Occitan Studies Day

11 May 2024

University of Oxford

TARASCON — Procession de la Tarasque

Photo Modern, Tarascon - Rep. int.

▲ La Tarasque, the terrible monster of Tarascon

BOOK OF ABSTRACTS

« Comment peut-on être occitanophone ? » Les locuteurs actuels de l'occitan et leurs parcours.

Grégoire Andreo Raynaud

(Laboratoire Parole et langage UMR CNRS 7903, Aix-Marseille Université)

Cette communication vise à explorer les caractéristiques biographiques et sociales d'une partie des occitanophones actuels à travers la notion de « carrière » empruntée à la sociologie interactionniste et définie comme une série « de statuts et d'emplois (…) , de suites typiques de positions, de réalisations, de responsabilités et même d'aventures (…), de changements dans la perspective selon laquelle la personne perçoit son existence comme une totalité et interprète la signification de ses diverses caractéristiques et actions, ainsi que tout ce qui lui arrive » (Hughes, 1937, 408). Je m'intéresserai particulièrement aux locuteurs les plus actifs, ceux et celles que l'on qualifie ordinairement d'occitanistes (Jeanjean, 1992) dans la mesure où ils associent à leurs compétences langagières une activité plus ou moins professionnalisée de transmission et de promotion de la langue occitane. Il s'agit spécifiquement d'analyser l'occitanophonie comme le résultat d'un processus de socialisation mettant en jeu la notion d'identité et comme un ensemble de choix conscients et de contingences biographiques qui entrent en relation avec l'évolution de la situation sociale de l'occitan au cours des 50 dernières années. L'analyse de ces parcours permet de mieux comprendre la façon dont se transmet aujourd'hui la pratique de l'occitan alors que la transmission familiale intergénérationnelle s'était presque complètement interrompue à la fin du 20ème siècle.

Dans une démarche qualitative, je ferai référence aux entretiens menés durant une enquête de terrain effectué en Région Occitanie lors de ma thèse de doctorat portant sur les *Calandretas*, écoles associatives et immersives occitanes (Andreo Raynaud, 2021, 2023). Le corpus comprend des entretiens avec des informateurs et informatrices de différentes générations qui sont des occitanophones engagés de manière associative ou professionnelle dans la transmission de la langue occitane (en tant qu'enseignant.es, artistes ou encore professionnel.les de la culture). Les entretiens ont permis de recueillir des informations sur leur rapport à la culture occitane et la place qu'elle occupe dans leurs parcours de vie et dans leur environnement social, sur leur découverte de la culture occitane, leur apprentissage de la langue, leur entrée dans l'action militante, l'usage de la langue et le rôle assigné à la langue dans la vie personnelle ou professionnelle. Mélange de discours autobiographiques, et de discours épilinguistiques, ce corpus permet d'appréhender certains mécanismes de la transmission et de la langue occitane aujourd'hui.

Références :

Andreo Raynaud, G. (2021). « La transmission de la langue occitane à travers le processus de patrimonialisation », dans *Trans-mission. Création et hybridation dans le domaine d'oc*, éd. par Fabio BARBERINI et Camilla TALFANI, Turnhout, 2021 (Publications de l'Association Internationale d'Études Occitanes, 14), p. 229-241.

Andreo Raynaud, G. (2023). « Apprendre l'occitan, promouvoir le patrimoine ». Paris : L'Harmattan.

Hughes E. C. (1937) « Institutional office and the person », *American Journal of Sociology*, vol. XLIII, p. 408-410.

Jeanjean, H. (1992). *De l'utopie au pragmatisme? Le mouvement occitan, 1976-1990* (1. éd). Llibres del Trabucaire.

