



HAL
open science

Documenting multilingual language practices and the erasure of language boundaries

Isabelle Léglise

► **To cite this version:**

Isabelle Léglise. Documenting multilingual language practices and the erasure of language boundaries. *Journal of Language Contact*, 2022, 15 (2), pp.404-437. 10.1163/19552629-15020005 . hal-03908274

HAL Id: hal-03908274

<https://cnrs.hal.science/hal-03908274>

Submitted on 21 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DOCUMENTING MULTILINGUAL LANGUAGE PRACTICES AND THE ERASURE OF
LANGUAGE BOUNDARIES¹

Isabelle Léglise

CNRS (French National Centre for Scientific Research), SeDyL Research Unit

isabelle.leglise@cnrs.fr

Abstract

Although we know multilingualism is the norm, most previous work has focused on languages as self-contained entities. Research on language contact mostly assumes bounded languages or repertoires: most studies presuppose contact between stable “communities” and the identifiability of specific languages in bilingual (sometimes plurilingual) corpora. Similarly, language annotation in corpus linguistics is based on the principle of univocity of items belonging to specific languages. In this paper, I address the notion of language boundaries, constructed both by linguists and by language users, and consider heterogeneity as a linguistic resource for speakers in their everyday multilingual language practices. First, there is a need for a shift in focus from linguistic systems toward language users. Second, there is a need for a solid methodology to reveal the heterogeneity of language practices through the annotation of plurilingual corpora. Ambivalence or play on boundaries is a common characteristic of communication in multilingual contexts that we can document.

Keywords:

Multilingualism, multilingual language practices, plurilingual corpora, language boundaries

1 Introduction

Although we know multilingualism is the norm and monolingualism an exception, linguistic description and the documentation of “minority” languages has long focused on languages as self-contained entities, assuming that languages are spoken within monolingual and monocultural communities. From the very early beginning of their disciplines, sociolinguists and linguistic anthropologists have criticized such bias, separating languages from their sociocultural context and ignoring questions of variation, multilingualism, and the social construction of language

¹ This paper has benefited from various comments and discussions. It is based on four invited lectures I gave in May 2016 at the Center for Research on Bilingualism (University of Stockholm), in June 2017 at the University of Chicago Center in Paris, in June 2018 at the 18th ACBLPE conference in Ziginchor, Senegal, and in January 2019 at the LSA in New York. I also thank two anonymous reviewers for their useful comments on a first draft.

(Hymes, 1967; Gal and Irvine, 1995). Minority languages – or more correctly minoritized languages because minoritization is not only about numbers but power relations – are rarely found alone but rather in contexts characterized by multilingualism. They are an intrinsic part of an ecology, or what has been recently called linguistic landscapes (Gorter, Marten, and Mensel, 2011) or multilingual peripheral sites (Pietikainen and Kelly-Holmes, 2013). Minoritized languages are spoken by plurilingual speakers with varying degrees of exposure to various languages, which may be languages of first socialization or lingua francas, and with different attitudes and ideologies with respect to them (O'Rourke, 2005) and with respect to multilingualism (Irvine and Gal, 2000). Within multilingual ecologies, power relations, language attitudes, and ideologies are always subject to (slight or important) change from one scale to another, one place to another, one interactional setting to another. Minoritization and deminorization processes (Léglise and Alby, 2006) are also part of the multiple ecologies that minoritized languages encounter. Heterogeneity, at all these levels, is thus the norm.

