Society for Pidgin and Creole Linguistics

The 24th Annual Conference of the Association of Portuguese and Spanish-Lexified Creoles (ACBLPE) and the Summer Conference of the Society for Pidgin and Creole Linguistics (SPCL) – Joint Summer Meeting,

Aarhus University, 26-28 August 2026

ABSTRACTS / RESUMOS

AARHUS UNIVERSITY

A diachronic perspective on Portuguese contact effects on Kimbundu

Alexander Yao Cobbinah (Universidade de São Paulo)
& Lilian Teixeira de Sousa (Universidade Federal da Bahia)

The Portuguese have a continued presence in West Central Africa, since Diogo Cão's landfall in the end of the 15th century and particularly after foundation of the colony of São Paulo de Luanda in 1575. Since then Portuguese and Kimbundu have been in contact trough multilingual speakers in the native population, amongst the settlers and their Euro-African offspring and also the priests of the Jesuit and other missions active in the area. Kimbundu was used in the economic domain but also as a language of catechization, as evidenced by Pedro Dias grammar of 1697 and Pacconi's catechism of 1642. The publication of grammars and dictionaries later in the colonial history by Héli Chatelaîn, Antonio de Assis Junior, Antonio da Silva Maia and others attests to the ongoing importance of Kimbundu. In the course of this centuries-long coexistence, Kimbundu has absorbed lexical items but also syntactic structures. The use of conjunctions such as conditional *se* (from Portuguese *se*), conjunction of purpose *pala* (from Portuguese *para*) are widespread in modern Kimbundu and the 19th century sources (Chatelaîn 1894).

1 *Ng-a-biluk-ile* *pala* *ku-tala* *se* *u* *mu-tu*
1Sg-TAM-turn-Past for Cl.ku-see if COP.2Sg Cl.mu-person
u-enda *o* *ngongo*
2Sg-go Det hardship
`I had transformed myself to see, whether thou art one to stand hardship. `
Chatelaîn 1894: 35

This talk aims to present incipient research comparing the more recent varieties of Kimbundu with the 17th century material in order to establish how subordination functioned at an incipient stage of contact with Portuguese that probably reflects pre-contact usage more closely. While the language of the catechism and modern Kimbundu are remarkably close in terms of morphology, the earlier varieties show a much less integrated use of Portuguese lexicon, except for religious technical vocabulary, but also in syntactic structure. These varieties make use of non-finite constructions, nominalizations and a set of different conjunctions which are not attested in later chronolects. The conjunction *ene* in ex. 2, completely unattested in later sources, is used in context that translate as `because` and also `in order to`, also common are *he* for conditional subordinates, modally marked subordination and constructions with the verb *uchi* `say` in order to introduce complements or indirect speech.

2 *ü-ila* *ngahim* *ene* *Nzambi* *a-mu-lòloqu-e*
3Sg-say how Conj God 3Sg-3Sg.O-absolve-Subj
a-mu-equelexi *o* *i-tuxi* *a-y-e-pe* *co* *eulu*
3Sg-3Sg.O-forgive Det Cl.i-sin 3Sg-go-Subj-also Cl.ku heaven

`(....)que hade fazer este homem, para que Deos lhe perdoe os peccados, e se salue?`
[What does this man have to so that God forgive his sins and he save himself.']
Pacconi, 1642, dialogus octavus

This suggests that language change triggered by extended contact with Portuguese has impacted the syntax of Kimbundu quite profoundly. The time depth of available written data on Kimbundu is a remarkable asset in this endeavour, and the analyses here presented are based on an interlinearised corpus that contains the complete catechism of Pacconi (1642), Chatelaîn's (1888) grammar and part of his folk-tales (1894) as well as contemporary recordings made in the Libolo region as part of an ongoing documentation project.

References

Chatelain, H., 1888. *Grammatica elementar do Kimbundu ou lingua de Angola*. Schuchardt.

Chatelain, H. ed., 1894. *Folk-tales of Angola: fifty tales, with Ki-mbundu text, literal English translation, introduction, and notes* (Vol. 1). American Folk-lore Society.

Dias, Pedro 1697. Arte da lingua de Angola, oeferecida [sic] a virgem Senhora N. do Rosario, mãy, & Senhora dos mesmos pretos. Lisboa : Na officina de Miguel Deslandes, impressor de Sua Magestade. Com todos as licenças necessárias

Pacconi, F. and Maria, A., 1642. *Gentilis Angollae fidei mysteriis Lusitano olim idiomate per RP Antonium de Covcto Soc. Iesu theologum; nunc autem Latino per FR. Antonium Mariam Prandomontanum, concionatorem Capucinum,... instructus, atque locupletatus*. typis S. Congreg. de Propaganda Fide.

Names of Creole languages as examples of sociolinguistic Othering

Alla Klimenkowa (Universität Göttingen)

Drawing on historical examples of labelling Creole varieties in Upper Guinea and the French Caribbean in French records, the contribution illustrates that the colonial enterprise of the naming and appropriation of newly discovered space also included the naming of people and their languages "by classificational logics of metonymy and exclusion, recognition and opposition" (Dirks 1992).

The Cape Verde Islands and Upper Guinea became the first location where Pt. name *crioulo* and its loan translations Fr. *créole*, Sp. *criollo,* and En. *creole* were used to designate pidgin or creole varieties in the 17th century. However, studies on naming conventions for pidgin and creole varieties in general and in this region in particular are scarce (Bakker 2026; Lang 2007; Léglise& Migge 2006; Ladhams 2006; Bartens 2002; Bakker accepted). Léglise& Migge (2006) and Bakker (accepted) problematize methodological necessity in these studies of considering the heterogeneity of local social actors, produced by the disparities in power and positionality. A language ideological perspective on naming of pidgin and creole varieties is particularly relevant because linguistic labelling appears to be representative for the semiotic dimension of identity formation processes that depend on sociolinguistic Othering (Irvine& Gal 2000; Kroskrity 2021).

In addition to focusing language ideological aspects, this contribution offers a new perspective through its comparative approach. After discussing the French records concerning the Upper Guinea Coast in the 17th-18th centuries, it addresses naming conventions in the French own contemporary colonies on the Antilles (Hazaël-Massieux 2008). The comparative perspective develops the core argument, which is the need to analyse the sociolinguistic Othering in the foundational moments of the Old and New World engagement, by seeing how this process was central in the European mindset regardless of geographical boundaries and socio-cultural contexts. To underpin the arguments, this study analyses several 17th- to 18th-century French accounts about the Upper Guinea Coast and the French Antilles compiled primarily by merchants and travelers (N. Villaut de Bellefonds, M. Jajolet de La Courbe, R. Challe, J. Barbot, M. Moreau de Saint Méry, J. Girod de Chantrans).

The central argument is that the historical experience of naming pidgin or creole varieties within local creole communities, on the one hand, and by European outsiders, on the other hand, does not only create different indexicalities, but, first of all, invokes the clash between underlying understandings of ethnocultural and linguistic identities. Examined in light of a political, economic, and cultural rivalry, ideological biases underlying outsiders' negative aesthetical judgments of creole varieties become particularly evident. Disdainful references to the speakers of local varieties were usually seconded by depreciative labelling of their language. Such labels demonstrate the overriding power that social factors still have in the face of linguistic ones.

Português(es) em Angola em ecologias multilíngues:
um estudo comparativo da expressão do objeto in(direto)

Ana Paulla Braga Mattos (Aarhus Universitet), Izequiel Bernardo Ngolayami ()
& Márcia Santos Duarte de Oliveira (Universidade de São Paulo)

As chamadas variedades de português constituem um conceito metalinguístico que, na prática, encobre um conjunto incontavelmente heterogêneo de práticas linguísticas situadas em ecologias multilíngues nas Américas, África, Ásia e Europa. Termos como monolinguismo, bilinguismo ou multilinguismo, bem como dicotomias simplificadoras como urbano/rural, escrito/falado ou vernacular/padronizado (Mattos & Oliveira 2022: 425–427), mostram-se insuficientes para descrever a complexidade das práticas efetivas. Ao mesmo tempo, a persistência de ideologias linguísticas que associam "língua" à homogeneidade de povo, cultura e nação (Levinsen 2024: 77) sustenta políticas linguísticas que reiteram hierarquias coloniais e modelos eurocêntricos de normatividade.

É fundamental ressaltar que, em Angola, o português é falado predominantemente por uma população de substrato linguístico bantu, marcada pela presença de dezenas de línguas autóctones, faladas por milhões de pessoas. Embora a norma reconhecida institucionalmente continue ancorada no português europeu — historicamente associado à norma ortográfica de 1945 —, recentemente, devido à grande quantidade de material acadêmico disponibilizado livremente na internet, a norma brasileira vem sendo divulgada. Porém, tanto a norma do português europeu, quanto a brasileira permanecem distantes das práticas linguísticas cotidianas dos falantes angolanos. Como argumentam Friederike Lüpke e Anne Storch (2013), a ideia de língua "dialeto com missionário e dicionário" evidencia como processos de padronização e descrição linguística estão historicamente ligados a projetos de poder, apagando ecologias multilíngues e formas locais de legitimação.

Este trabalho propõe investigar fenômenos linguísticos nas variedades angolanas do português, inspirando-se em abordagens comparativas e em perspectivas de contato linguístico (Mattos & Oliveira 2020), como as desenvolvidas no âmbito do Portal de Variedades Português (PVP). O foco recai sobre os fenômenos de expressão do objeto direto e indireto, analisados comparativamente entre diferentes contextos de Angola e em contraste com variedades do Brasil, Portugal, Moçambique e São Tomé e Principe.

Ao historicizar a formação e legitimação das normas e evidenciar padrões estruturais sistemáticos nas variedades angolanas, argumenta-se que a expressão do objeto direto e indireto podem ser compreendidos como efeitos indexicais de múltiplas territorialidades sobrepostas, no sentido de Haesbaert (2019): o território escolar/estatal; os territórios 'locais' e o contato bantu–português; os territórios transnacionais de fala portuguesa (Oliveira, no prelo).

Este estudo contribui para os estudos de contato linguístico ao analisar a sistematicidade das variedades angolanas do português. Ao discutir o português de Angola como realidade sociolinguística situada, pretende ampliar o debate sobre sua descrição e reconhecimento.

Espera-se que os resultados possam oferecer subsídios para reflexões sobre políticas linguísticas e educacionais mais sensíveis às práticas locais.

‘Língua de preto’ na história do contato linguístico

Ana Paulla Braga Mattos (Aarhus Universitet)
& Eduardo Ferreira dos Santos ()

Este trabalho investiga o conceito de “língua de preto” a partir de uma perspectiva linguística ancorada em documentação histórica. O objetivo é examinar os agentes formuladores dessa designação, relacionando os contextos histórico-linguísticos e a tipologia diversa de dados textuais, identificando quais características linguísticas lhe foram atribuídas ao longo do tempo. A partir dessa análise, discutem-se os limites empíricos do conceito e seu estatuto na linguística de contato. O corpus abrange textos dos séculos XV a XVIII, referidos como escritos em língua de preto, bem como textos de estudiosos do tema dos séculos XX e XXI e organiza-se em três conjuntos: (1) representações literárias e administrativas da chamada “língua de preto”; (2) textos redigidos por africanos ou afro-portugueses em português; (3) textos de referência na tradição linguística que contribuíram para consolidar o termo como categoria analítica.

No primeiro conjunto, incluem-se:

– cantigas atribuídas a Anrique da Mota (final do século XV);

– teatro de Gil Vicente (início do século XVI);

– poesia de Sá de Miranda;

– teatro satírico de António Ribeiro Chiado;

– poesia satírica de Gregório de Matos (século XVII);

– registros de cronistas e escribas (séculos XVI–XVIII), nos quais falas atribuídas a africanos aparecem citadas e mediadas.

No segundo conjunto, incluem-se:

– cartas diplomáticas de Afonso I do Kongo (início do século XVI);

– textos religiosos do clero congolês;

– documentos administrativos do Reino do Kongo (séculos XVI–XVII);

– cartas e petições de elites afro-portuguesas na África Ocidental;

– registros administrativos produzidos por intérpretes africanos.

O terceiro conjunto contempla análises propostas por linguistas como Kihm & Rougé (2013, 2016), entre outros, que discutem a possível natureza pidgin ou pré-crioula da chamada “língua de preto”.

A metodologia combina: (i) levantamento e anotação sistemática das ocorrências atribuídas à “língua de preto” no corpus; (ii) descrição linguística contrastiva dos fenômenos recorrentes; (iii) análise historiográfica da tradição que passou a tratar o termo como categoria linguística. O trabalho discute três pontos centrais: (1) quando e em que contexto a expressão “língua de preto” passa de rótulo textual a categoria linguística; (2) se os dados disponíveis permitem

caracterizá-la como variedade estruturada ou como representação social; (3) em que medida é metodologicamente adequado classificá-la como pidgin, considerando a natureza mediada e literária de grande parte da documentação. Ao reexaminar criticamente o material documental e a tradição interpretativa, o estudo contribui para uma discussão mais precisa sobre a relação entre categoria histórica, representação textual e evidência linguística nos estudos da linguística de contato.

Borrowing and Cliticization of the Copula *ʧʰe* in Khatri

Antonia Alvares (University of Hong Kong)

Khatri is an Indo-Aryan contact language spoken by the Savjis, a small Hindu community who live in different parts of India. The paper focuses on the language as spoken in Belagavi, Karnataka. Khatri is characterized by having significant lexical and grammatical influence from the Indo-Aryan languages Gujarati and Marathi. The Gujarati influence in the language is surprising given that the speakers a) do not speak Gujarati, and b) presently live in places where Gujarati is not widely spoken. Examples of Gujarati grammatical elements are the present tense copula, the dative and accusative case marker, the perfective aspect marker, possessive determiners and possessive pronouns, gender inflection markers, and number inflection markers.

The paper focuses on the present tense copula ʧʰe in Khatri. It will be argued that a) the present tense copula was likely borrowed from Gujarati, and b) that the copula has as an auxiliary undergone cliticization to -ʧ. Compare Khatri example (1) with Gujarati example (2):

(1) ḍ̱ev ʧʰe.
god COP.PRS
'There is a God.' (Khatri)

(2) bʰəgvɑṉ ʧʰ-e.
god COP.PRS-3SG
'There is a God.' (Gujarati)

In Gujarati, as seen in example (2), the present tense copula is variant. The form ʧʰe is for three cells in the present tense copula paradigm: the second person singular, third person singular, and third person plural. In Khatri, as seen in example (1), ʧʰe is the only invariant form for the whole paradigm. The other prominent difference between Gujarati and Khatri involves negation: in Gujarati, ʧ- is replaced by a negation marker fused with a different verb tʰəvũ 'happen' giving rise to the form nəṯ̱ʰi. In Khatri, in contrast, we have symmetric negation i.e., ʧʰe used along with the analytic negation marker nei. These observations suggest that the copula was likely borrowed into Khatri from Gujarati.

In the New Indo-Aryan languages, tense and mood are considered to be a single category (Masica 1993: 279), which is located in the auxiliary (Masica 1993: 282). Further, the copula is usually identical with the auxiliary (Masica 1993: 337). This is the case in Gujarati. It will be argued, however, that in Khatri the tense/mood functions are in the marker -ʧ which attaches to lexical verbs as well as to negation markers. It will be argued, therefore, that this is the copula

tʃʰe that has undergone cliticization as an auxiliary as part of the Verb-to-TAM (tense, aspect, mood) chain (Heine 1993: 53). The Verb to TAM chain has a 'verbal/lexical structure at one end' changing into a 'grammatical marker of tense, aspect, modality, etc. at the other.' (Heine 1993: 53). In addition, the presence of both the original copula form t͡ʃʰe versus that of the clitic -t͡ʃ can demonstrate the principle of divergence in grammaticalization i.e., having the original form still present as an autonomous form although it has undergone grammaticalization (Hopper 1991: 22).

References

Heine, Bernd. 1993. *Auxiliaries: cognitive forces and grammaticalization*. Oxford: Oxford University Press.

Hopper, Paul J. 1991. On some principles of grammaticization. In Elizabeth C Traugott & Bernd Heine (eds.), *Approaches to grammaticalization, 2 vols* (Typological Studies in Language), vol. 19. Amsterdam: John Benjamins, vol. 1: 17-35.

Masica, Colin P. 1993. *The Indo-Aryan languages* (Cambridge Language Surveys). Cambridge: Cambridge University Press.

“Ele mira sem-semi, tem pesi, sabi?”:
mistura de recursos multilíngues no garimpo surinamês ‘Vila Brasil’

Antonio Lorenzo Dormal Calleja (Universidade de São Paulo)

O escopo desta comunicação centra-se no contato linguístico resultante da interação entre matauais e brasileiros na comunidade garimpeira multilíngue de Vila Brasil (VB), na Amazônia surinamesa, focalizando a variedade do brasileiro falado no território maroon dos Matauai. Nessas interações, mobilizam-se repertórios que incluem o brasileiro falado no Suriname (BFS), o surinamês (língua franca nacional), o saramakan, o matauai e o espanhol caribenho (Dormal Calleja, 2019, 2025; Dormal Calleja & Oliveira, no prelo). Embora, no garimpo VB, o BFS e o surinamês atuem como línguas veiculares, ao menos outras 12 línguas são atestadas na comunidade garimpeira compondo um ecossistema multilíngue de alta densidade interacional.

Neste contexto, observa-se a mobilização de múltiplos recursos multilíngues que se misturam na materialização de uma competência comunicativa adaptada às exigências de um ambiente de contato intenso em um território reduzido, o território Matauai. No exemplo a seguir, apresenta-se uma fala de um saramakan-matauai que fala a variedade ‘brasileiro de VB’ (BFS-VB): Quadro 1 – Fala de um garimpeiro saamaka/matauai na variedade BFS-VB

(GS/M): Ele mira sem-semi, tem pesi, sabi?						
	ele	mira	sem-semi	tem	pesi	sabi?
brasileiro	3.P.S			ter. P 3P.S	peixe	
surinamês			muito atento		fisi	sabi.
saramakan			muito atento		fisi	sabi
espanhol		mirar. P. 3P.S				
Trad. literal	Ele mirar atentamente; ter peixe, saber					
Trad. livre	*Ele mira mesmo (atentamente), tem peixe, sabe?*					

Fonte: Dormal Calleja (2025, p. 258) renumerado

A glossa revela a copresença de recursos provenientes do brasileiro (ele, tem), do espanhol caribenho (mira < mirar), do surinamês e do saramakan (sem-semi ‘muito atento’; sabi ‘sabe’), além de forma híbrida em pesi (‘peixe’), que combina traços fonológicos do saramakan e/ou surinamês com base lexical do brasileiro.

No quadro (1), atesta-se: (i) o uso de empréstimos lexicais multilíngue (Matras, 2010, p. 67-78); (ii) hibridismo estrutural em ‘*pesi’* (Sanchez-Stockhammer, 2012, p.135);

(iii) alternância de códigos envolvendo quatro línguas, fenômeno recorrente no Suriname (Yakpo, 2015, p. 270); no entanto não há predominância da gramática de nenhuma das línguas em uso O que emerge é uma configuração dinâmica de múltiplos recursos misturados (MR-M), acionados a partir dos repertórios disponíveis aos falantes.

Os resultados dessas interações apontam para um grau elevado de convergência linguística intracomunitária (Winford, 2003 p.74-84), que configura a variedade do brasileiro falado em Vila Brasil (BFS-VB) e a distingue do brasileiro falado em Paramaribo (BFS-P), descrito em Dormal Calleja (2019). Tal convergência envolve reconfigurações sistemáticas nos níveis lexical, fonológico e discursivo.

Argumenta-se que o BFS-VB não constitui mera variedade diastrática do brasileiro, nem se explica por modelos de bilinguismo estável, mas resultado emergente de um regime de contato multilíngue intensivo, no qual recursos de línguas diferentes são recombinados de modo funcional e recorrente.

A análise contribui para os estudos de contato e para investigações sobre o brasileiro como língua de migração e diáspora, ampliando o debate sobre emergência varietal, convergência e dinâmicas migratórias em espaços multilíngues contemporâneos.

Grambank and contact languages

Aymeric Daval-Markussen (Aarhus Universitet)

In recent years, several linguistic databases such as WALS (Dryer & Haspelmath (eds.) 2013) have been developed and made available, allowing large-scale typological comparisons of languages of the world.

Grambank (Skirgård et al. 2023), a newly-released database, documents the presence or absence of 195 morphosyntactic features in almost 2500 languages of the world, inlcuding a number of contact varieties. The sample of contact languages included is not particularly representative, in that most of the varieties considered are based on English. However, creoles lexified by other languages are also present in the database and data on 20 creoles based on 8 lexifiers can be found, as well as on the main substrate languages and all the lexifiers. So far, this rich data set has not been exploited in the context of creole studies.

The work initiated by Bakker et al. (2011) showed systematic differences between clusters of creoles and non-creoles. Following work by these authors, and others, has provided similar results, but this strain of research has been heavily criticized, because, among several reasons, it relies on WALS data, which have been depicted as being "riddled with depenedencies" (Meakins 2022: 195). In this study, we will apply similar phylogenetic methods to assess the degree of similarity of various languages, including several contact varieties in Grambank. We will focus specifically on the issue of interdependency of features, which has been claimed to be a major drawback in such investigations (e.g. Meakins 2022).

To date, this study represents the most extensive investigation of its kind, taking into account almost half a million data points for the comparisons. It will be shown that the Grambank data provide reliable results and that contact languages do cluster together regardless of the languages that contributed to their emergence.

References

Bakker, Peter, Aymeric Daval-Markussen, Mikael Parkvall & Ingo Plag. 2011. Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages*, 26(1) : 5-42. Dryer, Matthew S. & Haspelmath, Martin (eds.). 2013. *WALS Online* (v2020.4) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.13950591 (Available online at https://wals.info, Accessed on 2026-02-12.)

Meakins, Felicity. 2022. Empiricism or imperialism. The science of Creole exceptionalism. Journal of Pidgin and Creole Languages, 37(1): 189-203.

Skirgård, Hedvig & the Grambank consortium. (2023). *Grambank v1.0* (v1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7740140

The grammar of Carriols in the dialogues of Oldendorp

Aymeric Daval-Markussen (Aarhus Universitet)

In this talk, I will present a grammatical description of Carriols (also known as Virgin Islands Creole Dutch or Negerhollands), the extinct Dutch creole spoken in the former Danish West Indies, based on dialogues written by the Moravian missionary Oldendorp (1777/2000-2002) in the late 18th century, roughly one hundred years after the initial settlement of the islands by the Danes.

A tentative phonology will be sketched and the major parts of speech will be reviewed, and their functions will be described according to the closed corpus of dialogues.

It will be shown that the nature of this doculect remains largely creole-like, despite obvious influences from the lexifier Dutch, thus falling well within what to expect for a creole language in general, but also more specifically for Carriols in its early stages of development.

References

Oldendorp, Georg Christian Andreas. 1777/2000-2002. *Historie der caribischen Inseln Sanct Thomas, Sanct Crux und Sanct Jan insbesondere der dasigen Neger und der Mission der evangelischen Brüder unter denselben. Kommentierte Ausgabe des vollständigen Manuskriptes aus dem Archiv der Evangelischen Brüder-Unität Herrnhut*. Harmut Beck, Gudrun Meier, Stephan Palmié, Peter Stein & Horst Ulbricht (eds.). 4 vols. Berlin: VWB-Verlag für Wissenschaft und Bildung.

Aspect markers in Upper Guinea Creoles

Bart Jacobs (Uniwersytet Jagielloński)
& Nicolas Quint (CNRS, Langage, Langues et Cultures d'Afrique, Villejuif)

The Upper Guinea Creoles (UGCs) are a family of closely-related Afro-Portuguese Creole languages, comprising three branches: continental (Guinea-Bissau and Casamance Creoles), insular (Capeverdean), and ABC (Papiamentu).

One of the most striking features of this language family, when compared with their Ibero-Romance lexifier, is the use of an array of mostly preverbal particles exclusively devoted to the expression of aspect : see (1), where the particle *ta* expresses the habitual aspect.

(1) Santiago Capeverdean (Insular, Sotavento)
E ***ta*** *kánta*
S3SG **HAB** sing
's/he sings'

Prior work dealing with aspect marking in UGCs includes, e.g., Pratas (2021), Quint (2000), Silva (1990), Suzuki (1994) and Thiele (1991) on Capeverdean, Jacobs (2011) on both Capeverdean and Papiamentu, Kihm (1994), Peck (1988) and Scantamburlo (1981, 1999) on Guinea-Bissau Creole, Biagui (2018) on Casamance Creole, and Lenz (1927) and Maurer (1988) on Papiamentu. Aspect is furthermore covered in the respective APICS surveys (Michaelis & al. 2013). However, there is no specific study focusing on aspect morphology across the whole UGC family from a comparative-historical perspective. In the present talk, we intend to fill this gap.

In part one, we will briefly introduce the general notion of aspect (as defined by authors such as Comrie 1978, Cohen 1989, Kortmann 1991) and the methodology we have followed (language selection, data provenience…). In part two, we will present the different aspect markers attested in what we consider to be the core UGC varieties, i.e. continental and Sotavento (Southern Capeverdean) varieties, discussing their semantics, syntax, and etymology. In part three, we will deal in turn with aspect markers in the two outlier groupings of the family, namely Barlavento (Northern Capeverdean) and Papiamentu. Finally, in part four, we will propose and evaluate possible historical scenarios and reconstructions accounting for present-day UGC aspect systems and emphasize the diachronic and typological perspectives offered by our study.

Contextual Morphology and the Syntax-Phonology Interface, from Bantu to French Creole

Christopher Legerme (Massachusetts Institute of Technology)

1 Introduction. *How does syntax and phonology influence allomorphy?*

Comparing (1) and (4) below, we can see that *verb syncopation* or *short–long* patterns of French-Lexified Creole Languages (FLCs) (Seuren 1990) and the *conjoint–disjoint* alternation in Bantu (van der Wal 2017) show a strikingly parallel distribution: the short or conjoint forms surface only when the verb is not final in a syntactic domain (e.g., VP-final in Syea 1992 for Mauritian Creole short forms as well as in Halpert 2016 for the Zulu conjoint forms), and the long or disjoint form obligatorily surfaces in the same constituent-final position. I show how the underlying representation of the short/conjoint form interacts with a phonological constraint against unlicensed empty categories to derive this categorical edge-conditioned alternation: that is, short/conjoint forms lexically encode a segmentally vacuous CV syllable containing a nucleus which cannot be licensed at certain prosodic boundaries. My analysis builds on a model of cyclic, interface-driven spellout that integrates Minimalism and CV Phonology (Newell 2021; Newell & Sailor to appear; Scheer 2004, 2009), where syntax and phonology are autonomous grammatical modules that operate independently, and "are entirely incommunicado, in both directions" (Scheer 2016). While the Bantu conjoint–disjoint alternation garners much systematic crosslinguistic attention (van der Wal 2017), the comparable pattern among FLCs is relatively less understood across different languages. This presentation offers new empirical data from FLCs to refine our characterization and theoretical interpretation of the short/long verb alternation in FLCs.

2 Verb Alternations.

Verbal morphology in FLCs tracks language-specific, syntactic and phonological structure that is constrained by universal grammatical principles. There is a class of verbs among FLCs that alternate between a Short Form (SF) and Long Form (LF), for example, in Mauritian Creole or Haitian Creole the SF ***gad*** or the LF ***gade*** can both be used to spellout the verb meaning, "to look". Generally, the two forms may be complementary or interchangeable depending on the language and the syntactic or information-structural context. However, SFs (e.g., *gad*) are categorically excluded from being VP-final (Henri 2021; Seuren 1990; Syea 1992, ✓ is good, and X bad):

(1) *The SF/LF Alternation in Haitian Creole*

a.	Malis	te	✓ **gad** / ✓ **gade**	fim		b.	Fim	ki	te	X **gad** / ✓ **gade**	sou Netflix
	Malis	PST	watch	film			Film	that	PST	watch	on Netflix

'Malis watched movies.' 'Movies that were watched on Netflix.'

The facts are complex and the SF/LF alternation plays out differently across FLCs: between Haitian Creole and Mauritian Creole, for example, the conditioning syntactic factors may diverge

significantly, for example below:

(2) *Haitian Creole SF/LF with Pronominal Complements*

Saika ap X gad / ✓ gade li
Saika PROG watch 3SG

"Saika is watching it."

(3) *Mauritian Creole SF/LF with Pronominal Complements* (Henri 2021: 467)

a. Mo ti ✓ **trouv** / X **trouve** li
1SG PST see 3SG

"I saw him/her."

b. Mo ti ✓ **trouv** / X **trouve** Zan
1SG PST see John

"I saw John."

Haitian Creole pronominal objects behave differently from full DPs with respect to the LF/SF alternation (1–2). In Mauritian Creole, by contrast, the SF is used whenever the verb selects an in-situ non-clausal complement (3), regardless of the syntactic category of that complement (Henri 2021: 467). There's some consensus that the SF/LF alternation reflects deeper facts about syntactic constituency (van der Wal 2017; van der Wal & Veenstra 2015). In this regard FLCs are likened to Bantu languages such as Zulu where the Conjoint (CJ) and Disjoint (DJ) alternation is also "constituency-based", and CJ forms likewise may not be VP-final (e.g., Halpert 2016: 87–89; cf. van der Wal & Veenstra 2015: 120):

(1) *Zulu CJ/DJ Alternation*

a. uMlungisi u- X ya / ✓ Ø -pheka iqanda
AUG.1Mlungisi 1SG- DJ CJ -cook AUG.5egg
"Mlungisi is cooking an egg."

b. uMlungisi u- ✓ ya / X Ø -pheka
AUG.1Mlungisi 1SG- DJ CJ -cook
"Mlungisi is cooking."