Studying the long term development of Lengadocian syntax (12th-21st century): negation, periphrasis, language contact

Xavier Bach

(CLLE CNRS & Université Toulouse Jean Jaurès)

A common problem in Romance diachronic linguistics is the tendency of some research to only look at the initial stage and the end point of linguistic changes. At the same time, a large number of researchers specialise in a given period of each language (the medieval period, for the most part), and do not examine the changes beyond that specific period. This is particularly problematic for Occitan, where a large number of studies concern the medieval period (understood as 11th-early 14th century), and another set of studies analyse the contemporary language since the first dialectal descriptions of the end of the 19th century. Very little is done about what lies in between these two extremes, although, at least for the late medieval period, there is no dearth of documentation. The early modern period proves more challenging in that respect.

My current project aims at examining (syntactic) change in Lengadocian Occitan in its long term development. For this, I build a sample corpus of texts from the 12th to the 20th century, all from the Lengadocian area, aiming at having in the corpus at least a long text for each half century. For some periods, more texts could be found. But for the early modern period, there are less texts, and from less varied areas within the dialect boundaries. This sampling method at least enables the researcher to anchor their analysis of syntactic changes in actual diachronic data.

For this presentation, I centre on two main syntactic changes: negation, and periphrasis. Sentential negation broadly shows an evolution similar to French, in that medieval Occitan only presented a preverbal negation *non* while modern Lengadocian varieties systematically present a postverbal negation *pas*. The details of the evolution are rather different though (Bach 2023). While until the 14th century *non* is sometimes reinforced by a postverbal particle *ges*, in the 15th century, with increased contact with French, the two particles *pas* and *ponch* quickly take over (Paoli & Bach 2020). The loss of preverbal negation occurs earlier than in French, and following different lines of sound change. Finally, regarding negative concord, the situation of Occitan is strikingly different from French, going 'beyond' by reinnovating negative concord with postverbal negation.

Regarding periphrasis, I analyse two major periphrases, one temporal and one aspectual. The periphrasis *anar* 'go' + infinitive is famous for expressing the past in Catalan and in medieval Occitan. There is evidence that the periphrasis originated in Occitan (Wheeler 2018). But in the contemporary language, as is the case in French and most languages where such a periphrasis exists, it expresses futurity. I examine the evidence for a reversal in meaning throughout the early modern period, questioning in particular if such reversal can be due to a single shared notion of 'posteriority' (Paoli & Wolfe 2022). Occitan also sees the development, during the early modern period, of an aspectual auxiliary *tornar* + infinitive which expresses a single reiteration of an event, from a verb that in the medieval period only expressed movement. This change cannot have been influenced by French, because this language does not grammaticalize this verb as an aspectual auxiliary. I trace this development during the early modern period.

Finally, examining linguistic changes in the syntax of Occitan leads to asking the question of a possible influence of French, at two crucial periods, the 16th (when most administration switches to French) and 19th century (with the rise in schooling and more advanced bilingualism, and the progressive disappearance of monolingual Occitan speakers). Some changes, for example regarding negation, can be shown to originate because of contact with French. But such influence is far from systematic. The development of an iterative aspectual periphrasis for example has nothing to do with contact, as such periphrasis never existed in French.

References

Paoli, Sandra & Bach, Xavier. (2020). Postverbal negators in Gallo-Romance: The view from Old Occitan. In Sam Wolfe & Martin Maiden (eds.), *Variation and change in Gallo-Romance grammar*, Oxford: OUP, 117-137.

Bach, Xavier. (2023). Negation in contact: French and Occitan. *Transactions of the Philological Society* 121(4).

Paoli, Sandra & Wolfe, Sam. (2022). The GO-future and GO-past periphrases in Gallo-Romance: a comparative investigation. In Adam Ledgeway, John Charles Smith & Nigel Vincent (eds.), *Periphrasis and inflexion in diachrony: a view from Romance*. Oxford: Oxford University Press, 123-144.

Wheeler, Max. (2018). Stages in the grammaticalization of the periphrastic preterite va + infinitive in Catalan and Occitan. Presentation at the *46th Romance Linguistics Seminar*, Trinity Hall, Cambridge, 5-6 January 2018.