New trends in documentary linguistics (Grenoble and Furbee-Losee, 2010) have been in recent years interested in documenting multilingualism,² and this special issue is devoted to this topic. Such an approach was called for a decade ago. Lüpke (2010) argued against documenting West African languages without the inclusion of variations in language use, involving the choice of languages, dialects, styles, or registers, and Grenoble (2007) argued for documenting pragmatics, including native speakers' metalinguistic and ethnographic knowledge pertaining to language use, and for placing the sociocultural setting at the center of documentation projects. Léglise and Migge (2006) argued that in order to understand and document multiethnic contact areas, it is not only vital to assume an emic or so-called "native" point of view but to broaden the current linguistic anthropological notion of "emic" (Pike, 1964; Mondada, 2002) to include the perspectives of the different social actors involved in the area and situation – including non-speakers or outsiders which also contribute to shape circulating ideologies. They devised a methodology for the documentation of languages with specific reference to multilingual areas and proposed a three-step procedure. First, analysis of naming conventions. Second, analysis of language attitudes and ideologies using a discursive method, and third, a linguistic analysis of language varieties. Migge and Léglise (2013) developed a more complete methodological approach to documenting language in multilingual ecologies, integrating interactional-sociolinguistic, anthropological-linguistic, discourse-analytical, and quantitative-sociolinguistic approaches with structural description to language speech. They proposed to apply a range of complementary data collection and analysis methods, and positioned issues of language ideologies, variation, contact, and interaction on center stage. Although some recent publications devoted to documenting rural or indigenous multilingual zones do take variation and language ideologies seriously (see for example Lüpke (2016) and Carlo (2018)), mainstream language documentation and field linguistics do not seem to have changed their modus operandi (Migge, 2020).

Many scholars in the field of the sociolinguistics of multilingualism with a background in the sociology of language or sociolinguistics have long documented the coexistence of languages on specific sites, either a territory (how languages are distributed in space), a linguistic repertoire at the level of a community or a speaker (how many languages and what kind of resources coexist), or a domain (who speaks what language to whom and when). Gumperz' research in northern India (1958) and the contributions of Fishman (1965) are probably the most influential work to have shaped those fields. Although it may have been argued that the contemporary sociolinguistics of multilingualism has focused on European contexts, where concepts such as superdiversity have arisen (Blommaert and Backus, 2011), or on contacts between a colonial language and indigenous ones, there is a long tradition of sociolinguistic and linguistic anthropological work within traditional or indigenous ecologies, in India, Australia, Africa, and South America. When working on multilingualism in postcolonial settings, we need to re-center on local voices, adapt our conceptual

² Workshops including "Language documentation and linguistic diversity," "Small-scale multilingualism" (SOAS, 2016), and the LSA Annual Meeting Satellite Workshop on Documenting Multilingualism (New York, 2019), among others.

tools (Léglise, 2017), and consider other ways of conceptualizing language in the various ecologies, following the critical assessments by colleagues from the Global South (Makoni and Pennycook, 2007; Bagga-Gupta and Dahlberg, 2018; Heugh and Stroud, 2019).

In a way, we could argue that while field and documentary linguistics do still focus on languages as self-contained entities, even the documentation of societal multilingualism by sociolinguists has long focused on the (external) coexistence of bounded languages. For example, illustrations of the coexistence of languages in a single space can be found online for French Guiana,³ both at the level of a village and of the entire territory, as well as at the level of linguistic repertoires and how individuals use or combine various languages depending on their interlocutors (Léglise, 2013). Even though many studies have been conducted under the label of codeswitching, what is still to be done is to document multilingual language practices that exhibit intrinsic heterogeneity because plurilingual corpora are not gaining visibility in fields such as corpus and contact linguistics.

In this paper, I will more specifically address the notion of language boundaries in multilingual practices, constructed both by linguists and by language users, and will view heterogeneity as a linguistic resource for speakers in their everyday multilingual language practices. This may sound unfamiliar to linguists who work on language documentation, contact or corpus linguistics, since – by referring to linguistic resources for speakers – I draw on current conceptions of the speaker as a language user and social actor in everyday interaction from an interactional sociolinguistic dynamic perspective (Gumperz, 1982) where social interactions produce social effects (Boutet, Fiala, and Simonin, 1975) and language users construct social meaning, including positionings, identity or political claims (Kroskrity, 2000; Bucholtz and Hall, 2005) through the use of linguistic resources linked to indexicality and language ideological processes like iconization, recursivity, or erasure (Irvine and Gal, 2000). In order to do so, as the history of 40 years of research conducted under the label of codeswitching has shown, there is first a need for a “shift in focus from linguistic systems toward language users [... and their] repertoires drawn from lived experiences that may disrupt presumed connections between language, community, and spaces” (Hall and Niley, 2015: 615). Second, there is a need for a solid methodology to reveal the heterogeneity of language practices within plurilingual corpora. I will show how ambivalence or play on boundaries is a common characteristic of communication in multilingual contexts. Language users make use of all kinds of linguistic resource in order to communicate, and when they share more or less the same language(s) background they sometimes use shared or ambivalent elements, attributable to various languages or varieties. In doing so, they do not perform language boundaries but instead “float” in mid-stream. This is particularly obvious in postcolonial contexts (involving, for example, close language varieties or a Creole and its lexifier in decreolization contexts). Woolard (1998) showed that the uses of what she calls bivalent features may be “politicized and controversial” or may represent a linguistic resource that is “strategically marshaled” by language users to position themselves in their everyday interactions. I will show that this is linked to processes of homogenization and differentiation in language practices themselves.

