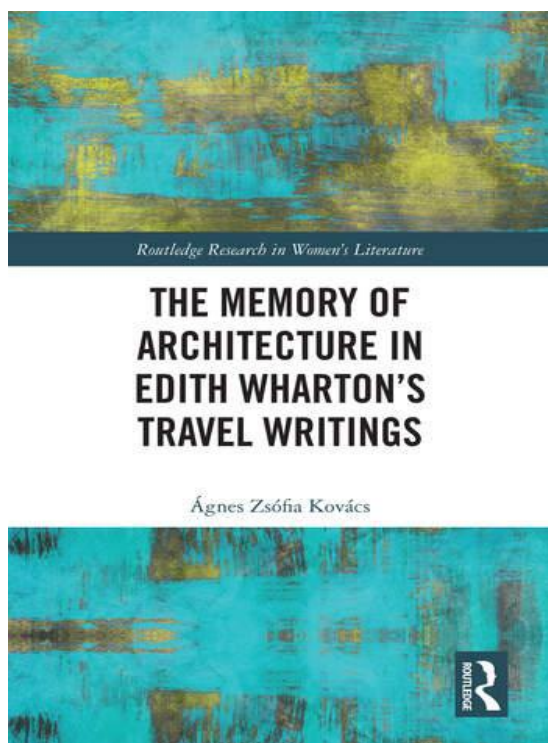
REFERENCES

- Acetos, Matia. 2017. Acetos. "Mythologies of Roland Barthes. Language is Never Innocent. Monographic issue in honor of Roland Barthes." *Romance eReview*, Vol. 21.
- Barry, C.L., et.al. 2009. "Obesity metaphors: How beliefs about the causes of obesity affect support for public policy." *Milbank Quarterly* 87: 7-47. doi:10.1111/j.1468-0009.2009.00546.x.
- Boero Natalie, C. 2007. "All the news that's fat to print: the American "obesity epidemic" and the media." *Qual. Sociol.* 30 (1): 41-60.
- Buckels, E. E., and Trapnell, P. D. 2013. "Disgust facilitates outgroup dehumanization." *Group Processes & Intergroup Relations*, 16(6):771-780. <https://doi.org/10.1177/1368430212471738>.
- Connected for Health project - https://connected4health.pixel-online.org/index_nogrant.php.
- Cotter, Colleen, Danniella Samos, and Deborah Swinglehurst. 2021. "Framing obesity in public discourse: Representation through metaphor across text type," *Journal of Pragmatics*, Vol. 174: 14-27, <https://doi.org/10.1016/j.pragma.2020.12.015>.
- Flint, Stuart W., Sophie Reale. 2014. Obesity stigmatisation from obesity researchers. *The Lancet*, Volume 384 (9958): 1925 – 1926.

- Foster, G.D., et al. 2003. "Primary care physicians' attitudes about obesity and its treatment." *Obesity Research*, 11(10):1168-1177.
- Harjunen, H. 2016. *Neoliberal Bodies and the Gendered Fat Body* (1st ed.), London: Routledge. <https://doi.org/10.4324/9781315583976>.
- Hilton, Shona, Chris Patterson, and Alison Teyhan. 2012. "Escalating coverage of obesity in UK newspapers: the evolution and framing of the obesity epidemic from 1996 to 2010", *Obesity* 20 (8): 1688-1695.
- Landau, M.J. 2010. B.P. Meier, L.A. Keefer. "A metaphor-enriched social cognition." *Psychological Bulletin* 136(6):1045-67. doi: 10.1037/a0020970.
- Landau, M.J. and L.A. Keefer. 2015. "The persuasive power of political metaphors." In *Social psychology and politics*, edited by J.P. Forgas, K. Fieldler and W. Crano, , 129-142. New York, NY: Psychology Press, Taylor and Francis Grp.
- Lieberman, D.L., J.M. Tybur, and J.D. Latner. 2012. "Disgust sensitivity, obesity stigma, and gender: contamination psychology predicts weight bias for women, not men." *Obesity* (Silver Spring) 20(9):1803-14. doi: 10.1038/oby.2011.247.
- Manning, Jimmie. 2017. "In vivo coding." In Matthes, J. (Ed.), *The international encyclopedia of communication research methods*. New York, NY: Wiley-Blackwell.
- Park, J.H., F. Van Leeuwen, and Y. Chochorelou. 2013. "Disease-Avoidance Processes and Stigmatization: Cues of Substandard Health Arouse Heightened Discomfort with Physical Contact." *The Journal of Social Psychology* 153(2): 212-228. <https://doi.org/10.1080/00224545.2012.721812>.
- Park, Justin H., Mark Schaller, and Christian S. Crandall. 2007. "Pathogen-avoidance mechanisms and the stigmatization of obese people." *Evolution and Human Behavior* 28(6): 410-414, <https://doi.org/10.1016/j.evolhumbehav.2007.05.008>.
- Pop, Anisoara. (2022). "Obesity and Fat Stigma versus Eating Disorders and Slenderness Idealization. The Humanities' Role." *Acta Marisiensis. Philologia* 4 (1): 19-25, 10.2478/amph-2022-0061.
- Puhl, R.M. and C.A. Heuer. 2010. "Obesity stigma: important considerations for public health." *American Journal of Public Health* 100(6):1019-1028.
- Puhl RM, et al. 2013. "Headless, hungry, and unhealthy: a video content analysis of obese persons portrayed in online news." *J Health Commun.* 18(6):686-702. doi: 10.1080/10810730.2012.743631.
- Scarry, E.1985. *The body in pain – the making and unmaking of the World*. New York: Oxford University Press.
- Semino, Elena, et al. 2020. *Metaphor. Cancer and the End of Life. A Corpus-Based Study*, New York: Routledge.
- Smith, R.A. 2007. "Language of the lost: An explication of stigma communication." *Communication Theory* 17: 462-485. doi:10.1111/j.1468-2885.2007.00307.x.
- Smith, R. A. 2012. "An Experimental Test of Stigma Communication Content with a Hypothetical Infectious Disease Alert." *Communication Monographs*, 79(4): 522-538. <https://doi.org/10.1080/03637751.2012.723811>.
- Sontag, Susan. 2002. *Illness as Metaphor & AIDS and its Metaphors*. London: Penguin Group.
- Woolf, Virginia. 1926. "On Being Ill." In: *The new criterion – a quarterly review*, edited by Eliot TS. London: Faber & Gwyer, Ltd.

BOOKS

Ágnes Zsófia Kovács, *The Memory of Architecture in Edith Wharton's Travel Writings*, New York: Routledge, 2025, 226 p.



Drawing on a post-theory definition of travel writing as social discourse, Ágnes Zsófia Kovács's book argues that a close analysis of architectural and natural spaces in Edith Wharton's travel writing repositions these texts within a larger cultural field shaped by aesthetic, social, and historical conventions and by the institutionalization of early twentieth-century publishing. Most readers today associate Wharton with major novels of manners such as *The Age of Innocence* (1920) and *The House of Mirth* (1905), even though, as Kovács notes, Wharton was actually better known during her lifetime for her travel writing. Recent criticism has done much to change this view by broadening the Wharton canon through archival discoveries and the publication of little-known documents. It has also brought an interdisciplinary lens to Wharton's engagements with ethnography, visual culture, and cultural history—

discursive practices that together form the nebulous generic category of her "travel writing." Kovács's book is part of this growing body of criticism, as it seeks to complement existing research, including the major book-length study on the topic, Sarah Bird Wright's *Edith Wharton's Travel Writing* (1992) as well as articles by Annette Benert, Nancy Bentley, Emily Orlando, and Gary Totten. It aims to show that an



analytical framework combining art-historical and ethnographic perspectives can reveal a more nuanced understanding of Wharton as a reflective thinker concerned with metaphysical questions such as the historical and cultural continuity encoded in her representation of architectural monuments, cultivated spaces, and natural landscapes.

To achieve this, the book maps the geographical areas covered in Wharton's travel writing and situates them within the context of nineteenth-century traditions of writing about travel and art history. In addition to the five major volumes of travel writing that Wharton published during her lifetime, Kovács also includes several other travel pieces brought to light by recent academic and archival research. Chapters 1 and 2 investigate Wharton's views of ancient and modern Italy, focusing respectively on her aim to write a handbook of garden architecture in *Italian Villas and Their Gardens* (1904) and on her theoretical reflections on art history in *Italian Backgrounds* (1905). The next two chapters explore Wharton's writings on France, where, as an American expatriate, she contemplates the sentimental feelings evoked by Gothic cathedrals in *A Motor-Flight Through France* (1908) and contemporary popular discourse about French civilization during the Great War in *Fighting for France* (1915). Chapter 5 examines *In Morocco* (1920), focusing on how Oriental stereotypes and French colonial discourse shape Wharton's portrayal of North African society and culture. The last two chapters are dedicated to more recently discovered travel writings documenting Wharton's travels in the Aegean – her 1888 diary *The Cruise of the Vanadis* and fragments about her 1926 trip collected in *Osprey Notes* – as well as accounts of her travels in Spain, including her "Spain Diary" ([1925] 2011), "Back to Compostela" ([1928] or [1930] 2011), and "A Motor-Flight Through Spain" (1928 or 1930).

Crucial to these arguments is a critical attention to Wharton's position as a female author in the contemporary literary marketplace, along with an examination of how her works refract both earlier traditions of travel writing and the dominant aesthetic theories of her time. Following Wright, Kovács argues that Wharton's travel writing occupies a position between the earlier belletristic tradition, focused on impressionistic sketches and the picturesque, and the genteel tradition of the Grand Tour, devoted to individual education and the cultivation of the nation. The former includes, for instance, Washington Irving's *Alhambra* (1832) and Nathaniel Hawthorne's *French and Italian Notebooks* (1864), and the latter is exemplified most notably by John Ruskin's ideal of precise and objective observation. While in her descriptions of architecture Wharton clearly prioritizes the Ruskinian model, Kovács shows how the business of publishing shaped Wharton's travel writing, as editors sometimes expected her to produce more impressionistic pieces aimed at a female readership, as in *Italian Villas and Their Gardens*. On the other hand, the book also explores how Wharton departs from Ruskin's legacy of aesthetic judgment and moral didacticism, adopting instead a cultural-historical mode of understanding in her architectural representations and the cultural continuity embedded within them.

The discursive definition of travel writing enables Kovács to situate Wharton's work within multiple interlocking contexts, including broader nineteenth-century U.S. cultural practices linked to technological progress, economic development, politics, and social change, as well as the institutionalization of the contemporary literary

marketplace. According to Kovács, Wharton navigates the belletristic, genteel, and objectifying traditions of travel writing, “moving away from the legacy of a Romantic sensitivity through an objectivizing continuity-oriented stance towards an agnostic aesthetic sensuousness of modernist taste” (24). Significantly, the democratization of travel in Wharton’s time went hand in hand with her development as a professional female travel writer. Starting in the 1840s, rapid industrialization and social change made travel possible for an increasingly large middle class population, whose interests and tastes also influenced, to some degree, the modes of publishing and the contemporary reception of various writings on travel. Kovács’s study is sensitive to this sophisticated process, which allowed Wharton to convert her cultural capital as a learned author into profit within the literary marketplace, while continually negotiating with magazine editors and book publishers over readers’ expectations, deadlines, marketing strategies, and sales revenues. In keeping with the custom of the age, Wharton’s poems, essays, short stories, and book chapters were first published in illustrated monthly magazines – both highbrow and middlebrow – including *Harper’s Bazaar*, *Scribner’s Magazine*, *The Century Magazine* as well as the *Pictorial Review*, the *Delineator*, and the *Ladies’ Home Journal*. Just how important publication history is to Wharton’s travel writing is evidenced by the key changes she made to the magazine pieces when adapting them to books for a changing audience.