Still, the historical connection between Bantu and FLCs, and the recurrence of the clause-final restriction for SFs and CJ forms remains a vibrant and growing area of ongoing investigation (e.g., van der Wal and Veenstra 2015: 106, and Henri 2021: 478).

3 Spellout and Empty CVs.

The SF/LF and CJ/DJ alternations show that allomorphy is shaped by the interaction of syntax and phonology, and reducible to no module alone. I adopt an intermodular model of spellout that integrates proposals from Minimalist theory as it has been applied to Syntax and Phonology, where each module operates autonomously over their own vocabularies (Newell & Sailor to appear; Scheer 2004, 2009, 2016). In a Strict CV framework, empty nuclei are silenced when they are immediately followed by a non-silenced V position, and domain-final empty nuclei are always silenced (Charette 1991). I propose that the clause-final restriction on SF and CJ forms results from a lexically specified CV containing an empty nucleus which fails to be properly governed and silenced within its phonological domain, linking allomorphy to computations over abstract suprasegmental structure informed by sonority (i.e., nuclei are local sonority peaks) (Scheer 2016: 345–346). Morphological sensitivity to prosodic vacuity at spellout is attested (e.g., Kandybowicz 2015), as is the link between the SF/LF alternation and the distinct lexical

representations of the two verb forms (Henri 2021). The present account, however, derives the clause-final restriction in FLCs and Bantu from principles of syntactic and phonological computation within a succinct and Minimalist architecture.

References

Charette, M. (1991). *Conditions on phonological government*. Cambridge University Press.

Halpert, C. (2016). *Argument licensing and agreement*. Oxford University Press.

Henri, V. (2021). Morphomic structure in Mauritian Kreol: On change, complexity and creolization. *Morphology*, *31*(3), 309–340. https://doi.org/10.1007/s11525-021-09383-9

Kandybowicz, J. (2015). On prosodic vacuity and verbal resumption in Asante Twi. *Linguistic Inquiry*, *46*(2), 243–272. https://www.jstor.org/stable/ 43695678

Newell, H. (2021). Deriving Level 1/Level 2 affix classes in English: Floating vowels, cyclic syntax. *Acta Linguistica Academica*, *68*(1-2). Newell, H., & Sailor, C. (to appear). Minimalism and the syntax–phonology interface [Draft available at https://craigsailor.net/]. In K. K. Grohmann & E. Leivada (Eds.), *The cambridge handbook of minimalism*. Cambridge University Press.

Scheer, T. (2004). *A lateral theory of phonology, vol. 1: What is CVCV and why should it be?* De Gruyter.

Scheer, T. (2009). Intermodular argumentation and the word-spell-out-mystery. In K. K. Grohmann (Ed.), *Explorations of Phase Theory: Interpretation at the Interfaces* (pp. 23–66). Mouton de Gruyter.

Scheer, T. (2016). Melody-free syntax and phonologically conditioned allomorphy. *Morphology*, *26*, 341–378. https://doi.org/10.1007/s11525-016-9283-6

Seuren, P. (1990). Verb syncopation and predicate raising in Mauritian Creole. *Theoretical Linguistics*, *13*(1), 804–844.

Syea, A. (1992). The short and long form of verbs in Mauritian creole: Functionalism versus formalism. *Theoretical Linguistics*, *18*, 61–97.

van der Wal, J. (2017). What is the conjoint/disjoint alternation? parameters of crosslinguistic variation. In J. van der Wal & L. M. Hyman (Eds.), *The conjoint/disjoint alternation in bantu* (pp. 14–60). De Gruyter Mouton. https://doi.org/10.1515/9783110490831-002

van der Wal, J., & Veenstra, T. (2015). The long and short of verb alternations in mauritian creole and bantu languages. *Folia Linguistica*, *49*(1), 85–116. https://doi.org/10.1515/flin-2015-0003

O emprego de artigo definido diante de possessivo no continuum afro-brasileiro de variedades do português: convergências e contrastes

Danielle Kely Gomes (Universidade Federal do Rio de Janeiro)

A presença ou ausência do artigo definido antes de pronomes possessivos — como nas alternâncias entre "***ø*** minha casa" e "***a*** minha casa" — tem sido amplamente discutida na literatura sociolinguística, especialmente em estudos voltados à descrição e comparação das variedades do português faladas no Brasil e em Portugal. Estudos pioneiros, como os desenvolvidos por Oliveira e Silva (1982, 1996), evidenciam que a realização dessa variável não é homogênea no conjunto das variedades do português, revelando comportamentos diferenciados no português brasileiro e no português europeu. No contexto das variedades africanas do português, investigações anteriores indicam que as normas urbanas faladas no continente africano se aproximam do português brasileiro, na medida em que o uso do artigo diante de possessivos não é categórico, mas sujeito à variação. Partindo desse cenário, o presente estudo propõe uma análise contrastiva entre normas do português faladas no estado do Rio de Janeiro e normas urbanas de São Tomé e de Maputo. A investigação se insere em uma perspectiva mais ampla de reflexão sobre o *continuum* de variedades luso-afro-brasileiras do português, conforme sugerido por Petter (2015) e Brandão (2021). Assume-se, como hipótese central, que as normas brasileiras e africanas compartilham condicionamentos estruturais semelhantes no uso do artigo com possessivos, mas se diferenciam quanto à influência de fatores extralinguísticos. No que diz respeito às variedades africanas, postulou-se que o contato intenso entre o português e as línguas locais — crioulos de base lexical portuguesa em São Tomé e línguas bantu em Moçambique — poderia atuar como fator de inibição para o emprego do artigo definido, uma vez que essas línguas não dispõem de um sistema de determinantes equivalente ao do português. Em contrapartida, considerou-se que, no português brasileiro, a escolha entre as variantes não estaria significativamente condicionada por fatores extralinguísticos, configurando-se como um fenômeno condicionado essencialmente por variáveis morfossintáticas e discursivas. A análise baseia-se em dados provenientes de três *corpora* representativos das normas urbanas investigadas, estratificados segundo sexo, nível de escolaridade e faixa etária. Para o português brasileiro, foram examinados 36 inquéritos do projeto *Comparaport,* relativos às comunidades de Copacabana e Nova Iguaçu. A norma são-tomense é representada por 17 entrevistas do corpus VAPOR, desenvolvido no âmbito do projeto Variedades do Português (CLUL/Universidade de Lisboa). Já a norma urbana de Moçambique foi analisada a partir de 18 entrevistas da amostra Moçambique-PORT. O tratamento estatístico dos dados foi realizado com o auxílio do software *R-Studio*. Os resultados iniciais da análise contrastiva indicam convergência entre as comunidades quanto à atuação de fatores estruturais na seleção das variantes. Por outro lado, observam-se contrastes substanciais entre as variedades brasileira e africanas no que se refere ao papel desempenhado por restrições extralinguísticas. A hipótese de que o contato linguístico seria um fator determinante na dinâmica da variação nas variedades africanas não se

confirmou, já que aspectos como o estatuto de aquisição do português e a circulação dos falantes em contextos multilíngues não apresentaram relevância estatística.

Documenting a Creole: Aligning community goals with research goals

Eeva Sippola (Helsingin yliopisto), Danae Perez (independent scholar)
& Jillian Melchor (Helsingin yliopisto)

This presentation examines the intersections and divergences between community and researcher perspectives on language documentation in a creole-speaking community with the aim of exploring new methodological approaches to the documentation of creole varieties. The focus lies on Cavite Chabacano, a severely endangered variety of Philippine Creole Spanish spoken in Cavite City by less than 3000 speakers of the grandparental generation in restricted domains (Lesho & Sippola 2013).

Community-based research emphasizes collaborative practice in which linguistic data are understood as expressions of people's lived languages that hold personal and communal meaning (Czaykowska-Higgins 2009; Leonard & Haynes 2010; Rice 2018). Such approaches recognize that definitions of "language" vary widely: while positivist linguistics prioritizes structural systems, speakers often situate language in relation to emotion, place, spirituality, art, and collective history. In plurilingual contexts such as Cavite City, fluid language boundaries further challenge what counts as a language and what researchers consider to be data (cf. Migge 2020). These differing orientations shape expectations about language documentation and its desired outcomes.

This study draws on interviews and community workshop discussions conducted between November 2025 and January 2026 in Cavite. Using discourse and nexus analysis that foreground intersecting practices and ideologies (Scollon & Scollon 2004), we examine how participants conceptualize the Chabacano-speaking community and what priorities for language documentation they name.

Preliminary findings reveal that, due to the language's endangered status, participants hesitated to identify an existing active speaker community; instead, they referred to a historical community that used Chabacano in churches, marketplaces, and at home. Participants identified culinary traditions and religious practices as key domains for documentation, reflecting their cultural and spiritual salience and the symbolic functions of the language. This stands in contrast to linguists' usual objectives focused on everyday multilingual dialogues and language practices. Importantly, community members viewed documentation as inseparable from revitalization. They emphasized the need for institutional involvement by local government bodies and the Department of Education to support broader revitalization and sustainability goals. This expectation highlights the community's desire for documentation to contribute directly to intergenerational transmission and language visibility, while the linguists' initial plan had been the documentary corpus and its archiving.

Overall, the results align with scholarship arguing that collaborative research extends beyond conventional academic products like grammars, dictionaries and texts.

Instead, they highlight the importance of relationship-building, co-developed materials responsive to community priorities, and research output that advances community well-being and linguistic continuity (cf. Akumbu 2020; Faraclas et al. 2019). The Cavite case shows the potential and the challenges of aligning researcher frameworks with community-driven goals for the future of endangered creole languages and minoritized languages in general.

References

Akumbu, Pius W. 2020. Reflections on a community-based approach to writing grammars of endangered African languages. *Studies in African Languages and Cultures*. 54: 71-96.

Czaykowska-Higgins, Ewa. 2009. Research models, community engagement, and linguistic fieldwork: Reflections on working within Canadian Indigenous communities. *Language documentation and conservation* 3(1). 15-50.

Faraclas, Nicholas, Ellen-Petra Kester, Eric Mijts. 2019. *Community Based Research in Language Policy and Planning*. (chapter 2: Community Based Research in Postcolonial Communities).

Lesho, M. & E. Sippola 2013. The sociolinguistic situations of Manila Bay Chabacano-speaking communities. *Language Documentation & Conservation* 7, 1–30.

Leonard, W. Y. & E. Haynes. 2010. Making "collaboration" collaborative: An examination of perspectives that frame linguistic field research. *Language documentation and conservation* 4, 268–293.

Migge, B. 2020. Researching endangered languages: Critical reflections on field and documentary linguistics. In Deumert, A., A. Storch, & N. Shepherd. (eds). *Colonial and Decolonial Linguistics: Knowledges and Epistemes*. OUP.

Rice, K. 2018. Collaborative research: Visions and realities. In S. Bischoff & C. Jany (eds.), *Insights from practices in community-based research: From theory to practice around the globe,* 13–37. Berlin/Boston: Mouton de Gruyter.

Scollon, R., & Scollon, S. W. 2004. *Nexus Analysis Discourse and the emerging Internet.* Routledge.

A core vocabulary database for Romance lexified Creoles

Fábio Barcellos Granja (Universiteit Utrecht)

Despite there being several studies analysing the phylogenetic and diachronic links between Romance-based Creoles and their lexifiers (Stein, 1984; Quint, 2000; Sippola, 2017), most of their approaches employ qualitative methods (Levisen et al., 2017) and focus on morpho-syntactic rather than phonological features (Smith, 2008; Silva, 2023). Furthermore, although there are large open-access vocabulary databases concerning many of the world's languages (e.g. WOLD, Haspelmath & Tadmor (2009); ASJP, Wichmann et al. (2025)), there is currently no such resource encompassing a significant amount of Creoles. In this study, we aimed to address these issues by constructing an open-access database of core vocabulary wordlists in 34 Romance lexified Creoles. Using this dataset, we have employed computational methods to estimate the phylogenetic dis-tances between Creoles and their lexifiers, as well as lexical-phonetic distances between Creoles deriving from related lexifiers.

Our database was formed by compiling 207-word Swadesh lists (Swadesh, 1955) for 20 Portuguese, 4 Spanish and 10 French-based Creoles; this also included Creole varieties (e.g. Zamboanga and Cavite Chabacano). This resulted in a .csv dataset of 6336 words collected from dictionaries and articles resulting from fieldwork (e.g. Carring-ton (1984)). Each list was phonetically transcribed following the original authors and standardised in accordance with PHOIBLE conventions (Moran & McCloy, 2019). For each Creole list, we have also added a ”Target” column proposing the likely origin of each Creole word, taking into account diachronic and diatopic variation internal to their lexifiers.

To test the hypothesis that the alignment between our sample languages was areal, we have analysed the phonetically transcribed lists from Creoles and their lexifiers using the AlineR R package (Downey et al., 2017) and the SplitsTree software (Huson & Bryant, 2024). This allowed us to quantify lexico-phonetic distances between each Creole and its lexifier, assessing both cognacy and diachronic phonological processes. Using these data, we could compare Creoles between themselves to verify their distribution in an unrooted NeighborNet that showed phylogenetic and lateral connections.

Our results show that the phylogenetic distribution of our sample languages is mostly geographical. Moreover, Creoles coming from genetically related lexifiers also form larger groups, with the Spanish-based Creoles being closer to the Portuguese than to the French ones. In addition, they also reveal lateral influences between Creoles that share socio-historical links (e.g. South Asian and Southeast Asian Portuguese-based Creoles), making it possible to verify and quantify previous qualitative claims about them (Chaudenson, 1979; Carvalho & Lucchesi, 2016).

With regards to phonology, this study innovates by allowing the ranked visualisation of the main sound changes affecting lexifiers during creolization. This provides new, useful insights concerning which phonological processes are the most frequent ones in Romance Creole diachrony. When comparing Creoles to their lexifiers (i.e. targets), contact conditions were important predictors, with fort Creoles being generally more stable than maroon ones. Lastly, the implications of having a unified basic vocabulary database of Romance-based Creoles for linguistic research are manifold. Besides aid-ing the documentation of often under-researched languages, it provides a new tool for phylogenetic research in diachronic typology as well as for studies assessing similarities between Creoles coming from genetically related languages.

References

Carrington, Lawrence D. 1984. *St. lucian creole : a descriptive analysis of its phonology and morpho-syntax* Kreolische Bibliothek, Bd. 6. Hamburg: Helmut Buske.

Carvalho, Ana Maria & Dante Lucchesi. 2016. Portuguese in contact. In W. Leo Wetzels, João Costa & Sergio Menuzzi (eds.), *The Handbook of Portuguese Linguistics*, 41–55. Oxford: John Wiley & Sons. doi:10.1002/9781118791844.ch3.

Chaudenson, Robert. 1979. *Les créoles français*. Fernand Nathan.

Downey, Sean S., Guowei Sun & Peter Norquest. 2017. alineR: an R package for optimizing feature-weighted alignments and linguistic distances. *The R Journal* 9. 138–152. doi:10.32614/RJ-2017-005. Https://doi.org/10.32614/RJ-2017-005.

Haspelmath, Martin & Uri Tadmor (eds.). 2009. *Wold*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wold.clld.org/.

Huson, D. H. & D. Bryant. 2024. The splitstree app: interactive analysis and visualization using phylogenetic trees and networks. *Nature Methods* 21. 1773–1774. https://doi.org/10.1038/s41592-024-02406-3.

Levisen, Carsten, Eeva M. Sippola & Peter Bakker. 2017. *Introduction* 1–4. John Benjamins Publishing Company. https://doi.org/10.1075/z.211.01lev.

Moran, Steven & Daniel McCloy (eds.). 2019. *Phoible 2.0*. Jena: Max Planck Institute for the Science of Human History. https://phoible.org/.

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

Silva, Carlos. 2023. *Consonant stability of portuguese-based creoles*: University of Porto Phd thesis. Available at https://hdl.handle.net/10216/157550.

Sippola, Eeva M. 2017. *Similarities and differences among Iberian creoles* 241–268. John Benjamins Publishing Company. doi:doi:10.1075/z.211.11sip. https://doi.org/10.1075/z.211.11sip.

Smith, Norval. 2008. Creole phonology. In *The Handbook of Pid-gin and Creole Studies*, chap. 5, 98–129. Oxford: John Wiley & Sons. https://doi.org/10.1002/9781444305982.ch5.

Stein, Peter. 1984. *Kreolisch und Französisch*. Berlin, Boston: Max Niemeyer Verlag. https://doi.org/10.1515/9783110963472.

Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21(2). 121–137.

Wichmann, Søren, Eric W. Holman, Cecil H. Brown, Matthew S. Dryer & Qibin Ran. 2025. The asjp database (version 21). doi:10.5281/zenodo.3835952. https://asjp.clld.org.

Linguistic Variation, Orthography and Identity in Caboverdean

Fernanda Pratas (Faculdade de Letras, Universidade de Lisboa, Centro de Linguística), Ana Karina Moreira (Universidade de Cabo Verde), Dominika Swolkien (Universidade de Cabo Verde), Eliane Vieira Semedo (Universidade de Cabo Verde) & Mariana Almeida (Faculdade de Letras, Universidade de Lisboa, Centro de Linguística)

This paper presents a study that investigates the relationship between internal linguistic variation, orthographic representation, and identity in Caboverdean (*Kriol(u)*), a language spoken by the entire population of Cabo Verde, but which has yet to attain much-desired official status. Adopting an interdisciplinary approach, the study combines detailed linguistic description with a sociolinguistic perspective, with the aim of understanding how speakers' identity-related attitudes influence the acceptance or rejection of proposals for orthographic standardization.

The Caboverdean linguistic landscape comprises varieties traditionally grouped into two broad geographical areas: Sotavento (the southern islands) and Barlavento (the northern islands). This division, widely recognized by speakers, has been described as socially salient, both within the archipelago and in the diaspora, and continues to play a central role in discussions of linguistic and national identity (Batalha 2004; Rego 2008; Baptista et al. 2010). Although international organizations emphasize that linguistic diversity should be valued in inclusive societies (UNESCO 2025), the full integration of Caboverdean into written formal domains – particularly the education system – requires the development of written resources, such as school textbooks and dictionaries, which in turn entails the adoption of a standardized orthographic system.

The central hypothesis of this study is that speakers' strong identification with local varieties, even more than with the traditional Sotavento/Barlavento groups, decisively shapes their attitudes towards orthographic proposals based exclusively on a single variety, and that resistance to orthographic standardization is therefore not only linguistic but also ideological and identity-based. This hypothesis directly engages with research showing that orthographic systems are not neutral instruments, but rather products of historical, social, and political processes, often associated with struggles over power and symbolic recognition (Sebba 2007).

Methodologically, the project develops along two complementary lines of inquiry. First, it draws on new fieldwork conducted on the islands of Brava, Boavista, Sal, and São Nicolau, whose varieties remain comparatively underdocumented, as well as on the digital annotation of a large body of data from all the islands (LUDViC), providing comprehensive coverage of the true extent of language variation at all levels and how this may interact with its written representation. Second, it examines metalinguistic discourse surrounding the current controversy over standardization through a mixed-methods approach: (i) qualitative analysis of both previous and new interviews with speakers, ethnographic observations available in the literature, historical texts, press articles, and social media posts; and (ii) quantitative analysis of a questionnaire specifically designed to assess linguistic attitudes and perceptions of identity. Some preliminary results of this questionnaire are already available at this point.

By bringing together linguistic description, discourse analysis, and quantitative data on attitudes and language use, this study significantly contributes to the crosslinguistic research on the role of identity in the standardization of pluricentric and postcolonial languages (e.g. Ardoino 2024; Bissoonauth et al. 2025; Lüpke 2018). The findings are therefore relevant not only to the debate on Caboverdean, but also to other contexts in which internal diversity constitutes, on the one hand, an immense resource at the linguistic and cultural levels, and, on the other hand, a political and educational challenge.

Introducing the *Oral Corpus of the Indo-Portuguese Creoles of Daman and Diu*

Hugo C. Cardoso (Universidade de Lisboa), J. Clancy Clements (Indiana University) & Pedro Jordão (Universidade de Lisboa)

The Indo-Portuguese Creoles in and around Diu and Daman (including the hinterland territory of Dadra and Nagar-Haveli) have traditionally been counted among the *Norteiro* ('northern') creoles, as they developed in the northernmost Portuguese colonial possession in India. They are currently spoken by relatively small and fragile communities in their original territories and among a diaspora which concentrates mostly in the United Kingdom and in Portugal, and their study has revealed a significant degree of similarity among the various regional varieties, but also some differences (including community-internal sociolinguistic variation, see Clements 2014).

The first initiatives to document them began in the late 19th century, with Hugo Schuchardt's (1883) article about the variety of Diu and Sebastião Dalgado's (1902-3) about Daman. In more recent times, further studies and descriptions were produced in the first 2 decades of the 21st century, (e.g. Clements & Koontz-Garboden 2002; Cardoso 2009, 2013; Clements 2009), although no annotated corpus was produced, unlike for several other Asian-Portuguese Creoles. A recent project developed at the University of Lisbon has been working on closing this documentation gap, producing *Cor-DD – Corpus Oral dos Crioulos Indo-Portugueses de Damão e Diu*, an open-access, morphosyntactically annotated corpus of oral speech – hosted by the Lisbon University Centre for Linguistics' TEITOK platform – combining legacy materials with new data collected in Daman, Diu, Dadra and Nagar-Haveli, as well as various locations in the United Kingdom. In presenting the *Cor-DD* corpus, its contents, features, and the workflow underlying its production, from data collection to data processing and annotation, we highlight the usefulness of this new resource and, more generally, underscore the potential impact that corpus-based approaches to the study of these languages have for comparative linguistic studies and how they can contribute to the preservation of an endangered linguistic heritage.

References

Cardoso, Hugo C. 2009. *The Indo-Portuguese Language of Diu*. Utrecht: LOT.

Cardoso, Hugo C. 2013. Diu Indo-Portuguese. In *The Survey of Pidgin and Creole Language*, vol. 2 [*Portuguese-based, Spanish-based, and French-based Languages*], S. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds.), 90-101. Oxford: Oxford University Press.

Clements, J. Clancy. 2009. Gradual vs. abrupt creolization and recent changes in Daman Creole Portuguese. In *Gradual Creolization: Studies Celebrating Jacques Arends*, R. Selbach, H.C. Cardoso & M. van den Berg, 55-75. Amsterdam: John Benjamins.

Clements, J. Clancy. 2014. Lectal differences in Daman Indo-Portuguese. *Revista de Crioulos de Base Lexical Portuguesa e Espanhola* 5, 115-156.

Clements, J. Clancy & Andrew Koontz-Garboden. 2002. Two Indo-Portuguese creoles in contrast. *Journal of Pidgin and Creole Languages* 17, 191-236.

Dalgado, Sebastião Rodolfo. 1902–1903. Dialecto indo-português de Damão. *Ta-Ssi-Yang Kuo, Archivos e annaes do Extremo Oriente Portuguez* 3, 350–367; 4, 515–523.

Schuchardt, Hugo. 1883. Kreolische Studien III. Über das Indoportugiesische von Diu. *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften zu Wien (Philosophisch-historische Klasse)* 103, 3–18.

Contato linguístico e prosodização de marcadores discursivos no português de Calulo

Isabella Matos Rodrigues & Flaviane Romani Fernandes Svartman (Universidade de São Paulo)

Esta pesquisa realiza uma descrição e análise sobre a prosodização dos marcadores discursivos no português falado no município de Calulo (doravante, PCL), em Angola. Com base em corpus de fala espontânea e informal, coletados pelo Projeto Libolo (cf. Figueiredo, Oliveira, 2016) para compor o "*corpus* mínimo" de um projeto intitulado C-Oral-Angola (Rocha, Mello, Raso, 2018), investigamos se há possíveis especificidades prosódicas dessa variedade na fala espontânea, devido ao contato linguístico da região.

A pesquisa pretende contribuir para: (i) os estudos comparativos do português, ao focalizar um tópico ainda inexplorado nesta variedade — os marcadores discursivos e suas prosodizações; (ii) a ampliação das descrições prosódicas do PCL, que ainda conta com poucos trabalhos na área (cf. Santos, 2020); e (iii) os estudos de contato linguístico

A fim de atingir esses objetivos, a metodologia do estudo envolve: (i) a identificação dos itens linguísticos que desempenham função de marcador discursivo no corpus, com base em trabalhos de referência sobre o português brasileiro (Risso, Silva e Urbano, 2006; Raso, 2014); (ii) a análise de sua prosodização à luz da Fonologia Entoacional (Pierrehumbert, 1980; Pierrehumbert & Beckman, 1988) e da Fonologia Prosódica (Selkirk, 1984; Nespor & Vogel, 2007 [1986]), buscando descrever padrões de contorno entoacional, fronteiras prosódicas e possíveis configurações recorrentes; e (iii) o estabelecimento de correlações entre esses padrões e o contexto de contato linguístico, caso sejam identificadas singularidades sistemáticas.

Partindo dos estudos prosódicos de Santos (2020), que analisa tipos frásicos do PCL com dados controlados, levantamos a hipótese de que a análise de produções orais espontâneas — sobretudo de um objeto particularmente sensível à interface prosódico-pragmática, como os marcadores discursivos — pode revelar singularidades da variedade ainda não descritas.

Um primeiro levantamento dos marcadores discursivos no PCL foi realizado por Oliveira, Zanoli e Andrade (2018), que analisaram quatro itens (*ya*, *eh pa*, *tás a ver* e *Júlia*) à luz da Teoria da Língua em Ato (L-AcT), proposta por Raso (2014). Na presente pesquisa, ampliamos esse escopo, adotando uma abordagem centrada na prosódia.

As análises parciais concentram-se em dois marcadores particularmente característicos do PCL, não atestados em variedades do português brasileiro ou europeu: *ya* e *wei*. O primeiro é descrito por Oliveira, Zanoli e Andrade (2018) como possivelmente relacionado ao contato histórico com alemães na região de Calulo. Abaixo, apresentamos dados como exemplo:

(1) você que tem conhecido / ya //
(2) wei / o cota mandou dá já um toque aqui //

Os resultados preliminares indicam que o *ya* tende a ocorrer com contorno nuclear ascendente, independentemente de sua posição no enunciado. Já o *wei* apresenta maior variação prosódica em posição inicial, mas revela tendência a contornos descendentes quando ocorre em posição final.

Referências bibliográficas

Figueiredo, C. F. G.; Oliveira, M. S. D. **'Projeto Libolo' – Município do Libolo, Kwanza-Sul, Angola: aspectos linguísticoeducacionais, histórico-culturais, antropológicos e sócio-identitários**. Lisboa: Chiado, 2016. v. 1.

Nespor, M.; Vogel, I. **Prosodic phonology**. Dordrecht: Foris Publications, 2007 [1986].

Oliveira, M. S. D; Zanoli, M. L.; Andrade, G; M.. Marcadores Discursivos no português falado em Angola, subvariedade Libolo: um estudo inicial de base prosódico-pragmática. **Filologia e Linguística Portuguesa**, v. 20, n. Especial, p. 159-186, 2018.

Pierrehumbert, J. **The phonology and phonetics of English intonation**. 1980. Tese (Doutorado) - M.I.T., Cambridge, Mass., 1980.

Pierrehumbert, J.; Beckman, M. **Japanese Tone Structure**. Cambridge, Mass.: M. I. T. Press, 1988.

Raso, T. **Prosodic constraints for discourse markers**. In: RASO, T; MELLO, H. (org.). Spoken corpora and linguistic studies. Amsterdam, John Benjamins Publishing Company, p. 411-467, 2014.

Risso, M. S.; Silva, G. M. O.; Urbano, H. Traços definidores dos marcadores discursivos. In: Jubran, C. C. A. S.; Koch, I. G. V. (org.). **Gramática do português culto falado no Brasil: construção do texto falado**, v. 1. Campinas: Editora da UNICAMP, 2006.

Rocha, B.; Mello, H.; Raso, T. Para a compilação do C-ORAL-ANGOLA: um corpus de fala espontânea informal do português angolano. **Filologia e Linguística Portuguesa**, v. 20, n. Especial, p. 139-157, 2018.

Santos, V. G. **Aspectos prosódicos do português angolano do Libolo: entoação e fraseamento** . 2020. Tese (Doutorado em Filologia e Língua Portuguesa) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2020.

Selkirk, E.O. Phonology and syntax: the relation between sound and structure. Cambridge, Mass. & London: MIT Press, 1984.

Constructing, sustaining, and transmitting Cameroonian identity through multilingual repertoires in diasporic context

Isolde Bonnet (Universität Freiburg)

The paper examines how first-generation Cameroonian immigrants construct, maintain, and transmit cultural and linguistic identity through everyday linguistic practices in the German, French, and British diasporas. Particular attention is paid to the roles of Cameroonian Pidgin and Camfranglais as indexically dense resources through which speakers negotiate belonging, racialization, and integration within postcolonial and contemporary European contexts.

The analysis draws on approximately 22 hours of semi-structured sociolinguistic interviews conducted across the three national contexts. The interviews were orthographically transcribed and annotated in ELAN, allowing for fine-grained analysis of lexical choice, code-switching, and metalinguistic commentary.

Grounded in scholarship on racial and language ideologies (Guillaumin 1995; Gal & Irvine 2019) and theories of the subject as discursively and ideologically constructed (Lacan 1975; Althusser 1976; Heath 1981), the study explores how speakers negotiate identity under conditions of linguistic discrimination, racialization, and foreignization. In sociolinguistic environments where the internalization of racial hierarchies is normalized and linguistic and behavioral alignments with whiteness are socially rewarded (Frankenberg 1994; Guillaumin 1995; El-Enany 2020), diasporic subjects experience intensified identity tensions (Hansen-Thomas 2007; Lukate 2024). These tensions often give rise to integration-oriented linguistic strategies that may involve forms of identity negation, self-hatred, or alienation (Fanon 1952; Memmi 1979; hooks 1992).