Framework to build and annotate a corpus of early modern Occitan

Gilles Couffignal[1], Laure-Anne Caraty[1], Lucence Ing[2]
[1](Sorbonne Université), [2](École nationale des Chartes)

Occitan has a rich textual tradition going back to the Middle Ages. This documentation allows us to observe the evolution of linguistic uses over the long-term. Scholars can rely on state-of-the-art corpora for the medieval period until 1550 (COM, TMAO, DocLing) as well as for modern varieties from 19th C. (Batelòc). Nevertheless, structured dataset is lacking between these two periods.

In this conference, we will present a framework to build an Early Modern Occitan corpus, from automatic data acquisition to lemma and POS tagging. We will present the tools developed (scripts and AI models) the choices made for text annotation and the first annotated dataset. We have developed our own reference lists to enable, in a second stage, comparison between medieval, early modern and modern sources.

Framework overview

OCR AI model for printed Early Modern Occitan.
The first OCR models trained by deep learning to recognize Old Occitan were developed in 2016 (Camps & Couffignal 2020). The performance of new tools and models now makes it possible to fine-tune generalist models to obtain exploitable results, particularly on modern print (Couffignal in press). We recently fine-tuned the French ocr4all library model to acquire literary corpora from the 16th to 18th centuries.

Tokenization. The first step in text enrichment is to split words and punctuation marks (tokenization). This step is crucial, as it determines the ability to compare similar forms over time. We will present the choices made, in the perspective of a large historical coprus of Occitan.

Lemma and POS tagging: from manual annotation to training an AI model
Lemmatization and POS tagging are dealt at the same time. For lemmas, we lack a consensual and coherent scientific reference. To facilitate machine learning, we decided to create our own guidelines. The aim is to create a list of lemmas that can be used in long diachrony. The POS tags are those of the CATTEX manual (Guillot et al.).

Establishing an initial corpus annotated by human enables us to train an AI. We are currently testing Pie (Manjavacas et al. 2019), a lemmatizer-tagger in Python developed to handle language states with high variation. The first training, with a small dataset (25k tokens) reached a 88 % accuracy rate. These first results are limited but encouraging given the limited data used. Since it is faster to correct bad data than to tag every token from scratch, the accuracy of the model will increase incrementally. We are aiming for a dataset of 100k tokens as a first step for a usable lemmatizer.

References

BaTelÒc Bras M. et al. Basa textuala per la lenga d'Òc, http://redac.univ-tlse2.fr/bateloc/.

Camps, J. B., & Couffignal, G. G. (2020). La production de corpus d'occitan médiéval et prémoderne. In *Actes du XIIe Congrès de l'Association internationale d'études occitanes*. Toulouse SFAIEO.

Gardar la tera emè son nom. La toponomastica en auta Val d'Estura

Francesco Dematteis

(Stura valley federation of mountainous municipalities – ATPM, Toponymous Atlas of mountainous Piedmont)

Liem l'Occitan sobretot al Sud de la França, mas quarque viatge se desoblia que decò en Italia i a un pichòt canton d'Occitania. En Piemont, dins las provincias (eng. "counties") de Coni e Turin, i a una desena de valadas entè se parla la Lenga d'Òc en la varietat Vivaroarpenc. L'Occitan Alpin es, per la lei italiana, una lenga de minorança ed es gardaa da la lei 482 del 1999 "*Norme in materia di tutela delle minoranze linguistiche-storiche*" (oc. "Reglas per gardar las minoranças languisticas-estoricas"); en fòrça d'aquesta lei es possible tirar sus divers trabalhs, mossò d'lhi quals naisson e creisson a l'aviron de l'Universitat de Turin.

L'Occitan de las Valadas es una lenga que rischa de murir, mas, bonur, se troba encara una bona vitalitat e son ren pauchi li joves que lo parlon o aquei que an recomençat a lo parlar apres que las familhas l'avion perdut. Beleu la tendença negatia a quitaa o a ralentaa e sembla estre eura una esperança de remontada. Aquesta es tutun una "novitat" perqué findi a lhi ans '80 del '900 semblava que tot lo monde occitan de las Valadas – lenga, tradicions, costums, trabalh sasonalhs e de la tera – devesse murir; en tot aquò desobliem ren la desmontada demografica que a colpit fortement las Alps piemontesas apres la II Guera Mondiala.