The scope and methodology of Kovács’s book are defined by a post-theory conception of travel writing as flexible, contextual, and interdisciplinary—pragmatically blending multiple approaches and shifting perspectives within frameworks that are spatially and historically localized, fragmented, and overlapping. The definition of travel writing as social discourse draws on Michel Foucault, whose work undergirds Edward Said’s account of Orientalism as the signifying practice through which Western culture has sought to manage and produce the Orient. Yet readers of Kovács will find that her interpretations of Wharton do not rely on any dominant, abstract theoretical framework such as poststructuralism, postcolonialism, feminism, or Marxism. Instead, her analysis emphasizes a loose definition of travel writing “as a group of discursive productions within a cultural field” (203), insisting on the genre’s fluidity and relying not on all-encompassing theories but on spatial metaphors such as “margins,” “borders,” and “extensions.” Doing so allows Kovács to trace the significance of travel writing within Wharton’s continually expanding oeuvre and to locate the topographies of her writing across a broad cultural field that blurs generic distinctions, including the distinction between fiction and nonfiction.

The implications of Kovács’s study extend well beyond Wharton scholarship. By focusing on Wharton’s popular work rather than her major novels, the book repositions literary history around notions of contemporary popularity rather than subsequent canonicity. Her definition of travel literature as a loosely assembled body of texts with permeable generic boundaries subverts traditional literary hierarchies and opens the way for further scholarly inquiry into cultural history and post-theory travel writing studies. A disaggregated theoretical stance is valuable because it seeks to move beyond the limitations of grand narratives and overarching theories. It may also explain the overall structure of the book, which reads more like a collection of

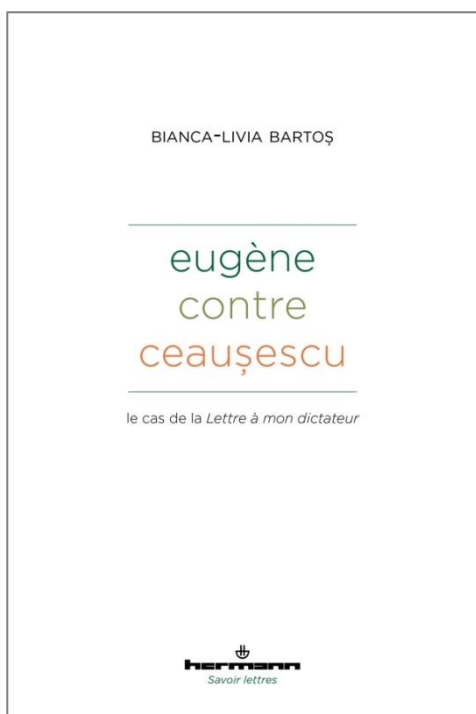
individual essays, each with its own introduction and conclusion, only tenuously related to the author's overall argument. The post-theory framing challenges readers to connect feminist and postcolonial readings of Wharton with the institutional contexts of publishing, which complement, but are not always fully integrated into, the book's thesis about architectural and natural topography. The book raises important questions that invite further research, including how Wharton's pursuit of cultural continuity might intersect with experiences of disruption, cultural destruction, and loss in her wartime travel writing. Conceived as "a textually oriented survey of Wharton's travel writings" (209), *The Memory of Architecture in Edith Wharton's Travel Writings* lays the groundwork for further inquiry into topics such as aesthetic representation and spatial description in the travel writings of Henry James, for instance, as well as Wharton's relationship to modernist authors including Marcel Proust. It will appeal to a broad academic readership while serving as an important resource for scholars and researchers working in the fields of travel literature, aesthetics, ethnography, visual culture, and cultural history.

Daniel DARVAY

Babes-Bolyai University, Cluj-Napoca, Romania
E-mail: daniel.darvay@ubbcluj.ro

BOOKS

Bianca-Livia Bartoș, *Eugène contre Ceașescu. Le cas de la Lettre à mon dictateur*, Paris : Hermann, 2024, 108 p.



L'essai *Eugène contre Ceașescu. Le cas de la Lettre à mon dictateur* écrit par Bianca-Livia Bartoș, enseignante-chercheuse à l'Université Babeș-Bolyai de Cluj-Napoca, est paru aux Éditions Hermann en 2024. L'auteure y analyse le récit de l'écrivain suisse en s'appuyant sur une approche psychanalytique, tout en suivant le parcours de celui-ci dans son dépassement du trauma jusqu'à la résilience. Dans ce cadre, elle examine notamment le lien entre Ceașescu et Eugène, ainsi que les conséquences de cette relation sur le devenir personnel et professionnel de l'écrivain.

L'étude est structurée en huit chapitres, dont les titres reflètent les étapes de la quête identitaire de l'écrivain. Pour commencer, la première séquence, « Meiltz-Ceașescu : un hiatus ontologique » présente les deux protagonistes de l'histoire, en esquisant les rapports entre eux. La deuxième séquence, « Seuils », s'appuie sur l'élément déclencheur de toute l'histoire : le départ pour la Suisse et la reconnaissance

qu'Eugène éprouvera plus tard pour le courage de ses parents. En outre, Bartoș met en avant l'ironie caractéristique aux Roumains qui distingue l'écrivain du *modus vivendi* suisse et témoigne de ses origines. Par la suite, les trois séquences qui s'enchaînent : « La "persona" », « La quête » et « La dette » analysent avec finesse l'effort d'Eugène de délivrer son *Soi* authentique de l'emprise des masques de la *persona*.

« Le procès » est le sixième chapitre de l'étude, qui se concentre sur une analyse critique de la pièce de théâtre insérée par Eugène dans son récit. Celle-ci met en scène le procès du couple Ceaușescu, stratégie privilégiée par l'écrivain afin de conférer à son destinataire le droit à la parole, en l'investissant comme protagoniste de cette séquence. Le chapitre suivant, intitulé « Le "Soi" et la résilience », marque l'aboutissement de la quête de l'écrivain qui, délivré de sa *persona*, peut désormais tourner la page de son obsession pour le dictateur et s'ouvrir à d'autres horizons. Enfin, la dernière séquence, « Le post-« ceaușisme » : un néocommunisme ? », retrace habilement la liaison entre la quête d'Eugène et celle de son peuple, qui, sans se rendre compte, change le communisme pour un néocommunisme. Bianca-Livia Bartoș souligne cette continuité en affirmant que la fin de la lettre et l'intention de l'écrivain de la déposer lui-même sur la tombe de Ceaușescu est, pourtant, loin de représenter la fin du régime totalitaire en Roumanie.

Ainsi, l'étude critique met en lumière, avec finesse, la construction identitaire d'Eugène à travers le prisme de la psychologie analytique de Jung. Les concepts mobilisés s'accordent parfaitement au contexte du récit, ce qui permet à l'auteure de souligner toute la profondeur psychologique de la *Lettre à mon dictateur*, en plaçant la quête identitaire d'Eugène sous les auspices de la résilience par l'écriture. La pertinence et la finesse de l'analyse se révèlent avec éclat dans le troisième chapitre, qui en constitue une illustration particulièrement éloquente : ici, l'autrice s'interroge sur les implications psychanalytiques du refus catégorique d'Eugène d'admettre la moindre possibilité de relation avec Nicolae Ceaușescu. En effet, Bartoș avance que, selon Jung, le refus de toute relation avec le dictateur relèverait d'une attitude propre à l'enfance, marquée par le rejet systématique de ce qui engendre le mécontentement. Cependant, elle soutient que cette hypothèse serait mise sous le signe de l'improbabilité, puisque l'acte d'écrire, entrepris de nombreuses années après les faits, ne saurait correspondre à de simples actions enfantines.

De surcroît, Bartoș réalise une analyse complexe et détaillée des aspects psychanalytiques liés au procès de la résilience par l'intermédiaire de l'écriture. Les arguments choisis pour appuyer les hypothèses initiales sont pertinents et habilement structurés afin de rendre l'ensemble de l'étude logique et homogène. De plus, les exemples choisis pour illustrer les aspects théoriques sont adéquats et insérés savamment. Finalement, elle reconnaît que son analyse ne se veut pas exhaustive et qu'il y a aussi d'autres pistes à exploiter dans une critique de la *Lettre à mon dictateur*, comme, par exemple, l'aspect esthétique, la *poiësis* (p. 91).

En fin de compte, *Eugène contre Ceaușescu. Le cas de la Lettre à mon dictateur* représente un outil indispensable pour qui veut découvrir les profondeurs psychanalytiques du roman d'Eugène. De plus, l'étude repose sur une bibliographie riche qui englobe, à part l'exégèse de l'écrivain et les références des théories littéraires, des titres importants dans le domaine de la psychologie et aussi de nombreux études sur le communisme. Ainsi, même si celui-ci s'adresse plutôt aux étudiants et aux chercheurs, il représente une lecture enrichissante pour toute personne intéressée aux études interdisciplinaires.

Andreea GÎNGA

Université Babeș-Bolyai, Cluj-Napoca, Roumanie
E-mail: andreea.ginga1@stud.ubbcluj.ro

BOOKS

Adrian Tudurachi (Ed.), *Dicționarul cronologic al romanului românesc de la origini până în 2000*, revised second edition, Cluj-Napoca: Presa Universitară Clujeană, 2023, 770p + 757p (2 volumes)

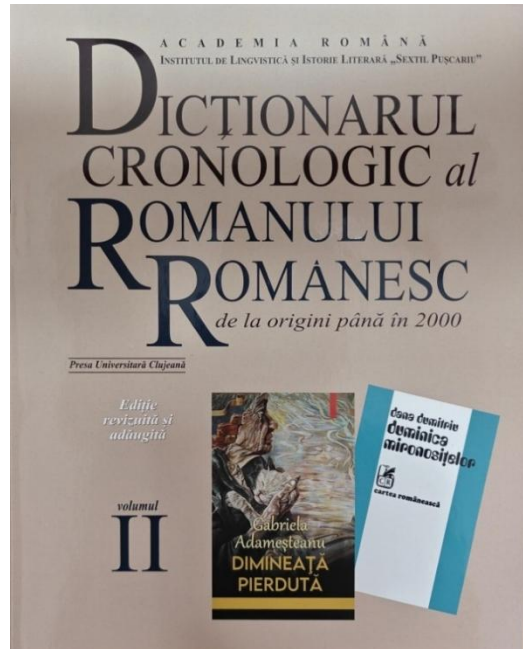
The new edition of *The Chronological Dictionary of the Novel from its Origins to the 2000s* [*Dicționarul cronologic al romanului românesc de la origini până în 2000*] is a large-scale, long-term undertaking, strengthened by an extensive research team of Sextil Pușcariu Institute from Cluj-Napoca. Intended as an “exhaustive bibliography”, the dictionary encompasses two chronological volumes, with the year of 1977 marking the end of the first tome and the beginning of the second. The work has a double aim, as mentioned in

the introduction: to illustrate the production and the literary reception of the novel and to highlight the social dynamics of literary forms. Although situated within literary studies, the project advances clearly articulated theoretical and sociological objectives.