Building on previous studies (Telep 2017; Fonkeu 2019; Lukate 2025), the findings show that first-generation Cameroonian immigrants deploy linguistic practices such as *whitisage* (i.e., speaking like a white person), lexical accommodation, and prosodic variation to facilitate sociolinguistic integration and develop an additional identity layer linked to the host nation. Simultaneously, they actively reclaim Cameroonian identity through language maintenance and heritage transmission. Pidgin emerges as an identity marker on numerous levels – regional, national, diasporic – particularly in the German and British contexts, while Camfranglais functions as an alternative informal register indexing modern urban Afrodiasporic identity in the French diaspora.

The paper argues that these competing orientations toward integration and ethnolinguistic maintenance generate complex identity dynamics, leading to the emergence of racialized labels and metalinguistic lexical innovations. While some of these practices are shared with other postcolonial diasporic contexts (Heyd 2015; Telep 2017; Fonkeu 2019; Ajagbe 2023), others are specific to the Cameroonian case. The study thus contributes to debates on multilingualism, racialization, and identity formation in contemporary Europe.

References

Ajagbe, Samsondeen. 2023. Afro-diasporan racial discourse in Germany. *Discourse and Society* 34(6), 673–690.

Althusser, Louis. 1976. Idéologie et appareils idéologiques d'état. *Positions*. Editions Sociales, Paris.

Fanon, Frantz. 1952. *Peau noire, masques blancs*. Paris : Les Éditions du Seuil.

Fonkeu, Bridget. 2019. *Sociolinguistic identity (re)construction in the German diaspora: The case of the use of address forms and kinship terms by Anglophone Cameroonian immigrants*. PhD thesis, TU Dortmund.

El-Enany, Nadine. 2020. *(B)ordering Britain: Law, race and empire*. Manchester: Manchester University Press.

Frankenberg, Ruth. 1994. *White women, race matters: the social construction of whiteness*.

University of Minnesota Press, Minneapolis.

Guillaumin, Colette. 1995. *Racism, sexism, power, and ideology*. Routledge. New York.

Gal, Susan, and Judith T Irvine. 2019. *Signs of difference: Language and ideology in social life*. Cambridge University Press.

Hansen-Thomas, Holly. 2007. Language ideology, citizenship, and identity: The case of modern Germany. *Journal of Language and Politics* 6(2). 249–264.

Heath, Stephen. 1981. *Questions of Cinema*. Basingstoke: Macmillan.

Heyd, Theresa. 2015. The metacommunicative lexicon of Nigerian Pidgin. *World Englishes* 34(4), 669-687.

hooks, Bell. 1992. *Black looks: race and representation*. Turnaround, London Lacan, Jacques. 1975. *Le séminaire livre XX*. Paris : Seuil.

Lukate, Johanna M. 2025. Becoming Black? African Migrants' Social Identity Constructions and Experiences of Racialization in Germany. *Identity* 25(1), 134–152.

Memmi, Albert. 1979. *Portrait du colonisé*. Payot. Paris.

Telep, Suzie. 2017. Pour une approche pluridisciplinaire du langage. Le cas du whitisage chez des migrants camerounais à Paris. *EspacesTemps.net*.

Comparing two South Asian contact languages: Korlai Indo-Portuguese and Nagamese

J. Clancy Clements (Indiana University)

Korlai Indo-Portuguese (KP) and Nagamese (NG) are contact languages that formed and developed endogenously on the Indian west coast and in the northeast of India, respectively. The paper seeks to offer an answer to the question as to why KP and NG retain suffixal verbal morphology from their respective superstrate languages (Portuguese and Assamese, respectively). It is argued that the presence of verbal morphology in these contact languages is not unexpected given that in both cases both substrate and superstrate languages possess rich suffixal morphology in their respective verbal systems.

In both the KP and NG cases, social and reproductive isolation was a significant factor in the creation of the contact languages in question. In addition to these factors, several linguistic factors also played a role in the formation of KP and NG. For KP, there were only two languages involved in the contact situation, Marathi (Indo-Aryan) and Portuguese (Romance). Both languages are of Indo-European origin and have rich suffixal inflectional verbal morphology. Thus, Marathi speakers targeting Portuguese would be sensitive to the ends of words they could parse. It is not surprising that superstrate suffixal morphology is retained in KP. Moreover, KP has also maintained allomorphy from the three verb classes (Clements 1996). Similarly, when the Assamese-speaking traders moved into and settled in Nagaland, there were already speakers of Tibeto-Burman ethnolinguistic groups residing in the area who generally avoided contact with one another, but interacted with the Assamese for purposes of trade. These Tibeto-Burman languages not only had suffixal verb morphology, but Subject-Object-Verb (SOV) order and postpositions, as well, just like Assamese, the language they were targeting (Bhattacharjya 2001, 2007). It is not surprising, then, that NG as an expanded pidgin and a creole would also possess suffixal verb morphology, postpositions, and SOV order. Thus, the Naga-language speakers who targeted Assamese were arguably sensitive to endings of Assamese verbs because in their native languages word endings were also important.

There are fairly straightforward reasons why in both cases speakers were able to incorporate verbal morphology into their developing grammars, namely, that the typological fit among the languages in both cases was good: in each contact situation the languages in contact had rich suffixal verb morphology. Thus, in learning the superstrate language, Marathi speakers on the one hand, and Naga-language speakers on the other, would be attentive to the alterations at the ends of words because of their own sensitivity to word endings in their respective native languages. In addition, in each of the cases, the substrate and superstrate languages existed in the same space for centuries Thus, it is argued that the outcome is not unexpected that speakers who targeted Portuguese and Assamese were able to detect and reproduce target-language inflected forms and incorporate them into the grammar of the new languages they were constructing.

References

Bhattacharjya, Dwijen. 2007. Nagamese (restructured Assamese), in *Comparative creole syntax: parallel outlines of 18 creole grammars*, edited by John Holm & Peter Patrick. London: Battlebridge Press, 237-254.

Bhattachariya, Dwijen. 2001. *The genesis and development of Nagamese:its social history and linguistic structure*. Unpublished dissertation. New York: City University of New York.

Clements, J. Clancy. 1996. *The genesis of a language: the formation and development of Korlai Portuguese.* Amsterdam: Benjamins.

Discourses on environmental and linguistic change in a coastal creole-speaking community

Jillian Loise Melchor (Helsingin yliopisto), Danae Perez (independent scholar) & Eeva Sippola (Helsingin yliopisto)

Social and geographic mobility can potentially impact the status of minoritised varieties when speakers increasingly abandon their language in contexts of social fragmentation and weakening speaker networks (UNESCO 2010). This dynamic is particularly pronounced in communities with long histories of trade and contact, such as the Cavite Chabacano community, whose Spanish-lexified creole emerged in Cavite City, approximately 30 kilometres southwest of Manila (Lesho & Sippola 2025). The city's proximity to the Philippine capital has been a key driver of hyper-urbanisation, accelerating demographic change, straining local infrastructure, and expanding the influence of Filipino, the nation's primary lingua franca, alongside English as the dominant language of education, administration, and wider communication (Lesho & Sippola 2013).

Beyond the social impact of its geography, Cavite City also contends with mounting environmental challenges due to its natural topography. Situated on a narrow, low-lying tip of a hook-shaped peninsula, this coastal community faces the constant threat of sea-level rise and its attendant risks, such as flooding, storm surges, and land subsidence (Sales 2009). This study explores the relevance of environmental factors in speaker perceptions of language shift to understand the extent to which language change and environmental change are intertwined (Maffi 2018). In our analysis of sociolinguistic interviews, ethnographic observations, and historical documentation, we ask whether narratives of language decline parallel those surrounding ecological vulnerability and landscape transformation. Within this context, language change linked to demographic movement may be understood not just in socioeconomic but also in ecological terms.

Preliminary results highlight recurring narratives about environmental and social change in the area in connection to language loss. First, flooding was described as a common problem, reflecting concerns about worsening environmental conditions against the backdrop of inadequate infrastructure. However, floods were rarely seen as disastrous in Cavite City as locals believed in the protection accorded by the local patron saint. Second, participants who had worked as fishermen mentioned the declining diversity of fish, linking it to pollution, ecological degradation, and the loss of traditional livelihoods. Third, issues such as trash accumulation and the presence of informal settlers were framed as signs of broader neglect, prompting discussions about social inequality and legitimacy by contrasting the present with the "cleaner" past when Chabacano was still spoken in the city. By analysing these intersecting discourses, the study contributes to broader discussions of how environmental precarity, urbanisation, and mobility shape linguistic practices and community perceptions of loss and change.

References

Lesho, M., & Sippola, E. (2013). The sociolinguistic situations of Manila Bay Chabacano-speaking communities. *Language documentation and conservation*, *7*, 1-30.

Lesho, M., & Sippola, E. (2025). Creole identity, transnational migration, and language endangerment in a Philippine context. *Journal of Pidgin and Creole Languages*. https://doi.org/10.1075/jpcl.24029.les

Maffi, L. (2018). Biocultural diversity. In H. Callan (Ed.), *The International Encyclopedia of Anthropology* (pp. 1–14). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118924396.wbiea1797

Sales Jr., R. F. M. (2009). Vulnerability and adaptation of coastal communities to climate variability and sea-level rise: Their implications for integrated coastal management in Cavite City, Philippines. *Ocean & Coastal Management*, *52*(7), 395-404.

UNESCO. (2010). *Atlas of the world's languages in danger* (C. Moseley, Ed.). UNESCO Publishing.

O expletivo *ele* no Português Brasileiro quilombola: indícios de mudança paramétrica

Joachim Steffen (Universität Augsburg)

Desde o século XIX, o Português Brasileiro (PB) vem passando por transformações estruturais significativas no domínio da expressão do sujeito, que o diferenciam progressivamente das demais línguas românicas de sujeito nulo e o aproximam de línguas de sujeito obrigatório. Entre essas transformações destacam-se a redução de sujeitos nulos referenciais e a restrição da inversão verbo–sujeito (Duarte 1995; Kato 1999). Esses fenômenos indicam uma transição do PB de um sistema [+NS] para um sistema parcial ou híbrido. Conforme demonstram estudos diacrônicos, os sujeitos de 1ª, 2ª e 3ª pessoas eram frequentemente nulos na escrita e na fala reconstruída do século XIX (Duarte 1993). No entanto, ao longo do século XX, o preenchimento do sujeito alcançou índices superiores a 70% (Duarte 1995), apontando para uma reanálise paramétrica.

Apesar dessa tendência, o PB manteve o expletivo nulo em construções impessoais, diferindo de línguas como o francês e o espanhol dominicano, que utilizam expletivos lexicais (Kato 2015). O presente trabalho investiga se o PB contemporâneo estaria desenvolvendo expletivos lexicais, a partir de dados coletados em 2019 na comunidade quilombola Baixio (Mato Grosso). O *corpus* revela ocorrências de *ele* como sujeito expletivo em contextos sem referente discursivo:

(1) Ele tinha escola sim
3SG ter-IMPF.3SG escola-SG sim
'Tinha escola sim'

(2) Ele te-m aí [no mato] bichinhos
3SG ter-PRES.3SG aí [no mato] bicho.DIM-PL
'Tem bichinhos aí [no mato]'

Esses dados sugerem um preenchimento sintático de natureza expletiva emergente nesta variedade rural afro-brasileira. Tal evidência é interpretada à luz da literatura sobre variação paramétrica e contato linguístico (Almeida Toribio 2000; Lucchesi 2009). Discute-se se esse fenômeno decorre de uma transmissão linguística irregular ou de processos de semicrioulização, em que o contato intenso entre o português e línguas africanas teria acelerado a fixação do parâmetro do sujeito preenchido. Assim, propõe-se que o uso de *ele* em Baixio represente um estágio avançado da mudança iniciada no século XIX. O estudo contribui para a pesquisa linguística ao demonstrar como variedades historicamente isoladas preservam traços inovadores que radicalizam tendências gerais do PB, reforçando o papel do contato na história interna da língua.

Referências

Almeida Toribio, Jacqueline. *Setting parametric limits on dialectal variation in Spanish.* *Lingua*, v. 110, n. 5, p. 315–341, 2000.

Bazengs, Aline Maria. *Subject expression in spoken varieties of portuguese. Diadorim*, v. 24, n. 1, 2022.

Duarte, Maria Eugênia Lamoglia. *A perda do princípio "evite pronome" no português brasileiro.* 1995. Tese (Doutorado em Linguística) – Universidade Estadual de Campinas, Campinas.

Duarte, Maria Eugênia Lamoglia. *Do pronome nulo ao pronome pleno: a trajetória do sujeito no português do Brasil.* In: Roberts, Ian; Kato, Mary A. (orgs.). *Português Brasileiro: Uma viagem diacrônica.* Campinas: Editora da Unicamp, 1993. p. 107–128.

Kato, Mary A. *Strong and Weak Pronominals and the Null Subject Parameter. Probus*, v. 11, n. 1, p. 1–31, 1999.

Kato, Mary A. *Expletivos nulos e construções de tópico/sujeito no português brasileiro. Cadernos de Estudos Linguísticos*, v. 57, n. 1, Campinas, jan./jun. 2015.

Lucchesi, Dante. *A realização do sujeito pronominal.* In: Lucchesi, Dante; Baxter, Alan; Ribeiro, Ilza (orgs.). *O Português Afro-Brasileiro.* Salvador: UFBA, 2009. p. 167–183.

Podem as atitudes linguísticas impulsionar a construção do bilinguismo ativo em São Tomé e Príncipe? Contributos a partir da experiência de Cabo Verde

Jorge Pinto () & Rita Gonçalves (Universidade de Lisboa)

Em São Tomé e Príncipe (STP), o português, língua oficial, encontra-se historicamente em contacto com o santome (crioulo dominante), o angolar e o lung'ie, além da língua cabo-verdiana (LCV). Apenas o português é ensinado formalmente no país. Nesta apresentação, iremos focar-nos apenas na interação entre o português e o santome, falados, respetivamente, por 98,4% e 36,2% da população (INE 2012).

Vários estudos têm discutido a perda progressiva do bilinguismo em STP (Hagemeijer, 2016, 2018, Bouchard, 2017, 2019), a nativização do português (Gonçalves, 2016) e o estatuto das línguas nacionais como línguas de herança (Hagemeijer, 2025). De facto, o bilinguismo manifesta-se de diversas formas numa comunidade: pode ocorrer a nível individual ou social e envolver competências distintas, desde o domínio ativo ao conhecimento passivo das línguas (e.g., Baker, 2001; Sebba, 2011). Atualmente, é consensual que o português é, na generalidade, a primeira língua adquirida em STP, sendo também amplamente reconhecido que as atitudes linguísticas desempenham um papel decisivo na manutenção ou erosão linguística. É, neste contexto, que o caso de Cabo Verde se revela particularmente relevante: a valorização social e simbólica da LCV, aliada a debates sobre a sua oficialização e presença no espaço público, oferece um quadro comparativo pertinente para refletir sobre os caminhos possíveis para a promoção de um bilinguismo ativo em STP.

Esta apresentação tem como objetivos: (i) discutir o bilinguismo em STP a nível social, correlacionando as atitudes linguísticas com variáveis sociolinguísticas, como a idade, o nível de escolaridade e a profissão; (ii) comparar os dados obtidos com os resultantes do estudo de Lopes (2011) para Cabo Verde, no qual a autora faz um mapeamento sociolinguístico do uso das duas línguas faladas no arquipélago – português e LCV.

No que respeita a STP, o estudo baseia-se em 77 entrevistas do corpus PALMA (Hagemeijer et al., 2022), recolhidas em São Tomé (2008–2012). Realizámos uma análise de conteúdo qualitativa, considerando: (i) juízos de valor sobre as línguas; (ii) contextos de uso; e (iii) revitalização do santome. Para o contexto de Cabo Verde, considerámos os resultados da análise de Lopes (2011) a 29 entrevistas, à luz das três categorias suprarreferidas.

A análise global dos dados de STP sugere um elevado valor atribuído ao português comparativamente ao santome (1), embora este esteja a ser (re)utilizado pelos jovens e/ou em *codeswitching* com o português, sendo considerada marca identitária nacional (2a). Não obstante, alguns falantes têm dela apenas conhecimento passivo (2b). Os dados remetem ainda para o santome como língua ameaçada e para o reconhecimento da aprendizagem formal da língua como determinante para a sua revitalização (3).

Quanto ao contexto cabo-verdiano, de acordo com Lopes (2011), o português é

valorizado enquanto língua de acesso ao conhecimento e à informação (4a), de comunicação internacional (4b), bem como língua predominante nos domínios formais (5a). A LCV predomina nos domínios íntimos e informais (5b), mas há uma ideia generalizada de que pode coabitar com o português em todas as funções sociais (escrita, administração e eventos oficiais), havendo um apoio, ainda que não consensual, quanto ao seu ensino na escola (6).

A comparação entre os dois contextos sugere, assim, que a passagem de um bilinguismo predominantemente simbólico/passivo para um bilinguismo efetivamente ativo depende não só das atitudes linguísticas dos falantes, mas da implementação de políticas educativas consistentes e sustentadas. Em Cabo Verde, a introdução, no ano letivo 2022–2023, da disciplina *Língua Cabo-Verdiana* no 10.º ano (opcional e experimental) ilustra como o reconhecimento institucional pode promover o desenvolvimento do conhecimento metalinguístico e contribuir para a valorização e diversidade linguística.

Exemplos:

(1) a. *Sempre houve muito preconceito em torno dessa questão de língua, não é? Durante muito tempo houve até famílias que proibiam seus filhos a falar forro.*" (STP-057-2008)

b. *Em muitas partes de São Tomé e Príncipe onde se vai, há lá pessoas que insistem em falar o crioulo forro ou mesmo outros crioulos (...) sobretudo quando chegam estrangeiros*. (STP-021-2008)

(2) a. *Não posso dar ao luxo de dizer que falo o crioulo forro com toda a fluidez, porque eu também tenho dificuldades, mesmo na escrita, (...) mas é verdade que compreendo o crioulo forro.* (STP-021-2008)

b. *Infelizmente, nós estamos muito atrasados com respeito a algum tratamento, alguma preocupação (...) com a revitalização da nossa língua, a sua conservação.* (STP-094-2008)

(3) *Dialeto é nosso direito, é nossa língua. Gente tem que aprender também a profundar e saber muito mais de dialeto*. (STP-002-2011)

(4) a. (...) *eu acredito que essencialmente que ele vai optar por usar o português como… para ler – para buscar informação - para estudar (...)* (Lopes, 2011, INF11)

b. *(...) a língua oficial é o português -- e não interessa ao país deixar de ter uma língua oficial com expressão internacional - o caso português (...)* (Lopes, 2011, INF03)

(5) a. *(...) quando se passa para uma situação mais formal - passa-se para o português - porque nós ligamos também o português à ideia de norma - à ideia de regra - à ideia de rigor (...)* (Lopes, 2011, INF13)

b. *(...) com a minha família utilizo normalmente o crioulo.* (Lopes, 2011, INF01)

(6) *(...) estou de acordo que o crioulo deve ser oficializado, que deve ser ensinado na escola (...)*
(Lopes, 2011, INF04)

Referências

Baker, C. (2001). *Foundations of bilingual education and bilingualism*. Multilingual Matters.

Bouchard, M-E. (2017). *Linguistic variation and change in the Portuguese of São Tomé.* [Tese de Doutoramento, New York University].

Bouchard M-E. (2019). Becoming Monolingual: The Impact of Language Ideologies on the Loss of Multilingualism on São Tomé Island. *Languages*, 4(3): 50. https://doi.org/10.3390/languages4030050

Gonçalves, R. (2016). *Construções ditransitivas no português de São Tomé* [Tese de Doutoramento, Universidade de Lisboa]. http://hdl.handle.net/10451/26409

Hagemeijer, T. et al. (2022). The PALMA corpora of African varieties of Portuguese. In N. Calzolari et al. (eds) *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, 5047-5053. European Language Resources Association.

Hagemeijer, T. (2016). O português em contacto em África. In A. M. Martins & E. Carrilho (Eds.), *Manual of Linguística Portuguesa* (pp. 43-67). Mouton de Gruyter.

Hagemeijer, T. (2018). From creoles to Portuguese: Language shift in São Tomé and Príncipe. In L. Álvarez López, P. Gonçalves, & J. Avelar (eds) *The Portuguese Language Continuum in Africa and Brazil* (pp. 169–184). John Benjamins. https://doi.org/10.1075/ihll.20.08hag

Hagemeijer, T. (2025). As línguas nativas de São Tomé e Príncipe como línguas de herança. *Revista Sêbê Non Linguagens*, 250-275.

INE (Instituto Nacional de Estatística). (2014). *III Recenseamento Geral da População e da Habitação 2012. Características educacionais da população.* INE.

Lopes, A. (2011). As Línguas de Cabo Verde: uma radiografia sociolinguística [Tese de Doutoramento, Universidade de Lisboa].

Sebba, M. (2011). Societal bilingualism. In *The SAGE Handbook of Sociolinguistics* (pp. 445-459). SAGE Publications Ltd, https://doi.org/10.4135/9781446200957

Continental West African Portuguese Creole areas:
when Glossonym meets autoglossomym

Joseph Jean François Nunez (Institut National des Langues et Civilisations Orientales, Paris)

Casamance (Southern Senegal) and Guinea-Bissau are multicultural and multilingual areas where Creoles languages are traditionally spoken (see Kohl 2018, Nunez 2015, Truppi 2021). This situation makes these continental African creoles special compared to the other Upper Guinea Creoles (UGC). These Creoles are local languages and do not have the status of official languages in each of the areas where they are traditionally spoken. In Guinea-Bissao, the Creole language is the lingua franca and represents the language of national unification (Truppi 2021, Kohl 2018). In Casamance, on the other hand, Creole played for a long time the role of lingua franca and unifying language in the local context. Nowadays it continues to play it but to lesser degrees in certain contexts.

In academic field, concepts are used both to designate the two Creole spoken in Guinea-Bissao (see Truppi 2021; Ie 2022, Scantamburlo 2018) and in Casamance (see Nunez 2022, Biagui, Nunez & Quint 2024). The Creole speakers of Casamance and those of Guinea-Bissau call their Creoles by the same autoglossonym *kriyol* (with alternatives pronunciation and writing) or even *liŋgu kristoŋ*, Christian language (Rougé, 1985; Kihm 1994; Nunez 2015 and Biagui 2018). However, the terms used by speakers of the other linguistic groups of the area to designate both the Creole language and speakers for whom Creole is the first language of socialization (see Nunez 2015 for the different groups who speak Casamance Creole) remain unknown as well for Casamance Creole and for Guinea-Bissau Creole.

The question of language names is generally a described subject (see Tabouret-Keller 1997; De Feral 2006) but there is little research in Creole languages (see Holm and Madeira, 2009; Léglise & Migge, 2006; Sippola, Selbach & Krämer, forthcoming). The aim of this Study in progress is, on the one hand, to describe the different way of naming practices linked to the both kriyol of Casamance and Guinea-Bissau from both an emic (insider) and etic (outsider) point of view. It also involves comparing terminologies used in academic world with those used by speakers to designate themselves or their languages. On the other hand, it necessary determine how the speakers of other local languages designate the speakers of these Creoles and show to what extent certain of these terms can be linked to ideology. This study is based on the methodology proposed by Léglise & Migge (2006) and constitutes a contribution to the description of the Creoles of Casamance and Guinea-Bissao in order to understand better Creole studies in general.

References

Biagui, N.B. 2018. *Description générale du créole afro-portugais de Ziguinchor (Sénégal)*. Paris: Karthala.

Biagui, N. B., Nunez,J.J.F. & Quint, N. 2024."Casamance Creole" In Fiederik Lüpke (Ed.). The Oxford Guide to the Atlantic Languages of West Africa. Oxford University Press : 131-152

De Feral, C. 2009. "introduction". In C. de Féral (dir.). Le nom des langues Ill. Le nom des langues en Afrique sub-saharienne pratiques, dénominations, catégorisations. Naming Languages in Sub-Saharan Africa: Practices, Names, Categorisations. Louvain-la-Neuve. Peeters. BCILL 124.

Ié, D. 2022. *Política linguística em língua guineense: caso implícito*. Universidade Federal de Paraná: MA dissertation.

Kihm, A. 1994. *Kriyol Syntax. The portuguese-base creole of Guinea-Bissau*. Amsterdam / Philadelphia: John Benjamins.

Kohl, C. 2018. *A creole nation. National integration in Guinea-Bissau*. Berghahn.

Léglise, I. & Migge B. 2006. "Language-naming practices, ideologies, and linguistic practices: Toward a comprehensive description of language varieties". Language in Society 35:3, pp. 313-339.

Nunez, J. J. F. 2015. *L'alternance entre créole afro-portugais de Casamance, français et wolof au Sénégal : une contribution trilingue à l'étude du contact de langues*. PHD, Paris / Dakar, INALCO / UCAD.

Nunez, J.J.F. 2022. « L'expression de la comparaison de supériorité en créole de Casamance ». Études créoles, 37, V. 1-2.

Rougé, J-L. 1985. *Formation et évolution du lexique du créole portugais de Guinée Bissau et de Casamance*. PHD, Lyon: Université Lyon 2.

Scantamburlo, L. 2018. *Dicionário do guineense: dicionário guineense-português*. vol. 2. Lisbon: Colibri.

Tabouret-Keller, A. 1997. *Le nom des langues I – Les enjeux de la nomination des langues*. Peters.

Truppi, C. 2021. "Kriol (guineense), línguas nacionais e multilinguismo: a grande riqueza linguística da Guiné-Bissau". In Museu virtual da arte e cultura da Guiné-Bissau.

The Arlene hypothesis:
Progressive marker variation in a northwestern dialect of Jamaican

Joseph T. Farquharson & Hubert Devonish (The University of the West Indies, Mona)

The extant literature on aspect marking in Jamaican generally assumes a neat geographical split between two forms used to mark the progressive: *de* used in northwestern dialects, and *a* (< *da*) used elsewhere. However, there is sufficient evidence to question the accuracy of this bifurcation given that many speakers from northwestern Jamaica (especially in the parish of Saint James) actually use both forms of the progressive. This variation in form raises questions such as whether the simultaneous use of *a* and *de* in this western dialect is a recent phenomenon, whether it represents the growing/hegemonic influence of the dialect of Kingston, the capital city, on this language variety, and whether the two forms exist in free variation or their use is conditioned by other factors. The paper conducts a historical exploration of development and change in the form and usage of the progressive markers, showing that the *a*-progressive has been supplanting the *de*-progressive in western dialects for several generations to the extent that the *de*-progressive has virtually disappeared in some constructions, for example, negative declarative progressive constructions (*Mi naa nyam* vs *Mi no de nyam* 'I am not eating.'). The latter still persists in varieties that are historically related to Jamaican (such as San Andrés Creole), but has been replaced by the former in northwestern varieties of Jamaican.

Although on the surface the alternation between the *a*-progressive and the *de*-progressive may appear to be in free variation, we reject this framing on the assumption that variation in form signals a difference in usage, no matter how small. Support for our hunch came from the 1980 dancehall song "Arlene" by General Echo in which both progressive forms are used but each is apparently associated with a different type of verb. On the basis of this preliminary evidence we propose The Arlene Hypothesis that says that for the northwestern Jamaican dialect under study, where variation exists in progressive marker form, the form used is determined by the type of verb. To test the hypothesis, we have assembled a small corpus (currently at approximately 20,000 words) comprising interviews by 5 young male dancehall artistes from the parish of Saint James. Our preliminary analysis demonstrates that the alternation between *a*-progressive and the *de*-progressive is not random, and suggests that the *de*-progressive is chiefly used with verbs that refer to some type of psychological state.

Most studies on the progressive in Creole languages focus on the description of this aspect of the TMA system, very often with a view to identifying substrate models. As far as we are aware, this is the first study that looks at the phenomenon from the perspective of dialectal variation and change.

Mecanismos de distinção referencial na fala secreta do Cafundó

Juanito Ornelas de Avelar (University of Campinas)

Neste trabalho, analiso estratégias de referenciação em terceira pessoa na fala secreta da comunidade quilombola do Cafundó (Salto de Pirapora, São Paulo, Brasil), conhecida como cupópia.

A comunidade do Cafundó começou a ser estudada nos anos 70 do século passado e chamou a atenção por abrigar uma suposta língua africana, que teria sido preservada desde a escravidão. O material gravado por pesquisadores da Universidade Estadual de Campinas, com cerca de 45 horas, revelou que a cupópia era realmente formada por vocábulos de origem africana, mas tinha o português como base gramatical (Vogt & Fry 1996). Dessa perspectiva, a cupópia pode ser entendida como uma "variedade" do português brasileiro rural, apesar de suas origens remontarem a uma possível língua franca com elementos do quimbundo, do quicongo e do umbundo (Slenes 1992).

Para este trabalho, foram analisadas duas situações de conversa em cupópia, totalizando cerca de 40 minutos. A amostra revelou quatro tipos de distinção referencial, evidenciados por itens vocabulares específicos: (i) **por gênero**: *tata* 'ele, homem' *vs*. *angutu* 'ela, mulher'; (ii) **etária**: *camanaco/angutuzinha* 'menino/menina' *vs*. *tata/angutu* 'adulto homem/mulher' *vs*. *jocorocoto* 'velho(a)'; (iii) **étnico-racial**: *vimbundo* 'preto' *vs*. *cafombe* 'branco'; (iv) **por definitude**: *tata/angutu* 'ele/ela' (definido) vs. *nhemanhara* 'alguém' (indefinido). O enunciado *Tata cuendou vavuro no ingombe do andaru* 'Ele andou bastante de carro', por exemplo, mostra o uso de *tata* para se referir a uma terceira pessoa contextualmente definida por meio da distinção referencial por gênero, enquanto em *Vem cuendando nhemanhara* 'Vem vindo alguém', *nhemanhara* é de referência indefinida.