Ad aquesta trista situacion s'es cercat de butar un argin emè de recherchas, una d'aquestas avia la mira de salvar la toponimia locala – lhi noms de lhi pòst, de las teras – que, apres l'abandon de la montanha – donca de lhi champs e de las bestias –, avia lo destin de desparir. Dins lo 1983 lo Professor Arturo Genre de l'Universitat de Turin a pensat ad un grand proget, lo "Atlante Toponomastico del Piemonte Montano" – ATPM ("Atlant toponomastic del Piemont montan") entè lhi estudios o lo monde dal pòst chalavon culhir tuchi lhi topònim, partits comuna per comuna. L'ATPM es un trabalh pensat per totas las comunas montanas del Piemont, indipendentement da la lenga parlaa e donca es compresa tota l'area occitanòfona. Lo proget es encara viu encuei e sus esquasi 200 comunas i son estat culhidas (e publicadas) 62.

Dins lo 2018 ai començat lo proget que m'a menat a la laurea magistrala (eng. "postgraduate research project") entè ai trabalhat sus la toponomastica de l'Argentiera, la mai auta comuna de la Val d'Estura; apres ai trabalhat sus autras comunas de la mesma valada. Ai menat la recercha travers de entrevistas emè lo monde del pòst que, en fòrça del contat emè la tera, pòl referir lhi nom de pras, champs, arbo o ròchas. D'aquest trabalh ai polgut veire coma la toponomastica se lia a la lenga del monde e coma lo monde dona un nom a la tera que viu tuchi jorns. La toponomastica pòl studiar la lenga e gardar, almeno dins la memòria, las velhas tradicions.

The Curious Case of an Occitan Sonnet: BdT 96,7

Barbara Francioni

(OVI – CNR, Florence)

The lyrical text classified by Pillet-Carstens under the code *96,7* is found today in only one chansonnier, *f,* known as the Giraud Chansonnier, at f. 19r. Structurally a sonnet (rhyme scheme: *abba abba cde cde*), the poem appears to be spuriously attributed to Blacasset, a troubadour active between 1220 and 1240, following another sonnet, attributed to Jacme Mote d'Arles by the compiler of the manuscript. The sonnet, a metrical form which was not practiced by the troubadours and will not make its first "official" appearance before the so-called *Scuola Siciliana* – with Giacomo da Lentini above all –, will enter the list of French literary genres only much later, in the 16th century (cf. Gendre 1996). It does not belong either to the canonical genre repertoire of Occitan lyrics, even if the *BEdT* gives three examples of poems classified as "sonnets", each written by Tuscan authors from the second half of the 13th century (Paul Lanfranc de Pistoja and Dante da Maiano).

The study aims at presenting a first attempt of analysis of this curious text, that has received scarce attention in Occitan studies until now – either because of its apocryphal nature, or of its metrical structure, or even because of the difficulty in deciphering the chronotopical references it contains; Meyer 1871 was the only one who published an interpretative edition of the text, without going into detail on the problems related to authorship or dating; after his edition, the text was totally forgotten by researchers. Nonetheless, the poem offers itself as a particularly interesting sample of the last influences of the troubadours' "outreach" in Italy, in a chansonnier, such as *f,* assembled in Arles during the 14th century (cf. Zufferey 1987). Furthermore, the text under analysis belongs to a typical Italian genre written in the *langue d'oc*, a circumstance that makes it even more fascinating. The highly hybrid language of the text itself, not to mention the use of the term *rassa* – so relevant in the lyrical production of one of the most important troubadours of the golden age of French *Midi*, Bertran de Born –, deserves particular attention considering both its strange poetical form and the geographical references it includes.