In section two of this paper, I will show how heterogeneity is treated in various fields, such as contact linguistics and corpus linguistics; I will address more specifically the question of codeswitching and other labels that have been recently proposed in the sociolinguistics of multilingualism, such as (poly/trans) languaging as linked to heterogeneity. Section three is devoted to the annotation method we propose to apply to reveal the heterogeneity of language practices within plurilingual corpora. I will illustrate some solutions for annotating and analyzing corpora of spontaneous speech by means of a multilayered annotation system based on JAXE. I will also show how languages or varieties may overlap in these corpora, making it irrelevant to draw arbitrary lines and boundaries between linguistic resources. This has of course an impact on the way we as linguists consider the attribution of language (or variety) labels to linguistic forms when annotating or analyzing data. In section four, I will describe a tendency, among bilinguals sharing the same languages background, to use unmarked or bivalent features that can belong to two or more languages or varieties, focusing on simultaneities that are part of plurilingual communication.

³ <https://leglise.cnrs.fr>

Section five will show how social actors make use of linguistic resources, marking or unmarking language boundaries in multilingual interactions. I will show how using bivalent features as linguistic resources may perform fuzziness and challenge language and social boundaries, making it possible to avoid traditional ethnolinguistic affiliations and to perform new identities, such as urbanity and masculinity. Finally, I will show that it is always possible to use specific linguistic features to mark dialectal or language boundaries by way of dis-alignment, differentiation, and dis-affiliation, which create social meaning.

2 Heterogeneity in language practices from multilingual settings

Although multilingualism and linguistic bricolage are far from being uncommon, heterogeneity is not gaining visibility in fields such as corpus linguistics, nor, to a lesser extent, contact linguistics. I understand here heterogeneity to mean internal linguistic diversity in terms of linguistic resources – not only languages but also features associated with dialects, styles, or registers sequentially or simultaneously used in language practices. Heterogeneous language practices are still treated as exceptions and few heterogeneous corpora are available.⁴ In corpus linguistics, what are generally called “multilingual corpora” include texts in different languages but in which each text is monolingual (Botley, McEnery, and Wilson, 2000). So-called “comparable multilingual corpora” are comparable in number and type, such as the genre of the texts in each language. So-called “comparable parallel corpora” display relations of equivalence and translation between the items that make up the various texts and their translations. As a consequence, corpus linguistics is still essentially restricted to comparison of stable, monolithic objects, either languages or genres.

With the focus on heterogeneity and hybridity in the humanities, language contact emerged as an important area of research. Research in contact linguistics has provided many insights into the types of contact phenomena, settings, and conditioning factors (Winford, 2003). However, investigation mostly deals with cases of diachronic contact and completed changes. Within the field of contact linguistics, two trends can be identified but both assume that languages are bounded. Research on contact-induced language change mostly focuses on one language and the reorganization within that language due to processes and outcomes of contact with another language (Thomason, 2001; Heine and Kuteva, 2005), leading sometimes to complete reorganization or to the rise of mix languages (Matras and Bakker, 2003). 15 years ago, language contacts were generally viewed as marginal phenomena, arising when “normal” monolingual situations become more complex (Nicolai, 2007). Contact linguistics also mostly deals with completed changes in a diachronic perspective, although some authors proposed to apply the same methodology on ongoing processes (Léglise and Migge, 2005) or on the relation between ongoing language variation and contact (Léglise and Chamoreau, 2013). Occasionally contact linguistics has focused on language change in multilingual areas, essentially through the concepts of convergence and “Sprachbund,” adopting the perspective of areal typology (see for example Friedman and Joseph 2017). Long-term linguistic convergence has been shown for various languages in the Balkans since the work of Trubetzkoy in the 1930s, and more recently in Suriname viewed as a linguistic area (Yakpo and Muyskens, 2020). A number of studies challenge the unidirectionality of language change in multilingual settings. Migge and Léglise (2011) showed that various social forces may give rise to contradictory linguistic development: on the one hand a reduction of diversity (through the reduction of differences between varieties, i.e. leveling, or through contact-induced diffusion and