Diferentes mecanismos de referenciação podem aparecer combinados, como em *Tata cafombe picopiou com tata vimbundo* 'O homem preto conversou com o homem branco', em que a distinção por gênero se combina com a étnico-racial, e *Tata jocoroto vimbundo curima vavuro* 'O homem preto velho trabalha muito', com a étnico-racial se combinando com a etária. Em tais enunciados, *tata cafombe*, *tata vimbundo* e *tata jocorocoto vimbundo* dizem respeito a indivíduos discursivamente bem demarcados na conversação que poderiam, em português, ser referidos por *ele*.

Adicionalmente a essas distinções referenciais, é recorrente na cupópia o emprego de sintagmas preposicionados, formados geralmente pela indicação de uma característica do indivíduo referenciado ou de um elemento a ele relacionado. Vemos esse mecanismo em ocorrências como *Tata cafombe picopiou com o tata do injó da anguara* 'O homem branco conversou com o homem do armazém' e *Tata cafombe no mussuruco do pelotão vavuro coçumba a picópia* 'O homem branco com muito cabelo na cabeça entende a conversa', nos quais os termos *do injó da anguara* "armazém" e *no mussuruco do pelotão vavuro* "com muito cabelo na cabeça" são um reforço

referencial adicional às distinções étnico-racial e por gênero.

Os dados indicam que o uso das formas em questão em vez de pronomes portugueses contribui para complexificar a interlocução, dificultando o seu entendimento por quem não pertence à comunidade. Os mecanismos de referenciação da cupópia se constituem, assim, como estratégias léxico-gramaticais e semântico-discursivas que ampliam o caráter secreto da fala do Cafundó.

Variação da Concordância Verbal e Construção Identitária no Recôncavo Afro-Brasileiro

Juliete Bastos Macêdo (Universidade Estadual De Feira)

Este trabalho investiga a variação na concordância verbal de terceira pessoa do plural em comunidades quilombolas do Recôncavo baiano, argumentando que alternâncias como *eles fala ~ eles falam* constituem não apenas variabilidade morfossintática, mas práticas estilísticas socialmente indexicais. Ancorado na Sociolinguística Variacionista (Labov, 1966, 1972, 2001) e nos desdobramentos da terceira onda dos estudos da variação (Eckert, 2000, 2012), o estudo dialoga com a perspectiva sociocultural da identidade proposta por Bucholtz & Hall (2005), segundo a qual identidades são produzidas interacionalmente por meio de processos de indexicalidade, posicionamento e relacionalidade. No que se refere à história da concordância verbal no português brasileiro, a análise considera as contribuições de Naro e Scherre (2007), que demonstram o caráter sistemático da variação e, em Macêdo e Araújo (2023), para a compreensão da persistência histórica de padrões variáveis na morfossintaxe do PB, afastando interpretações deficitárias. Parte-se, assim, da hipótese de que a ausência de marcação explícita de plural no verbo, historicamente associada ao estigma do "português popular", funciona simultaneamente como variável socialmente avaliada e como índice performativo de pertencimento local em territórios afrodescendentes marcados por heranças coloniais de racialização e desigualdade. Em consonância com debates contemporâneos sobre linguagem e pós-colonialidade, sustenta-se que a avaliação negativa dessas formas não pode ser dissociada das hierarquias raciais, educacionais e simbólicas produzidas no contexto da formação social brasileira. Nesse sentido, mobiliza-se a noção de capital linguístico (Bourdieu, 2008) para compreender como determinadas variantes adquirem maior legitimidade institucional, enquanto outras são desvalorizadas, ainda que estruturalmente sistemáticas. A variação é, portanto, interpretada não como déficit estrutural, mas como padrão regular inserido em ecologias sociolinguísticas específicas, nas quais normas escolares, ideologias linguísticas e valores comunitários coexistem e se tensionam. Os dados provêm de entrevistas sociolinguísticas realizadas com participantes de diferentes faixas etárias e níveis de escolaridade. A análise multivariada identifica os condicionamentos linguísticos e sociais da variável. Resultados provisórios indicam que a não marcação de plural não se distribui aleatoriamente, mas associa-se a perfis sociolinguísticos e a contextos interacionais distintos. Observa-se que falantes mais jovens e com maior escolarização demonstram maior sensibilidade ao monitoramento estilístico, alternando formas conforme a situação comunicativa e os regimes de autoridade simbólica em jogo. Os achados contribuem para debates internacionais sobre variação, desigualdade e agência linguística em contextos pós-coloniais, demonstrando que a concordância verbal participa ativamente da negociação de pertencimento, legitimidade e capital simbólico. A análise reforça um modelo de prática estilística no qual normas coletivas e projetos identitários individuais se articulam dinamicamente, evidenciando que a variação é ação social situada e historicamente enraizada.

Referências

Bourdieu, Pierre. *O que falar quer dizer: a economia das trocas linguísticas*. São Paulo: Edusp, 2008.

Bucholtz, Mary; Hall, Kira. Identity and interaction: a sociocultural linguistic approach. *Discourse Studies*, v. 7, n. 4–5, p. 585–614, 2005.

Eckert, Penelope. *Linguistic variation as social practice: the linguistic construction of identity in Belten High*. Oxford: Blackwell, 2000.

Eckert, Penelope. Three waves of variation study: the emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, v. 41, p. 87–100, 2012.

Labov, William. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics, 1966.

Labov, William. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press, 1972.

Labov, William. *Principles of linguistic change. Volume 2: Social factors*. Oxford: Blackwell, 2001.

Macêdo, Juliete Bastos; Araújo, Silvana Silva de Farias. A concordância verbal no português brasileiro: uma abordagem sócio-histórica em estudos variacionistas. In: Gonçalves, Eliana Correia Brandão; pereira, Norma Suely da Silva; Almeida, Isabela Santos de; Carmo, Tereza Pereira. (Org.). *História e funcionamento das línguas naturais:* raízes e sementes da pesquisa acadêmica em Letras. Salvador: EDUFBA, 2023, p. 255-270.

Naro, Anthony Julius; Scherre, Maria Marta Pereira. *Origens do português brasileiro*. São Paulo: Parábola Editorial, 2007.

Melchior Schmidt and the linguistic situation on the west side of St. Croix, 1770s–1780s

Kristoffer Friis Bøegh (Universiteit Utrecht) & John Balz (University of Wisconsin-Madison)

For decades, scholars of Carriols, or Virgin Islands Dutch Creole, have presumed it was spoken on Danish St. Croix during the 18th century based on a metalinguistic set of observations by the Moravian missionary and historian C.G.A. Oldendorp (2000–2002). Writing around 1770 about the linguistic situation on St. Croix, Oldendorp (2000: 357–358), however, also gave the following assessment: "The English [planters] mostly retain only their own language and do not learn Creole [i.e. Carriols], and on their plantations the slaves must adjust to them in this respect" (our translation). He further remarks (e.g. 2000: 862) on the prevalence of English – which should be understood as referring to English but especially to early Virgin Islands English Creole – specifically on the west side of St. Croix. At the same time, he adds that Carriols was in use as well (e.g. 2000: 608). Few of the missionaries active in the Danish West Indies at the time knew English, and in communication with members of the enslaved population, they generally used Carriols. Knowledge about the language situation on the west side of St. Croix in the 18th century, including the distribution, levels of interaction, and cross-influences between Carriols and English Creole, has until now been based almost exclusively on Oldendorp's observations, rather than on direct linguistic evidence.

Our presentation will focus on an almost unknown missionary, Melchior Schmidt (1722–1784), whose writings in Moravian mission station diaries shed new light on St. Croix's linguistic situation. Schmidt was of German-speaking background but had also lived many years in Pennsylvania before he traveled to the Danish West Indies in 1765 and in 1771 was tasked with starting a church on the less densely settled west side of St. Croix. He led that congregation for 13 years and died there in 1784 at age 63. Given Schmidt's relative isolation on the west side of St. Croix and his assigned task to build a nascent congregation from scratch, he probably spent considerable time trying to learn Carriols through self-study and social immersion, while at the same time being in contact with English Creole speakers.

Schmidt's station diary presents insights into the linguistic situation on the west side of St. Croix in the late 18th century, and features, among other things, samples of Carriols he recorded there. We explore Carriols as documented by Schmidt and compare his writings with other Moravian authors active around the same time and with the Carriols corpus as a whole. Among other things, we look at whether there was more English/English Creole influence than is reflected in other contemporary sources.

Reference

Oldendorp, Christian Georg Andreas. 2000–2002. *Historie der caribischen Inseln Sanct Thomas, Sanct Crux und Sanct Jan insbesondere der dasigen Neger und der Mission der evangelischen Brüder unter denselben. Kommentierte Ausgabe des vollständigen Manuskriptes aus dem Archiv der Evangelischen Brüder-Unität Herrnhut.* Edited by Hartmut Beck, Gudrun Meier, Stephan Palmié, Peter Stein & Horst Ulbricht. 4 vols. Berlin: VWB-Verlag für Wissenschaft und Bildung.

Noun phrase structure of Yilan Creole & the effects from its source languages

Sin Yee Melody Li (National University of Singapore)

Yilan Creole (YC) is a unique Japanese-based creole with Austronesian and Chinese influences, spoken in four Atayal villages in Yilan County, Taiwan. Chien and Sanada (2010) suggested YC emerged from the mixed population of Atayal and Seediq people during the Japanese colonization of Taiwan, whereas Tan (2023) argued the Japanese education played a more crucial role. The current population of four villages is around 3100, but the estimated number of YC speakers varies between 550-1400, depending on their YC proficiency. Because of no intergenerational transmission and lack of institutional support, YC is considered "severely endangered" on the Language Endangerment Index (Van Way & Lee, 2013).

I have conducted fieldwork since mid-2023 and lived in Tungyueh village for over 1.5 years. My project aims to provide a comprehensive description of YC, but this talk focuses on the noun phrases (NP). While several papers concern aspects of the NP, including case (Chien, 2016), pronouns (Chien, 2018) and demonstratives (Chien, 2019), no study has discussed its NP structure entirely. This talk presents my fieldwork data and discusses how the NP structure of YC has been influenced by the source languages and universal constraints.

(1)	wasi	age-ta	Aki	ni	**[utux**	**omiage]NP.**
	1.SG	give-PST	Aki	DAT	**[one**	**present]**

"I gave Aki a/one present."

(2a)	kyo	no	bang	**[nan**	**ge**	**nigen]NP**	gohan	tabe-ru?
	today	GEN	night	**[how**	**CL**	**person]**	rice	eat-non.PST

Q: "How many people are having dinner tonight?"

(2b) **[hachi mei]NP.**

[eight CL]

A: "Eight people."

Numerals in YC can be derived from Atayal (*utux*) and Japanese (*hachi*), as in (1) and (2b) respectively. Since classifiers are optional in Atayal, the NP is [numeral + head noun] with Atayal numerals in (1). Conversely, Japanese and Mandarin require classifiers, so the structure becomes [numeral/quantifier + classifier + head noun] with Japanese quantifiers (*nan*), as in (2a). It is interesting the classifier *ge* (2a) and *mei* (2b) come from Mandarin and Japanese respectively, though the head noun is supposedly the same ("people"). The classifier *mei* (2b) seems to function

as a “proform” to replace the omitted head noun, which is common in Mandarin.

(3)	didi	kuri-ta	**[[wasi**	**mi-toru]RC**	**no**	**hon]NP**
	younger brother	steal-PST	**[[1.SG**	**see-PROG]**	**GEN**	**book]**

“My brother stole the book which I was reading.”

Example (3) is an elicited relative clause (RC). The relativized head noun (*hon*) is linked to the RC by the Japanese-derived genitive particle *no*, aka. the “nominalizer” (Iwasaki, 2002). However, the SVO word order of the main clause and the RC resembles the Mandarin structure, differing substantially from Japanese.

In short, the numeral-head word order in YC may be attributed to the congruent effect of the same structure in Japanese, Atayal and Mandarin, but also to the universal tendency of having prenominal numerals in creoles (Bobyleva, 2012). While RC tends to occur post-nominally across contact languages on APiCS, the prenominal RC in YC may reflect the overgeneralization of Japanese NP structure by the YC-Mandarin bilingual speakers. Hence, universal features of creoles and second language acquisition also contribute to the NP structure of YC. (503 words)

Abbreviations

1.SG = first person singular; CL = classifier; DAT = dative particle; GEN = genitive particle; non.PST = non-past tense; PROG = progressive aspect; PST = past tense.

References

Bobyleva, E. (2013). *The development of the nominal domain in creole languages: A comparative-typological approach.* Netherlands Graduate School of Linguistics.

Chien, Y. (2016). Yilan Creole Case Marking. 国立国語研究所論集 *[NINJAL Research Papers], 10*, 1-17.

Chien, Y. (2018). 日本語を上層とする宜蘭クレオールの人称代名詞 [The Personal Pronominal System of Yilan Creole]. 日本語の研究 *[Studies in the Japanese Language], 14*(4), 31-47. [Japanese source]

Chien, Y. (2019). 日本語を上層とする宜蘭クレオールの指示詞 [The demonstrative system of Yilan Creole]. 社会言語科学 *[The Japanese Journal of Language in Society], 21*(2), 50-65. [Japanese source]

Chien, Y. C., & Sanada, S. (2010). Yilan Creole in Taiwan. *Journal of Pidgin and Creole*

Language, 25(2), 350-357.

Iwasaki, S. (2002). *Japanese (London Oriental and African language library; v. 5)*. John Benjamins Publishing Company.

Tan, G. L. (2023). A new view on ‘Yilan Creole’. *Journal of Pidgin and Creole Languages, 38*(2), 320-388.

Van Way, J., & Lee, N. H. (2013, February 28-March 3). *Assessing levels of endangerment in the Catalogue of Endangered Languages (ELCat)* [Poster session]. The 3rd International Conference on Language Documentation and Conservation, University of Hawai‘i at Mānoa, HI, United States.

Tracing Negation: Pragmatic Innovation in Afro-Brazilian Portuguese

Luana Lamberti (Iowa State University)

Brazilian Portuguese presents three negative strategies known in the literature as neg1, or canonical negation, with a pre-verbal negative marker, *n ̃a o* 'no'; neg2, with both pre- and post-verbal negative markers; and neg3, which surfaces as a post-verbal and/or clause-final negative particle. This phenomenon has been widely studied (Schwenter, 2005; Lamberti, 2014; Petry et al., 2021). How-ever, its historical origins in Brazilian Portuguese, and more specifically, Afro-Brazilian Portuguese, still require further exploration. While non-canonical strategies (neg2 and neg3) were historically available in 16th-century European Portuguese and present in early 18th-century Brazilian Portuguese (Seixas & Alkmim, 2013), neg3 was infrequent and pragmatically restricted. Considering this linguistic panorama, the present study investigates the role of pragmatic specialization in the development of alternative negative strategies during the contact period between Portuguese and Niger-Congo languages in 19th-century Bahia. The analysis relies on synchronic data from the Afro-Brazilian Por-tuguese variety spoken in Helv´ecia, Bahia. The dataset comprises 3,794 tokens drawn from sociolinguistic interviews conducted between 1994 and 2019, ana-lyzed statistically including both linguistic and social factors. The former factors such as contextual activation (i.e., the need for propositional content to be acti-vated; see (Schwenter, 2005)) and sentence complexity (favoring simple clauses) emerged as significant predictors for the use of neg2 and neg3. Social factors, such as age and lifelong residence status, also proved significant: older speak-ers who have never left the community statistically employ the non-canonical negation strategies more frequently than other speakers. In addition to iden-tifying linguistic constraints, this study includes a crosslinguistic comparison between these alternative negative strategies in Afro-Brazilian Portuguese and the most common negation patterns found in Niger-Congo languages. Bantu languages and Yorùbá, known to have been in contact with Portuguese during the formation of Helv´ecian Portuguese in the 19th century, are particularly rel-evant to this analysis (Lamberti, 2022, 2025). Findings show that post-verbal and clause-final negation is widespread across the Niger-Congo phylum (Idiatov, 2019; Cyffer et al., 2009), often governed by complex pragmatic constraints that shape speakers' choices among available strategies. The main contribution of this research is to demonstrate the role of pragmatics in the emergence of inno-vative negative strategies in Afro-Brazilian Portuguese. Based on these results I argue that, although double

negation and post-verbal negation were already available in restricted contexts for second-language learners, speakers creatively enriched the Portuguese negation system by assigning new pragmatic functions to neg2 and neg3. This process reflects imposition from native languages (Winford, 2003), where similar strategies were pragmatically motivated. Ultimately, this study shows that pragmatically constrained phenomena are amplified in contact situations, offering new insights into the dynamic interplay between morphosyntax, pragmatics, and the sociolinguistic history of contact languages.

References

Cyffer, Norbert, Erwin Ebermann & Georg Ziegelmeyer (eds.). 2009. *Negation Patterns in West African Languages and Beyond*. John Benjamins Publishing Company. doi:doi:10.1075/tsl.87.

Idiatov, Dmitry. 2019. *An areal typology of clause-final negation in Africa* 115–164. Berlin, Boston: De Gruyter Mouton. doi:doi:10.1515/9783110607963-005.

Lamberti, Luana. 2014. *Motivaçõ es pragmáticas para o uso de dupla negaçã o: um estudo do fenômeno em portugu ̂es europeu*. Universidade Federal do Rio Grande do Sul MA thesis.

Lamberti, Luana. 2022. *The Contribution of Niger-Congo Languages to Loca-tion Marking in Afro-Yungueñ o Spanish and Helv ́ecia Portuguese*: The Ohio State University dissertation.

Lamberti, Luana. 2025. Location marking in helv ́ecian portuguese: A case of contact-induced change. *Journal of Language Contact* 18. 1–49.

Petry, Paloma, Marcos Goldnadel & Luana Lamberti. 2021. Funçõ es pragmáticas de enunciados com dupla negaçã o em porto alegre no final do s ́eculo xx: uma análise qualitativa de dados extra ́ıdos de entrevistas soci-olingu ́ısticas. *Revista da ABRALIN* 1–38.

Schwenter, Scott A. 2005. The pragmatics of negation in Brazilian Portuguese. *Lingua* 115(10). 1427–1456.

Seixas, Vivian Canella & Mônica GR de Alkmim. 2013. A negaçã o sentencial em textos de autores brasileiros dos s ́eculos xviii e xix: consideraçõ es sobre implementação, transição e origem da estrutura [nãovnão]. *Revista Veredas* 17(2). 83–113.

Winford, Donald. 2003. *An Introduction to Contact Linguistics*. Blackwell Publishing.

Afro-Iberian Languages in Africa and in the Americas: Contact and Sociohistory

Luana Lamberti (Iowa State University)
& Ana Lívia Agostinho (Universidade Federal de Santa Catarina)

This book project (AUTHORS, in press) examines the sociohistorical consequences of European colonization in Africa and the Americas, focusing on the linguistic outcomes that emerged from contact between Indo-European and Niger-Congo languages. Enslaved Africans gradually adopted varieties of Portuguese and Spanish. In these contexts, new grammars arose through restructuring triggered by intense and untutored language contact (Thomason & Kaufman 1988). Beyond internal restructuring, Winford (2020) proposes that second-language learners also employed creative strategies not directly derived from either their target or native languages. His naturalistic second-language acquisition model integrates macro-level factors, such as social organization and power structures, with micro-level factors tied to individual learners' linguistic backgrounds. Schwegler (2010) and Winford (2020) note that contact varieties and creoles remain understudied within Hispanic and Lusophone Linguistics. They argue that post-colonial Spanish-and Portuguese-contact languages arose in comparable sociolinguistic settings and through similar processes of naturalistic second-language acquisition. Consequently, they call for a closer relationship among these fields. Responding to this call, this project proposes a unified framework for studying Afro-Iberian languages, defined as varieties shaped by the sociolinguistic and historical conditions produced by the transatlantic slave trade and colonialism. Across two volumes, the project brings together perspectives from Creole Studies, contact linguistics, and Hispanic and Lusophone linguistics to analyze nineteen languages and varieties spoken in Africa and the Americas. We adopt the term "Afro-Iberian languages" to encompass these creoles and contact varieties, highlighting the central role of social and historical forces in their development. The project outlines the theoretical foundations guiding the volumes, drawing on frameworks from language contact and Creole Studies. The handbook provides both typological and sociohistorical overviews of Portuguese- and Spanish-lexified creoles, as well as Afro-Portuguese and Afro-Hispanic contact varieties. The chapters illustrate a wide spectrum of linguistic outcomes, from highly restructured creoles to indigenized varieties shaped by varying demographic and sociolinguistic conditions. The languages and varieties examined include the Portuguese-based creoles of the Gulf of Guinea and Upper Guinea; the Spanish-based creoles Papiamentu and Palenquero; Afro-Brazilian Portuguese varieties spoken in Bahia and Maranhão; the mixed languages Kalunga and Cupópia; African Portuguese varieties from São Tomé e Príncipe, Guiné-Bissau, Angola (Libolo and Cabinda), and Mozambique; and several Afro-Hispanic varieties, including Afroyungueño, Afro-Mexican Spanish, Afro-Puerto Rican Spanish, Afro-Colombian Spanish, and Equatorial Guinean Spanish. The project produced a multilingual (English, Spanish, and Portuguese) open-access handbook published by Language Science Press. Intended for introductory courses in linguistics, language contact, Creole

Studies, and Hispanic and Lusophone linguistics, the handbook specifically aims to reach undergraduate and graduate students in the Global South, where access to major linguistic publications is often limited.

References

Authors. In press. Afro-Iberian Languages in Africa: Contact and Sociohistory. Vol. 1. Language Science Press.

Authors. In press. Afro-Iberian Languages in the Americas: Contact and Sociohistory. Vol. 2. Language Science Press.

Schwegler, Armin. 2010. Pidgin and Creole studies: Their Interface with Hispanic and Lusophone Linguistics. Studies in Hispanic & Lusophone Linguistics 3(2).

Thomason, Sarah Grey & Terrence Kaufman. 1988. Language Contact, Creolization, and Genetic Linguistics. University of California Press.

Winford, Donald. 2020. New Spanishes in the Context of Contact Linguistics 11–41. John Benjamins Publishing Company.

Diglossia or Continuum? Reconfiguring Inherited Models of Creole Language Contact

Ludovic Mompelat (University of Miami)

Classical models of diglossia have historically long shaped analyses of Creole-speaking societies by presupposing a stable alignment among structural differentiation (i.e. basilect, mesolect, acrolect), functional compartmentalization (High vs. Low functions), and ideological hierarchy (High vs. Low prestige) (Ferguson 1959; Fishman 1967). While such alignment remains descriptively adequate in certain contexts, empirical work in the French Caribbean has repeatedly documented cases in which these dimensions partially diverge (Prudent 1981; Bernabé 1982; Hazaël-Massieux 2006; Managan 2016; Bellonie & Pustka 2018).

This project proposes a multidimensional framework for analyzing language contact that treats structural positioning, functional span, and ideological valuation as analytically independent yet empirically interacting dimensions. Rather than determining whether a community "is" or "is not" diglossic, the framework conceptualizes diglossia and bilingualism as idealized configurations—one of many—within a broader analytical space. This approach makes it possible to visualize and compare contact ecologies in terms of degrees of convergence, divergence, and reconfiguration across dimensions, without multiplying typologies or flattening Creole variation into binary oppositions.

The framework is empirically populated using original survey data from Martinique, which are explicitly mapped onto its three dimensions. Structural positioning is modeled along a French–Creole continuum that includes basilectal, mesolectal/interlectal, and acrolectal practices. Functional span is operationalized probabilistically through reported language use across social domains, allowing varieties to occupy overlapping vertical ranges rather than discrete High/Low slots. Ideological valuation is modeled independently through attitudinal structure, revealing internally differentiated prestige representations within each language rather than a single cross-linguistic prestige hierarchy. This mapping shows that in Martinique, functional expansion of Creole is largely mediated by stabilized intermediate varieties and coexists with ideological differentiation that does not mirror functional reach: basilectal and standardized Creole index identity, affect, and authenticity despite restricted institutional access, while French retains institutional authority without monopolizing symbolic legitimacy.

The contrastive value of the framework is demonstrated by projecting Haiti into the same analytical space. In this case, structural differentiation, functional stratification, and ideological hierarchy remain largely aligned. (Valdman 2015; Dejean 1993; Schieffelin & Doucet 1994) Although intermediate forms such as Kreyòl swa mediate certain registers, they do not substantially disrupt the overarching stratification between Haitian Creole and French. French continues to dominate high-prestige institutional domains, and

ideological representations largely reinforce this hierarchy despite the universal use and strong affective valuation of Haitian Creole. Mapping both territories within the same framework makes visible where diglossia remains descriptively robust and where its generalization becomes analytically limiting.

By explicitly disentangling structure, function, and ideology, the proposed framework facilitates cross-territorial comparison of language contact situations by allowing researchers to map how different societies diverge or converge along each dimension independently. At the same time, it provides a basis for diachronic comparison within a single territory by tracking how alignments between typological structure, functional allocation, and ideological valuation shift over time.

References:

Bellonie, J. D., & Pustka, E. (2019). Représentations des «mélanges» linguistiques en Martinique: des créolismes au français régional. *Études créoles. Cultures, langues, sociétés*, (36| 1-2).

Bernabé, J. (1982). Contribution à une approche glottocritique de l'espace littéraire antillais. *La linguistique, vol. 18, 1*, pp. 85-109.

Dejean, Y. (1993). An overview of the language situation in Haiti. *International journal of the sociology of language*, *1993*(102).

Ferguson, C. A. (1959). Diglossia. *word*, *15*(2), 325-340.

Fishman, J. A. (2020). Bilingualism with and without diglossia; diglossia with and without bilingualism. In *The bilingualism reader* (pp. 47-54). Routledge.

Hazaël-Massieux, M. C. (1996). Du français, du créole et de quelques situations plurilingues: données linguistiques et sociolinguistiques. *Francophonie. Mythes, masques et réalités. Enjeux politiques et culturels*, 127-157.

Managan, K. (2016). The sociolinguistic situation in Guadeloupe: Diglossia reconsidered. *Journal of Pidgin and Creole Languages*, *31*(2), 253-287. Prudent, L. F. (1981). Diglossie et interlecte. *Langages*, 13-38.

Schieffelin, B. B., & Doucet, R. C. (1994). The "real" Haitian Creole: Ideology, metalinguistics, and orthographic choice. *American ethnologist*, *21*(1), 176-200. Valdman, A. (2015). *Haitian Creole: structure, variation, status, origin*. Equinox Publishing Limited.

Despalatalização das fricativas pós-alvelares /ʃ, ʒ/ no guineense de Bissau e Gabu

Mamadu Baciro Balde (Universidade de São Paulo)

O guineense é o crioulo da Guiné-Bissau (CGB), desenvolvido na região de Alta Guiné (Scantamburlo, 2013; Costa, 2014). Este trabalho tem como objetivo analisar a produção e a percepção das consoantes fricativas coronais /s, ʃ, z, ʒ/ dessa língua. Para tanto, partimos da produção dos guineenses residentes no Estado de São Paulo e provenientes de dois lugares da Guiné-Bissau: Bissau, a capital, e Gabu, uma região localizada na província leste do país. A pesquisa parte de hipótese de que as duas variedades são demarcadas fonologicamente por comportamento destes segmentos por meio de despalatalização (/ʃ/ → [s]) ou (/ʒ/ → [z]). Consideramos em nossa análise fatores estruturais e influências de línguas em contato com o guineense nos referidos locais. Três das consoantes em questão (/ʃ, z, ʒ/) são marcadas em maioria de línguas africanas faladas no país, a exemplo de fula, pepel, mandjaco etc., sendo o português a única língua em que elas gozam de status fonêmico (Parkvall, 2012; Chapouto, 2014). Com base na fonologia de Laboratório (Pierrehumbert; Beckman; Ladd, 2012), desenhamos experimentos usando itens lexicais ou logatomas contendo as fricativas em causa (/saˈpatu/ "sapato"; /ʃiˈnɛlu/ "chinelo"; /zɛˈɾadu/ "zerado"; /ˈʒɛlu/ "gelo". Para a análise do processo de despalatalização, levaram-se em conta a qualidade da vogal, a tonicidade da sílaba e a posição desta na palavra, para serem lidos em frase-veículo: "N'ta fala x tris bias[1]" Para dados de percepção, foram adotados pares mínimos, permutando as fricativas pós-alveolares /ʃ, ʒ/ e as suas contrapartes alveolares /s, z/ (sestu vs chestu; zeru vs jeru; lichu vs lisu; jelu vs zelu. A coleta de dados foi por meio de entrevistas com três falantes de cada variedade. Não se obteve quantidades iguais de dados dos dois grupos de falantes. Do total de 1.044 produções, 64.3% (672/1.044) são de Bissau e 35.7% (372/1.044) de Gabu. No que diz respeito à despalatalização das fricativas pós-alveolares, registrou-se 7 casos de despalatalização (/ʃ/ → [s]) ou (/ʒ/ → [z]) nas produções de Bissau e 98 casos nas produções de Gabu. Uma análise baseada em contextos fonológicos de produção mostra que estes não são relevantes no desencadeamento da despalatalização, tendo o processo ocorrido quase em todos os contextos analisados, com exceção da posição medial diante da vogal abaixa /a/. Em relação teste à percepção, constatou-se uma tendência de os falantes de Gabu alternarem os segmentos que compõem os dois pares [s, ʃ] e [z, ʒ] tomando um por outro, com ênfase para as formas com as alveolares [s, z], correspondentes ao processo de despalatalização. Assim, considera-se que processo não se correlaciona a nenhum condicionante e atribui ao fator fonológico o comportamento contrastante das fricativas coronais nestas variedades do guineense, que, inseridas em contextos sociolinguísticos diferentes, são moldadas por línguas fonologicamente divergentes.