Is the poem to be considered as a local text or rather as the sophisticated attempt of an Italian author who enjoyed composing a sonnet in the foreign language traditionally linked to lyrical production? And why was it mistakenly attributed to a minor troubadour as Blacasset? The proposed study will try to answer these questions by offering a new edition of the sonnet *BdT 96,7*, including a linguistic, historical and codicological analysis of this peculiar Occitan text.

References

Antonelli, R. (1989). *L'invenzione del sonetto*, "Miscellanea Roncaglia": 35-75.

Canettieri, P. (2011). *Appunti per la classificazione dei generi trobadorici*, "Cognitive Philology", 4.

Gendre, A. (1996). *Évolution du sonnet français*, Paris, Presses Universitaires de France.

Magro-A. Soldani, F. (2017). *Il sonetto italiano. Dalle origini a oggi*, Roma, Carocci.

Meyer, P. (1871). *Les derniers troubadours de la Provence […]*, Paris, Librairie A. Franck.

Schulze, J. (1989). *Sizilianische Kontrafakturen*, Tübingen, Niemeyer.

A missing piece in the Medieval Romance puzzle:
Old Catalan, Old Occitan, and a pesky particle

Afra Pujol i Campeny

(British Academy/University of Oxford)

Aim and findings: In this talk, I explore the syntactic distribution and function of the particle *si/sí* in Old Catalan and Old Occitan, with the aim of better understanding its role in these two varieties and uncovering what its distribution can tell us about the evolution of their syntax. It will be shown that in very early Old Catalan texts (10th-11th c) the distribution of *si* is connected to the presence of V2 word order, that is, the requirement for the verb to appear at least in the second position of the clause, preceded by any constituent (and crucially, not necessarily the subject). In later texts (12th-15th c) its distribution is to be linked either to the expression of positive polarity (to express that a statement is indeed the case) or to register and style. Old Occitan *si* has only been described as a formal device to satisfy V2, when no other suitable candidate can fill in the preverbal position. However, a closer look at the data shows that it operates as a register and stylistic device. This sets Catalan and Occitan aside from contemporary Langues d'Oil and Medieval Italian dialects, and calls for the reassessment of the Medieval Gallo-Romance dialectal continuum at syntactic level, offering a further argument for an Occitano-Catalan group based on syntactic criteria during the period under study (11th-15th c).

Background: The particle *si* has been analysed as a last resort means to satisfy the V2 requirement in Gallo- and Italo-Romance (Marchello-Nizia 1985; Reenen & Schløsler 2000; Wolfe 2018 for Old French, Poletto 2005 and Ledgeway 2008 for Old Italian varieties, Donaldson 2016 for Old Occitan), and it has been connected to the expression of positive polarity in Medieval Ibero-Romance (Batllori & Hernanz 2013 and Rodríguez Molina 2014 for Old Spanish, Pujol i Campeny 2019; 2020 for Old Catalan). To this date, no comparative account of its evolution and uses across the Romance varieties exists.

The data: For Old Catalan: Borrell's diploma (10th century) and texts contained in CICA (Torruella et al. 2009); for Old Occitan, Occitan Medieval Chartes (11th-13th c), *Cançó de Santa Fe* (11th c), *Roman de Flamenca* (13th century), *Vidas* (13th c), *Douceline* (13th c).

***Si* in Old Catalan:** the use of *si* appears parallel to that of *si* in contemporary Old French (linked to the satisfaction of V2 word order) in early texts written in Latin and very early legal documents written in Catalan (10th/11th c).

This use then disappears from the written record, but resurfaces in learned texts throughout the 13th, 14th and 15th centuries (*Crònica de Bernat Desclot*, *Crònica de Ramon Muntaner*, 14th century Balearic legal texts and learned 15th c texts including *Tirant lo Blanch*), all under Italian humanist influence.