⁴ To mention a few: from section 3 onward I will draw on the Clapoty corpora (this first step of the research was funded by the ANR-09-JCJC-0121-01, 2009-2014), featuring 40 languages in various combinations (from 2 to 10) and settings: <http://clapoty.vjf.cnrs.fr/index.html> ; there is also the Crossroad corpus which explores a situation in which 4 languages are at play in Casamance, Senegal (SOAS, 2014-2018): <https://soascrossroads.org/>. Corpora involving 2 languages are more frequent although rarely available but see the <https://biling.talkbank.org/> and CHILDES project for child language acquisition: <https://childes.talkbank.org/>

convergence), and on the other hand an increase of diversity (through the process of construction of differences), leading either to koineization and lingua francas, for example, or to the emergence of new varieties, as in the case of Maroon languages in French Guiana and Suriname. Similar tensions have been exemplified in the case of indigenous multilingualism in the Vaupès region when comparing pre- and postcolonial contacts (Epps, 2018), in Northern Australia (Vaughan, 2018), and in West Africa, a zone in which fragmentation and high linguistic diversity coexist with convergence areas (Lüpke and Watson, 2020).

In contrast, research on codeswitching, focusing either on its linguistic structure (Myers-Scotton, 1993; Poplack, 1981) or on its social meaning (Auer, 1998; 2005), deals with synchronic spontaneous speech recordings, identifying clearly the various alternating “codes.” Terms such as “matrix language” and “code-alternation,” commonly used in these two subfields, are good examples of the fact that they both assume bounded languages or bounded repertoires (see Légise, 2018 for a discussion). Codeswitching has been prototypically the locus of analysis of linguistic heterogeneity, that is to say the presence of (generally) two languages in the same corpus. While the language pairs first studied were almost always European languages, by now a large number of language pairs around the world have been studied under the umbrella of codeswitching. Auer’s (1999) typology, which is a way to categorize the various forms and functions bilingual speech may adopt (alternational or insertional codeswitching, discourse-related codeswitching, code-mixing, fused lects, etc.), has been applied to various peripheral sites from South America (Alby and Migge, 2007) to Australia (O’Shanessy, 2006), from the Balkans and Thrace (Adamou, 2010) to South Africa (Stell, 2015). However, the term codeswitching by itself has come to be seen as either too narrow or too broad to describe the reality of experiences of linguistic heterogeneity, such as crossing (Rampton, 1998; 2005), transidiomatic practices (Jacquemet, 2005), polylinguaging or languaging (Jørgensen et al., 2011), translanguaging (Creese and Blackledge, 2010; García and Wei, 2013), and translingual practice(s) (Canagarajah, 2012; 2017). All these terms have been coined in the last 15 years in the fields of the sociolinguistics of multilingualism, linguistic anthropology or applied linguistics, to better grasp the reality they try to describe, in the attempt (as I see it) to give a voice to heterogeneity. Whether codeswitching is an old term that should be thrown away, given its association with code-alternation, meaning the alternation of named languages (Otheguy, García, and Reid, 2015) or “doing codes”(see Auer forthcoming for a discussion), is another question. Looking back at 40 years of research on codeswitching, Hall and Nilep (2015) note that there has been a shift from the first studies of how bounded groups or communities come into contact and what the consequences are, to more recent work on how social actors reproduce, construct, and negotiate positions, relations, groups, social meanings, and the construction of boundaries through play on boundaries.