The first volume opens with an overview of the contributors responsible for the dictionary entries. After that, the introduction reconstructs the contextual framework of the first edition’s publication, focusing on the sociopolitical influences on the literary field. As such, the authors of this section – Cosmin Borza, Alex Goldiș and Adrian Tudurachi – underline the dynamics between cultural legislation and literary interest. The year 1974 represents a pivotal moment for Romanian scholarship in this area, a period when cultural institutions transitioned from state subordination to the authority of universities.



With this shift, literary projects transcended the framework of national cultural history to engage with the transnational circulation of literary theory. The introduction traces several turning points in Romanian literary criticism, noting that the dictionary's first edition [2004] was initially read through an aesthetic and hermeneutic lens. Only after 2016, with the Romanian reissuing of Pascale Casanova's *The World Republic of Letters*, was it reinterpreted within a broader systemic framework. As the researchers point out, a paradox emerges at this point: the dictionary, which would ordinarily "serve a cultural heritage function, preserving and organizing existing knowledge, is now called upon to occupy a central position within the shifts and explorations of contemporary research" (p. XX). In this regard, the 2023 republication marks a stage in Romanian literary criticism's maturity, aligning it with world literary discourses.



First of all, it is important to acknowledge the merit of this new edition, which provides a significant update to the first. In one of the prefatory texts, Adrian Tudurachi, director of the Sextil Pușcariu Institute's Department of Literary History, provides an overview of the specific contributions of this new edition under his coordination: 300 articles were completed and rewritten, 100 titles initially omitted were identified, and a total of 116 novels have been recovered through the databases developed in recent years within projects led by Roxana Patraș, Ștefan Baghiu, and Vlad Pojoga. Although both the first and the second edition of the dictionary were undertaken by the team of the Sextil Pușcariu Institute, the project has involved a significantly larger research team in recent years. This lexicographical instrument was designed to reflect a broad perspective on the Romanian novel, tracing its development within the local context. In this sense, the research team has taken all forms of novelistic production into account, paying particular attention to serialised work (*feuilleton*). Furthermore, erroneous translations of novels present in the first edition were removed, as they could cause inaccuracies in studies on the reception of Romanian literature beyond national borders. It is notable that the dictionary also takes novels authored by Romanians in the diaspora or in exile during the 1990s into consideration, assessing these contributions as an integral component of Romanian literature. After highlighting the outstanding merits of the dictionary for this field, Tudurachi acknowledges the contributions of all the researchers involved in this five-year process of examining and reevaluating the

entire material of the previous edition. Without making claim to any local prestige, the dictionary is presented as a working project that is both intergenerational and national, having fostered the participation of scholars from multiple Romanian research centers. At the same time, the merit of libraries nationwide is recognized through interlibrary loans and the digitalization of certain texts, procedures that provided researchers with access to materials otherwise unavailable in their region. Despite not claiming to be exhaustive, the shortcomings of this reedition are noted in the preface. One such shortcoming refers to the awards sections and to details pertaining to the circulation of the novels, both of which remain incomplete.

The next section of the dictionary outlines the terminology of the subgenres employed in the taxonomy of the novels, a process that represents, arguably, the most remarkable contribution. In other words, the new edition of the dictionary doesn't only cover the morphology of the Romanian novel, but connects the development of this genre to a larger international context. Therefore, the project serves as an instrument for *World Literature* methodologies, which open up Romanian literature to dynamics beyond its national borders. More than being a qualitative hermeneutic undertaking, this lexicographic tool is intended as a quantitative framework designed to provide a foundation for studies investigating the dynamics of subgenres within the Romanian literary field. The terminological table placed at the beginning makes a list of Romanian subgenres together with their corresponding Anglo-American and French equivalents. In addition to widely circulated novelistic terminology, the dictionary also incorporates local formulas for which there are no direct foreign correspondents. Some examples would be *the novel of the „obsessive decade”* [*romanul „obsedatului deceniu”*], a subversive subgenre in regards to the communist regime, *the novel of uprising* [*romanul răscoalei*], centred on Romanian peasant revolts, *the novel of the revolution* [*romanul revoluției*], centred on the fall of the communism in 1989, and *the textualist novel* [*romanul textualist*], which highlights the late 20th century Romanian movement of the same name. In several instances, international terminology overrides attempts to introduce local subgenres. This is illustrated by Mircea Eliade's *Maitreyi* (1933), framed as an *existential novel*, but containing the formula of *the novel of experience* [*romanul „experiențialist”*], which refers to the interwar Romanian intellectual current bearing the same name. In the case of some entries from the first edition, the authors attempted, with the resources at their disposal, to provide a subgenre categorization. However, this classification remains tentative and frequently lacks precision. For instance, it applies broad formulations, such as that of the *erotic novel*, or assigns a novel to multiple subgenres, as is the case for the 1961 novel *A Lost Rider in the Steppe. From the Time of Petru the Great* [*Un călăreț pierdut în stepă. Din vremea lui Petru cel mare*] by Felix Aderca, which, in the first edition, is classified as a *socialist realist, historical and adventure novel*, but in the second it is only classified as an *adventure novel*. In numerous instances, novels were later subject to a reevaluation of their subgenre, even when the original taxonomy reflected widely used terminology. This is exemplified by *I came from darkness* [*Veneam din întineric*] by Nicolae Jianu, which transitions from the category of *memoir* to that of *autobiography*. The researchers' effort was to classify each novel within a subgenre, a task that in some cases required negotiation, since several

novelistic formulas coexisted in the same text. Simultaneously, certain initial classifications were based largely on thematic criteria, without considering additional factors in the assignment of labels. The second edition, however, provides a comprehensive view of both the publication context of the novels and the reception of these works. This can be seen, for example, in the case of Ury Benador's 1934 novel *Ghetto 20th Century* [*Ghetto Veac XX*], which shifts from "novel about the Jewish community and milieu" to *ghetto novel*, a category that reflects the subgenre's profound political implications. Given the terminological clarifications, subgenres may function as interpretative keys for the novels, directing the reader to the text's core idea.

Even though the dictionary is not physically available yet, it is digitally accessible on the Institute's website and has already been referenced in scholarly studies. Beyond its significant impact on literary scholarship, a desirable outcome of this new edition would be for it to find application in Romanian pre-university education, where the high school curriculum addresses the issue of subgenres. For instance, Mihail Sadoveanu's 1930 novel *The Hatchet* [*Baltagul*] is studied as a "traditional, objective, mythical novel", while actually being a *rural novel*. Similarly, G. Călinescu's 1938 novel *Otilia's Enigma* [*Enigma Otiliei*] is studied as a novel of "Balzacian realism" and as a "objective, urban *Bildungsroman*", whereas the new dictionary frames it as a *social novel*. Failing to integrate lexicographic projects into the school curriculum leaves students trapped between outdated pre-university literature courses and the realities of contemporary literary theory and criticism. Aligning the curriculum with such projects is essential to fostering a precise and nuanced understanding of literature.

Continuing this line of thought, *The Chronological Dictionary of the Novel from its Origins to the 2000s* seeks to function as a catalyst for the formation of a new generation of researchers. It stands as a project that clearly marks a new stage in Romanian literary studies, one that manages to integrate this field within the broader network of *World Literature*.

Irina GORGAN 

Babeş-Bolyai University, Cluj-Napoca, Romania
E-mail: irina.gorgan@stud.ubbcluj.ro

BOOKS

Alex Ciorogar, *Ascensiunea autorului în epoca globalizării digitale*, Cluj-Napoca: Presa Universitară Clujeană, 2025, 418 p.



Alex Ciorogar's book, published in 2025, takes on the complex task of sifting through theories of authorship that have emerged on the academic scene over the past fifty years. Ciorogar notices the strange place that the author occupies today in the literary field. After the death of the author, consecrated in the essays of Roland Barthes and Michel Foucault, the academic world seems to have gotten stuck, unable to go further and propose new theories. Although the two philosophers' works are crucial, they have been "more quoted than read" (2), Ciorogar says, arguing that a new approach is needed, which requires "the death of the myth of the author's return" (12).

The first chapter is an exploration of the controversies that the idea of the individual subject has gone through due to the radical changes registered by the humanist tradition. Ciorogar agrees with Romanian critic Mihaela Ursa's remarks that the author and subjectivity are tied to one another. He identifies here

multiple attitudes: critics who claim there is no way to ignore the speaker and their intention, even if we claim anonymity (Stanley Fish), critics who believe that anti-author readings are meant to validate interpretations (Sophie Rabau) and critics who believe that authorship is already complex enough and its nature allows for it to be interpreted (Richard Shusterman). Ciorogar validates Shusterman's ideas and appreciates the "fan-concept", which refers to ideas that are complex and interpretable. Another reason to contest authorship concerns, of course, its ties to ideology and structures

of power. He also mentions what Cheryl Walker names “persona criticism”, seeing and examining the author in relation with multiple study areas.

Chapter two is split into two parts. The first one looks at authorship theories that appeared in the wake of post-structuralism, posing the question of what happened between the death of the author and the return thereof. Right after the sixties, the author’s intention was no longer relevant as the focus shifted to the language used and its relationship to existent power structures. Another interesting perspective that Ciorogar investigates in this part is the idea that authorship does not belong to the author, but to discourse itself. As such, “the text constructs the figure of the author” (47). He specifies that after poststructuralism, authorship became easy to relativize and must be studied in relation to the new economic situation, global digitality and the new institutions. While some critics observe that the notion of the modern author is a result of capitalist ideology, several neomarxist thinkers note that the gifted author appeared exactly as a reaction to the emergence of the book market. Therefore, the author’s relationship with capitalism is conflicted. Regarding Barthes’ famous essay, Ciorogar argues that the French thinker was concerned not so much with hermeneutics, as people seem to believe, but with finding pleasure in the reading experience itself. That is why the concept of intertextuality should take the blame for any ‘suppression’ of the author. Andrew Bennet remarks that Barthes and others failed to notice that the romantic model of authorship was already complex; the creator is formed by ideas that appear during the process of creation. In the second part of this chapter, Ciorogar engages with methods of studying authorship, including epistemological ones. To begin with, there is no escaping the author. The reader tends to construct an image of the author, even while lacking any empirical knowledge. Julie Smith defines the author as a “rhetorical choir” (98) in the new digital age, but Ciorogar mentions that this term can be applied outside the field of technology. Still, the preferred methodological solution is Atsushi Akera’s ecology of knowledge, which he adapts to authorship studies while keeping in mind the social dimension of authorship, the information that makes up the field and an analysis of the phenomenon’s material practices.