***Si* in Old Occitan:** Old Occitan *si* has traditionally been described as behaving parallelly to Old French's *si*, connected to the satisfaction of V2. This conclusion was reached on the basis of one text: the *Vidas*. However, no cases of this use of *si* occur in any of the other texts included in this study, indicating that in Old Occitan *si*'s use was connected to register and style, once more, influenced by Medieval Italo-Romance (the *Vidas* have been described as displaying high levels of Italian influence, Boutière (1964)).

References

Batllori, Montserrat & Hernanz, Maria Lluïsa. (2013). Emphatic polarity particles in Spanish and Catalan. *Lingua* 128. 9–30. (doi:10.1016/j.lingua.2012.11.010)

Boutière, Jean. (1964). Biographies des troubadours: Textes provençaux des XIIIe et XIVe siècles. Paris: A.-G. Nizet.

Donaldson, Bryan. (2016). Preverbal subjects, information structure, and object clitic position in Old Occitan. *Journal of Linguistics* 52(1). 37–69. (doi:10.1017/S0022226714000619)

Ledgeway, Adam. (2008). Satisfying V2 in early Romance: Merge vs. Move. *Journal of Linguistics*. Cambridge University Press 44(2). 437–470. (doi:10.1017/S0022226708005173)

Marchello-Nizia, Christiane. (1985). *Dire le vrai: L'adverbe "si" en français médiévale. Essai de linguistique historique* (Publications Romanes et Françaises 168). Geneva: Droz.

Poletto, Cecilia. (2005). Sì and e as Expletives in Old Italian. In Batllori, Montserrat & Hernanz, Maria Lluïsa & Picallo, Carme & Roca, Francesc (eds.), *Grammaticalization and Parametric Variation*, 206–235. Oxford: Oxford University Press.

Pujol i Campeny, Afra. (2019). 'E sí la hoïren tots': sí and emphatic positive polarity in Old Catalan. *Isogloss. Open Journal of Romance Linguistics* 5. 1–32. (doi:10.5565/rev/isogloss.73)

Pujol i Campeny, Afra. (2020). Sí and hoc in the history of Catalan: From the 13th to the 17th century. *Revue Romane* (Online-First Articles). (doi:https://doi.org/10.1075/rro.19019.puj) (https://benjamins.com/catalog/rro.19019.puj)

Reenen, Pieter van & Schløsler, Lene. (2000). The pragmatic functions of the Old French particles ainz, apres, donc, lors, or, puis and si. In Herring, Susan C. (ed.), *Textual Parameters in Older Languages*. Amsterdam: John Benjamins Publishing Company.

Rodríguez Molina, Javier. (2014). La gramática oculta de la polaridad positiva en español antiguo. *Rilce: Revista de Filología Hispánica* 30(3). 861–915.

Torruella, Joan & Pérez Saldanya, Manuel & Martines, Josep. (2009). Corpus Informatitzat del Català Antic. *Corpus Informatitzat del Català Antic*. (http://cica.cat/) (Accessed October 11, 2020.)

Wolfe, Sam. (2018). Probing the Syntax of a Problematic Particle: Old French 'si' Revisited. *Transactions of the Philological Society* 116(3). 332–362. (doi:10.1111/1467-968X.12123)

Oralité au 21ème siècle dans le cas des langues minorisées : analyse de l'occitan et du poitevin-saintongeais parlé

Marianne Vergez-Couret

(Université de Poitiers)

Le projet développé ci-dessous trace les grandes lignes en vue du dépôt d'un projet European Research Grant à l'université de Poitiers. Le projet Oralité21 propose une exploration du concept d'« oralité » dans le cas particulier de deux langues minorisées, l'occitan et le poitevin-saintongeais. Nous visons une analyse sémantique et pragmatique de données langagières orales de ces deux langues minorisées. Dans ce contexte particulier, il nous semble nécessaire de redéfinir le concept d'« oralité » et de contribuer à l'élaboration de nouveaux modèles conceptuels qui prennent en compte deux sources d'influence :

La moindre influence de l'écrit avec une riche tradition orale nous offre la possibilité de considérer une variété de facteurs pertinents comme le canal de transmission, les sources d'apprentissage et la pratique des locuteurs. L'occitan et le poitevin-saintongeais sont pratiquées dans un contexte sociétal oscillant entre « oralité mixte », où l'influence d'un système d'écriture n'est que partielle pour des raisons sociologiques et éducationnelles et d'« oralité seconde » selon les contextes et les locuteurs (Zumtor, 1983).