First is a shift in focus from linguistic systems toward language users. The earliest research in the field viewed languages as discrete systems in contact. Studies under the heading of code-switching or related terms shifted analysis toward the people at the edges of communities and languages and then to discourse practices straddling such edges. Much recent work centers on repertoires drawn from lived experiences that may disrupt presumed connections between language, community, and spaces (Hall and Nilep, 2015: 615).

In my view, whether we ought to forget the term codeswitching in light of those more recent studies is not the most important question here. The most important is how to understand multilingual language practices, that is, practices primarily made of language (the activity of language) and not of languages as bounded entities (as could evoke the term linguistic practices). I draw here on the French tradition of work on language practices. “Pratiques langagières” (Boutet, Fiala and Simonin-Grumbach, 1976) is a term coined 45 years ago to convey the fact that language practices are determined and constrained by a social order and at the same time construct social meaning, produce social effects, and contribute to change the social order. They are a product of the activity of language (Boutet, 1995). “Languaging” might be a good translation. As Jørgensen and Juffermans (2011: 2) put it:

A languaging perspective sees language in actual practice not as bounded, countable entities that are given in the natural world, but as dynamic, creative potential to speak. It

emphasises that people do not primarily use “a language,” or “some languages,” but use language, linguistic resources. Bilinguals are not seen as “speaking two languages,” but as languagers making use of resources that are recognized by the speakers or others as belonging to two sets of resources.

In a monolingual context, there are various social evaluations of language practices (Bakhtin/Voloshinov, 1973), and language practices are socially heterogeneous. In a multilingual environment, these language practices are also heterogeneous because they are made up of linguistic resources attributable to various sources and resulting from the diverse linguistic repertoires of the social actors. They thus have their own internal heterogeneity in addition to the social heterogeneity of their evaluations (Léglise, 2013). They can be made up of all sorts of features which are either attributable to various languages or indexical to various varieties, styles, or registers. These heterogeneous practices might illustrate linguistic bricolage (Lüdi, 1994; Mondada, 2012) or ways by which social actors use linguistic resources in order to create new meanings. They are performed by plurilingual speakers with varying degrees of exposure to various languages (with superdiverse repertoires, as Blommaert and Backus (2013) put it), and consequently include a lot of non-standard forms.

Such ordinary heterogeneous practices are in many cases unremarkable for ordinary plurilingual speakers (see Stroud, 1998; Makoni and Pennycook, 2007, for example), when plurilingual social actors are discussing among themselves (i.e. in endolingual bi/multilingual settings if we follow de Pietro’s 1988 typology),⁵ or when plurilingual social actors are trying to communicate in exolingual settings where they do not share the same linguistic backgrounds or repertoires as their interlocutors. Sometimes, however, such practices are remarked on and viewed as highly political; they may be prohibited or described as “broken language” either by monolingual institutional regimes such as schools (García, 2009; Creese and Blackledge, 2011) or by elders nostalgic for a vanished golden age seen as monolingual (Léglise and Alby, 2006). They may simply be routinely and consistently devalued (Kroskrity, 2000). If long-term ethnography, macro-sociological evidence, and discourse analysis are necessary, especially in non-western settings, to understand “how people perceive their lives as well as [to provide] an understanding of societal dynamics” (Stroud, 1998: 323), they make it possible to identify speakers’ different attitudes and ideologies with respect to multilingualism (Irvine and Gal, 2000) and to their environment, including the various languages and varieties available (Léglise and Migge, 2006). But language ideologies are not fixed; they are always subject to change from one discourse to another (Canut, 2001) and from one interactional setting to another: here a certain wording might be indexical of a certain divergent positioning, there the same wording might be used by another speaker to justify a position contrary to the previous one (Migge and Léglise: 2013, ch. 4). So, while emic perceptions are essential, it is not possible to rely only on speakers’ intuitions⁶ or on metalinguistic comments on their practices and attitudes; the conflicting ideologies at play (Kroskrity, 2000) must always be addressed with a precise analysis of heterogeneous language practices (Migge and Léglise: 2013, ch. 9). Finally, language ideologies as well as minoritization and deminoritization processes are also dependent on spaces of multilingualism (Blommaert, Collins, and Slembrouck, 2005), as are the multiple identities and positions entangled on different sociolinguistic scales such as community, region, and nation, enabling contradictory discourses and perspectives to simultaneously overlap (Léglise and Migge,

⁵ In de Pietro’s typology, bi/multilingual-endolingual settings would involve interlocutors sharing the same languages and communicating in a sort of fluid manner whereas bi/multilingual-exolingual settings involve interlocutors that do not share the same repertoires and show some sort of adjustments and asymmetry within communication. Monolingual-exolingual settings typically would for example typically involve a so-called ‘native’ speaker and a language learner and the interaction would be full of traces of this asymmetry.