The third chapter is also divided into two parts. It begins with examining the return of biographies in contemporary literary historiography. This debate unites the fields of aesthetic theories and literary studies due to their common involvement in aesthetics and the subsequent crisis. Looking at multiple definitions, Ciorogar arrives at Jean Marie Schaeffer, who defines the aesthetic experience as part of everyday life. Turning to biography, the author explains its evolution, from its honorary position in the Romantic period to its downfall at the time of modernism, when the power of the word itself prevailed. Nowadays, biography takes a central place in political readings which seek to prove why studying literature is still important and profitable in the modern capitalist landscape. According to Ciorogar, “biographical studies have the advantage of having been inspired from the long tradition of representing real life” (127).

The second part of this chapter is dedicated to the era of digital globalization. As authorship is tackled from multiple angles that have emerged due to globalization (ecological issues, eurocentrism critique, translation studies), Ciorogar believes this should not be the object of a single field of studies. Mads Rosendahl Thompson argues that societies are no longer restricted by geopolitical borders, and the only paths forward are either isolation or cosmopolitan states. The latter is preferable. In the era

of “cognitive capitalism” (MacKenzie Wark), immaterial products, like intellectual property, have become profitable. Therefore, the relevance of today’s author depends on the new means of production. Ciorogar explains that globalization is a polarizing topic among thinkers, whom he groups into “pessimists and optimists” (143). Optimist thinkers believe the new era has democratized cultural production, offering transparency, adaptability and alternative communities. Art becomes the trading coin of businesses, picturing the artist as the exemplary entrepreneur. Ciorogar notes that many critics seem to struggle with conceptualizing the global era, but he validates the perspective of German thinker Ulrich Beck. He distinguishes between globalism, which is the worldwide market, and globalization, which refers to the changes institutions must undergo on various levels. He desires a new, cosmopolitan methodology that can change the role of nations forever and combat the dangers of nationalism.

The fourth chapter is dedicated to French thinker Roland Barthes. Ciorogar begins with an interesting taxonomy of authorship studies, structured around the following categories: the Author and Text (for the mimetic, pragmatic and ideological approaches), the Author of the Text (for theories focused on the intention, hermeneutics, biography and psychoanalysis), the Author inside of the Text (for formalism), the Author as the Text (Ciorogar invokes Carla Benedetti here) and the Author together with the Text (for modern theories which rely on relations). Agreeing with Sarah Brouillette, who identifies the representatives of New Criticism and poststructuralism as assailants of the author, Ciorogar sets out to reread the famous Barthesian essay, proving that he was concerned with ethics, politics, and, most interestingly, inspired by Chomsky’s generative grammar. The Romanian critic explains that the goals of this essay were as much political as they were aesthetic. Barthes himself admitted that it was a strategic move, and quite the successful one, managing to “replace the sorbonards with the program of the May ’68 movement” (147). Barthes proved the author to be no hero, but a pawn that helped maintain the power structures already governing literature. The established literary institutions were therefore able to control the narratives of interpretation. The author, as a “modern scribe” (147), had to return to a Homeric state. In Ciorogar’s view, Barthes could be thought of as a pragmatic critic, due to his preoccupation with the moral duty of authors. The most interesting contribution of this chapter is the generative angle, grounded in the belief that there must be some inherent and independent literary rules inscribed within the human mind, dismantling the narrative of the genius.

The fifth chapter tackles the other player in the author’s demise, Michel Foucault. Ciorogar aims to prove that following several conflicts, a new definition of authorship began developing at the end of the sixties. With the author dead, something had to fill the function of authorship. This was Foucault’s interest, whilst also loosening the restrictions the author had upon interpretations. He understood that authors used performative strategies to construct themselves as subjects in relation to structures of power. For Ciorogar, Foucault is the founder of posturality, opening the analysis of authorship with his ideas regarding the author’s function. Here the chapter splits in two subdivisions, one dedicated to somaesthetics and another to the author’s function from a historical perspective. It is interesting how a connection is made between Richard Shusterman’s pragmatic project which became somaesthetics and Michel Foucault’s turn to pragmatic methodology in the last years of his career.

Chapter six begins with a critical overview of Dario Compagno. Ciorogar points out that his synthesis ends up disoriented and confusing. Continuing his discussion on the return of the author, he shows that during the nineties, the relationship between literature and ethics became important, shining a light on the creator's ethics. Critics like John Trimbur and Mihaela Ursa argue for the return of the author, with Trimbur supporting the idea because "the literary experience cannot be imagined outside dialogical theory" (223). Ursa remarks that there is no return to the pre-structuralist status of the author. As such, the answer could be "authorial markers" (225), a concept that can serve as the right tool to manage this complex phenomenon. Carla Bendetti questions the disappearance of the creator but is wrong to say that authorship has not changed. Moreover, Ciorogar critiques her confusion between the literary landscape and the book market. In her view, the existence of an author means the possibility of creating the new. Jane Gallop presents an interesting queer reading and a "necrophiliac" nuance, arguing for a new type of relationship between the reader, writer and author. Another important critic for Ciorogar is Sean Burke, who, like Gallop, adopts the method of close reading for theoretical discourse. For Burke, neither Barthes nor Foucault managed to create new forms of knowledge, despite their innovations. The return of the author means the transformation of the normative-transcendental subjectivity into a biological, material one. The way Burke reads Barthes and Foucault proves that the return was already anticipated by the two thinkers.

The final chapter, titled "The Ascension of the Author", investigates the role that the author's function needs to fulfill in the "extreme contemporaneity" (287). The author may become a tool to promote the book, and eventually a product in their own name. Following the statements of Andrew Wernick, Ciorogar warns that we should not confuse success on the literary market with canonization in literary criticism, although the latter influences the former. The academic world seems to be on a completely different path than the commercial market, which is another argument for what Ciorogar calls the "ecological theory of authorship" (302), a model which allows for the existence of previous contributions from various fields of study, without remaining restricted by them. Each exists in its context with its dynamics. Ciorogar prefers the term ascent rather than return, as it encompasses multiple nuances of current authorship: the verticality of relations, neoliberal economic ascension, or social ascendancy. Only an ecological structure can contain the complexities of the paradox of authorship.

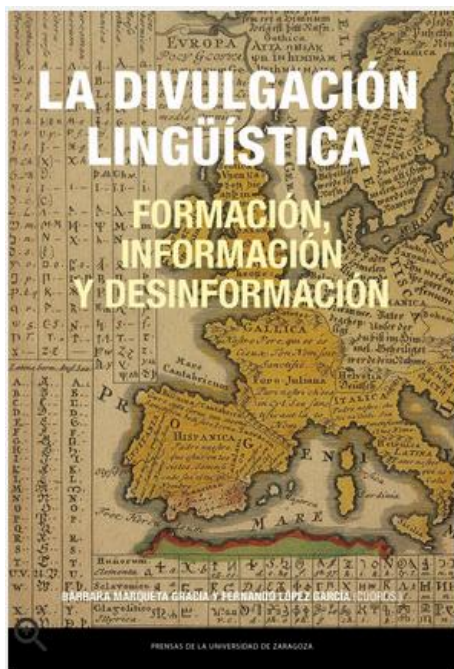
In Ciorogar's own words, his book provides the academic world with a coherent image of metatheoretical theories and methodology, waiting for a case study. Without a doubt, the book is comprehensive, and as exhaustive as possible, with an impressive bibliography. Any prospective reader should be aware, though, that it is a book written for those who have already made forays into the field of authorship. For those readers, it does not only provide the solution to the crisis of the author nowadays, but it offers an inventory of theories, paired with a polemical approach resulting in analytical and evaluative commentary.

Andreea-Diana HOSU

*Babeş-Bolyai University, Cluj-Napoca, Romania
E-mail: andreea.diana.hosu@stud.ubbcluj.ro*

BOOKS

Bárbara Marqueta Gracia y Fernando López García (coords.),
La divulgación lingüística: formación, información y desinformación,
2025, Prensas de la Universidad de Zaragoza, 278 p.



La divulgación lingüística es una actividad necesaria, no solo porque las leyes educativas actuales insisten más en la comunicación social de la ciencia, sino porque esta actividad contribuye a reorientar los imaginarios colectivos sobre las lenguas y a legitimar las disciplinas orientadas a su estudio, cuya condición científica suele denostarse.

Con este espíritu alentador se presenta la obra *La divulgación lingüística. Formación, información y desinformación*: un volumen que reúne a dieciocho especialistas heterogéneos con una dilatada experiencia y cuyo contenido puede resultar de interés para investigadores de perfiles diversos. El manual ofrece un panorama actualizado de la divulgación lingüística en la era digital, combinando la presentación de recursos y ejemplos de proyectos inspiradores de divulgación lingüística con el análisis y la reflexión contrastada sobre la pertinencia de estas actividades.

El volumen consta de nueve capítulos, más un prólogo y un epílogo de los coordinadores, que se distribuyen en tres bloques: “La información y las lenguas”, con tres capítulos dedicados a tratar las iniciativas de divulgación lingüística en los medios de comunicación y a nivel institucional; “La formación y las lenguas”, con dos capítulos orientados a la enseñanza y la formación; y “La desinformación y las lenguas”, con cuatro capítulos centrados en la naturaleza de las lenguas y la diversidad lingüística.

El prólogo, de Bárbara Marqueta Gracia, no se limita a presentar la organización interna de la obra, el perfil heterogéneo de los autores y la orientación temática de sus

capítulos, sino que explica también tres cuestiones importantes que permiten comprender el sentido de la obra. En primer lugar, se recuerda que la Universidad mantiene un compromiso social histórico con la divulgación del conocimiento, si bien este se ha ido reforzando en cada nueva ley educativa, lo cual justifica su protagonismo actual. En segundo lugar, se destaca una diferencia clave entre la idea de *divulgación* que persigue este volumen, entendida como una práctica desinteresada de apropiación de los adelantos científicos por parte de los no expertos, y la noción de *transferencia*, popularizada en el marco de la Universidad empresa y más vinculada a la voluntad de rentabilizar económica o curricularmente las actividades dirigidas a comunicar la ciencia en sociedad. En tercer lugar, se defiende en particular la divulgación lingüística porque existe un problema adicional que afecta a este ámbito: la existencia de una tendencia social generalizada a opinar sobre lenguas sin una base científica, lo cual genera una necesidad de intervenir.

El primer bloque arranca con el capítulo “La divulgación lingüística a través de la lengua escrita”, de Francisco M. Carriscondo Esquivel. El autor explica el sentido etimológico del término *divulgar* y ofrece una definición exhaustiva de la *divulgación lingüística* a partir de la respuesta a las siguientes preguntas: *qué es, cómo se elabora, dónde se publica, quiénes la realizan y para qué se hace*. De este recorrido, se desprende que el divulgador ha de reunir tres características: perspicacia, profesionalidad y exquísitez formal acorde a la formación filológica. Además, el autor incluye ejemplos de buenas prácticas y añade dos recomendaciones: la actividad de divulgación ha de ser pedagógica y lúdica, para enseñar mientras se entretiene, y no conviene confundir la divulgación lingüística con la publicidad comercial de obras lingüísticas, pues esta debe orientarse a la creación de un espacio compartido solidario, comprometido con el entorno cercano.