L'influence d'une langue majoritaire est aussi un facteur à étudier dans le cas de ces deux langues en contact avec le français. Même si tous les locuteurs contemporains parlent couramment le français depuis que l'instruction dans cette langue est devenue obligatoire (Lois de Jules Ferry en 1881 et 1882), ils ont des trajectoires variables : pour certains, la langue a été transmise en famille ; pour d'autres elle a été apprise comme langue seconde.

Comparer conjointement l'occitan et le poitevin-saintongeais est justifié linguistiquement par la proximité étymologique de ces deux langues romanes. L'occitan (incluant les parlers intermédiaires du croissant) appartient à l'ensemble gallo-roman méridional tandis que le poitevin-saintongeais appartient comme le français à l'ensemble gallo-roman septentrional (langues d'oïl) avec un substrat occitan. Ainsi est-il intéressant d'étudier comparativement ce continuum dialectal de l'occitan au français en passant par les parlers du croissant et le poitevin-saintongeais. Par ailleurs, l'occitan et le poitevin-saintongeais évoluent différemment d'un point de vue sociolinguistique, même si toutes deux n'ont aucun statut officiel et sont non standardisées, l'occitan profite en particulier d'un système éducatif qui n'existe pas pour le poitevin-saintongeais.

Actuellement de nombreuses études (Koch & Osterreicher 2001, Biber & Conrad 2009) ont dépassé la conception dichotomique écrit/oral pour décrire ce que Koch et Osterreicher (2001) appelle un « continuum communicatif » avec une gamme de possibilités langagières identifiées selon des critères situationnels et contextuels. Plusieurs facteurs sont incorporés dans ce type de modèle, tels que le contexte (privée vs. public), le type d'interlocuteur (intime vs. public), la spontanéité vs. la préparation…

Dans cette présentation, nous souhaitons présenter nos travaux les plus récents (Carruthers et Vergez-Couret 2021, Carruthers et Vergez-Couret 2023) sur lesquels nous nous appuyons pour proposer une étude plus large de l'oralité : l'analyse de la temporalité dans un corpus de narration orale (contes écrits et oraux) en occitan[1] a conduit à prendre en considération des

---

'degrés d'oralité' qui tiennent à la fois compte des facteurs des modèles ci-dessous mais également de facteurs émergeant de la situation de langue minorisée, comme les sources et la pratique du conte, l'origine des compétences linguistiques, la variation diatopique (langue peu standardisée) et le contact entre l'occitan et le français.

Références :

Biber, Douglas et Conrad, Susan (2009). *Register, Genre and Style*, Cambridge, CUP.

Carruthers, J. & Vergez-Couret, M. (2023). 'Temporal patterning and 'Degrees of Orality' in Occitan and French Oral Narrative', Oral Tradition, 36/1, 91-122.

Carruthers, J. & Vergez-Couret, M. (2021). ʿTemporal structures in Occitan oral narrative: the role of frames and connectivesʾ, Lingvisticae Investigationes, 44(1), 1-36.

Koch, Peter et Oesterreicher, Wolf (2001). 'Gesprochene Sprache und geschriebene Spracheʿ, in: Günther Holtus/Michael Metzelti,/Christian Schmitt (edd.), *Lexikon der romanischen Linguistik* I.2, Tübingen, Max Niemeyer Verlag, 584-627.

Zumthor, Paul (1983). *Introduction à la poésie orale,* Paris, Seuil.