⁶ As Gumperz (1982) already argued, so-called native speakers generally have few intuitions about or recognition of their own conversational code switches (see also Nilep, 2006).

2019). Heterogeneity is thus the rule at all levels of production and interpretation of multilingual practices.

We will now turn to the methodology chosen for work on heterogeneous corpora made up of multilingual practices.

3 A methodology for work on plurilingual or heterogeneous corpora

According to the discussion of heterogeneity in Section 2, plurilingual or heterogeneous corpora are corpora composed of heterogeneous language practices. Their internal heterogeneity is the result of features attributable to various languages or to variation and non-standard forms. These forms are harder to classify and pose formidable problems, with respect not only to the identification of forms but also their transcription and annotation. These plurilingual corpora are still few in number, not readily available to the community of linguists, and with very few specific annotation tools.

Based on these observations, the Clapoty project was developed a decade ago with the following series of goals: 1) to put together a common plurilingual corpus – each participant has to put some of their data at the project's disposal, 2) to develop standards of annotation (including linguistic, ideological, and social metadata) and methodology for analysis and a multiple-factor explanation of the linguistic consequences of multilingualism, 3) to set up a database displaying the linguistic consequences of language contact together with suggestions for analysis, and 4) to develop computing tools specifically designed for plurilingual and heterogeneous corpora. Here I will only present briefly two scientific and technical innovations from stages (2) and (4) that could be valuable for documenting multilingual practices. Both are based on a JAXE-based annotation system that enables a multilayered analysis of corpora and contact phenomena at morphosyntactic, discursive, and interactional levels. JAXE is an open-source text editor in Java configurable with XML schemas. The XML schema named “Corpus-Contact”⁷ (Vaillant, 2010) has been developed within the Clapoty project to standardize the annotation and visualization of plurilingual corpora based on a non-aprioristic approach to contact phenomena that can be evaluated a posteriori in the light of different theoretical frameworks (see Vaillant and Léglise, 2014). The Text Encoding Initiative (TEI) proposes a set of standards for annotating corpora commonly used in corpus linguistics. Here is a small section of the schema developed using TEI: it represents the identification of speech turns, speakers, and languages. Along with a presentation of the interactional context and categorization of the contact setting through three typologies,⁸ the header of the corpus comprises both languages and speakers' inventories and specifically the speakers' linguistic repertoires, which is essential for the allocation of languages.

⁷ The “Corpus-Contact” annotation schema follows most of the TEI standards. It adopts the Unicode encoding for transcribing texts, ISO-639 codes for identifying languages, and an XML document markup for exporting documents. It is therefore interoperable with other annotation tools such as ELAN / TOOLBOX. After having transcribed, segmented, and annotated the plurilingual corpus, a transform style sheet allows any browser compliant with XSLT to display the corpus as a sequence of aligned speech-turns.

⁸ Each corpus is categorized following three contact typologies of multilingual settings: Auer (1999) on the types of codeswitching/mixing, Winford (2003) on the types of contact scenario, and de Pietro (1988) on the multilingual interactional setting and language background of the interlocutors.