[1] Digo x três vezes.

Políticas de oficialização e ensino bilíngue da língua guineense face ao decreto 7/2007

Marcelino Issa da Cunha (Universidade Federal da Paraíba)

A Guiné-Bissau constitui um "autêntico mosaico étnico-cultural" (EMBALÓ, 2008), onde a Língua Portuguesa (LP), embora seja o único idioma com estatuto oficial, é falada como língua materna por uma fração reduzida da população. Em contrapartida, a Língua Guineense (LG), também designada por *Kriol*, consolidou-se como a verdadeira língua de unidade nacional e fator de identidade supra-étnica. Este trabalho analisa as políticas linguísticas guineenses sob a ótica da tríade de Spolsky (2004), composta por práticas, crenças e gestão linguística, confrontando a legislação vigente com a realidade sociolinguística das salas de aula.

O cerne da discussão reside no conflito entre a gestão linguística explícita, materializada no Decreto-lei nº 7/2007, que impõe a obrigatoriedade da LP nos departamentos de Estado e na comunicação social, e o "caso implícito" (IÉ, 2022) das práticas sociais. Argumentamos que a manutenção de um sistema educativo exclusivamente monolíngue em LP atua como uma barreira pedagógica, resultando em altos índices de insucesso escolar, conforme observado por Cunha (2024) na análise dos manuais "Comunicar em Português 6". A hipótese central sustenta que a oficialização da LG e a implementação de um modelo de Educação Bilíngue (EB) são imperativos para garantir os direitos linguísticos e a eficácia do ensino.

Para ilustrar a função da LG como amálgama identitário e a sua estrutura linguística, apresentamos o seguinte dado de oratura popular (BULL, 1989), essencial para a compreensão da coesão nacional: *(1) Anós i fidjus di um mamé.* nós somos filhos de uma mãe (‘Nós somos todos irmãos/filhos da mesma pátria.’)

A análise demonstra que, enquanto a LP é percebida como língua de prestígio administrativo, a LG carrega a "filosofia e sabedoria" (BULL, 1989) do povo. A pesquisa fundamenta-se ainda no contraste histórico: se o passado colonial foi marcado pelo "Estatuto do Indigenato" (1954), que estratificava cidadãos pela competência linguística e cultural, o "futuro" da Guiné-Bissau depende da descolonização curricular. A análise detalhada das crenças de professores e alunos revela que, embora a LP seja desejada para ascensão social, a LG é a língua em que o conhecimento é efetivamente processado e construído no quotidiano escolar.

Este estudo contribui para a linguística de contacto ao propor que o futuro das línguas crioulas não deve ser o da marginalidade institucional, mas o da cooficialidade. Conclui-se que o reconhecimento jurídico da LG é o passo necessário para alinhar a "gestão" à "prática", transformando o legado histórico de uma língua de contacto num instrumento de soberania e desenvolvimento humano.

Reverberações da ancestralidade negra no discurso religioso no Boi dos Papudos – MA

Marcelo Nicomedes dos Reis Silva Filho ()

O Bumba meu boi do Maranhão é diverso, plural e multicultural, tendo bases sólidas na integração, nada pacífica, das três etnias que compõem a constituição do povo brasileiro e de sua cultura popular. No Maranhão, sob a perspectiva de Azevedo Neto (2010), o bumba meu boi tem essa constituição, onde o autor divide a manifestação como tendo três grupos principais: branco, indígena e negro; cada um deles consegue abarcar grupos que apresentem e representem características das etnias. Para isso, o autor também apresenta a noção de sotaque, que é a forma como cada grupo de grupo de bumba meu boi, de características semelhantes, organiza seus grupos. O bumba meu boi do Maranhão apresenta cinco sotaques principais; são eles: Baixada, Costa de mão, Matraca, Orquestra e Zabumba. Seguindo na perspectiva apontada por Azevedo Neto, os grupos de bumba meu boi dos sotaques da Baixada e Matraca compõem o grupo indígena; o sotaque de Zabumba e Costa de mão, o grupo negro; e o grupo branco abarcaria os grupos do sotaque de Orquestra. Tendo em vista que o contexto do bumba meu boi do Maranhão passa pelas questões de hibridismo cultural, embora os bumbas meus bois do sotaque de Pindaré e Matraca tenham características marcantes de indígenas, em termos de religiosidade, ele é atravessado por memórias ameríndias e afrodescendentes, como a religião da Umbanda e do Tambor de Mina. É com base nessa relação de tensão entre as culturas do branco, negro e indígena, em especial dos dois últimos, que constituímos a pergunta problema desta pesquisa, que é: Que funcionamentos do discurso religioso atravessam, atualizam e estabilizam sentidos presentes no léxico utilizado no discurso do Bumba meu boi dos Papudos – MA, e como ela reverbera sentidos das ancestralidades afro-indígenas? Frente aos questionamentos supracitados, constituímos o objetivo da pesquisa que se enseja: Analisar de que modo o léxico do Bumba meu boi dos Papudos (MA) incorpora e reconfigura elementos do discurso religioso afrodescendente, problematizando os efeitos ideológicos envolvidos nos processos de atravessamento, atualização e estabilização de sentidos. Para tanto, montaremos um arquivo que é composto de registros fotográfico, fonográficos e transcrições de letras das toadas que apresentam relação com as temáticas indígenas e afrodescendentes. Partindo dos estudos de Orlandi (1987), onde a autora propõe, para o processo analítico, dois gestos cruciais para o trato do corpus: recortar e segmentar. É a partir dos recortes que será possível trabalhar de forma mais profunda no corpus, esmiuçando-o em partes menores, proporcionando um batimento teórico consistente. Como teoria base para as análises, trabalhamos com os postulados de Pêcheux, sobre ideologia, sujeito, memória, interdiscurso e formações discursivas (Pêcheux, 2010, 2015); Orlandi, sobre questões metodológicas, além de noções como jogo de forças, segmentar, recortar e silenciamento (1987, 2007, 2012 e 2015); Cardoso (2004), sobre o léxico do bumba meu boi; e, sobre Bumba meu boi, Silva Filho (2021), Azevedo Neto (2019), Borralho (2015) e Carvalho (1995). Espera-se que as discussões suscitadas por este estudo colaborem para a ampliação do espectro de pesquisa sobre linguística, religião,

sociolinguística e discurso, pois é com a junção de ferramentas de um campo interdisciplinar como o proposto que se consegue verticalizar em estudos de linguagem, trabalhando de forma recursiva, partindo desses estudos.

Beyond 'already': The functional range of *dja* in Caboverdean

Marco Favaro (Universidade de Lisboa)

Typological research has shown that expressions meaning 'already' may develop into tense and aspect markers through processes of grammaticalization (Dahl & Velupillai 2013; Dahl 2022). Several instances of this functional development have been identified in Portuguese-related creoles, albeit following different trajectories (Holm 1989: 267; Winford 2018: 200–202). This study examines one such expression, the marker *dja* (rendered as *ja* in some varieties) in Caboverdean, derived from the Portuguese adverb *já* 'already' and now exhibiting a range of functions. Notably, it can be used both as a temporal adverb (1) and as a marker expressing different temporal and aspectual meanings (here generically glossed as TAM), such as completion (2) and the continuing relevance of an event to the present situation (3):

(1) **Ja** e ka ta mutu matrata-m.
now it NEG TAM much ill-treat-me

'Now, it does not ill-treat me too much.'

(Baptista 2002: 86)

(2) Nha mai more, nha pai **dja** more.
my mother die my father TAM die

'My mother died, my father died.'

(Baptista 2002: 84)

(3) Xobinho fra-l, **ja** tem tres dia qui'm ca
Xobinho say-him TAM have three day COMP-I NEG
come nada.
eat nothing

'Xobinho told him, it has been three days (already) since I've eaten anything.'

(Baptista 2002: 85)

The general properties of *dja*/*ja* have been addressed in various accounts of Caboverdean grammar (Baptista 2002; Pratas 2007; Swolkien 2015); however, no dedicated study has yet provided a systematic analysis of its features. In particular, both the boundary between its adverbial use and its use as a tense–aspect marker, as well as the precise functions of the latter, remain a matter of debate (Silva 1990; Pratas 2007: 107).

This paper reassesses the properties of *dja*/*ja* through a corpus-based analysis. The dataset

consists of approximately 300 occurrences extracted from the LUDViC corpus (Pratas 2020), a digital open-access collection of semi-informal interviews representing different varieties of Caboverdean. Using corpus-linguistic methods, the study describes the main distributional features of *dja*/*ja*, including (i) frequency of occurrence, (ii) sentence position and relevant co-text, and (iii) adjacency to the verb phrase. Additional factors, such as the lexical properties of the verb phrase and co-occurrence with other TAM markers, are also taken into account.

This approach makes it possible to delineate the functional range of constructions with *dja*/*ja* in Caboverdean and to identify prototypical instances of each category. The analysis shows that, while most uses of *dja*/*ja* are adverbial, the marker also occurs in a limited set of contexts expressing meanings that are cross-linguistically associated with perfect constructions, including resultative, existential, and recent past readings (Comrie 1976: 56–61; Dahl 2000; Ritz 2012). These findings are discussed in light of recent work on the mapping of meanings within the semantic domain of 'already' expressions, the perfect and related categories (Olsson 2013; Dahl & Wälchli 2016; Krajinović, Vander Klok & von Prince 2025).

Overall, this study provides a comprehensive account of the formal and functional properties of constructions featuring *dja*/*ja* in Caboverdean. More generally, it contributes to cross-linguistic research on the grammaticalization of 'already' expressions in creole languages and beyond.

References

Baptista, Marlyse. 2002. *The syntax of Cape Verdean Creole: The Sotavento varieties*. Amsterdam: John Benjamins.

Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.

Dahl, Östen. 2000. The tense-aspect systems of European languages in a typological perspective. In Östen Dahl (ed.), *Tense and aspect in the languages of Europe*, 3– 25. Berlin: De Gruyter.

Dahl, Östen. 2022. Perfects across languages. *Annual Review of Linguistics* 8: 279–297. Dahl, Östen & Viveka Velupillai. 2013. The Perfect. In Matthew S. Dryer & Martin Haspelmath, (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Accessed online at http://wals.info /chapter/68; 11 Feb. 2026).

Dahl, Östen & Bernhard Wälchli. 2016. Perfects and iamitives: two gram types in one grammatical space. *Letras de hoje* 51 (3): 325–348.

Holm, John. 1989. *Pidgins and Creoles*, 2 vol. Cambridge: Cambridge University Press. Krajinović, Ana, Jozina Vander Klok & Kilu von Prince. 2025. When (not) to establish a new category. The case of perfect, 'already', and iamitives. *Studies in Language* 49 (3): 546–586.

Olsson, Bruno. 2013. *Iamitives: perfects in Southeast Asia and beyond*. MA Thesis: Stockholm

University.

Pratas, Fernanda. 2007. *Tense features and argument structure in Capeverdean predicates*. Ph.D. Dissertation: Universidade Nova de Lisboa.

Pratas, Fernanda. 2020. LUDViC: An oral corpus of Caboverdean. http://teitok.clul.ul.pt/ludvic/

Ritz, Marie-Eve. 2012. Perfect Tense and Aspect. In Robert I. Binnick (ed.), *The Oxford Handbook of Tense and Aspect*, 881–907. Oxford: Oxford University Press.

Silva, Izione S. 1990. Tense and Aspect in Capeverdean Crioulo. In John Victor Singler (ed.), *Pidgin and Creole Tense-Mood-Aspect Systems*, 143–168. Amsterdam: John Benjamins.

Swolkien, Dominika. 2015. *The Cape Verdean Creole of São Vicente: Its genesis and structure*. Ph.D. Dissertation: Universidade de Coimbra.

Winford, Donald. 2018. Creole Tense–Mood–Aspect Systems. *Annual Review of Linguistics* 4: 193–212.

Infraestrutura de Inteligência Artificial para o Cabo-verdiano

Miguel Dias, Larissa Marinho Bonani, Erik Daskalyuk (Universidade de Lisboa), Maria Fonseca (Universidade de Évora), Ana Landim, Camila Monteiro, Ana Karina Moreira, Loribela Silva, Leandra Lopes (Universidade de Cabo Verde), Ricardo Pereira, António Raimundo, Mariana Sobral & Dominika Swolkien (Universidade de Lisboa)

Este artigo apresenta resultados preliminares do projeto OEI ORAL – kriOl(u) laRge lAnguage modeLs, que tem como objetivo desenvolver e disponibilizar, em código aberto, recursos linguísticos e ferramentas tecnológicas para a língua cabo-verdiana (ISO kea). O projeto inclui plataformas e aplicações inovadoras que permitem aos falantes de kea, tanto em Cabo Verde como na diáspora, interagir de forma natural, escrita e falada, na sua língua materna ou de herança, promovendo a projeção internacional do cabo-verdiano.

Apesar de o kea estar classificado como língua "ascendente" em termos de suporte digital (Eberhard *et al*., 2025), existem importantes lacunas no desenvolvimento de tecnologias linguísticas para o kea, particularmente no reconhecimento automático de fala (ASR) e na síntese de fala (TTS), uma realidade comum a muitas línguas de baixos recursos (Adelani *et al*., 2025; Prado, 2024). Recentemente, surgiram modelos multilíngua de tradução que incluem o cabo-verdiano (Robinson *et al*., 2024; Medina, 2025; Rowe *et al*., 2025), mas estes não apresentam soluções para o português. Simultaneamente, modelos de linguagem em grande escala (LLM), como GPT-4o, Grok 3 Beta, o3-mini, DeepSeek R1 e Gemini, conseguem gerar texto em kea via generalização "one-shot" (Duan *et al*., 2017). No entanto, a qualidade gerada é inconsistente, com pouca estabilização ortográfica e ausência de validação linguística sistemática.

Com base nos estudos piloto de Raimundo *et al.* (2025) e Daskalyuk & Pereira (2025), o projeto está a desenvolver dez artefactos linguísticos e computacionais, disponibilizados em acesso aberto na plataforma *Hugging Face*. A uniformização ortográfica interdialetal e intradialetal, fundamentada no Alfabeto Cabo-verdiano (2009), é um dos focos principais do projeto daí que será apresentada a metodologia de criação, curadoria e normalização de corpora textuais para as variedades de Santiago (kea-ST) e São Vicente (kea-SV), recolhidos por doação, licenciamento, parcerias institucionais e *web scraping*. Apresentaremos, ainda, os dados referentes à resolução das áreas linguísticas críticas na elaboração de corpora paralelos kea ↔ português europeu (pt-pt), nomeadamente parâmetros *pro-drop*, frases resumptivas, arquitetura aspeto-temporal e extensões semânticas.

Serão demonstrados os primeiros resultados do treino e afinação de dois modelos LLM: um para geração de texto e diálogo interativo em kea-ST e kea-SV, baseado na continuação do treino com um corpus de 913 183 frases, do modelo Llama-3.2-1B (Meta 2025), e outro para tradução bidirecional kea ↔ pt-pt, pela afinação com um corpus paralelo de 12 242 frases, do modelo mBART-large-50-many-to-manymmt (Tang *et al*., 2020), desenvolvido pela Facebook especialmente para tarefas de tradução automática multilíngua**.**

No domínio da fala, o projeto desenvolve o primeiro sistema ASR para kea, através da adaptação de modelos multilíngua adequados para línguas de baixos recursos, com

abordagens como Wav2Vec2-BERT (Baevski et. al., 2020; Lacombe, 2024) e arquiteturas baseadas em *Connectionist Temporal Classification* (CTC) (Hannun *et al.,* 2014). O treino usa cerca de 70 horas de gravações em kea-ST e kea-SV, complementadas por novas recolhas. Paralelamente, são criados dois léxicos fonéticos para as duas variedades, anotados com informação fonética e prosódica, cujos resultados preliminares também serão apresentados.

Ao conjugar desenvolvimento tecnológico, investigação em linguística e ciência aberta, os resultados deste trabalho contribuem para a redução de desigualdade digital do cabo-verdiano, promovem a sua padronização e disponibilizam recursos com potencial educativo (Swolkien *et al.*, 2025), fomentando a soberania digital de uma língua crioula de baixos recursos no ecossistema digital global.

Referências

Adelani, D. I., Dogruöz, A. S., Shode, I., & Aremu, A. (2025). Does generative AI speak Nigerian-Pidgin? Issues about representativeness and bias for multilingualism in LLMs. *Findings of the Association for Computational Linguistics: NAACL 2025*, 1571–1583.

Alfabétu Kabuvérdianu (AK). Conselho de Ministros [do Governo da República de Cabo Verde]. (2009). Decreto-Lei No. 8/2009. Boletim Oficial da República de Cabo Verde, Praia 19.03.2009, I Série, No.11, 74–76.

Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. https://arxiv.org/abs/2006.11477

Daskalyuk, E., & Pereira, R. (2025, October 27). First steps towards the development of an LLM for the Cape Verdean Creole. In Proceedings of the 1st International Undergraduate Research Symposium at Iscte (SIL-Iscte 2025) https://zenodo.org/records/17334220

Duan, Y., Andrychowicz, M., Stadie, B., Ho, J., Schneider, J., Sutskever, I., Abbeel, P., & Zaremba, W. (2017). *One-shot imitation learning*. Advances in Neural Information Processing Systems, 30. https://arxiv.org/abs/1703.07326

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2025). Ethnologue: Languages of the world (28th ed.). SIL International. http://www.ethnologue.com

Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., Prenger, R., Satheesh, S., Sengupta, S., Coates, A., & Ng, A. Y. (2014). *Deep speech: Scaling up end-to-end speech recognition*. arXiv. https://arxiv.org/abs/1412.5567

Hugging Face. https://huggingface.co/blog/fine-tune-w2v2-bert.

Medina, R. C. (2025). Creating NLP resources for Cape Verdean Creole: Challenges and solutions for underrepresented languages (Dissertação de mestrado, ISCTE – Instituto Universitário de Lisboa).

Meta (2024). Llama-3.2-1B (Large language model). Hugging Face. https://huggingface.co/meta-llama/Llama-3.2-1B

Lacombe, Y. (2024). wav2vec2-bert-CV16-en-libri (Pre-trained model). Hugging Face. https://huggingface.co/ylacombe/wav2vec2-bert-CV16-en-libri

Prado, D. (2024). *Línguas indígenas no mundo digital: Inventário de recursos e lacunas*.

Organização dos Estados Ibero Americanos (OEI).

Raimundo, A., Dias, M. S., Daskalyuk, E., Pereira, R. (2025), Advances in the development of a Kriolu LLM, world's first Large Language Model for the Cape Verdean Language. In *1st Atlantic Crossing International Symposium USA – Portugal*, 15th January 2025, Iscte.

Robinson, N., Dabre, R., Shurtz, A., Dent, R., Onesi, O., Monroc, C., Grobol, L., Muhammad, H., Garg, A., Etori, N., Tiyyala, V. M., Samuel, O., Stutzman, M., Odoom, B., Khudanpur, S., Richardson, S., & Murray, K. (2024). Kreyòl-MT: Building MT for Latin American, Caribbean and Colonial African Creole languages. In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers) (pp. 3083–3110). Association for Computational Linguistics.

Rowe, J., de Gibert, O., Klimaszewski, M., Haley, C., Birch, A., & Scherrer, Y. (2025). *EdinHelsOW WMT2025 CreoleMT system description: Improving Lusophone Creole translation through data augmentation, model merging and LLM post-editing*. In *Proceedings of the Tenth Conference on Machine Translation (WMT), Volume 2: Shared Task Papers* (pp. 1166–1182). Association for Computational Linguistics.

Swolkien D., Semedo, E., Reis, E., Baptista, M do C., Moreira, A.K.T., & Lopes, A. de M (2025). *Manual da disciplina Língua e Cultura Cabo-verdianas 10.º ano* (versão experimental). Ministério de Educação de Cabo Verde/Porto Editora.

Tang, Y., Tran, C., Li, X., Chen, P.-J., Goyal, N., Chaudhary, V., Gu, J., & Fan, A. (2020). Multilingual translation with extensible multilingual pretraining and finetuning (Preprint). https://arxiv.org/abs/2008.00401

Rethinking "Traditional Palenquero" and Tradition in San Basilio de Palenque

Miguel Gutiérrez Maté (Universität Augsburg)

It is widely acknowledged that the revitalization process of Palenquero Creole —which by the end of the last century appeared to be close to completing a process of language shift in favor of Spanish— is being somewhat successful. This is independent of the fact that the varieties spoken by new speakers (or *New Palenquero*, Lipski 2012) may differ considerably from the traditional varieties, and it does not intrinsically lose legitimacy if one acknowledges that, beyond obvious identity-related motivations, tourism —and the "local flavor" produced by a language distinct from Spanish— has also accompanied this revitalization process.

In present-day San Basilio de Palenque, several varieties coexist in frequent interaction (Lipski 2020): (i) the traditional varieties used by elderly speakers who grew up with regular exposure to both Palenquero and Spanish (cases of individuals who grew up with substantially more Palenquero than Spanish are rare, though attested; see Maglia/Moñino 2015); (ii) middle-aged and some additional elderly individuals who passively understand Palenquero but do not speak it; (iii) a generation of cultural and linguistic activists (often originally belonging to the previous group) who, since the late 1990s, have actively engaged in revitalization efforts; and (iv) children and younger speakers who acquired Palenquero as a second language in school.

My approach to the study of Palenquero has consistently involved working with subcorpora of interviews with traditional speakers, rather than simply treating age (and, by extension, traditionality) as an independent or *explanatory* variable in broad sociolinguistic analyses of the Palenquero community as a whole. Such an approach would assume a single (albeit internally variable) Palenquero and would, in effect, place speakers of different generations on the same analytical level, despite the fact that their command of the language, their acquisition contexts, and their linguistic ideologies and expectations differ profoundly.

The aim of this presentation is to discuss the possible arguments for considering *traditional Palenquero* separately from *New Palenquero*. A central argument concerns the interrelationship between language and its ecology: many aspects of the traditional world of San Basilio de Palenque, within which the use of Palenquero found its niche, have disappeared. A secondary argument is purely linguistic: certain variants of the traditional varieties — particularly those that are difficult for non-linguists, including cultural activists, to perceive—are likely to be lost without ever being transmitted to the varieties of new speakers. Let us consider the following example, in which an 82-year-old woman expresses her discomfort when urged to use a cell phone and comments metalinguistically on the word *aló*:

(1) *y'* *a* *konosé* *"aló"* *h[w]é* *pa* *kumé*
1P.SG CMPL know rice FP for eat
'I know *aló* [=rice] TO EAT'

(female, 82 y.o.; fieldwork by the author, July 2017)

First, and most importantly, it is clear that the informant was not only unsettled by the device itself, but even more so by the communicative practices associated with non-face-to-face interaction —a situation that is largely foreign to the use of (traditional) Palenquero. Phone communication typically involves the greeting and phatic marker *aló* ('hello'), a global word that has entered the speech of younger generations in the village and that happens to be homonymous with the traditional Palenquero word *aló* (< Spanish *arroz* 'rice'). Second, the speaker employs the invariable copula and focus marker *hwe* (< Sp. *fue*), a feature that is sporadically attested among traditional Palenquero speakers. In the speech of younger generations, however, this form has been replaced by an inflected copula verb, parallel to the general use of the so-called *ser focalizador* in Colombian Spanish (Gutiérrez Maté 2017, 2020).

References

Gutiérrez Maté, Miguel. 2017. La partícula focal *jue* (< español *fue*) en el criollo palenquero: ¿gramaticalización y/o sustrato?. In: *Revista Internacional de Lingüística Iberoamericana* 15 (30), S. 5-46

Gutiérrez Maté, Miguel (2020): Neues zur Entstehung und Ausbreitung der Konstruktion mit SER FOCALIZADOR: ein Plädoyer für die afrohispanische Hypothese. In: Carolin Patzelt, Elton Prifti (Hg.): Diachrone Varietätenlinguistik: Theorie, Methoden, Anwendungen. Berlin: Peter Lang (Studia Romanica et Linguistica ; 59), S. 175-203

Lipski, John M. 2012. The *New Palenquero*: Revitalization and re-creolization. In Rafael Orozco & Richard J. File-Muriel (eds.), *Colombian Varieties of Spanish*, 21–41. Frankfurt a.M., Madrid: Vervuert, Iberoamericana.

Lipski, John M. 2020. *Palenquero and Spanish in Contact*: *Exploring the Interface.* Amsterdam/Philadelphia: John Benjamins.

Maglia, Graciela & Yves Moñino. 2015. *Kondálo pa bibí mejó / Contarlo para vivir mejor*: *Oratura y oralitura en San Basilio de Palenque.* Bogotá: Editorial Pontifica Universidad Javeriana

Bare Nouns in Cape Verdean Portuguese: Transfer or an innovative pattern?

Nélia Alexandre (Universidade de Lisboa)

This study examines the use of nouns without overt determiners (bare nouns) in Cape Verdean Portuguese (CVP), exploring whether this feature results from transfer from Cape Verdean (CV) or constitutes an emergent grammatical innovation within CVP. Situated within a context of intense and long-standing language contact (Alexandre & Swolkien, 2024, Lopes, 2011), CVP provides a fertile ground for investigating how bilingual grammars interact and reshape morphosyntactic patterns.

In European Portuguese (EP), bare singular count nouns (BNPs) are generally restricted in argument position and typically require a determiner (for a comprehensive study on this topic in EP, see Verdial Soares, 2018) (1). In contrast, Cape Verdean allows BNPs freely in subject and object positions, with definiteness and specificity often inferred pragmatically or through word order and context rather than through overt articles (Alexandre & Verdial Soares, 2005, Pires de Oliveira & Martins, 2017), as in (2).

The data analysed in this study were extracted from a 9,600-word interview from the COrU-PCV corpus (an ongoing collection of 60 hours of recordings made in the cities of Praia, Mindelo and Espargos in Cabo Verde in 2019 and 2022 – cf. Masotti, 2024). This sample of CVP demonstrates that this variety of Portuguese exhibits a higher frequency and wider distribution of BNPs than EP, even in syntactic environments where standard Portuguese would require a definite or indefinite article (as in (3)). According to Masotti (2024: 48), omission of the article is more frequent with common nouns (up to 69%).

Two main hypotheses are evaluated. The transfer hypothesis attributes the pattern to substrate influence from Cape Verdean, suggesting that bilingual speakers map CV nominal structures onto Portuguese, leading to determiner omission. The innovation hypothesis**,** however, proposes that the phenomenon reflects an internally motivated restructuring of the Portuguese determiner system under contact conditions. Rather than simple transfer, CVP may be developing a reduced or pragmatically driven article system, where definiteness is more frequently encoded through discourse context than morphology. From this perspective, BNPs are not merely deviations to the EP grammar or direct replicas of CV grammar but part of an emergent, contact-induced variety with its own grammatical norms.

The analysis of the data suggests that both mechanisms likely interact. While initial expansion of BNPs can plausibly be traced to CV influence, their stabilization and systematic distribution in CVP point toward grammatical reanalysis and standardisation within the Portuguese variety itself. Thus, BNPs in CVP represent not only contact-induced transfer but also the emergence of an innovative nominal pattern shaped by bilingual competence, sociolinguistic factors (to assess in the future), and ongoing language change.

(1) a. ***Cão*** *está* *em vias de* *ex,nção*. dog
is in.way.of exancaon
'Dog is becoming exanct.'

(adapted from Verdial Soares, 2018: 139)

b. *Eu* *vi* ***aligatores*** *na* *Florida*.
I see alligators in.the Florida 'I saw alligators in Florida.'

(adapted from Verdial Soares, 2018: 146)

(2) ***Gatu*** *morde* *rixu*. cat
bite strong
'The cat bit severely.'

(CV, adapted from Alexandre & Verdial Soares, 2005: 344)

(3) a. *estou a falar de Santa Cruz / de São Lourenço dos Órgãos / não estou nem a falar da capital Praia / estou a falar mais* ***de[EP: do]*** *interior /* [I'm talking about Santa Cruz, São Lourenço dos Órgãos. I'm not even talking about the capital, Praia; I'm talking more about **(the)** interior.]

(COrU-PCV, 5_220322F2E15SC, Masotti, 2024)

b. *há departamentos / há quem esteja a par &da/da/das redes sociais / do jornal / de jornal on-line / então há toda uma estrutura como se fosse realmente* ***Ø[EP: um]*** *jornal //* [There are departments... there are those who keep up with social media, newspapers, online newspapers. So, there is a whole structure as if it were **(a)** real newspaper.]

(COrU-PCV, 3_041219M1PSI, Masof, 2024)

Referências bibliográficas

Alexandre, N. & Swolkien, D. (2024). Cape Verde. In U. Reutner (ed.). *Manual of Romance Languages in Africa*. pp. 357-389. de Gruyter.

Alexandre, N. & Verdial Soares, N. (2005). O domínio nominal em Crioulo de Cabo Verde: o puzzle dos Bare Nouns. In I. Duarte & I. Leiria (orgs.), *XX Encontro Nacional da APL*. 335-350. Colibri.

Lopes, A. (2011). *As Línguas de Cabo Verde: uma radiografia sociolinguís,ca*. Diss. Doutoramento. Universidade de Lisboa.

Masof, M. (2024). *Corpus Oral Urbano do Português de Cabo Verde: transcrição e análise de aspectos morfossintá,cos*. Trabalho de Projeto de Mestrado. Lisboa: FLUL.

Pires de Oliveira, R. & Marans, J. (2017). Preliminary remarks on the nominal phrase in Cape Verdean: The semanacs of bare nouns cross-linguisacally. *Glossa: a Journal of General Linguis,cs* 2(1): 100. 1–22, DOI: hjps://doi.org/10.5334/gjgl.157

Verdial Soares, N. (2018). *Bare Nouns in European Portuguese*. Ph.D. Diss. Universidade de Lisboa.