En el segundo capítulo, “La divulgación lingüística en las redes sociales. El proyecto #Twitter Para Lingüistas”, María del Carmen Horno Chéliz expone la complejidad de divulgar en las redes sociales (RRSS), pues implica un conocimiento de su funcionamiento para lograr empatizar con los usuarios y un cambio en la interpretación de la divulgación. En esta tarea, además, es preciso mediar entre el fin y el medio, dado que no se debe renunciar a la exactitud de la ciencia, pero, a su vez, es preciso adaptar la actividad a un modelo digital cuyo discurso implica la necesidad de una simplificación. En todo caso, pese al reto que plantea el uso de las RRSS, estas son concebidas como un altavoz interesante para multiplicar el efecto de la divulgación en la sociedad, bien sea utilizándolas para compartir la divulgación que se realiza en otros medios o como un lugar de divulgación en sí mismo. En este segundo sentido, la autora aporta la experiencia del proyecto #Twitter Para Lingüistas, cuyos fines se dirigen, entre otros, a fomentar la reflexión metalingüística y la reflexión sobre el uso del lenguaje en los medios de comunicación y a transmitir una imagen adecuada del lenguaje, superando los prejuicios.

En el capítulo tres, “La divulgación del español en el extranjero”, Julio Villa-García, Imanol Suárez-Palma y Hugo Sánchez-Llana destacan la presencia del español entre las lenguas más habladas del mundo y presentan un panorama actualizado de las actividades que realizan las diversas instituciones que favorecen la difusión del español en el extranjero, como el Instituto Cervantes, los centros educativos de titularidad del Estado español en el exterior o las Agrupaciones de Lengua y Cultura Españolas (ALCE),

entre otras. Sobre esta actividad de difusión, los autores destacan que el énfasis actual está en la digitalización, teniendo en cuenta el cambio de paradigma social y educativo.

El segundo bloque, sobre la formación y las lenguas, comienza con el capítulo “La divulgación del lenguaje entre el público infanto-juvenil. Una propuesta de enseñanza de la lengua en la educación primaria”. En él se advierte que la legislación educativa española presenta la lengua a nivel curricular como un instrumento que el usuario debe dominar a partir de la asunción de una serie de reglas, reduciendo el estudio de la lengua a aprender dichas pautas y sin pretender que el estudiante averigüe o razone algo sobre ella. Del mismo modo, las prácticas docentes, las propuestas didácticas y las actividades de los libros de texto de referencia se encuentran orientadas a la comprensión e interiorización de estructuras lingüísticas fijas. Sin embargo, esta perspectiva dista tanto de la finalidad comunicativa de la lengua como de las actitudes que se esperarían del alumnado en el contexto de otras materias. Por tanto, la consideración menos dinámica del estudio de la lengua evidencia el peso que tienen los imaginarios sociales sobre su valor, lo cual justifica aún más la necesidad de estimular la divulgación lingüística.

En el segundo capítulo de este bloque, “La divulgación de las lenguas clásicas”, de Francisco J. Rubio Orecilla, se repasa la presencia de las lenguas clásicas y la cultura grecolatina en diferentes medios de difusión, tales como los libros, el cine, la radio o las webs, incluyendo tanto referencias de autores que han contribuido a divulgar el mundo clásico como las que lo incluyen en cierto modo como escenario para contextualizar o ambientar sus discursos. En su exposición, el autor sostiene que los filólogos y los lingüistas suelen pecar de acomplejados al realizar sus actividades de divulgación, lo cual tiene mucho que ver con la respuesta social esperada, mientras que otros investigadores lo tienen mucho más fácil para acercar los resultados de su debate académico al gran público, dado que se prevé un interés y una aceptación general de sus avances.

El tercer bloque, sobre la desinformación y las lenguas, comienza con el capítulo “El español normativo y su difusión”, de Irene Areces, Antonio Luis Marín Benedicto e Isabel Pérez Jiménez. Los autores definen el *español normativo* como el conjunto de usos lingüísticos que se consideran socialmente prestigiosos en una comunidad de hablantes. Además, alertan sobre la necesidad de no confundir el concepto social de *normativamente correcto/incorrecto* con otros conceptos puramente lingüísticos como *gramatical/a gramatical* o *formal/coloquial*. También abordan cuestiones ligadas a la difusión explícita e implícita de la norma. Para los autores, la normatividad y los estereotipos lingüísticos tienen un gran impacto social, pues al hablante medio de una lengua le preocupa si su forma de hablar y escribir implica que se le caracterice como hablante culto de su lengua.

En el segundo capítulo, “La divulgación científica de la lingüística”, José Luis Mendívil-Giró explica de manera extensa por qué la lingüística es una ciencia en sentido estricto y defiende su dignidad y la pertinencia de su divulgación científica, planteando un análisis profundo del propio concepto de *ciencia* y de los criterios que la definen. Según el autor, la lingüística es una disciplina empírica porque asumimos que su objeto de estudio (el lenguaje y las lenguas) es un objeto natural, que existe como parte de la realidad y es susceptible de estudio científico. Además, esta ciencia presenta un carácter diferencial, pues opera con una información adicional de carácter complejo que no es abordable por otras disciplinas. Asumiendo entonces la lingüística como ciencia, la

divulgación lingüística tiene sentido en tanto el conocimiento logrado con los avances de la lingüística ha de ser público y trascender a la sociedad.

En el tercer capítulo, “Los tópicos lingüísticos. Las lenguas como escenario ideológico”, Pedro Fernández Riquelme y Francisco Joaquín García Marcos rescatan el concepto de *imaginario social* al abordar la construcción histórica e ideológica de diversos tópicos sobre las lenguas. Para los autores, los hechos del lenguaje pueden ser componentes del imaginario social, es decir, formar parte de los clichés y tópicos colectivos, con sus correspondientes roles entre los parámetros ideológicos que maneja una sociedad. Teniendo en cuenta esto, la divulgación lingüística es una actividad muy relevante porque favorece un cambio en el imaginario social: “La transformación en la hegemonía ideológica de una sociedad comporta la inevitable mudanza de su imaginario social, incluidos los componentes lingüísticos” (p. 204). Por otro lado, la divulgación lingüística ha adquirido aún más valor en nuestro tiempo porque el multilingüismo se sigue intensificando en el planeta y “los individuos necesitan más lenguas, además de la materna, para circular de manera eficiente dentro de la aldea global” (p. 213).

Finalmente, en el cuarto capítulo, “La divulgación de las lenguas minoritarias”, Carla Amorós y Henrique Monteagudo Romero proporcionan una descripción actualizada y muy relevante de la situación objetiva de las lenguas minoritarias, la convivencia de lenguas en un mismo territorio, los derechos lingüísticos de los hablantes y la protección legal y cultural de las lenguas vulnerables. Los autores recuerdan que el estado natural de las lenguas en la mayoría de las regiones del mundo no responde a un patrón de homogeneidad y monolingüismo sino de heteroglosia y multilingüismo. Sin embargo, históricamente entraron en juego distintos capitales simbólicos muy desiguales y se desarrollaron políticas colonialistas que incentivaron procesos de minorización. Además, aunque hoy ya podamos ver que el multilingüismo está siendo paulatinamente legitimado, las medidas de promoción de las lenguas minoritarias siguen adoleciendo de una implementación ineficaz o no logran resultados satisfactorios. En todo caso, el fomento del multilingüismo es una actividad esencial para visibilizar y legitimar la competencia plurilingüe de los hablantes, la defensa de una mayor justicia lingüística y social y el logro de un entendimiento mutuo. En relación con esta práctica, los autores nos ofrecen una interesante descripción sobre la divulgación de las lenguas minoritarias en España, perfectamente aprovechable como material didáctico para el aula.

En el epílogo, Fernando López García reconoce la calidad y pertinencia de las aportaciones de los autores y reivindica la necesidad de entender la divulgación como método digno para la transmisión de conocimiento hacia la comunidad lingüística.

En definitiva, estamos ante un volumen novedoso y muy necesario, que abrirá sin duda la puerta a nuevas iniciativas en esta línea. Sin esconder las dificultades y retos que ha de enfrentar el divulgador en la actualidad, más aún si se es lingüista, se anima a los especialistas en el ámbito a aprovechar las ventajas que existen hoy para compartir y comunicar su labor científica, pues cada pequeña contribución, debidamente contrastada, ayuda a dinamizar y dignificar el estudio de las lenguas y su proyección social.

Gonzalo LLAMEDO-PANDIELLA 

Universidad de Oviedo, España
E-mail: llamedogonzalo@uniovi.es

BOOKS

G. Herrera Peña, *Sinfonía de acentos. Rasgos de la variante guatemalteca del español*, Ciudad de Guatemala: Editorial La Pepita, 2025, 93 p.



Si desde el punto de vista dialectal Guatemala fue durante muchos años, en certeras palabras de Lope Blanch (1968: 89), "una de las modalidades peor conocidas hasta ahora" de la América hispánica, no cabe duda de que, por fortuna, tal situación de desconocimiento se ha dejado definitivamente atrás, pues – sin negar la necesidad de mayor investigación sobre cuestiones puntuales como, por ejemplo, la variación dialectal y sociolingüística – la aparición de un conjunto no menor de estudios de muy diverso enfoque y distinta naturaleza permite obtener ya datos de indudable interés sobre la(s) forma(s) como los guatemaltecos emplean el español. Aunque son muchas las investigaciones que se podrían citar al respecto, quizá no esté de más señalar la especial relevancia de aquellas que se deben a Guillermina Herrera Peña, quien en estos momentos constituye la más profunda conocedora de las variedades lingüísticas que conviven en el país, y de ahí que sea de agradecer la aparición de un trabajo como el que se reseña en estas

páginas, en el que se compendian y se exponen de forma detallada y accesible las principales características que las identifican.

En este sentido, es de destacar que *Sinfonía de acentos* – que parte de los datos de algunos de los estudios previos de la autora, a los que se añaden otros novedosos – constituye el primer libro de conjunto sobre las hablas hispánicas que se emplean hoy

en el territorio guatemalteco, y esta voluntad de exponer de forma completa y a la vez sucinta una temática tan amplia y tan compleja se refleja claramente en su organización: así, tras el prólogo y un primer apartado introductorio sobre la unidad y variación que identifican al español (pp. 1-5), se presenta un interesante capítulo sobre las características fundamentales que ofrece este idioma en Guatemala (pp. 7-15), para pasar posteriormente a una descripción ordenada por niveles, tanto de la fonética y fonología (pp. 17-26) como de algunos aspectos de la morfosintaxis (pp. 27-48) y – con mayor extensión – del léxico (pp. 49-83); el libro se cierra con unas breves conclusiones (pp. 85-87) que reafirman las principales ideas expuestas en la obra y en las que, además, se resalta explícitamente el hecho fundamental de que "la variante guatemalteca del español tiene la capacidad de reforzar la identidad de la sociedad que la emplea, lo cual promueve la cohesión social" y sirve a sus usuarios "para distinguirse entre los hispanohablantes por medio de especificidades en el uso de la lengua, que reflejan el origen y la historia de la variante, así como la realidad del propio mundo cultural de sus hablantes" (p. 87). A todo lo anterior se añade, como era de esperar, una bibliografía que, pese a su relativa brevedad (pp. 89-93), tiene el mérito indudable de constituir una auténtica guía de lectura para quien pretenda indagar acerca del español guatemalteco, lo que sin duda supone un aporte no menor del trabajo.