Variation in grammatical gender in Old Provençal and Old Languedocien

Marinus Wiedner
(University of Freiburg)

Even though Old Occitan was one of the first languages studied in the romance philology, some aspects have not yet been studied, for instance the grammatical gender: what happened to the Latin neuter in the transition from Latin to modern Occitan? Which gender was assigned to neologisms, e.g. nouns derived from a verb due to conversion? How are gender doublets distributed in time and space?

This work will explore the variation in gender in Old Occitan with a focus on Languedocien and Provençal. It is well known that a reduction in gender occurred from Latin – a tripartite system (cf. e.g. Pinkster. 2015. 24–50) – to modern Occitan – a binary system (cf. Oliviéri & Sauzet. 2016. 328–329). The starting point in Latin is (more or less) well known as is the result in today's Occitan (cf. e.g., Oliviéri & Sauzet. 2016. 328–329), but the transition from a three-gender system to the two-gender system somewhere in the medieval times is not yet well studied.

We presume that the transition manifests itself as a period of strong variation in gender assignment, as there is much evidence in the medieval text corpus. In the *Leys d'amors*, an essay about grammar and rhetoric written in the 14th century by the *Consistori del Gay Saber*, a group of poets in Toulouse, we find a commentary about gender:

per que se peco cil que dizo […]
la propheta la papa. quar devon dir
tug li orde le propheta. le papa
(Leys d'amors p. 125 rᵒ, column 1, l. 13–16)

Figure 1 Screenshot from the manuscript of the Leys d'amors, p. 125 rᵒ

'because one makes a mistake who says […]
la propheta la papa. Because they shall say
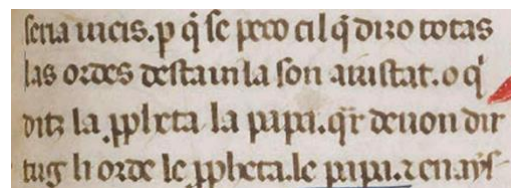all the time *le propheta. le papa*' (our translation)

This shows that there was definitively variation in gender in the 14th century in Old Occitan, and we aim to understand regularities in gender variation with our work.

As a precondition of such an analysis, a large corpus of Old Occitan texts is needed. Many editions of troubadour texts exist; however, most of them do not correspond to the letter in the manuscript. The editors usually made several changes (e.g. we find _la pantecosta_ in Meyer's edition of the *Roman de Flamenca* (Meyer. 1901. 96) yet _li pantecosta_ is written in the manuscript (p. 45 rᵒ)). This led to a sort of normalisation that excluded most of the variation (especially in gender). As such normalised editions are of no use for research in gender variation, we had to create a corpus ourselves. Therefore, we used Transkribus, where we trained a PyLaia HTR model for Old Occitan from scratch (our model has a CER of 3.61% on the validation set). So far, we've transcribed 12 manuscripts with a total of around 1500 pages and 350.000 tokens. The manuscripts we used and plan on using are mostly stored in the *Bibliothèque nationale de France* and in the *British Library*, and mostly written from the 12th to the 14th century in the language areas of Languedocien and Provençal.

References

Loporcaro, Michele. (2018). *Gender from Latin to Romance: history, geography, typology.* Oxford: University Press.

Meyer, Paul, 2nd edn. (1901). *Le Roman de Flamenca : publié d'après le manuscrit unique de Carcassonne. Traduit et accompagné d'un vocabulaire.* Paris: Librairie Emile Bouillon.

Oliviéri, Michèle & Patrick Sauzet. (2016). Southern Gallo-Romance (Occitan). In Ledgeway, Adam & Martin Maiden (ed.), *The Oxford Guide to the Romance Languages*, 319–349. Oxford: University Press.

Pinkster, Harm (2015). *The Oxford Latin Syntax Volume I. The simple clause*. Oxford: University Press.

Manuscripts:

Roman de Flamenca (Bibliothèque municipale de Carcassonne, cote 34).

Leys d'amors (Bibliothèque municipale de Toulouse, cote 2883/-4).