Haitian Creole, Standard English and Classroom Practice in the Turks & Caicos Islands

Nicole Scott, Nadine Logan Ashley & Desmine Bonner Edwards (Mico University College)

This paper investigates the negotiation of Haitian Creole and Standard English in primary and secondary classrooms in the Turks & Caicos Islands, a multilingual context where Turks & Caicos English Creole, Haitian Creole, Spanish and other Caribbean English-lexifier varieties co-exist alongside Standard English - the official language of instruction. In settings where students' everyday communicative repertoires differ significantly from the language ideologically privileged in schooling, language becomes not only a pedagogical medium but a site of identity construction, legitimacy, power and inequality (Baker, 2011; Pennycook, 2010). Teachers, positioned at this intersection, navigate complex linguistic ecologies that influence both classroom interaction and broader sociolinguistic ideologies.

Guided by Translanguaging Theory (Li Wei, 2018; García & Li Wei, 2014) and Communicative Language Teaching (Savignon, 2002; Hornberger, 2003), this study examines how teachers in both primary and secondary classrooms negotiate these linguistic realities in everyday practice among L1 Haitian Creole speakers. Through interviews and classroom observations, the paper explores teachers' language choices, their responses to Haitian Creole use by students and their beliefs about Standard English, code-switching and correcting language errors. The research is informed by extant work on Creole-speaking classrooms which recognizes that language attitudes and pedagogical practices often reflect broader sociopolitical hierarchies (Siegel, 2006; Hornberger & McKay, 2010) and that teacher language ideologies shape classroom norms (Bilbrough, 2009).

Preliminary findings reveal a multifaceted set of strategies among teachers in dealing with language variation. These include matching students with peers who already have some proficiency in English (cf. Creese & Blackledge, 2010), using gestures and scaffolding techniques to support comprehension, recommending extra classes focused on basic communicative English and selective code-switching to facilitate understanding. At the primary level, teachers invest more in extra instructional support and adaptive practices. Conversely, some teachers, particularly in secondary classrooms, report a sense of pedagogical constraint: despite awareness of students' limited comprehension, they continue to teach predominantly in Standard English, operating under the assumption that immersion will yield eventual proficiency (cf. Romaine, 2013). This range of approaches reflects varying degrees of ideological alignment with Standard English norms as well as differing conceptions of Creole varieties as either pedagogical resources or obstacles to language development (cf. Jaffe, 1999; Lippi-Green, 2012).

Challenges identified include teachers' limited training for linguistically diverse classrooms, persistent stigma toward non-Standard varieties and inconsistent correction practices that oscillate between explicit feedback and implicit reformulation. Observations indicate that moments of code-switching often perform critical pedagogical functions such as repair, clarification and affective support yet these practices are rarely acknowledged in formal policy. The contrast between teachers' stated beliefs and observed practices underscores the

disjuncture between language ideologies and classroom realities (cf. Li Wei, 2018).

The comparative dimension of the analysis (primary vs. secondary; urban vs. rural) highlights systemic patterns that extend beyond individual classrooms, pointing to structural tensions in language education policy and practice. Since the study situates local classroom dynamics within broader discourses on linguistic legitimacy, identity and educational equity, it contributes empirical insight to Caribbean sociolinguistics and Creole studies and offers implications for teacher education and language policy reform.

References

Baker, C. (2011). *Foundations of bilingual education and bilingualism* (5th ed.). Multilingual Matters.

Bilbrough, N. (2009). *Teachers' linguistic and literacy beliefs and practices in Creole-speaking communities*. Journal of Caribbean Language and Education.

Creese, A., & Blackledge, A. (2010). *Translanguaging in the bilingual classroom: A pedagogy for learning and teaching?* The Modern Language Journal, 94(1), 103–115.

Hornberger, N. H., & McKay, S. L. (2010). *Sociolinguistics and language education*. Multilingual Matters.

Jaffe, A. (1999). *Ideologies in action: Language politics on Corsica*. Mouton de Gruyter.

Lippi-Green, R. (2012). *English with an accent: Language, ideology and discrimination in the United States* (2nd ed.). Routledge.

Li Wei. (2018). *Translanguaging as a practical theory of language*. Applied Linguistics, 39(1), 9–30.

Pennycook, A. (2010). *Language as a local practice*. Routledge.

Romaine, S. (2013). *Sociolinguistic precepts and language teaching*. In *The Cambridge Handbook of Language Policy*. Cambridge University Press.

Siegel, J. (2006). *Roots of Creole structures: Weighing the contribution of substrates and superstrates*. John Benjamins.

Louisiana Creole as spoken in New Orleans:
Newly 'discovered' recordings from 1935

Oliver Mayeux (University of Cambridge)

All but very sparse descriptions exist of the variety of Louisiana Creole once spoken in New Orleans, despite the city's historical importance as a center of Creole language and culture (Mercier 1880; Lane 1935; Harrison 1882). In all likelihood no speakers of this variety are still living today—or, if they are, they are extremely elderly and can be counted in the single digits—meaning that historical recordings of natural speech would be the most promising source for documentation and description. Until now, no such materials had been found.

This paper presents the first detailed documentation and description of Louisiana Creole as spoken in New Orleans from the mid-19th to mid-20th century. The description is based on a hitherto unknown collection of phonograph recordings made in the 1930s. Unlike the much better-known Lomax phonograph recordings, the materials described in this paper were collected for linguistic purposes.They feature a surprisingly wide range of genres and registers as well as multilingual practices, and provide fascinating insights into this variety of Louisiana Creole in its local urban context. A full description of the collection is given and I will share the story of their 'discovery'. Using genealogical and local archival records, it has been possible for me to recover the identity and biography of the individuals recorded which, in turn, will provide new sociolinguistic insights.

The morphosyntactic and phonological characteristics of variety of Louisiana Creole as spoken in New Orleans which can be recovered from these recordings provide new insights on the diachronic development and historical sociolinguistics of this now critically-endangered language. The recordings also hold implications for language revitalisation, which will briefly be discussed to propose some new ways for Creolists to understand the 'social lives of legacy materials' (Dobrin & Schwartz 2021).

References

Dobrin, Lise & Saul Schwartz. 2021. The social lives of linguistic legacy materials. *Language Documentation and Description*. Language

Documentation and Description 1-36 Pages.
https://doi.org/10.25894/LDD12.

Harrison, J. A. 1882. The Creole Patois of Louisiana. *The American Journal of Philology* 3(11). 285. https://doi.org/10.2307/287099.

Lane, George S. 1935. Notes on Louisiana-French II: The Negro-French Dialect. *Language* 11(1). 5. https://doi.org/10.2307/408911.

Mercier, Alfred. 1880. Étude sur la Langue Créole en Louisiane. *Comptes-Rendus de l'Athénée Louisianais* 5. 378–383.

Jaffna Portuguese at present

Patrícia Pardal da Costa (Universidade de Lisboa)

Sri Lanka Portuguese (SLP) is a Portuguese-lexified creole that emerged during the period of Portuguese colonial rule in Sri Lanka. Once spoken across diverse regions and communities, SLP has undergone substantial contraction over the past century and is now spoken exclusively within Portuguese Burgher communities, primarily in the Ampara, Batticaloa, and Trincomalee districts of Eastern Sri Lanka. By all accounts, SLP is classified as endangered, with approximately 1,300 remaining speakers (Cardoso et al. 2019; Pereira 2022).

The Northern city of Jaffna was assumed to no longer host active speakers, yet recent exploratory fieldwork has identified the presence of a small number of speakers. This paper therefore reports on the synchronic status, distribution, and degree of vitality of Jaffna Portuguese.

Based on sociolinguistic survey data and language samples (narratives, conversation, and elicitation sessions) collected between 2020 and 2024, we provide an overview of this community, including its geographic and demographic distribution, intergenerational transmission, and domains of use. We also highlight linguistic features distinguishing Jaffna Portuguese from eastern varieties, including lexical and morphosyntactic patterns influenced by the dominant regional language, Tamil.

Developed in isolation from the wider SLP-speaking community, Jaffna Portuguese appears to be severely endangered: of the 51 individuals surveyed, 16 reported active use of the language, predominantly among older adults, with no evidence of transmission to younger generations. These findings underscore the urgent need for further documentation and revitalization efforts to preserve this variety of Sri Lanka Portuguese.

References

Cardoso, Hugo C, Radhakrishnan, Mahesh, Costa, Patrícia & Pereira, Rui. 2019. Documenting modern Sri Lanka Portuguese. In *Documentation and Maintenance of Contact Languages from South Asia to East Asia: issues on theory and practice*., Mário Pinharanda Nunes & Hugo C. Cardoso (eds), 1-33.

Pereira, Rui. 2022. *O Crioulo português do Sri Lanka: Uma Perspetiva de Conservação Linguística*. Master thesis, Universidade de Lisboa.

Philippine Spanish and its relationship to Chabacano

Patrick O. Steinkrüger (Universität Göttingen)

In my talk I will mainly argue that Chabacano (C) and Philippine Spanish (PS) are linguistically and historically closely related and that the traditional Mardika-story is just a folk legend, probably from the 19th century, without any historical or linguistic basis. It tells us, in a nutshell, that freed slaves (mainly christianised) speaking Ternate Malay (a Malay-based creole influenced by Papuan) came over from the Moluccas to the Philippines and established the first linguistic community of Chabacano in the Bay of Manila (e.g. explained in Whinnom 1956).

Instead, we have have a lot of evidence - directly and indirectly - that C is the result of an in situ creolization of PS at the earliest during the late 17th century in certain urban areas (especially ports and fortresses) with the influence of Spanish pidgins (*español del parián*; see also Bakker & Parkvall, upcoming). Geographically, where PS had and has its hotspots are currently the hotspots of C (own fieldwork experience & data from the literature). But the lack of documentation before the 19th century makes it difficult to reconstruct the linguistic situation of the time between the 16th and 18th centuries. Documentation both of C and PS starts very lately in the 19th century.

Consequently, we relay also on linguistic arguments: phonology (e.g. stage of the phonetic development of *f-* and the sibilants), grammatical features (as an example will be chosen the object-marking) and lexical items (identical in C, PS and also the pidgins).

Furthermore, the loans in the different Philippine languages let us permit to look back to a complex cultural and linguistic scenario envolving Amerindian, Spanish, Chinese, Japanese, Philippine and even Indian and African people in Early Modern Times (see e.g. Tremml-Werner 2020) with the influence of Portuguese and Portuguese-based creoles and pidgins (which explains some similarities with them and C; see Cardoso 2012) in the Philippines. Slavery played an important role (contra McWhorter‘s claim “total dissociation from slavery“).

References (small selection)

Bakker, Peter & Mikael Parkvall (2021): "Spanish pidgins", in: *Contact varieties of Spanish and Spanish-lexified contact varieties*, ed. by Leonardo Cerno, Hans-Jörg Döhl, Miguel Gutiérrez Maté, Robert Hesselbach, Joachim Steffen (HSK: *Handbooks for Linguistics and Communication Science / Handbücher zur Sprach- und Kommunikationswissenschaft*) (with Mikael Parkvall). Submitted September 1, 2021. Revised version sent Jan. 26 2022 (to be sent out to reviewers).

Blumentritt, Ferdinand (1882/1885): *Vocabular einzelner Ausdrücke und Redensarten, welche dem Spanischen der philippinischen Inseln eigenthümlich sind* [zweibändiger Separatdruck aus dem XVI. Jahresbericht der Communal-Ober-Realschule in Leitmeritz]. Leitmeritz (Böhmen). Cardoso, Hugo et al. (eds.) (2012): *Ibero-Asian Creoles. Comparative perspectives*. Amsterdam/Philadelphia: Benjamins.

Fernández, Mauro & Sippola. Eeva (2018): "On the chronology of of the formation of the Chabacano varieties: a reply to Parkvall & Jacobs 2028", *Journal of Ibero-Romance Creoles* 8, 38-56.

Litamahuputty, Betty H. J. (2012): *Ternate Malay : grammar and texts*. (PhD-thesis). Leiden University.

Perce, Donna & Ronald Otsuka (eds.) (2009): *Asia and Spanish America* (1500-1850). Denver: Art Museum.

Tremml-Werner, Birgit (2020): „The Global and the Local in Early Modern Manila`s Communication Spaces", in: Gommans, Jos & Ariel Lopez (eds.), *Philippine Confluences. Iberian, Chinese and Islamic Currents, c. 1500-1850*. Leiden: U.P., 191-215.

Whinnom, Keith (1956): *Spanish Contact Vernaculars in the Philippine Islands*. Hong Kong: Hong Kong University Press.

Carriols Creole: slow or sudden genesis?

Peter Bakker & Joost Robbe (Aarhus Universitet)

There is an ongoing discussion about whether creoles develop quickly, i.e. within a generation or two, or gradually, i.e. over a course of, say 50 to 100 years. Carriols (also known as Virgin Islands Dutch Creole) is the creole with the highest quantity of historical printed works, in addition to a considerable quantity of manuscripts. It is an ideal language to study creole genesis and development. In the Carriols project, we have studied virtually all documents in the language, from the handwritten letters in the 1730s to printed resources from the 18th and 19th century to sound recordings of the 1980s.

The null hypothesis of the investigation was that Carriols had developed gradually. Early texts (e.g. many of the letters of the enslaved, written between the 1730s and the 1760s), appeared clearly closer to Dutch, and the texts collected throughout the 20th century looked furthest removed from Dutch. However, closer scrutiny leads to the inescapable conclusion that the genesis of Carriols happened quickly, and at one location.

The proof includes historical facts and linguistic arguments. First, the three islands were populated in 1672 (St. Thomas), and from there St. John in 1718 and Saint Croix in 1733, over a period of 61 years. This means that the Creole must have jelled on St. Thomas, before being transplanted. Second, the missionaries had chosen for Dutch orthography, which gives the written texts a more conservative flavour. The study of rhyme pairs in hymn books, however, shows that only five or seven vowels were used from the beginning, and that many final consonants were dropped, especially dentals, indicating that pronunciation deviated considerably from Dutch despite being written like Dutch. Third, there are no dialectal differences between the three islands where the language is spoken, suggesting a single birth of the language on St. Thomas. Fourth, the set of non-Dutch words (notably Spanish/Papiamentu, African languages, dialectal Dutch) is identical on all islands, and pretty constant from the 18th to the 20th century. Fifth, the study of the hundreds of proverbs documented throughout a period of over 120 years show only conservative creole traits, suggesting that they were transmitted from the earliest period. Sixth, preverbal TMA markers are typical for all creoles (also other markers of tense, mood and aspect can be found in creoles), and all of those that are attested in the 20th century can already be found in the earliest letters. Seventh, African substrate traits are much more frequent in the 20th century texts than in the historical record, like verb fronting, reduplication, serial verbs and perhaps ideophones. However,

there is evidence from the 18th century that they were already present in the earliest sources. Finally, metalinguistic remarks point to the existence several registers of the language, from the speakers of European descent who used more Dutch-like constructions, to the more basilectal speech of the local creoles as attested in early colloquial dialogues. The conclusion of a sudden genesis is inescapable.

On a contact Malay inhibitory effect on L3 phonological sensitivity

Peter Slomanson, Mikko Pajunen & Heini Kallio (Tampereen yliopisto)

The contact Malay varieties of Sri Lanka, over almost four centuries on the island, have incorporated a range of new morphological and syntactic features into their grammar(s). At the same time, they have also adopted a range of areal phonological features, both segmental and prosodic. One of the features that stands out strongly is apparent quantity contrasts, which are most evident in the many intervocalic geminates, including both obstruents and sonorants. The source language for this feature was debated in Nordhoff (2012) and Smith (2012), partly with reference to Tapovanaye (1995). We will not address that question however. In this paper, intervocalic gemination is simply treated as areal, since it is absent from all known Malay varieties spoken outside Sri Lanka. We claim that there is a typological continuum with respect to how and whether gemination is instantiated in a language, and several of the languages of Sri Lanka occupy points on that continuum in which gemination is phonetically present.

In a much more recent contact scenario, large numbers of Sinhala speakers have been moving to Finnish cities for education and employment opportunities. We will compare experimental results measuring sensitivity to singleton-geminate contrasts in Finnish with Sinhala speakers in both countries and speakers of contact Malay in Sri Lanka as the participants. The work is part of a larger second language phonetics project in which recent Sri Lankan migrants are tested for their receptive and productive abilities with respect to quantity contrasts in Finnish, using automated lexical discrimination testing and acoustic analysis with Praat. One significant finding of that project is that L1 Sinhala speakers perform strikingly better than L1 English speakers do.

We will discuss the finding that the contact Malay speakers who we tested similarly performed worse than those Sinhala speakers who lack knowledge of Malay, including a control group in Sri Lanka with no exposure to Finnish. We attribute this to the fact that the contact Malay singleton-geminate contrasts are phonologically-conditioned and hence allophonic, whereas they are clearly phonemic in Sinhala, as they are in Finnish. The difference is consistent with an established effect in the second language acquisition literature (for example Curtin et al 1998) in which contrasts that may be phonetically present in the L1, but not phonemic, are difficult for learners to perceive and acquire in an L2 in which they are phonemically contrastive.

The majority of contact Malay speakers in Sri Lanka are also high proficiency

speakers of Sinhala, the source of our initial hypothesis that they would perform similarly to the Sinhala participants without Malay. Consequently, our most significant contact linguistic finding is that acquisition of Sri Lankan Malay, including by fluent L2 speakers, appears to have an inhibitory effect on the Sinhala advantage that we identified in the perception of quantity contrasts in unknown Finnish words. Regardless of order of acquisition of the respective languages, fluency in Sinhala did not enhance performance on the Finnish tests for most Malay speakers. The fact that such an effect would be visible even in late acquirers of the inhibiting phonology is significant as a third language acquisition finding. This highlights the empirical value of investigating the cross-linguistic interaction of donor language and contact language phonologies in bilingual speech communities and provides us with a window into how the interaction could influence the course of additional language acquisition at an early stage in the process.

References

Curtin, S., Goad, H. and Pater, J. 1998. Phonological transfers and levels of representation: the perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research*, 14, 389-405.

Nordhoff, S. 2012. Establishing and Dating Sinhala Influence in Sri Lanka Malay. *Journal of Language Contact*, 5, 23-57.

Smith, I. 2012. Comments on Nordhoff's "Establishing and Dating Sinhala Influence in Sri Lanka Malay", *Journal of Language Contact*, 5, 58-72.

Tapovanaye, S. 1995. Vowel Length in Sri Lankan Malay. Unpublished University of Hawai`i MA thesis.

Naming in and for creole languages

Rachel Selbach (independent scholar), Eeva Sippola (Helsingin yliopisto)
& Philipp Krämer (Vrije Universiteit Brussel)

This talk examines naming practices in creole languages and in scholarly discourse, focusing on how names function both *in* creole-speaking communities and *for* creole languages within academic analysis. It explores what constitutes a "creole name", how input languages shape creole onomastics, and under which conditions *creole* operates as a glossonym versus a class noun referring to a group of languages.

While onomastics has a robust research tradition, also within contact linguistics (Sandnes 2016), and although there is substantial work on glossonyms more broadly (e.g. Cysouw & Good 2013; Ainiala & Östman 2017), comparatively little research has examined naming practices within Creole Studies. Existing scholarship on the toponymy of territories where creole or pidgin languages are spoken remains largely case-based. Other branches of onomastics, such as personal names or product and brand names, have received even less attention. Only very few studies have investigated how names and labels circulate within creole-speaking societies themselves (cf. Léglise & Migge 2006; Migge & Léglise 2013, 2015; Krämer et al. 2022). Until recently, limited attention has been given to how the labels chosen by linguists compare with those preferred by speakers, or how naming practices affect linguistic ideologies, authority, and identity beyond mere categorization (see Bakker 2026, Levisen 2026, Melchor 2026, Klimenkowa 2026 for important contributions to these questions).

We will provide an overview of central questions surrounding the use and significance of creoles and pidgins in onomastic research. Drawing on examples from diverse creoles, we will offer a meta-analysis of existing research on place names, person names, commercial names, and group names. The results show that while many creole names originate rather uniformly in the superstrate languages, they subsequently continue to undergo varying independent phonological and semantic adaptations. The discussion of language names further illustrates that these do more than identify languages: they perform social and ideological work, contributing to categorization, identity formation, and the re-negotiation of social and sociohistorical conditions. Naming thus emerges as a constitutive act that shapes linguistic conceptualization. Recognizing this performative dimension is essential for understanding processes of Language Making and for interrogating the assumptions embedded in Creole Studies.

References

Bakker, Peter. 2026. Names for contact languages: An historical overview of the evolution of terms in the field of pidgin and creole languages. *Journal of Pidgin and Creole Languages* 41(1).

Ainiala, Terhi & Jan-Ola Östman. 2017. *Socio-onomastics and pragmatics*. Amsterdam: John Benjamins.

Cysouw, Michael & Jeff Good. 2013. Languoid, doculect and glossonym: Formalizing the notion 'language'. *Language Documentation and Conservation* 7. 331–359.

Krämer, Philipp, Eric Mijts & Angela Bartens. 2022. Language making of creoles in multilingual postcolonial societies. *International Journal of the Sociology of Language* 274. 51–82. https://doi.org/10.1515/ijsl-2021-0018

Klimenkowa, Alla. 2026. Naming creole varieties on the Cape Verde Islands and in Upper Guinea from the perspective of language ideologies. *Journal of Pidgin and Creole Languages* 41(1).

Léglise, Isabelle & Bettina Migge. 2006. Language-naming practices, ideologies, and linguistic practices: Toward a comprehensive description of language varieties. *Language in Society* 35(3). 313–339. https://www.jstor.org/stable/4169502

Levisen, Carsten. 2026. Terminologies in Crisis: Challenges and Insights from Multipolar Metalinguistics. *Journal of Pidgin and Creole Languages* 41(1).

Melchor, Jillian. 2026. What do we talk about when we talk about Chavacano? Language names in the archaeology of knowledge. *Journal of Pidgin and Creole Languages* 41(1).

Migge, Bettina & Isabelle Léglise. 2013. *Exploring language in a multilingual context: Variation, interaction and ideology in language documentation*. Cambridge, UK: Cambridge University Press.

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

Sandnes, Berit. 2016. Names and language contact. In Carole Hough (ed.), *The Oxford handbook of names and naming,* 540–553. Oxford: Oxford University Press.

Revisiting Romance Pseudo-Relatives in Lesser Antillean Creole and Beyond

Rasul Dent, Thibault Clérice, Pedro Ortiz Suarez & Benoît Sagot (Inria Paris Centre)

Among French Creoles, imperfective and future markers are among the most salient isoglosses. Most of these languages fall into one of two groups. Lesser Antillean South American Creoles use *ka* to mark imperfectivity (habitual and progressive) and *ké* for the future, as well as variants of *ka* + *alé* 'going' for the near future. In contrast, Louisiana, Haitian, and Mauritian Creoles use variants of *apé* as a progressive marker that in some cases extends into prospective and/or generic aspects.

In contrast to *apé*, which is easily traceable to the locative periphrastic *être après (à) faire* 'to be near (at) doing', the etymology of *ka* and *ké* remains an open question. Due in part to the geographic distance between the Lesser Antilles and French Guiana, the shared use of these markers has been at the center of several debates regarding possible genetic relations between French Creoles (see Jennings & Pfander (2018) and Parkvall & Jacobs, (2022)).

It has long been hypothesized that *ka* and *ké* are in some way connected to the *dèt k*i progressive used in Saint-Barthélemy French, which has in turn been linked to the *lété ki* progressive of Réunion Creole (Corne, 1999). Focusing specifically on Martinican Creole, rather than the comparative question, Damoiseau (2008) proposes an etymology that is nevertheless compatible with such a link. More specifically, he traces *ka* back to a fusion of another locative progressive, *être à faire* 'to be at doing' and deictic pseudo-relative clauses. The latter, which are well attested in Romance, have the surface structure of relative clauses, but typically serve to characterize actions rather than specify nominal relations.

In this work, we support Damoiseau's explanation of *ka* as the fusion of a pseudo-relative linker *k* and second marker *a* with new comparative data. First, we note that in Sri Lanka Portuguese Creole, *ki,* believed to be derived from the Portuguese relative pronoun *que,* serves as an optional marker of habitual aspect, as well as deverbal adjectives and nouns (Smith 1985, 2013). Then, we highlight that a progressive *estar que* which closely parallels the *dèt/lété ki* constructions has been observed in Northern Peru and Southern Ecuador (Tiegs, 2022).

Building on these observations, we propose that *ké* is a second case of the relative pronoun *k* and an originally separable progressive-future marker, namely *é*. Specifically, we consider Klingler's (2003) observation of a puzzling coincidence with the Pointe Coupée dialect of Louisiana Creole, which is the only French Creole known

to use such an *é*. Namely, in Lesser Antillean (and variably Guianese) Creoles, the negative particle *pa* shifts to *pé* before *ké*, and only before *ké.* When the *k* is elided in rapid speech, this yields the sequence *pé é,* and thus directly parallels the typical negative future of Pointe Coupée Creole. In our view, the specificity of the *pa* to *pé* change suggests the morphologicalization of an earlier structure in which *pa* and *é* were directly adjacent.

Confidence Scores for Reliable Linguistic Analysis of Large Automatically Transcribed Oral Corpora: Haitian Creole

Rayan Ziane (Université de Caen Normandie)

The rise of automatic speech processing technologies now enables the large-scale transcription of oral archives for historically under-resourced languages. However, this abundance of new data poses a major challenge for linguists: how can these raw transcriptions be reliably exploited when their quality is heterogeneous and exhaustive manual annotation remains out of reach? This paper addresses this issue through an analysis of part-of-speech (POS) annotation strategies applied to the Radio Haïti-Inter (RHI) corpus — 1,300 hours of Haitian Creole (HC) speech automatically transcribed (Havard, Govain, Lecouteux, & Schang, 2025) — and demonstrates how the systematic integration of confidence scores can enhance the reliability of linguistic analyses based on such resources (Havard et al., 2026).

Our contribution is twofold. First, we adapt a POS tagging model to the oral domain (Grobol et al., 2022; Ramponi & Plank, 2020): starting from the multilingual XLM-RoBERTa model (Conneau et al., 2020, XLM), we perform continued pre-training on the entire RHI transcriptions (13 million tokens), followed by fine-tuning on existing Universal Depen-dencies treebanks for HC (Zeman et al., 2025) UD-Haitian-Creole-Adolphe (1) and UD-Haitian-Creole-Autogramm (2) (Kahane, Pierre-Louis, Jagodzińska, & Savary, 2024), and finally adapt the tagger to the target data using a manually annotated subset. This adaptation proves crucial, yielding a +4.3 point accuracy gain over a non-adapted multilingual model and achieving an overall accuracy of up to 92.9%. Second, we systematically generate, for each annotated token, a confidence score derived from the normalized entropy of the probability distribution over POS labels inspired by Laptev and Ginsburg (2022).

Manual validation of a sample of 562 utterances (1,728 tokens) confirms the relevance of these scores as reliability indicators. A significant positive correlation is found between confidence scores and prediction accuracy ($r = 0.40$, $p < 0.001$), with accuracy increasing steadily across deciles: from 54.3% for the least confident tokens to 96.1% for the most confident ones. A significant negative correlation is also observed between confidence scores and automatic transcription character error rates (Pearson $r = -0.48$), confirming that higher scores are associated with more reliable transcriptions.

These findings pave the way for a principled exploitation of the corpus by the linguistic

community. Rather than treating annotations as a homogeneous dataset, researchers can now filter data according to their objectives: a high confidence threshold (≥ 0.95) will support fine-grained analyses of well-documented syntactic phenomena (e.g., verb phrase struc-ture, TMA markers), while a lower threshold may be acceptable for large-scale micro-syntactic or lexicometric studies. This approach is all the more relevant as error sources can now be clearly identified: beyond inherently ambiguous categories, the tagger proves sensitive to orthographic variations specific to oral speech (e.g., contractions such as *gon* for *gen yon*, as transcribed by the ASR system), non-standardized French borrowings, and disfluencies (tagged as X, with an accuracy of only 40%).

Our demonstration is based on the 50-hour publicly accessible subset of the RHI corpus (Havard et al., 2026), which includes not only POS annotations but also confidence scores for each token. This transparency regarding annotation reliability represents a methodological advance for Haitian Creole studies and, more broadly, for the processing of low-resource languages: it reconciles the scale of automatically processed data with the robustness requirements of linguistic analysis. Finally, we show how these metrics can guide future manual annotation efforts toward the most uncertain segments, thereby optimizing the trade-off between human investment and quality gains.

References

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., . . . Stoyanov, V. (2020, April). *Unsupervised Cross-lingual Represen-tation Learning at Scale*. arXiv. Retrieved 2024-04-19, from http://arxiv.org/abs/1911.02116 (arXiv:1911.02116 [cs]) doi: 10.48550/arXiv.1911.02116

Grobol, L., Regnault, M., Ortiz Suarez, P., Sagot, B., Romary, L., & Crabbé, B. (2022, June). BERTrade: Using contextual embeddings to parse Old French. In N. Calzolari et al. (Eds.), *Proceedings of the thirteenth language resources and evaluation conference* (pp. 1104–1113). Mar-seille, France: European Language Resources Association. Retrieved from https://aclanthology.org/2022.lrec-1.119/

Havard, W., Govain, R., Lecouteux, B., & Schang, E. (2025). Self-Supervised Models of Speech Processing for Haitian Creole. In *Interspeech 2025* (pp. 4018–4022). doi: 10.21437/Interspeech.2025-1852

Havard, W., Ziane, R., Menclé, M., Coavoux, M., Lecouteux, B., & Emmanuel, S. (2026). Radio haiti-inter: a large-scale annotated corpus of spoken haitian creole.