Pasando ya a los capítulos en sí – y más allá de los de carácter introductorio –, se ha indicado arriba que el primero pretende dar cuenta de los principales factores que moldean el español hablado en el país, y para ello se centra en dos que poseen una relevancia fundamental: por un lado, la posición periférica de Guatemala en la época virreinal, que explica el carácter si se quiere *arcaizante* de la variedad; por otro, la importante aportación de las lenguas amerindias a su conformación, muy especialmente – aunque no solo – en lo que se refiere a los préstamos de las lenguas mayas con las que el español entra en contacto, los cuales aparecen tanto en el léxico como en otros niveles más profundos de la lengua, sea el fónico o sea –según se muestra más adelante – la morfosintaxis. Esta primera constatación lleva a la autora a dedicar unas páginas de enorme trascendencia a lo que denomina la *variedad mestiza* del español guatemalteco, entendida como aquella en la que los resultados del contacto con las lenguas mayas se hacen más notorios; a este respecto, además de describir algunos de los fenómenos gramaticales que la identifican (la ausencia de concordancia, el doble posesivo, la simplificación del sistema de pronombres átonos, p. 14), Herrera hace hincapié en su origen – que ubica en "el habla de mayahablantes que aprenden español", pero se extiende posteriormente a "hispanohablantes monolingües que habitan en regiones pobladas mayoritariamente por mayas" (p. 13) – y en la aparente paradoja que supone su escaso prestigio social y su vigencia cada vez mayor, algo que explica por su presencia en el habla de (algunos de) los docentes y en los medios de comunicación, así como por su transmisión en "las familias indígenas, donde los niños aprenden esta variedad del español, aprendida antes por sus padres y familiares" (p.14).

Con todo, frente a este primer acercamiento al habla de los bilingües, el resto del libro se va a dedicar en general a la que emplean los hispanohablantes monolingües, habida cuenta de la escasez de estudios que se le han dedicado (p. 14). Así, en el capítulo consagrado al nivel fonético-fonológico se discuten fenómenos que esta variedad

comparte con otras del mundo hispánico como el seseo, el yeísmo, la distribución regional de la aspiración de /-s/, la nasalización vocálica o la velarización de la /-n/, a lo que se añaden rasgos más restringidos – y, por ello, de mayor peso dialectal – como son el debilitamiento vocálico, la pronunciación del grupo [tl], la incorporación de los fonemas mayas /ç/ y /ʃ/ o los resultados asibilados de la vibrante /r/, en una descripción que, además de presentar los fenómenos, se enriquece con numerosas notas acerca de su realización fonética específica, su distribución y valoración social o su profundidad histórica. De manera semejante, el capítulo de la morfosintaxis se centra en diversos aspectos que sirven para identificar – de manera total o muchas veces relativa, al presentarse también en otras variedades – el español de los guatemaltecos, bien sean del sistema verbal (por ejemplo, los valores de los pretéritos de indicativo, la pluralización del *haber existencial*, ciertos empleos del subjuntivo o algunas cuestiones sobre las perífrasis) o del sintagma nominal, donde se describen determinados usos preposicionales, el sistema pronominal átono de tercera persona y la característica estructura *un mi amigo*; junto a esto, es especialmente destacable la atención que se dedica al voseo, cuya distribución histórica y sociolingüística se detalla con gran atención, matizando – y aclarando – muchas de las confusas y contradictorias informaciones que presenta al respecto la bibliografía.

Finalmente, el capítulo del vocabulario constituye el más extenso de los que se dedican a la descripción del sistema, y en él –tras adelantar algunas claves que explican la conformación léxica de la variedad, tales como la presencia de indigenismos o la dialectalización hispánica del siglo XVI – se hace un recorrido por las voces que utilizan, en diferentes momentos históricos, autores literarios como, entre otros, sor Juana Maldonado y Paz o los decimonónicos María Josefa García Granados, Pepe Batres y Rafael García Goyena; en todos los casos, la autora señala y comenta algunos de los elementos presentes en sus textos, haciendo hincapié en su origen etimológico –con mucha atención a los indigenismos – o en su carácter dialectal, bien sea por su forma o con más frecuencia por su significado. Este recorrido histórico por el vocabulario termina con un último apartado sobre los cuentos contemporáneos de Luis Alfredo Arango y Víctor Muñoz, en los que "los escritores buscan incorporar una imagen de la oralidad de grupos de hispanohablantes urbanos" (p. 73); precisamente, esta circunstancia le sirve a la autora para describir no solo el componente léxico presente en ellos, sino también todos aquellos aspectos con los que se reproduce el registro oral propio de esta época, completando así una visión muy sugerente del vocabulario del país y mostrando, al mismo tiempo, las muchas posibilidades que ofrece la literatura para el estudio de la dialectología del español.

En relación con esta última cuestión, es importante resaltar que el volumen que se está reseñando presenta otra serie de características que suponen también un claro acierto para un libro como este: por un lado, la utilización de un corpus tan amplio como variado – compuesto por prensa, obras literarias e incluso las observaciones personales de la autora (pp. 12-13, 35-36, 52-54, 57-58)- para ejemplificar los fenómenos que interesa mostrar; por otro, las constantes referencias a otras variedades diatópicas americanas(especialmente las de los países limítrofes) para poner en contexto los fenómenos presentes en Guatemala (pp. 11, 23, 25, 27); además, la abundante

incorporación de datos históricos como marco desde el cual entender los resultados actuales que se describen (pp. 8-9, 18-19, 36-38). Al mismo tiempo, es preciso destacar también la minuciosidad que se despliega en el análisis de los fenómenos, que aúna el conocimiento lingüístico de la autora con su condición de hablante nativa de la variedad guatemalteca, lo que le permite no solo valorar con indudable solvencia usos muy específicos y sutiles (p. 39), sino también corregir y matizar, tal y como se señaló más arriba, algunas informaciones inexactas sobre estas hablas que se repiten en los estudios previos (pp. 40-41).

En definitiva, parece evidente a la luz de todo lo expuesto hasta el momento lo oportuno que resulta la aparición de un volumen como el que ahora se reseña, que viene a cubrir un notable *hueco en blanco* en la bibliografía especializada y en el que convive de manera armoniosa un claro afán divulgativo -al presentar una descripción amplia y fácilmente entendible, a la par que rigurosa, del español guatemalteco- con una serie de ideas y cuestiones que interpelan más directamente al investigador y le muestran posibles líneas de trabajo futuro. No queda, pues, sino aprovechar todas las potencialidades que *Sinfonía de acentos* encierra para, desde las sólidas bases que establece, seguir profundizando en el conocimiento de la dialectología del español de Guatemala, pues no cabe duda de que es este y no otro el camino que permitirá, en un plazo más o menos breve, dar por superada – afortunadamente – la cita de Lope Blanch con que dan inicio estas páginas.

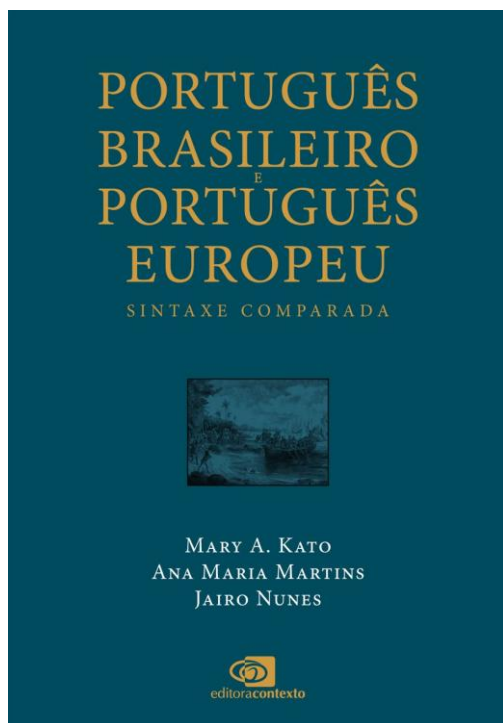
José Luis Ramírez LUENGO 

ILLA – CSIC, Madrid, España

E-mail: joseluis.ramirezluengo@gmail.com

BOOKS

Mary A. Kato, Ana Maria Martins, Jairo Nunes, *Português Brasileiro e Português Europeu: Sintaxe Comparada*, São Paulo: Contexto, 2023, 383 p.



Publicado pela Contexto, editora brasileira que se tornou uma referência incontornável para os especialistas em linguística brasileira e portuguesa, o volume *Português Brasileiro e Português Europeu: Sintaxe Comparada* é a versão em português da obra *The Syntax of Portuguese*, publicada no mesmo ano na prestigiada Cambridge University Press. De acordo com os autores, dois linguistas brasileiros, Mary Aizawa Kato e Jairo Nunes, e uma linguista portuguesa, Ana Maria Martins, este volume não é uma mera tradução da obra em inglês, sendo adaptado para o público lusófono, que mostra interesse pela língua portuguesa: “A tradução foi adaptada tomando como leitores potenciais falantes nativos interessados em questões de linguagem (mas não necessariamente especialistas) [...] deixando de lado a descrição de fatos que são do conhecimento dos falantes nativos, e faz uso bastante parcimonioso (e com as devidas explicações) de

termos técnicos que não são familiares a não especialistas” (12).

A pergunta a que se procura responder ao longo deste volume não é se o português brasileiro (doravante PB) e o português europeu (doravante PE) são a mesma língua ou se são duas línguas distintas, tomando em consideração as



particularidades sintáticas de cada variedade, mas sim se alguns dos mitos em torno da língua portuguesa têm algum fundamento, depois de uma minuciosa análise comparativa da sintaxe dessas duas variedades. Aliás, o dilema *língua única vs. duas línguas* é considerado pelos autores “uma questão muito espinhosa, pois existem inúmeros significados atribuídos à palavra *língua* e as respostas podem diferir drasticamente dependendo do que se assume como definição de *língua*” (355). E, continuaríamos nós, porque a história nos ensinou que os que decidem o que é língua e o que é dialeto são os políticos, e não os linguistas.

Para comparar a sintaxe – ou a gramática – das duas variedades do português, os autores utilizam uma analogia com a classificação empregue na biologia, quando são analisadas espécies que pertencem à mesma classe ou à mesma ordem. Em termos mais mundanos, partido da velha pergunta se os jabutis e as tartarugas são a mesma coisa ou não – e, além do mais, o que significa neste caso concreto ser igual ou ser diferente –, os autores tentam responder comparando a sintaxe do PB e a do PE em seis das suas características essenciais: sistema pronominal (21-56), concordância (57-128), sujeitos nulos (129-194), ordem das palavras (195-264), objetos nulos e possessivos nulos (265-294) e afirmação, negação, interrogação *sim/não* e respostas curtas (295-348).

O primeiro capítulo analisa o sistema pronominal, uma das áreas que apresenta mais divergências nas duas variedades do português, com consequências para a concordância, o licenciamento de sujeito nulo e a ordem das palavras. A reorganização do sistema pronominal do PB (com fenómenos já amplamente estudados, como a integração de *a gente*, a eliminação de *tu* e a expansão do uso de *você*) determinou uma simplificação das distinções de caso (por exemplo, as formas dativas *lhe* e *lhes* tendem a ser substituídas em PB por formas perifrásticas, *pra ele*, *pra eles*), o que não ocorre no PE. Em contrapartida, o PE mantém uma morfologia pronominal bastante rica, que permite a distinção de caso. As diferenças são significativas para os falantes nativos, uma vez que certas ocorrências (como, por exemplo, *A Maria só pensa em si.*) são possíveis só em PE, e não em PB.