In *Proceedings of the thirteenth language resources and evaluation conference*. Palma, Spain. (Accepté)

Kahane, S., Pierre-Louis, C., Jagodzińska, S., & Savary, A. (2024, February). *The first Haitian Creole treebank*. Retrieved 2026-01-26, from https:// hal.science/hal-04667550 (Published: Peer reviewed poster in the 2nd UniDive Workshop)

Laptev, A., & Ginsburg, B. (2022). Fast entropy-based methods of word-level confidence estimation for end-to-end automatic speech recognition. 2022 *IEEE Spoken Language Technology Workshop* (SLT), 152-159. Retrieved from https://api.semanticscholar.org/CorpusID:254853835

Ramponi, A., & Plank, B. (2020, December). Neural unsupervised domain adaptation in NLP—A survey. In D. Scott, N. Bel, & C. Zong (Eds.), *Proceedings of the 28th international conference on computational linguistics* (pp. 6838–6855). Barcelona, Spain (Online): International Committee on Computational Linguistics. Retrieved from https://aclanthology.org/ 2020.coling-main.603/ doi: 10.18653/v1/2020.coling-main.603

Zeman, D., et al. (2025). *Universal dependencies 2.17*. Retrieved from http:// hdl.handle.net/11234/1-6036 (LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL))

(1) https://github.com/UniversalDependencies/UD_Haitian_Creole-Adolphe.git

(2) https://github.com/UniversalDependencies/UD_Haitian_Creole-Autogramm.git

Body Part Semantics and the Form–Meaning Interface in Kabuverdianu

Rebecca Arkell (Universität Göttingen)

The body, a universal aspect of human experience, offers a rich domain for exploring the relationship between language, culture, and cognition. While universal patterns in body part terminology have been proposed (Brown, 1976; Andersen, 1978), a central methodological challenge in cross-linguistic research lies in determining precisely what linguistic forms denote, requiring careful attention to both intensional and extensional aspects of meaning (Enfield, 2006). Considerable cross-linguistic variation exists in terms of which parts of the body are singled out for reference (Enfield et al., 2006). Around one third of the world's languages employ one single term referring to both the concepts of 'arm' and 'hand', and a similar proportion exists for the lexical separation of the (English) concepts of 'leg' and 'foot' (Brown, 2008; Witkowski & Brown, 1985).

Body part semantics thus offers a productive site for investigating the relationship between form and meaning in Kabuverdianu and, more broadly, for understanding processes of creolisation. Research on the semantic structure of creoles has seen relatively little attention (Huttar, 2008), and substrate influence in semantics is still underexploited (Parkvall, 2000). Indeed, it seems that creoles differ most clearly from their superstrates not in their lexical stock but in their syntactic, morphological and semantic structures (Huttar, 2008). For instance, *mó* (<Portuguese 'mão') does not only refer to 'hand' but can extend to the whole upper limb (depending on the context). Similarly, Wolof and Mandinka, two major substrate languages for Kabuverdianu, use one term to refer to the whole upper limb, *loxo* and *buloo*. Comparable patterns are attested in other creoles, such as *han* in Jamaican Creole and *an* in Sierra Leone Krio, referring to both 'hand' and 'arm'.

This study investigates body part categorisation, conceptualisation and the semantic extension and grammaticalisation of body parts in Badiu, the Santiago variety of Kabuverdianu. The analysis draws on data collected through twelve weeks of field-based elicitation on Santiago Island. Body part terms participate in grammaticalisation processes: for example, *kabésa* ('head') functions as the most common reflexive strategy (Baptista, 2003). The grammaticalisation of body part terms towards reflexive markers and concepts of 'self' is especially widespread in Africa, though not restricted to it (Heine 2014; Schladt 2000). While extensions of terms such as 'mouth' and 'head' are cross-linguistically frequent (e.g., Portuguese *boca* 'mouth' → *bocado* 'a small quantity', 'a short amount of time'; Farsi *dahan* 'mouth' → *dahane* 'opening, aperture'), Kabuverdianu also displays less typical developments. For instance, the quantifier *pitada* ('a pinch of something', < Portuguese) has extended to mean 'vagina', motivated by the iconic resemblance between the pinching gesture and female genital anatomy. Such meaning extensions are grounded in embodied experience (Lakoff & Johnson, 1980). Similarly, the metonymic extension 'mother' → 'womb' is reflected in *madri*, which denotes both 'mother' and 'womb'.

Given the particular socio-historical settings of linguistic and cultural contact in which creoles emerge, the study of body part semantics in Kabuverdianu offers fertile ground for understanding the relationship between form and meaning through cognitive and cultural understandings of meaning (Lakoff & Johnson, 1980; Kraska-Szlenk, 2020; Wierzbicka,

2007; Enfield et al., 2006). Lexical semantic typology shows how meaning patterns are shaped both by language structure and by broader cognitive and cultural factors, helping to explain the diversity of lexical systems across languages (Evans 2010). The findings of this ongoing research align with the view that semantic restructuring and innovation are expected outcomes of language contact (Mufwene 2008) and contributes new empirical evidence for discussions of substrate and superstrate influence in creoles as well as research on embodiment.

(Pseudo-)Reflexivity in Santomean Portuguese: Evidence from an unstable grammar

Rita Gonçalves, Ana Espírito Santo & Nélia Alexandre (Universidade de Lisboa)

Several studies have discussed whether the clitic SE is preserved or omitted in African varieties of Portuguese (AVPs), particularly in (pseudo-)reflexive structures. Based on CORPUS-AFRICA data (Bacelar do Nascimento et al., 2006), Mendes and Estrela (2008) assert that SE omission is common in all AVPs, especially in Santomean Portuguese (STP), which has historically been in contact with Gulf of Guinea Creoles (GGCs). Madureira (2023) analysed the spoken PALMA-PST corpus (Hagemeijer et al., 2022), showing that inherent clitics are omitted more frequently than reflexive clitics. Level of education appears to be a factor, with SE omission decreasing as level of education increases.

Reflexive pronouns in European Portuguese (EP) are intrinsically anaphoric forms that establish coreference between two arguments in the same sentence, usually the subject and a complement, although playing distinct thematic roles (1). In contrast, inherent or pseudo-reflexive pronouns occur with certain verbs without true semantic reflexivity: the pronoun does not refer to the antecedent or duplicate thematic roles (2). They differ from proper reflexive pronouns, in which referential identity and compositional interpretation of the action on oneself are present. In inherent pronouns, however, SE cannot be interpreted autonomously. For this reason, only reflexive pronouns can be paraphrased by the form *a si mesmo* (to oneself) (e.g., Brito et al. 2003; Lobo, 2013).

Reflexive strategies in GGCs differ structurally from those in EP. Santome, the dominant creole in São Tomé and Príncipe, among other strategies, exhibits reflexivity without any overt marking (Hagemeijer, 2007, 2011) (3).

The aim of this talk is to answer the following research question: Does the experimental data confirm the trends identified in corpus data, namely greater omission of pseudo-reflexive clitics than reflexive clitics among undergraduate informants? We hypothesize that the experimental data converge with the corpus findings, namely that: (i) omission rates for both reflexive and pseudo-reflexive structures are equivalent across experimental and corpus data among educated speakers; (ii) pseudo-reflexive structures exhibit higher SE omission rates than reflexive structures, mirroring corpus findings and reflecting the fact that, in these cases, the clitic does not contribute semantically to the sentence.

We have run a written production task (cloze test) composed of 101 items, including 56 items with SE structures and 45 fillers. From the 56 SE items, only 24 correspond to the target structures of this study. The task was run in pencil and paper, at the University of São Tomé, to 23 participants.

The findings show a balanced distribution of SE omission and realisation, both with pseudo-reflexive constructions (52.2% and 44.6%, respectively) and reflexive structures (47.8% vs. 46.7%). These rates are higher than those reported by Madureira (2023), infirming our first hypothesis. Moreover, although the rates of pseudo-reflexive SE omission are slightly higher than that of reflexive SE omission, we cannot confirm with a high degree of certainty our second

hypothesis, given the minimum difference between the two structures. In fact, the data highlight that the use of reflexive and inherent clitics in STP is unstable, even at higher levels of education. This is not surprising when we consider that STP has been historically in contact with Santome, a language that exhibits a strategy of reflexivity without overt marking.

(1) *A* *criança* *magoou-se* *(a si mesma).*
DET.FSG child hurt-PST.3SG-REFL herself
'The child hurt herself'.

(Brito et al. 2003, p. 807)

(2) *A* *Maria* *admirou-se* *(*a si mesma)*
DET.FSG Maria surprise-PST.3SG-PSEUDOREFL (*herself)

com *o* *preço* *das* *casas.*
with DET.M.SG price of-DET.F.PL houses
'Maria was surprised by the price of the houses.'

(3) *N* *ga* *ba* *kenta.*
1SG ASP go warm
'I am going to warm myself.'

(Ferraz, 1979, *apud* Hagemeijer, 2007, p. 40)

References

Bacelar do Nascimento, M. F.; Pereira, L.; Estrela, A.; Bettencourt, J.; Oliveira, S.; Santos, R. (2006). The African Varieties of Portuguese: Compiling Comparable Corpora and Analyzing Data-derived Lexicon. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation – LREC 2006*, 24-26 May, Genoa, Italy, 1791–1794. https://aclanthology.org/L06-1407/

Brito, A. M., Duarte, I., & Matos, G. (2003). Tipologia e distribuição das expressões nominais. In M. H. M. Mateus, A. M. Brito, I. Duarte, I. H. Faria, S. Frota, G. Matos, F. Oliveira, M. Vigário, & A. Villalva *Gramática da Língua Portuguesa* (pp. 795-869), Caminho. Hagemeijer, T. (2007). *Clause structure in Santome* [Doctoral Dissertation, University of Lisbon]. http://hdl.handle.net/10451/551

Hagemeijer, T. (2011). The Gulf of Guinea Creoles. *Journal of Pidgin and Creole Languages 26* (1), 111–154. https://doi.org/10.1075/jpcl.26.1.05hag

Hagemeijer, T.; Mendes, A.; Gonçalves, R.; Cornejo, C.; Madureira, R.; Généreux, M. (2022). The PALMA Corpora of African Varieties of Portuguese. In *Proceedings of the 13th Conference on Language Resources and Evaluation – LREC 2022,* 20-25 June, Marseille, France, 5047–5053. https://aclanthology.org/2022.lrec-1.539/

Lobo, M. (2013). Dependências referenciais. In E. B. P. Raposo, M. F. B. Nascimento, M. A.

C. Mota, L. Segura, & A. Mendes (Eds.), *Gramática do Português. Vol. II* (pp. 2177-2227). Fundação Calouste Gulbenkian.

Madureira, R. (2023). Clíticos reflexos e inerentes na variedade de português de São Tomé e Príncipe. *RAPL*, 10, 182-199. https://doi.org/10.26334/2183-9077/rapln10ano2023a10.

Mendes, A., & Estrela, A. (2008) Constructions with SE in African varieties of Portuguese. *Phrasis2*, 83–107.

From Canton Pidgin English to Shanghai Pidgin English in the Mid-19th Century in China --- The Language Features in *The Annotation of English*（1865）

Rongrong Xie (Ningbo Education College) & Peter Bakker (Aarhus Universitet)

Following the Opium War (1840-1842) and the shift of China's major commercial center from Canton to Shanghai, Chinese Pidgin English (CPE) underwent significant transformation. This study examines the booklet Yīng huà zhù jiě《英话注解》*The Annotation of English*, 1865), compiled by five Ningbo merchants and employing Ningbo dialect-based phonetic transcriptions to facilitate commercial interaction in Shanghai. Whereas previous scholarship has focused heavily on Canton Pidgin English, the Shanghai variety has remained comparatively underexplored. By situating *The Annotation of English* within its historical and sociolinguistic context, this paper addresses this gap by tracing the diachronic development and structural characteristics of CPE during this transitional period.

Building on the work of Wu (2001) regarding the role of *The Annotation of English* in linguistic transformation, this study shows that the grammatical core of *The Annotation of English* remains predominantly Sinitic, while WH-interrogatives reflect partial convergence toward English syntax, despite continued divergence from standard English norms.

Lexical Composition of the Two-word Chinese Phrase Section in *The Annotation*:

Phrases (compound words) unique to Sinitic formation account for 34%.

phrases (standard English) unique to English word formation account for 8%.

Phrases with word structures common to both English and Sinitic patterns account for 19.1%.

Phrases containing Cantonese and Portuguese elements account for 3%.

Notably, 36% of the two-word Chinese phrases are translated into single-word English equivalents.

Syntactic Categorization and Dominance (Four-word Chinese phrases to longer Chinese sentences) in *The Annotation*:

Syntactic structures type I (Sinitic alignment only) constitute 76% of the corpus,

confirming Sinitic dominance.

Syntactic structures type II (structures found in neither language) constitute 5% of the corpus.

Syntactic structures type III (English alignment only) constitute 7% of the corpus.

Syntactic structures type IV (common to both) constitute 12% of the corpus.

Transitional Interrogatives: WH-interrogative sentences show a unique shift, with 29% adhering to English syntax compared to 28% following Sinitic structures (Structures common to both languages make up the remaining 43%), suggesting an increasing but incomplete alignment with English norms.

The results also indicate that the dominant grammatical features of CPE, including the reduction of copula and auxiliary verbs, remain fundamentally Sinitic. However, the data from *The Annotation of English* reveals a "blended" stage of evolution, where the Shanghai variety began incorporating more standard English features than its predecessors.

This study contributes to linguistic research by providing a quantitative foundation for the transition from Canton to Shanghai Pidgin English. It underscores the significance of CPE as a linguistic bridge during China's transformative 19th century, highlighting how sociolinguistic environments dictate pidgin development before its eventual decline in the 20th century due to increased language contact and English-medium instruction.

Reference

Wu, Yixiong. 2001. Canton English and Sino-Western exchanges before the mid-19th century (广州英语与 19 世纪中叶以前的中西交往). Modern Chinese History Research 123(3). 172–202.

Orality transcribed: a source for Martinican Creole interactives (and more)

Sarah Melker (Universität Graz)

In over 80 handwritten pages contained in 7 letters sent to Hugo Schuchardt between 1882 and 1885 (Hurch, 2022), the Martinican writer Eleuthère St Prix Roné (?-1902) gives insight into daily life in François as well as sending poems and proverbs from Martinique, and – most significantly for the purposes of linguistic documentation – lengthy transcriptions of Br'er Rabbit tales. Roné attributes the stories to a griot named Laribote and states that his method of collection is on-the-spot transcription spanning over several performances.

The material discussed in this presentation consists of unpublished oral-register texts in Martinican Creole of the 1880s. In the wider category of French-lexified Creoles, similar narrative texts from the same century are documented thanks to the published works of Paul Baudot (Guadeloupe), François Chrestien (Île Maurice), Louis Héry (La Réunion), François-Achille Marbot (Martinique), and Georges Sylvain (Haiti). However, among the known texts, the earliest work that is considered properly "oral" and reflective of local Creole culture (see Ludwig, 2016), rather than translations or adaptations from the fables of La Fontaine, is that of Charles Baissac (Île Maurice). Thus, the dearth of linguistic documentation from this period and concerning this text-type suggests that the data these unpublished manuscripts contain can contribute to knowledge of earlier stages of the Creole in Martinique.

In particular, the texts provide many examples across a range of lexical items to examine the environments in which the typical features of Martinican Creole appear, such as nasal insertion (e.g. *manman* for *maman*) or absence of final <r> (e.g. *palé* for *parler*) (Hazaël-Massieux, 2002). In fact, Roné's texts show variation in these features, which will be briefly discussed, including potential explanations for this divergence from current documented forms ranging from the effects of an erudite transcriber (as Roné was) tending towards standard French, to possible influences of certain phonological environments.

The texts are also an unusually rich source of what Heine terms *interactives*. As the items in Roné's texts do not tend to align with standard lexical items, and have no standard French equivalent, they are useful in documenting language influences and phonotactics. Moreover, the presentation seeks to contribute to the languages surveyed in Heine 2023 to test his assertions concerning these items. The presentation discusses these items, categorizing them according to Heine's 10 types of interactives. There are those which are common to storytelling in a great number of Creoles (e.g. the call-and-answer *cric! crac!*), and those that come from standard French (e.g. *sapristi*); these borrowings appear stable and undergo no adaptation. In contrast, an example of an exclamation from the text, *ya yayi*, fits into Heine's category *primary interjection* and patterns with those he

has described in terms of their phonological content and their function in expressing the state of the speaker.

References

Hazaël-Massieux, M. C. (2002). Les créoles à base française: une introduction. *Travaux Interdisciplinaires du Laboratoire Parole et Langage*, *21*, 63-86.

Heine, B. (2023). *The grammar of interactives*. Oxford University Press.

Hurch, B. ed. (2022-). Eleuthère St Prix Roné to Hugo Schuchardt, letters 09027-09033. *Hugo Schuchardt Archiv*. Department of Linguistics, University of Graz. http://schuchardt.uni-graz.at.

Ludwig, R. (2016). Littératures des mondes créoles–des débuts aux questionnements actuels. *Études créoles. Cultures, langues, sociétés*, (33),1.

The Hawai‘i Language Contact Corpus (1789-1975):
An Update on Current Research

Sarah Roberts (independent scholar)

For the past 35 years, the present author has engaged in archival research to develop an exhaustive corpus of material documenting language contact in Hawai‘i to about 1975. Although about a third of the Hawai‘i Language Contact Corpus (henceforth HLCC) consists of pidgin/creole texts, the rest deals with an unlimited range of topics arising from the sociolinguistic setting.

The corpus includes descriptions of specific linguistic encounters between people, general statements about the overall linguistic situation, and information revealing the nature of social contact between speakers of different linguistic backgrounds. The data was gleaned from an array of different written genres, including published nonfiction like travel accounts and memoirs, testimony in court records, newspaper articles and editorials, novels, short stories, poetry, academic fieldwork, audio recordings, and manuscript sources like diaries, nautical journals, and correspondence.

The principal HLCC database resides in a Word document began in 1991 and is currently 1,341 pages long and a few weeks ago, the millionth word was added to the corpus. The last time the document was fully formatted and finalized was in June 2001 when it was only 788 pages in length. The current goal in 2026-2027 is to finalize the corpus a second time so it would form the basis of subsequent quantitative work. For three weeks in February 2026, the present author will be engaged in extensive fieldwork in Hawai‘i to extract as much data as possible that remain in old court records, visiting the four circuit courts on the islands of O‘ahu, Kaua‘i, Maui, and the Big Island. This will require the use of high-speed photography to digitize scrolling microfilm in order to cover a greater number of reels than otherwise possible. Once this material has been examined and utilized, the HLCC will be considered largely complete. Further work will require translations of some Hawaiian-language materials and transcription of some rare audio sources. Once formatted, the present author hopes to print a limited run of hardcover volumes for public distribution, and eventually as a PDF file on the Hawai‘i Linguistic Archive at archive.org.

In concert with the finalization of the above compilation will be the construction of a new corpus database of early pidgin/creole texts from Hawai‘i. Previously, the author has used an Excel file to code the texts but this has proven to be inadequate. A major problem in the linguistic analysis in prior work was the unevenness of data of varying quality in different sources, with exaggeration and inaccuracies more common in fictional composed texts like pseudonymous letters. The new database will allow one to define subcorpora, including one of high-quality data from real linguistic encounters (such as found in court records, newspaper articles, and academic fieldwork), coded with newly added sociobiographical details on

uniquely-identifiable speakers obtained via historical and genealogical research. This will

permit unprecedented analysis of sociolinguistic variation of lexical and grammatical structures, yielding insight into processes of creole formation in a variety of contexts.

Thematic analysis and classification of nigerian pidgin proverbs

Shalom Oladele (University at Buffalo)

The paper explores how Nigerian Pidgin English (NPE) proverbs store common socio-cultural knowledge and can serve as a resourceful tool in daily communication among Nigerians. The study is based on a qualitative, thematic analysis of a selective corpus of 50 high-frequency proverbs used in NPE, drawn from oral speech, spontaneous conversation, and literature. The study's approach was informed by previous research on Nigerian and African proverbs, the conceptualization of cultural phenomena, and NPE as a contact language (e.g. Mieder 2018; Mensah 2013; Balogun 2013; Callies 2024). The hypothesis addressed in the analysis is as follows: recurring themes and metaphorical mappings in NPE proverbs constitute a systematic indexing of Nigerian cultural values and, at the same time, a structural and pragmatic indexing of NPE discourse. The proverbs were categorized into 14 thematic areas based on previous works on proverbial classification and personal intuition (e.g. Mensah 2013; Callies 2024)

Animal metaphors e.g:

1. Monki dey work, baboon dey chop
   monkey PROG work baboon PROG eat
   'The monkey works, the baboon eats'

alongside food-based metaphors of the environment

Na condition make crayfish bend

FOC condition make crayfish bend
'It is circumstances that cause the crayfish to bend.'

recur as tropes of thought. Evaluative and epistemic stance is always encoded in the grammatical features of progressive aspect marking (*dey*), negation (*nor),* and focus marker constructions *(na*) across the corpus. These tendencies provide empirical evidence for the claim that NPE proverbs are a cultural index of Nigerian values. Preliminary results support the conclusion that NPE proverbs are an evolving, self-renewing source that preserves traditional wisdom, embodies the modern realities of Nigerian cities, and serves as an effective conversational genre and cultural repository for Nigerians.

Although earlier literature (e.g. Mensah 2013; Callies 2024) has investigated the metaphorical and pragmatic functions in a subset of Nigerian Pidgin proverbs, this research contributes to the literature by providing a thematic classification of 50 high-frequency proverbs and showing how metaphorical domains of interest can be categorized by recurring grammatical constructions. For example, contrary to previous literature, which emphasizes semantic interpretation, the paper demonstrates how aspectual marking (*dey*

work, dey talk) and serial verb constructions function in proverb discourse to encode habituality and moral inevitability. Urban proverbs have recently developed

> Person wey get phone get power
> person REL have phone have power
> 'Who has a phone has power.'

hence transforming the traditional proverbial reasoning into digital modernity, it equally shows how NPE proverb formulation is changing into a postcolonial and technological reality.

Brasil–Angola e a recombinação gramatical e lexical no continuum afro-atlântico do português

Silvana Araújo (Universidade do Minho)

Este trabalho examina convergências morfossintáticas e lexicais entre o português do Brasil (PB) e o português de Angola (PA), propondo que paralelismos estruturais observados nas duas variedades refletem ecologias de contato afro-atlânticas historicamente interconectadas. O estudo insere-se na perspectiva da Sociolinguística de Contato, conforme sistematizada em Savedra et al. (2021) e Azevedo, Araujo e Macêdo (2025), que articula variação, mudança e ecologia sociolinguística como dimensões inseparáveis da análise. Dialogando com a abordagem ecológica de Mufwene (2001, 2008), com a tipologia da mudança induzida por contato de Thomason (2001) e Winford (2003), com a noção de recombinação de traços de Aboh e DeGraff (2014; Aboh, 2015) e com discussões sobre variedades híbridas (Holm, 2004), sustenta-se que PB e PA devem ser analisados em um mesmo quadro comparativo transatlântico.

A pesquisa tem por base dados recolhidos de 18 entrevistas pertencentes ao *Corpus do Projeto Em busca das raízes do português brasileiro: estudos morfossintáticos*, sediado no Núcleo de Estudos em Língua Portuguesa da Universidade Estadual de Feira de Santana (NELP/UEFS), articulados a dados orais de variedades urbanas de Luanda. A metodologia combina Sociolinguística Variacionista e Linguística do Contato, com análise quantitativa de fatores condicionadores e interpretação ecológica de contextos multilíngues.

No plano morfossintático, observam-se estratégias de objeto nulo discursivamente condicionado:

(1) Eu vi Ø ontem.
1SG ver.PST Ø ontem
'Eu vi (ele/ela) ontem.'

Estruturas análogas são atestadas no PA urbano, sugerindo convergência funcional em contextos de contato com línguas bantas.

No plano lexical, identificam-se paralelismos tanto na preservação de itens de origem africana quanto em processos de expansão semântica e especialização pragmática. No PB, unidades como *caçula* (< quimbundo *kasula*), *moleque* (< quimbundo *muléke*) e *quitanda* (< quimbundo *kitanda*) mostram integração morfológica plena e reconfiguração semântica em diferentes domínios socioculturais. No PA, observam-se itens como *cubata* (< quimbundo *kubata*) e usos específicos de *mais-velho* como categoria sociorelacional, cuja estrutura ecoa padrões nominais bantos. Tais fenômenos não se restringem ao empréstimo lexical simples (borrowing), mas envolvem reinterpretações categóricas e reanálises pragmáticas, compatíveis com os modelos de replicação e recombinação de traços discutidos por Thomason (2001), Heine e Kuteva (2005) e Aboh (2015).

Parte-se da hipótese de que as convergências gramaticais e lexicais resultam de configurações sócio-históricas comparáveis no espaço atlântico, marcadas por multilíngues

intensivos e por circulação transcontinental de práticas linguísticas. Resultados preliminares indicam paralelismos significativos na frequência, na distribuição discursiva e na produtividade desses fenômenos em PB e PA.

Ao explicitar o eixo Brasil–Angola como unidade analítica, o estudo contribui para consolidar um modelo afro-atlântico de formação do português, ampliando o debate sobre mudança induzida por contato e sobre a circulação histórica de estruturas e léxicos no mundo pós-colonial.

The Eastern Ijo element in Berbice Dutch Creole and the Eastern Niger Delta slave trade

Silvia Kouwenberg (The University of the West Indies, Mona)

This paper's starting point is the dominant presence of forms of Eastern Ijo (EI) origin in the lexicon of Berbice Dutch Creole (BD), including basic and high frequency content vocabulary, forms central to plantation life, function words and bound grammatical morphemes. Like Smith et al. (1987), where these forms were first documented, and Kouwenberg (2009, 2015), which builds on that work with more fine-grained analyses of the EI component in BD, this paper takes these facts as solid evidence of the early presence of EI speakers in the Dutch-owned Berbice colony. A detailed discussion of the trade in enslaved Africans in the Eastern Niger Delta over the course of the seventeenth century is presented in support of that view. In so doing, the paper disproves Parkvall & Jacobs's (2023) claim that the first arrival of enslaved EI speakers in Berbice dates to 1713 – almost 90 years after the establishment of the Dutch colony in 1627 – at which time P&J presume EI speakers to have impacted a pre-existing Dutch-lexicon creole. While there is no doubt that the 1713 shipment identified by Parkvall & Jacobs carried enslaved Africans from trading posts in the Eastern Niger Delta, the paper will argue that the rise of city-states under the control of powerful king-merchants supported by councils of War Canoe Chiefs transformed the slave trade in the Eastern Niger Delta in the course of the seventeenth century. Based on oral and documented sources, historians have shown that coastal communities in the Eastern Niger Delta with established trading networks connecting them with inland communities were well situated to benefit from trade with the Europeans from the first time of contact, but also to become its victims: A tradition of rivalry over fishing grounds and trade routes which predates the European presence (Alagoa 1971, Horton 1969) meant that reciprocal raids became opportunities for the capture of individuals to be rendered into the slave trade. However, towards the end of the seventeenth century, the consolidation of villages into powerful conglomerates under single, well-defended leadership ushered in a time of peace in the Eastern Delta (Jones 1963).

By the time of the 1713 shipment, the historical context precluded the enslavement of EI speakers, who, instead, were involved in the trade in enslaved people from hinterland markets, both to render people into the European slave trade and to bolster the numbers of the War Canoe Houses. The paper concludes that neither the historical context nor the nature of the EI elements in BD supports the late arrival of EI speakers.

References

Alagoa, Ebiegberi J. 1971. The development of institutions in the states of the Eastern Niger Delta. *Journal of African History* XII(2), 269-278.

Alagoa, Ebiegberi J. 2005. *A history of the Niger Delta. An historical interpretation of Ijo oral tradition*. Port Harcourt, Nigeria: Onyoma Research Publications [Reprint; first published 1972 by Ibadan University Press].

Horton, Robin. 1969. From fishing village to city-state: A social history of New Calabar. In Mary Douglas & Phyllis M. Kaberry (eds.) *Man in Africa*, 37-58. London: Tavistock Publications.

Jones, G.I. 2000 [1963] *The trading states of the Oil Rivers. A study of political development in Eastern Nigeria.* [First publication by Oxford University Press.] Hamburg & Oxford: Lit & James Currey.

Kouwenberg, Silvia. 2009. The invisible hand in creole genesis: Reanalysis in the formation of Berbice Dutch. In Enoch Aboh & Norval Smith (eds.), *Complex processes in new languages*, 115–158. Amsterdam: John Benjamins.

Kouwenberg, Silvia. 2015. Early morphology in Berbice Dutch and source language access in creolisation. *Word Structure* 8(2). 138–159.

Parkvall, Mikael & Bart Jacobs. 2023. Returning a maverick creole to the fold: the Berbice Dutch enigma revisited. *Folia Linguistica* 57(1): 177–203.

Smith, Norval, Ian E. Robertson & Kay Williamson. 1987. The Ịjọ element in Berbice Dutch. *Language in Society* 16(1): 49–90.

On the genesis of two French contact varieties

Stéphane Goyette (Université Acadia University)

A number of French contact varieties have been referred to as "semi-creoles" in the literature: This is the case for Réunionnais (Chaudenson 1992, Holm 2007) as well as the West Indian variety first described by Highfield (1979) (a transplanted variety spoken in the Virgin Islands: hereinafter: VIF).

Chaudenson regards both varieties as "missing links" between other French creoles (Haitian, Mauritian…) on the one hand and the colloquial colonial-era (seventeenth century) spoken French varieties ancestral to all overseas French varieties (via L2 acquisition and dialect koïneization).