Os autores se debruçam sobre a concordância, evidenciando algumas divergências entre as duas variedades. No domínio nominal, o PB apresenta singulares nus com substantivos contáveis (por exemplo, *O João nunca compra livro usado*), a marcação do plural só no determinante (*os livro, as criança*) fenómenos que não ocorrem em PE. No que diz respeito à concordância verbal, nota-se o uso de cinco formas verbais em PE e de quatro em PB, mas deve-se acrescentar o uso prevalente de formas desprovidas de concordância em PB informal, o que na prática cria mais divergências em relação ao PE. Em PB observam-se três padrões de concordância de *a gente* (com verbo na 1ª pessoa PL., na 3ª pessoa do SG. ou na 3ª pessoa do PL.), o que assinala uma hesitação dos falantes brasileiros em relação à integração desta forma pronominal numa categoria morfológica de pessoa e/ou número.

O licenciamento diferente de sujeitos nulos em PE e PB decorre da redução da morfologia verbal em PB, por sua vez, uma consequência da reconfiguração do sistema pronominal, já amplamente analisada no primeiro capítulo. Enquanto o PE tem um

comportamento de língua *pro-drop* prototípica, em PB são mais comuns as frases com pronomes foneticamente realizados. Porém, o licenciamento nulo em PB depende de um parâmetro que os autores designam por “condição de valoração do traço proeminente”, que mostra a importância da saliência das formas verbais: “Uma dada flexão verbal Infl pode licenciar morfologicamente a elipse de um sujeito pronominal definido em português se o traço de Infl mais proeminente na escala *pessoa>número>gênero>caso* for valorado” (190). Essa abordagem explica a razão pela qual é possível ter contraparte nula do pronome *nós* em PB (a flexão verbal da 1ª pessoa do PL. é valorada), ao passo que no caso dos pronomes *você, ele/ela* e *a gente* não pode haver elipse.

A ordem canônica das palavras em frases declarativas simples em PE e PB e SVO. Outras ordens são também possíveis, mostrando o PB mais rigidez do que o PE no que diz respeito às estruturas VS, por exemplo. Em interrogativas e em frases com topicalização, o PE prefere a ordem VS, ao passo que o PB mantém na maioria dos casos a ordem canônica SVO, sendo este parâmetro “um divisor de águas entre as duas variedades” (257). Num grande número de contextos analisados (frases finitas ou não finitas), o PB mostra-se mais rígido quanto à ordem dos constituintes do que o PE. Outro fenómeno de divergência, nomeadamente a colocação de clíticos, salienta que em PE a ênclise é o uso *default*, ao passo que em PB o *default* é a próclise; os clíticos fundidos do PE, como *lho(s), lha(s), no-lo(s), vo-lo(s)*, não ocorrem em PB.

O PB apresenta uma distribuição mais ampla de objetos nulos, ao passo que o PE mostra uma maior possibilidade de licenciamento de possessivos nulos, revela o capítulo dedicado a estes fenómenos linguísticos. No que diz respeito aos objetos oblíquos nulos, o PB apresenta mais casos que permitem este tipo de licenciamento.

Tanto o PE, como o PB permitem o uso de verbos e de advérbios como respostas a perguntas. Ambas as variedades apresentam variação na expressão da polaridade, tendo o PE um maior leque de possibilidades de expressão da afirmação, ao passo que o PB mostra um espectro mais reduzido de expressão da negação. Nota-se uma divergência no caso das interrogativas *tag*, uma vez que o PE prefere a âncora negativa, ao passo que no PB é muito mais comum a retoma do verbo conjugado (por exemplo, *A Maria não fala bem inglês, pois não?* em PE vs. *A Maria não fala bem inglês, fala?* em PB) (325). Destacamos a explicação muito clara dos significados opostos da expressão *pois não* em PB, em função da entoação (344): a entoação ascendente é usada para convidar o interlocutor a intervir na conversação, ao passo que a entoação descendente é uma resposta a uma solicitação de serviço.

O volume *Português Brasileiro e Português Europeu: Sintaxe Comparada* é sem dúvida uma obra necessária, na medida em que se apresenta como uma primeira tentativa de analisar de forma sistemática e coerente a sintaxe do PE e do PE em alguns dos seus aspetos onde se observam divergências. Baseado em análise de corpora, o volume apresenta de forma clara – e por vezes divertida, sobretudo nas analogias com a taxonomia biológica, usada pelos autores ao longo dos capítulos – e sem excesso de metalinguagem fenómenos linguísticos que interessam tanto os especialistas, como os falantes. Destacamos a ampla bibliografia atualizada, que dá conta de estudos mais

recentes sobre a sintaxe de ambas as variedades, sobretudo na ótica da gramática generativa (ou gerativa). A obra pode ser útil aos investigadores, mas também aos professores de português (L1 ou L2) que queiram trazer para a sala de aula informações atualizadas e explicações pertinentes sobre a sintaxe da língua portuguesa, sobretudo no que diz respeito às duas das suas variedades mais estudadas, o português europeu e o português brasileiro.

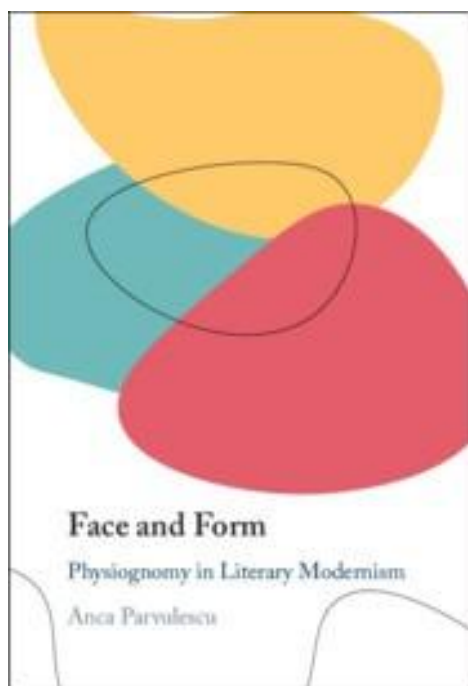
Veronica MANOLE 

Universidade Babeş-Bolyai, Cluj-Napoca, Roménia

E-mail: veronica.manole@ubbcluj.ro

BOOKS

Anca Pârvolescu. *Face and Form. Physiognomy in Literary Modernism*. Cambridge: Cambridge University Press, 2025, 196 p.



Anca Pârvolescu's latest study takes on an ambitious stake: unveiling a "history of the present" through the lenses of what she calls the "modern faciality machine" (17) – the apparatus of subject production described by Deleuze and Guattari in *A Thousand Plateaus*. The current obsession with technologies of the face (illustrated by the ubiquity of cosmetic adjustments, celebrity culture, "Instagram" faces or their equally commodified variants, facial recognition devices, or by the intensity of the debate over mask wearing by responsible citizens during the Covid pandemic, etc.) is invoked as the justification for a foray into the modernist effervescent, albeit ambivalent, reconfiguration of earlier regimes of faciality, itself drawing on the age-old identification of faces with texts to be deciphered. The book's anchorage in our collective enchantment with human faces sounds familiar but opportune alarm bells for analysts of modernity's appetite for

hyperconsumerist homogenization, mass surveillance and disinformation. Its substantive success, however, is ensured by the historicization of literary modernism's efforts to both abolish mimetic modes and counteract the drive towards anonymization through minimalist tactics of physical representation that culminate in avant-gardist efforts of self-effacement, even if they remain unable to extricate themselves from the discursive assemblages of race, ethnicity, gender, or class.

Pârvulescu situates her research within the paradigm of New Modernist Studies, replicating its extension of modernism's thematic and formal purvey, as well as its temporal and geographical reach. She traces the coagulation of the literary trope of the face, turned into a set of subjectivation strategies, and its traveling across cultures, continents (European, transatlantic, Japanese) and historical periods overlapping the span of modernity and the imperial project. The ideological lynchpin in the metamorphosis of this rich and ambiguous figure is provided by its fluctuating relationship with the lingering tradition of physiognomy. Physiognomy had found its most famous expression in the work of the eighteenth-century Swiss scholar Johann Caspar Lavater, and had made its way into the popular imaginary during the Victorian era. As Pârvulescu points out, Lavater purported to have founded no less than a "science of sciences" [that] claimed to offer a visual semiotics of the natural world," (2) decoding a correspondence between appearance and the inner significance of things. The modernist regime of faciality is approached via the artists' ambivalent positioning towards debunked physiognomic theories, simultaneously reinforcing and contesting their premises. Intellectual and aesthetic attitudes and practices of "facialization" are analyzed in the light of two very different essays produced during the age of canonical modernism – Georg Simmel's "The Aesthetic Significance of the Face" (1901) and Mina Loy's "Auto-Facial-Construction" (1919). Simmel's influential sociological appraisal tackled the human face as a privileged terrain of urban interaction and recognition via aesthetic categories such as unity or symmetry, with their respective deviations. In contrast, Loy's text elevates it as a site for creative and defiant self-fashioning through playful intermedial manipulation of social norms and aesthetic standards. As Pârvulescu concludes, Loy's ironic problematization of the sources of identity mobilized in the gesture of "auto-facial-construction" points towards both the persistence and the limitations of agency in modernity.

The book's chapters provide a series of resourceful, erudite and carefully woven close readings, mapping the itinerary of facial tropes across global modes of identity interpellation. The first section brings to the fore the shift in the "politics of the face" (23) staged by the writing of Thomas Mann between the publication of *Death in Venice* (1912) and that of the 1939 essay "Brother Hitler," which registers an anxious awareness of the dangers of physiognomic categorization. The discussion of Mann's novella, so central to the traveling modernist canon, exposes the mechanisms of domesticating exotic or unruly faces by casting them as ethnically or sexually aberrant, or by ekphrastic absorption within Western norms of beauty, as illustrated by Aschenbach's aestheticizing gaze. However, modernism's combative energies are salvaged by the analysis of Aschenbach's eventual "dandification" (39), which destabilizes predominant subject-object hierarchies through cosmetic practices of partial "auto-facial-construction." The next section, dedicated to the affordances of facial semiotics for Proust's investigation of memory, continues to zoom in on the modernist project of agential recuperation in the figure of Odette, who escapes Swann's physiognomically-inspired attempts at objectification by undergoing a process of material "refacialization" (62) that rallies the poses of others (including technological others, in the guise of photographic snapshots) in the service of aesthetic and social self-fashioning.

Perhaps the most interesting part, chapter 3 reinterprets Virginia Woolf's *Orlando* as an ironic deconstruction of the reciprocal summoning of texts and faces, investing its sequence of "scenes in the history of the face" (72) with far-reaching implications for "the temporal arc of the modern" (82). The successive encounters between Orlando and Elizabeth I, the Archduchess Harriet Griselda of Finster-Aarhon and Scand-op-Boom, Nicholas Greene, the Roma travelers she joins after her transition to femininity, random faces in London crowds, as we draw nearer to the twentieth-century, and her husband Marmaduke Bonthrop Shelmerdine, are subsumed into the novel's (mostly subversive) interventions in imperial, gender, racial and social structures, reinforced by the acts of auto-facial-construction identifiable in the reproduction and manipulation of the famous Vita Sackville-West photographs. The chapter's coda explores *Orlando*'s continuous relevance for the present by focusing on the dialogue between Woolf's satirical critique of the imperial tradition of pictorial and literary portraiture and the work of photographer Paul Mpagi Sepuya, occasioned by an exhibition dedicated to the novel by the Aperture Foundation in 2019 (77-82). According to Pârvulescu, Sepuya's engagement with the latter's opening scene, which has young Orlando playing with the skull of a Moor brought home as a trophy by one of his ancestors complicates Woolf's thematization of racial violence by supplementing the absence of the other's face through snapshots of Western representations of Arab figures, themselves multiply mediated and submitted to a critical viewer's gaze.

The book's next section pursues the intermedial entanglements staged by the exchange of gazes among subjects and objects of representation in Gertrude Stein's attempt to oppose the modern reification of faces and individuals through polemical reversals of portraiture conventions and radical tactics of self-effacement. Her expert curation of her position as a modernist author, including the mythologization of her Picasso portrait, the staging of her Man Ray photograph or her genre-crossing "(auto)biographical" narratives are shown to aim at achieving "mastery through the control of face" (96). Nevertheless, Pârvulescu's close scrutiny of modernism's imbrications with capitalism and historical modernity unveils the inescapability of the face, as Stein's experiments with defacement and even effacement are inevitably haunted by her authorial presence. Still, Stein's "exercises in facialization" (92) fomented creative responses to the convention of facial legibility, as witnessed by their impact on Nella Larsen's *Passing* and its performative auto-facial-constructions. Finally, chapter 5 deals with the circulation of culturally translated physiognomic tropes into transcontinental and postwar territories, turning to an interpretation of Kōbō Abe's novel, *The Face of Another* and its interrogation of the global technologization and commodification of the face.

What makes it difficult to summarize the value of Anca Pârvulescu's study is its generous address to the interests of multiple audiences. As a contribution to the growing archive of New Modernist Studies, it provides a fresh and complex take on a host of topics: intermediality, the intersections of race, gender and colonialism in the construction of the subject, the enduring legacy of modernism in our troubled present, its ability to furnish us with instruments and strategies for resistance, as well as to direct our awareness to the persistence of ossified conventions of representation.

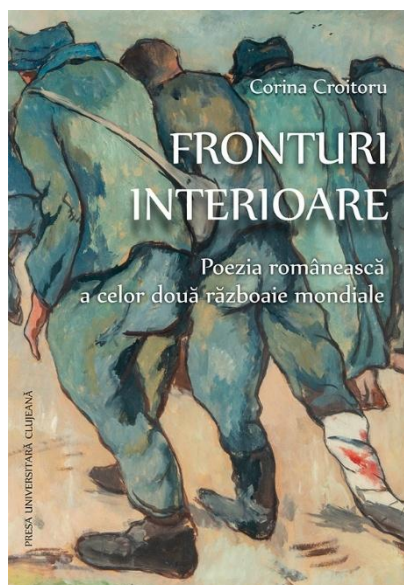
Although Pârvulescu does not neglect the ethically paramount modernist attention to the intractability of other people's lives, best illustrated by Woolf's meditations in "Mr. Bennett and Mrs. Brown" and "An Unwritten Novel," she intentionally focuses on "modernism's temptation to think otherwise" (126) – that is, to succumb to the lure of knowability. However, by constantly reminding us that the face is always "the face of an other, a stranger" (119), the book continues to perform modernism's ethical work.

Petronia POPA-PETRAR

*Babes-Bolyai University, Cluj-Napoca, Romania
E-mail: petronia.petrar@ubbcluj.ro*

BOOKS

Corina Croitoru, *Fronturi interioare. Poezia românească a celor două războaie mondiale*, Cluj-Napoca: Presa Universitară Clujeană, 2024, 273 p.



Corina Croitoru's *Fronturi interioare* is a most welcome addition to Romanian literary history and criticism, as evidenced by its receipt of the 2024-2025 Bac-Fest Award for Poetry Criticism. Just like the author's first book, *Politica ironiei în poezia românească sub comunism* (*The Politics of Irony in Romanian Poetry under Communism*), this volume examines the complex interplay between literary expression and historical reality. *Fronturi interioare* looks into an area of Romanian literature that has benefited from little scholarly attention, namely the poetry of the two World Wars. The volume, as clearly set forth in the introduction, revisits the poetry inspired by the war experience with a view to offering a critical perspective on how poetic expression, conceived as a *lieu de memoire* (via Pierre Nora), grapples with historical events while teasing the boundaries of its own aesthetic potential (21).

The first section of the volume is dedicated to World War I, with its ironic (via Paul Fussell) or cynical (Peter Sloterdijk) view of the world. After engaging with key scholarship on the distinctiveness of this war and its representations, the author provides a thoughtful addition by pointing out that "[t]he First World War is not about death [...], but about the *act of dying*, about the exploration of the various modern ways in which human life comes to an end"¹ (30-31). The new forms of dying, characteristic of this war (with its introduction of a new weapon, the shell, that disintegrates the body; with its impact on the soldiers' minds, disintegrating them on the inside) extend, Croitoru argues, beyond

¹ In the original: „Primul Război Mondial nu este unul al morții [...] ci al *muririi*, al explorării multiplelor forme moderne ale dispariției umane” (31-32).



the individual, triggering “the spiritual disintegration of society as a whole”, which is, in her view, “the inevitable consequence of the bodily and psychological disintegration of the human being in modern warfare”² (32-33). A similarly interesting point, albeit on a different issue, is made later in the chapter. Croitoru reflects on the neglect of war poetry in Romanian literary histories, suggesting that its cause is not *amnesia*, but rather *aphasia*, understood as the inability to articulate something that is actually remembered. The author claims that “the fading of war poetry into oblivion is, in fact, a form of *silence*”³ (142).

This first chapter acknowledges the plurality of voices that make the eclecticism of the Romanian poetry of WWI (42-43). It highlights the particularly complicated ways in which Romanians experienced the war, shaped by a historical context that divided them among three different armies and made them fight on opposite sides. It further analyzes the extent to which political and ideological divides influenced the Romanian poetic discourse about the war.

The chapter examines verses written by conscripts (particularly the conscription songs of Transylvanian soldiers, which reflect their conflicted feelings about fighting fellow Romanians on behalf of a foreign emperor), alongside poetry by career soldiers, often influenced by the tradition of earlier patriotic poetry that evokes the memory of past wars and contributes to the consolidation of past notions of heroism. It also includes works by established poets who were enlisted and sent to the frontlines (such as Vasile Voiculescu, Ion Pillat, Demostene Botez, George Topîrceanu, Camil Petrescu, Perpessicius and others), as well as poems by civilian poets who experienced the war either from home (Nicolae Iorga, Alexandru Vlahuță, George Coșbuc, Alexandru Obedenaru, George Ranetti) or from self-exile (Tristan Tzara). The lucid critical analysis of each author's literary contribution is often accompanied by colorful or bizarre anecdotes from their wartime experiences. We thus learn, for instance, that career officer Nicolae Vulovici was killed in exactly the manner he had described in one of his poems, by a bullet to the forehead, or that the enlisted writer Camil Petrescu went through the war wearing shoes because the military boots caused him discomfort. For all its diversity, Romanian poetry of the Great War seems to largely subscribe to the irony and cynicism theses discussed in the introduction to this chapter.

The poetic representation of the experience of the Second World War, as the author observes, is considerably different from that of The Great War. There is a qualitative shift, due primarily to the fact that “the most significant poetic achievements shaped by the atmosphere of the Second World War were authored by the civilian poets of the *war generation*”⁴ (212). Whereas World War I poetry retains the lyrical and meditative qualities of the 19th century poetic tradition, the Second World War brings forth an anti-lyrical, transitive form of representation (176).

² In the original: „dezagregarea spirituală a întregii societăți nefiind decât consecința inevitabilă a dezintegrării corporale și psihologice a ființei umane în războiul modern” (32-33).

³ In the original: „uitarea care atinge poezia de război nu este, de fapt, decât o formă de *tăcere*” (143).

⁴ In the original: „cele mai mari reușite poetice marcate de atmosfera celei de-a doua conflagrații sunt semnate de poeții civili ai *generației războiului*” (212).

Continuing the approach adopted in the chapter on the First World War, the author distinguishes between the poetry written by established poets who were enlisted – either as a form of disciplinary punishment for ideological transgressions (Radu Gyr, Sergiu Filerot, Al. Cerna-Rădulescu), or because they had completed military service or were career soldiers (Mihnea Gheorghiu, Ion Șugariu, Călin Gruia, Horia Agarici, Ion Petrache, Mircea Popovici, George Dan) – and the civilian poets of the *war generation*, who experienced the war far from the front lines (Geo Bogza, Dimitrie Stelaru, Ion Caraion, Constant Tonegaru, Mihail Crama, Ben Corlaci, Victor Todynopol). A special section is dedicated to the group of Romanian avant-garde poets of Jewish origin who self-exiled in France and joined the efforts of the French Resistance (Tristan Tzara, B Fundoianu, Ilarie Voronca, Mihail Cosma). The discussion on the poetry of World War II examines how the trauma of war shapes aesthetics, memory, and forms of engagement. It emphasizes the tension between ethical and aesthetic concerns, between forms of silence and forms of engagement.

The experience of the war, the author shows, is rewritten after the establishment of the communist regime with a view to supporting governmental propaganda. Thus, poetic engagement turns toward glorifying the Soviet liberator soldier and the peace ensured by communist rule. Two authors are the focus of the case study for this category: Mihai Beniuc and Eugen Jebeleanu.

The last chapter of the book is dedicated to women poets who contributed to the inscription of the war experience into literary memory. The work of poets such as Carmen Sylva, Matilda Cugler-Poni, Maria Cunțan, Elena Văcărescu, Ana de Noailles, Livia Rebreanu-Hulea (World War I), Anișoara Odeanu, Magda Isanos, Maria Banuș, Veronica Porumbacu, or Nina Cassian (World War II) is used by the author to dismantle the reductionist patriarchal prejudices that have consigned women writers to neglect for far too long. Although the plea for peace – with its propagandistic engagement after World War II – seems to guide the writing of these women authors, Croitoru notes that their work is more complex and varied than that. Elena Văcărescu, for instance, argues for entering the war to achieve national unification, Livia Rebreanu-Hulea reworks the epistolary form to bridge the emotional distance between home and the frontline, and Anișoara Odeanu explores love, both romantic and fraternal, in times of war. The nuanced readings in this chapter effectively invite further scholarly engagement with women's literary productions, extending beyond war poetry.

Grounded in conceptually rich reflections and a well-established scholarly bibliography, the volume stands out through the author's sharp critical voice and subtle, well-articulated irony. *Fronturi Interioare* is both insightful and informative and will offer a most interesting and delightful reading experience to a wide readership.

Amelia PRECUP 

Babes-Bolyai University, Cluj-Napoca, Romania
E-mail: amelia.precup@ubbcluj.ro