The goal of the proposed presentation is to argue that, far from being conservative, Réunionnais and VIF are both products of secondary language contact, and as a result cannot be used as evidence in favor of creolization as a gradual process.

A core piece of evidence relates to the fact that, while Réunionnais and VIF are indeed each much more French-like than other French-lexifier creoles, they are utterly unlike one another. It is quite unclear -and unexplained in the scholarly literature- why these two French "semi- creoles" should be so divergent from one another. The mystery deepens when one considers that other French creoles do not diverge from one another to such a degree. On the contrary, they share a large number of common features.

Furthermore, the diversity of the lexifier (Colonial French), and the claim that dialect mixture played in core role in the genesis of overseas French, has been shown to be a myth in the case of Laurentian French (Morin 2002): this claim of dialect mixture has been known to also be mythical in the case of Acadian ever since the pioneering work of Massignon (1962).

The explanation proposed in this talk involves rejecting the entire scenario sketched above. What will be assumed instead is that creolization is indeed a highly specific type of language change (whether this specificity is due to the intensity of language contact, as per Aboh (2019), or to the nativization of a pidgin, as was the dominant belief a generation ago- see McWhorter (2018) for the most recent restatement of this position- is immaterial heuristically for the purposes of this presentation).

It will be assumed that, from a common ancestor (Colonial French), creolization yielded most French creoles (Haitian, Mauritian…) on the one hand, and on the other that non- creole language change gave birth to such non-creole varieties of overseas French as Laurentian and Acadian French in Canada. The sharply divergent structures of VIF and Réunionnais can be accounted for, it will be argued, by assuming that the two varieties are in fact mirror images of one another: the former as a product of normal language change which was heavily influenced by (a) French creole(s), and the latter as a product of creolization which was heavily influenced by (a) non-creolized offspring of colonial

French. In the case of Réunionnais this conclusion regarding its genesis was already foreshadowed by the work of Ramassamy (1985).

Some thoughts on creole genesis and the three factors in language design

Stéphane Térosier (University at Buffalo)

Chomsky (2005) identifies three factors in language design: (i) the genetic endowment), (ii) experience, and (iii) principles that are not exclusive to language. The interaction of these three factors plays a crucial role in language acquisition and is, therefore, relevant to creole genesis. In fact, it is reasonable to argue that the goal of creole genetics should be to develop a principled account of creole genesis based on the interplay of these factors. This paper aims to lay the ground the groundwork for such an account by offering some conjectures on the roles of Factors 2 and 3 and suggests that claims of creole similarity should be more nuanced.

On the standard view that the main task of the language learner is to establish the values of the parameters in their native/target language, Biberauer (2019) proposes that Factors 2 and 3 are especially important in guiding the language learner through that process. She identifies Maximize Minimal Means—a third factor principle related to domain-general cognition—as a major influence on parameter setting. Specifically, it results in a NONE>ALL>SOME learning path. Given a feature F, unless evidence suggests its existence, the language learner will conclude that it is not present in the target language. If that evidence is found, the language learner will then assume that F is present on all the syntactic heads of a relevant set. If, however, F is actually only present on a subset of these heads in the target grammar, the language learner needs additional evidence to identify the members of that subset. Hence, the more idiosyncratic a feature is, the more data is needed for its successful acquisition.

On the view that creole genesis involves restrictions in the access to the target language (the lexifier in most cases), then the creole creator's challenge is to figure out the parameter values of that language with limited evidence. More idiosyncratic properties are thus likely to not be inherited. Less idiosyncratic features, on the other hand, have a greater chance of being successfully identified and transmitted to the creole. These are the predictions made by the NONE>ALL>SOME learning path. However, I argue that the identification of a feature is a necessary but not sufficient condition for successful parameter setting. If the evidence is not conclusive (e.g., because of morphological opacity), the creole creator may opt to not associate the feature with any of the relevant syntactic heads. This claim is substantiated by evidence from the analytical nature of French creoles. Building on Huang (2013), I argue that the initial loss of verb movement in their formation led to the emergence of various properties, including analyticity.

Thus, this study suggests that the interplay of Factor 2 and 3 may be responsible for the frequency of some properties across creoles and, critically, that these properties may result from a single "failure" in parameter setting. It follows that claims of creole similarity should take into consideration the interrelatedness of these properties to avoid the pitfall of overestimation.

An abundance of aspectual affixes: Investigating claims about grammaticalisation for Creoles and sign languages

Susanne Maria Michaelis, Felix Stilzebach, Lilith Stein (Universität Leipzig), Luigi Lerose & Nick Palfreyman (University of Central Lancashire)

Several claims are made in the literature on grammaticalisation in Creole languages (CLs) and sign languages (SLs). It is said that CLs are analytic or isolating languages lacking inflectional markers as opposed to their synthetic inflectional marking lexifier languages (McWhorter 2005; Parkvall 2008; Velupillai 2015). Meanwhile, SLs are said to have poor sequential morphology and few sequential affixes (Aronoff, Meir & Sandler 2005; Pfau & Steinbach 2006). Indeed, some have commented on cross-modal similarities between CLs and SLs (Woodward 1973, Fischer 1978, Adone 2012, Gil 2014).

We focus on aspect markers due to their prominence in CLs and SLs (Adone 2012) and the amount of available research. Given the claims above, we hypothesised that aspect markers in CLs and SLs exhibit more instances of early-stage grammaticalisation (auxiliaries, free morphemes) than non-creole spoken languages, and fewer examples of late-stage grammaticalisation (affixes, stem change).

Evidence from our cross-modal survey of aspect markers in 50 Creole languages and 25 sign languages goes some way to disproving this hypothesis. Applying four of the grammaticalisation parameters from the MAGRAM Project (Bisang & Malchukov 2020) – semantic reduction, phonetic reduction, bondedness and allomorphy – to our data, we find an abundance of previously unattested affixes for CLs and SLs.

We use Haspelmath's (2021) definition of an affix as 'a bound morph that is not a root, that must occur on a root, and that cannot occur on different root classes' (p18), and we find that spelling systems (for CLs) and transcription practices (for SLs) have rendered affixal morphology all but invisible. Imperfective markers in Nengee (1a) and Santome (1b) are not typically analysed as affixes, yet they are bound and host-selective, and other material cannot intervene. There is a strong case for analysing these as affixes (2a, 2b).

For sign languages, affixing markers have routinely been analysed as clitics (Pfau & Steinbach 2006; Palfreyman 2019); however, this does not take into account the reduction (in handedness, location and movement, for example) that renders an affix as a bound form (i.e. it cannot occur on its own). The South Korean Sign Language marker FINISH (3b), for instance, derives from a free lexical source FINISH (3a) (in Japanese Sign Language) but is now highly reduced, host-dependent, and restricted to action roots. We further identify aspectual superfixes in SLs: mouth gestures articulated coextensively with manual action roots, such as <bn> in Turkish Sign Language (4), which is a reduced form of its source construction Turkish *bitti* 'finished' (Dikyuva 2011).

We will present these findings in the context of the world's non-Creole spoken

languages so as to respond to claims regarding the similarities between CLs and SLs. All examples can be viewed in our cross-modal grammaticalisation (CrossMoGram) open database.

Examples

(1a) Santome (Portuguese-based), Hagemeijer (2013)

San ***ska*** *be* she
PROG go 'She is going.'

(1b) our analysis

San ***ska****be* she
PROG.go 'She is going.'

(2a) Nengee (English-based), Migge (2013)

Pe *a* ***e*** *tan*?
where 3SG IPFV stay
'Where does she stay?'
she stay?'

(2b) our analysis

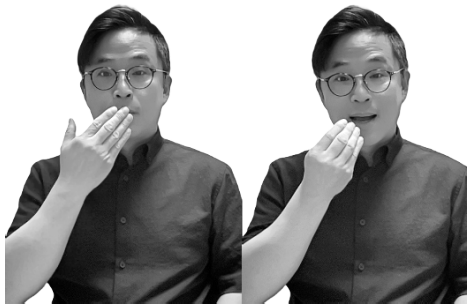
Pe *a* ***e****tan?*
where 3SG IPFV.stay
'Where does

(3a) Japanese Sign Language

FINISH

perfect marker

(3b) South Korean Sign Language



EAT (AFF)FINISH

'have eaten'

Examples (3a) and (3b) from Sagara, Byun, Chen and Palfreyman (in press)

(4) Turkish Sign Language (TID), Dikyuva (2011: 47)

mouth channel:			bn
manual channel:	BEN	INGILTERE	GITMEK
English gloss:	ME	ENGLAND	GO
English translation:	'I have been to England.'		

References

Adone, Dany (2012) http://doi.org/10.1515/9783110261325.862

Aronoff, Mark, Irit Meir & Wendy Sandler (2005) https://doi.org/10.1353/lan.2005.0043

Bisang & Malchukov (2020) https://doi.org/10.1515/9783110563146

Dikyuva, Hasan (2011) http://knowledge.lancashire.ac.uk/id/eprint/18988/

Fischer, Susan (1978) *Sign language and creoles*. In: Siple (ed.), 309–331. Gil, David (2014) https://doi.org/10.1093/acprof:oso/9780199685301.003.0003

Hagemeijer, Tjerk (2013) Santome. In: Michaelis et al. (eds.), 50-58. Oxford: OUP. Haspelmath, Martin (2021). Bound forms, welded forms, and affixes: Basic concepts for morphological comparison. *Voprosy Jazykoznanija*, 1: 7–28.

McWhorter, John (2005) https://doi.org/10.1093/oso/9780195166699.001.0001

Migge, Bettina (2013) Nengee. In: Michaelis et al. (eds.), 39-48. Oxford: OUP. Palfreyman, Nick (2019) https://doi.org/10.1515/9781501504822

Parkvall, Mikael (2008) https://doi.org/10.1075/slcs.94.17par

Pfau, Roland & Markus Steinbach (2006) Modality-independent and modality-specific aspects of grammaticalization in sign languages. *Linguistics in Potsdam* 24.

Sagara, Keiko, Kang-Suk Byun, Yu Yi Chen & Nick Palfreyman (in press) Grammaticalisation and contact-induced semantic change in JSL, SKSL and TSL. *Sign Language Studies* 26:1, Fall 2025. Velupillai, Viveka (2015) https://doi.org/10.1075/cll.48

Woodward, James (1973) https://doi.org/10.1353/sls.1973.0006

Too afraid to say the wrong thing:
Phonological development in the Jamaican setting

Tanyia Joy Wilkins Miller (The University of the West Indies, St. Augustine)

This paper investigates the frequency with which 2 groups of primary school children in Jamaica produce the voiced dental fricative [ð] and the voiceless dental fricative [θ] which were embedded in nonsense and real words. The dental fricatives are phonemes of Jamaican English (JE) but not of Jamaican Creole (JC).

This study is part of a wider investigation examining how school-age children in Jamaica produce selected segments that contrast in JE but do not in JC. The research that is the focus of this paper was guided by the following research questions: (1) What are the rates at which a group of grade 2 children and a group of grade 5 children produce the voiced dental fricative [ð] which is present in JE but not in JC? (2) What are the rates at which a group of grade 2 children and a group of grade 5 children produce the voiceless dental fricative [θ] which is present in JE but not in JC?

At the beginning of formal schooling, most Jamaican children speak some form of JC as their only language (Pollard 1998, Christie 2003, Bryan 2004). The goal of the Ministry of Education in Jamaica, as laid out in its 2001 Language Education Policy, is to ensure that children are competent in the use of Standard Jamaican English (SJE) by grade 4. Despite this goal, the performance of Jamaican students in Language Arts examinations at the primary and secondary levels is relatively poor and is a major concern for educators and policy makers.

To explore the research questions, a quantitative approach was taken. One hundred children from grades 2 and grade 5 participated in the study. They were selected from 4 government-run primary schools in Jamaica. The children ranged in age from 7;2 – 9;0 for grade 2 participants and 10;3 to 12;2 for those from grade 5. The participants completed a repetition task involving mostly nonsense words and a spelling task involving both nonsense and real words.

The results show that for nonsense words, participants either produced dental fricatives or the acoustically similar labiodental fricatives [v] or [f]. For real words, participants were likely to produce either the voiced dental fricative [ð] or its JC reflex [d] which is not highly stigmatized. This pattern was not the same for real words containing the voiceless dental fricative [θ]. If the voiceless fricative was not produced, the acoustically similar voiceless labiodental fricative [f] was produced resulting in a non-word. There appears to be avoidance of the voiceless alveolar stop [t], even if it would have resulted in a real JC word. The voiceless alveolar stop [t] is a highly stigmatized reflex of the JE dental fricative [θ]. This suggests a level of linguistic insecurity which may stem from the negative perceptions associated with JC, the native language of many Jamaicans.

Register variation and variable complementizer expression in Mauritian Creole: Omission or Insertion

Tonjes Veenstra (Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin) & Nele Arnold (Humboldt-Universität zu Berlin)

Variable complementizer expression is one of the classic topics in variation and register studies since the seminal works of Sankoff and Cedergren (1972b; 1972a) and Biber (1999). While standard French does not allow complementizer omis-sion, it is common in non-standard varieties and in French outside France (Liang et al., 2021). Alleesaib et al. (2024) find that omission of *que* in Mauritian French is even more frequent than in the Caribbean and Canadian contexts. Their study reveals that spoken data show a lot more omission, and that certain grammatical factors (negation on the clause-embedding predicate (CEP), subject of the embed-ded clause other than personal pronoun) disfavour omission. They briefly compare it to Mauritian Creole (MC), where they find very high rates of omission (92 % in a spoken corpus) and hypothesize an impact of negation.

The question, however, is whether in MC the complementizer *ki* is omitted or inserted in embedded contexts. Our study examines the use of *ki* in MC in more detail on the basis of the *Spoken Morisien Corpus*, the first large corpus of a creole language with a special focus on register variation (Veenstra, 2016) . This corpus contains 30 hours of recordings of speakers in different situations making it suit-able to study situation-specific variation. We compare our findings to the results of Alleesaib et al. (2024) on Mauritian French, taking different grammatical factors and social factors into account.

Three grammatical factors influence the presence of *ki*: (i) nature of the CEPs; (ii) presence of stage-setting adverbials; (iii) full DP subject vs pronominal subject in the subordinate clause. Following recent work by Syea (2024) suggesting that MC can be characterized as a topic prominent language, full DP subjects are in the C-domain, just like stage-setting adverbials. These elements require that the embedded C-domain is syntactically activated.

Furthermore, we show that the insertion of *ki* takes place more often with spe-cific verbs (*realize, remarke, dekouver, montre*: > 50 %) than with prototypical ones (*dir, kone, krwar*: < 10 %) and argue that this is related to register and points to-wards influence from French. This implies that MC has basically the same sub-ordination system with two complementizers (∅ vs *pou*) that we find in most of the French-related creoles in the Caribbean, as has been argued for Haitian Creole (Lefebvre, 1998) and Guadeloupean Creole (Kezerian et al., 2025).

We propose creoles have a diachronically layered lexicon of CEPs, consisting of an older layer with more prototypical CEPs without *ki* and a newer layer with more specific CEPs that take *ki*.

References

Alleesaib, Muhsina, Shimeen Chady, Guilhem Florigny, Guillaume Fon Sing, Shrita Hassamal, and Anne Abeillé (2024). L'omission de que en Français Mauricien. Talk presented at Congrès Mondial de Linguistique Française (CMLF).

Biber, Douglas (1999). A register perspective on grammar and discourse. *Discourse Studies 1*, 131–150.

Kezerian, William, Vasily Tselioudis, Béatrice Jeannot-Fourcaud, Isabelle Barrière, and Renauld Govain (2025). Testing clause boundaries: Focus and complementa-tion in Guadeloupean Creole. Talk presented at Séminaire Grammaires créoles, October 20, 2025.

Lefebvre, Claire (1998). *Creole genesis and the acquisition of grammar: the case of Haitian Creole*, Volume 88 of *Cambridge studies in linguistics*. Cambridge: Cambridge University Press.

Liang, Yiming, Pascal Amsili, and Heather Burnett (2021). New ways of analyz-ing complementizer drop in Montréal French: Exploration of cognitive factors. *Language Variation and Change 33*(3), 359–385.

Sankoff, Gillian and Henrietta Cedergren (1972a). Sociolinguistic research on French in Montreal. *Language in Society 1*(1), 173–174.

Sankoff, Gillian and Henrietta Cedergren (1972b). Sociolinguistic research on French in Montréal. *Language in Society 1*(1), 173–174.

Syea, Anand (2024). Absence of syntactic passive in creoles: Evidence from French-based Mauritian Creole. *Canadian Journal of Linguistics/Revue canadienne de linguistique 69*(2), 174–202.

Veenstra, Tonjes (2016). *Spoken Morisien Corpus*. Collection MFE01 at cata-log.paradisec.org.au. URL https://dx.doi.org/10.26278/SZB3-WE75, Ac-cess: 30.10.2025.

A new approach to understanding the Haitian determiner

Tristan Chopinez (Universiteit Leiden)

The Haitian Creole determiner morpheme is known for its anti-markedness allomorphy, surfacing as [la] following consonant-final nouns, and as [a] following vowel-final nouns. Less discussed is its nasal allomorphy, which can be contextually motivated, as in (1), or stylistic, as in (2).

[papa-a]	[malad-la]	[laʒã-ã]	[bãk-lã]	[madãm-nã]
dad-DET	book-DET	money-DET	bank-DET	woman-DET

(1) : phonologically conditioned allomorphs (Cadely, 2002)

[papa-ã]	[malad-lã]	[madãm-lã]
dad-DET	book-DET	woman-DET

(2) : stylistically conditioned allomorphs (Cadely, 2002)

When the determiner is expressed as a single vowel, the hiatus context can be resolved by a glide insertion if the noun-final vowel is (mid)-high, as in (3).

[vɛ-a]	[papje-ja]	[lapli-ja]	[bɔkɔ-a]	[bato-wa]	[tu-wa]
glass-DET	paper-DET	rain-DET	wizard-DET	boat-DET	hole-DET

(3) : hiatus resolution (Klein, 2003)

The morpheme also features a particular interaction with word-final /r/. Haitian /r/ is deleted word-finally, but resurfaces when a vowel-initial suffix is added. When followed by a determiner, one could expect underlying /r/ to trigger the [la] allomorph, or else the [a] allomorph to cause /r/ to resurface. Instead, the /r/ does not surface, and the [a] allomorph is chosen, as in (4).

[flɛ]	[flɛʁ-i]	[flɛ-a]
flower	flower-PST.PTCP	flower-DET

(4) : rhoticity (Nikiema & Bhatt, 2003)

While the morpho-phonological complexity of this morpheme has been discussed in many studies, no analysis has been conducted which considers all its various properties. Some (Klein, 2003; Bonet Lloret & Mascaró, 2007) look only at the anti-marked nature of its [a] – [la] alternation and at glide insertion. Others (Nikiema & Bhatt, 2003; Lahrouchi & Ulfsbjorninn, 2024) also look at its interaction with the rhotic, but do not analyze its nasal allomorphs. Cadely (2002) discusses every allomorph but does not acknowledge the interaction with the rhotic. Furthermore, most studies gather their data from previous analogous studies, such that direct reference to recent spoken data cannot be found in any of these analyses. This study aims to solve these issues.

The spoken corpus "Radio Haïti" (Havard et al. 2025) is used to verify the descriptive claims made

during prior research on the Haitian determiner, and to gather further details about its morphological, phonological, and phonetic properties. Annotations are used to gather all examples of definite determiners in the corpus, as well as their preceding nouns. These can then be categorized based on determiner forms, and generalizations for each form can be made about the phonological and phonetic properties of the preceding nouns. Speaker identity is also considered, to investigate whether the variation in determiner nasalization described by Cadely (2002) is speaker dependent or stylistic/contextual.

The corpus is also used to gather data on the broader phonological and phonetic properties of Haitian. Vowel hiatuses are phonetically analyzed to see if glide insertion is systematic and clearly distinct from other vowel transitions. Segments neighboring nasal segments are measured for nasal residue, to investigate potential phonetic causes for nasal spread and to verify whether documented examples of nasal spread is an active phonological process or a phonetic by-product. Words with a deleted underlying /r/ are analyzed to verify whether deletion is complete or if there is any phonetic residue, and whether all environments are equivalent in this regard. This data is then used to create a complete morpho-phonological analysis of the Haitian determiner morpheme, in the context of a broader Haitian phonological system.

References

Bonet, E., Lloret, M-R., Mascaró, J (2007). "Allomorph selection and lexical preferences: Two case studies". *Lingua* 117, pp. 903-927.

Cadely, J-R. (2002). "Le statut des voyelles nasales en Créole haïtien". *Lingua* 112, pp. 435-464.

Havard, W. N., Govain, R., Lecouteux, B., Schang, E. (2025). "Self-Supervised Models of Speech Processing for Haitian Creole". Presented at: Interspeech 2025.

Klein, T. (2003). "Syllable structure and lexical markedness in creole morphophonology: Determiner allomorphy in Haitian and elsewhere". *Phonology and Morphology of Creole languages*, ed. Plag, I., pp. 209-228.

Lahrouchi, M., Ulfsbjorninn, S. (2024). "Nasal Assimilation Counterfeeding and Allomorphy in Haitian: Nothing is Still Something!". *Linguistic Inquiry* 55(2), pp. 255-286.

Nikiema, E., Bhatt, P. (2003). "Two types of R deletion in Haitian Creole". *Phonology and Morphology of Creole languages*, ed. Plag, I., pp. 43-69.

Estudo preliminar sobre as estratégias de complementização no kriol

Vinícius Albuquerque de Souza (Universidade de São Paulo)

Esta pesquisa busca observar as estratégias de complementização do kriol, língua de contato nativa da Guiné-Bissau, na tentativa de verificar as motivações e restrições sintático-semânticas dos usos de seus complementizadores. Os trabalhos de Kihm (1994) e Intumbo (2007) estão entre os poucos que fazem referência a complementizadores nessa língua sem, contudo, sistematizar o comportamento gramatical desses itens. Ambos os autores mencionam três itens do kriol que se configuram como possíveis complementizadores – ku, pa e kuma –, uma vez que os três introduzem orações encaixadas. Kihm (1994) e Intumbo (2007) destacam que o estatuto de pa e kuma como legítimos complementizadores não é claro, dado que há contextos em que pa se comporta como uma preposição e kuma como um verbo. Fica evidente, portanto, que, embora haja descrições acerca de constituintes que introduzem orações encaixadas no kriol, a literatura ainda carece de trabalhos que estabeleçam um quadro formal que permita explicar as propriedades gramaticais dos complementizadores do kriol. A fim de sanar esta lacuna, este trabalho tem como norte teórico preceitos da teoria da Gramática Gerativa, mais especificamente em sua versão reformulada desde o Programa Minimalista (Chomsky, 1995). Para a presente pesquisa, nos pautamos nas propostas de Chomsky (2008) e Ouali (2008), que versam sobre as noções de herança de traços no complexo C-T. A descrição e análise das estratégias de complementização do kriol a partir da Teoria de Princípios e Parâmetros permitirá situar, em termos comparativos, essa língua minorizada no conjunto das línguas naturais, em particular no que diz respeito ao modo como o sistema de complementização determina (ou é determinado por) outros elementos gramaticais. Os resultados alcançados pelo estudo também contribuirão para ampliar, no plano teórico, a nossa compreensão sobre a função dos complementizadores na Faculdade Humana da Linguagem. Considerando que não tivemos ainda a oportunidade de ir a campo coletar dados primários, utilizamos em nossa análise os dados dispostos em Kihm (1994), Intumbo (2007) e Danfá (2021). Como hipótese inicial, propomos que, no kriol, pode haver retenção ou transferência de determinados traços de C para T, como previsto no modelo de Ouali (2008), e isso se expressa na variabilidade de complementizadores no kriol. Quando ku e kuma são selecionados como complementizadores, C pode compartilhar traços-φ e/ou modo-temporais, uma vez que é possível a realização fonológica dos morfemas modo-temporais em sentenças introduzidas por ku e kuma. No tocante ao estatuto morfossintático de kuma, os dados indicam que há duas entradas lexicais distintas: uma em que ele é complementizador e outra em que é verbo, visto que eles não co-ocorrem no mesmo contexto sintático. No caso de pa, este pode reter traços modo-temporais – resultando em uma oração reduzida ao infinitivo – ou traços de Caso – resultando em uma oração com Marcação Excepcional de Caso (MEC), em que o sujeito da oração encaixada assume a forma

de Caso oblíquo, em vez da de Caso nominativo. A MEC é, portanto, uma evidência favorável à hipótese de que o pa, no kriol, é, de fato, um complementizador.

A miratividade no caboverdiano

Wânia Miranda (Universidade de São Paulo)

Na língua caboverdiana, os nomes são, em geral, nus (Miranda, 2020). A língua possui estratégias semânticas para veicular definitude e indefinitude que prescindem da presença de determinantes. Embora exista, na língua, o determinante *un(s)* que pode veicular, entre outras coisas, indefinitude, sua presença não é obrigatória para que um dado sintagma nominal (NP) seja interpretado como indefinido (Miranda, 2013; 2015).

Em relação à partícula *uma*, dentro dos estudos sobre o sintagma nominal do caboverdiano, esta tem sido pouco explorada e, quando citada, é, em geral, tomada como a contraparte feminina do determinante *un* (Veiga, 2002; Delgado, 2008; 2009; Baptista, 2002; entre outros). Baptista (2002), além de considerá-la como tal, afirma que os informantes caboverdianos reservam seu uso a raras ocasiões, sendo sua ocorrência tão escassa em seu *corpus* que seria possível tomá-la como meramente acidental e não sistemática.

Em nossos estudos, contudo, observamos, em dados coletados na ilha de Santiago, que a utilização de *uma* se dá de forma sistemática na língua, com padrões de ocorrência próprios, além de prosódia específica, e não poderia ser considerada como a contraparte feminina do determinante *un*. Em que pese a coincidência de sua posição sintática por relação aos determinantes, essa expressão não compartilha das características semânticas e pragmáticas destes, como mostraremos no presente trabalho. Importante, ainda, é o fato de a língua caboverdiana não realizar concordância de gênero, exceto para seres animados, neste sentido, *uma* não poderia ser a contraparte feminina de *un* dado que não realiza sempre concordância de gênero com os nomes que modifica (como em *UMA pexi* (cf. 1), por exemplo).

Quint (2000: 187) analisa *un* e *uma* como adjetivos indefinidos e observa o caráter enfático desses adjetivos na língua. Delgado (2009: 283), em sua análise sobre a variedade de Barlavento, entende que *un* e *uma* são artigos indefinidos, e chama atenção para o fato de a interpretação trazer conotações desiguais, pois *uma* "só aparece em situações específicas, ou seja, quando se quer dar ênfase a um substantivo num determinado contexto de comunicação". Veiga (2012: 455), por sua vez, classifica *uma,* entre outras coisas, como um advérbio, considerando-a como uma partícula aumentativa relacionada à grau. No entanto, diferentemente do que apontamos em Miranda et al. (2010), *uma* também não pode ser considerado um modificador de grau, nos termos de Kennedy e McNally (2005), por não veicular grau do mesmo modo que outros modificadores da língua, tal como *txeu* (muito), por exemplo.

Neste trabalho, propomos que *uma* expressa miratividade no caboverdiano. A categoria da miratividade tem como função marcar sentenças que reportam uma informação que é nova ou

surpreendente para o falante, a proposição modificada pelo mirativo é nova, ainda não integrada ao conhecimento de mundo geral de quem a profere (DeLancey, 1997). Um dos significados mais consistentes e mais comumente observados nas línguas do mundo expresso pela categoria da miratividade é a surpresa do falante em relação a um evento ou uma ação (DeLancey, 1997; Aikhenvald 2012) e está ligada à noção de algo inesperado ou não usual. Indica que a mente do falante não estava preparada para a experiência narrada por ele e poderia, de acordo com Aikhenvald (2012), ser considerado o significado central expresso pela miratividade.

De acordo com nossas análises, a ocorrência da partícula *uma,* no caboverdiano*,* está estreitamente ligada, na grande maioria dos contextos, para não dizer em todos, à reação de surpresa do falante frente ao enunciado, atrelada à experiência pessoal, ou seja, o objeto ou evento modificado por *uma* é surpreendente ou excepcional a partir do ponto de vista do falante, evidenciando o caráter pragmático de seu uso. Em (1), ao modificar o nome *pexi* com o mirativo *uma*, o falante expressa a sua surpresa em relação ao peixe que foi pescado, evidencia a reação do falante diante do tamanho e/ou qualidade do peixe em questão:

1. Ami Ø peska UMA pexi!
   1SG.T PFV pescar MIR peixe[1]
   Eu pesquei o maior peixão!

A utilização de *uma* indicaria, deste modo, a reação do falante diante de um objeto ou evento que superaria a sua expectativa. A qualidade ou especificidade do objeto ou evento modificado por *uma* o torna exclusivo ou excepcional diante da expectativa do falante, do conhecimento de mundo que ele possuía até então. O uso de *uma* não seria adequado para descrever situações ordinárias e essas características o aproximam dos mirativos.

Ademais, embora ainda pouco estudada nas línguas africanas, a categoria da miratividade é atestada em línguas como !xun, hone, kisuaíli (Aikhenvald, 2012; Kihara, 2023), entre outras, assim como a da evidencialidade (Gibson; van der Wal, 2023), e pode ser um traço das línguas em contato que deram origem ao caboverdiano. Neste trabalho, portanto, apresentaremos a análise dos dados e os contextos de uso de *uma* a fim de fundamentar nossa hipótese a respeito da expressão de miratividade na língua caboverdiana.

1 1SG: primeira pessoa do singular; T: partícula tônica; PFV: aspecto perfectivo; MIR: mirativo