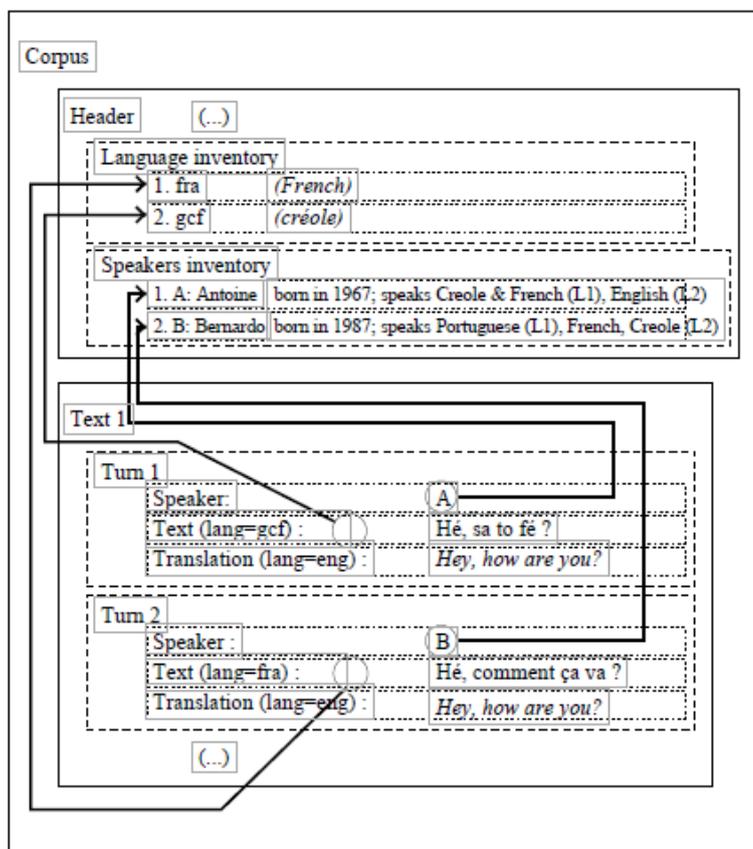


Figure 1: Inventory and identification of languages and speakers (Vaillant: 2010, 12)

The TEI P5 Guidelines (8.4.5) recommend identifying the basic language of each sentence, and noting in angle brackets when an item from another language is introduced, for instance as <foreign item belonging to language x>, as in the case of the sentence: “I proposed that wir können vielleicht go to Warsaw and Vienna” where the basic language is English and the “foreign” insertion comes from German:

Finally, phenomena such as code-switching, where a speaker switches from one language to another, may easily be represented in transcript by using the <foreign> element provided by the core tagset:

(1)

<u who= “#P1”> I proposed that <foreign wml:lang= “de”> wir können <pause dur= “PT1S”/> vielleicht </foreign> go to warsaw and <emph>vienna</emph></u>

These TEI standards are based on the tagging of languages and follow a strict utterance-based model or matrix-based model (Jake, Myers-Scotton, and Gross, 2005). Such standards are usually used for the documentation of less-described languages (Mettouchi, Vanhove, and Caubet, 2015) and for instances of borrowing or classical “insertions” such as codeswitching (Manfredi, Simeone-Senelle, and Tosco, 2015). However, an initial problem in the case of heterogeneous practices is that the systematic allocation of one basic language for each speech turn is not possible (see Légise and Alby, 2016, for a fuller discussion).

For example, the following extract is made up of French and French Guianese Creole. The interaction takes place at the office of the national electricity company in Cayenne, French Guiana, where a customer has an appointment with a company agent in order to solve a problem. After some turns in French and discussion using some Creole words, the employee addresses the customer, alternating elements in French and Guianese Creole. Instead of deciding arbitrarily if this speech turn is produced in one language (French because it begins and ends in French) with an insertion in another (Creole), or the reverse (Creole because most verbs and syntactic material are in Creole with

insertions and discourse markers in French), we prefer to annotate the language of the utterance as <multiple> (i. e. multilingual) both in French and French Guianese Creole. This has of course many implications for the way we look at our data and hence for the statistics⁹ we can derive from our corpora. As in most cases, we observe several linguistic resources in the same speech turn produced by the same speaker.

Here is a screenshot of the visualization of this line of the corpus with multilayered glosses¹⁰:

AGT :	090.					rendez-vous	
	-01.	bon	écoutez ↗	nou	ka	roupran	rendévous
		bon.INTJ	écouter.IMP	1PL	AS.NPRF	reprendre	rendez-vous
		PRT	V	PRN	PRT	V	N
		<i>bon écoutez on va reprendre un rendez-vous</i>					
		so listen we will make again an appointment					
	090.						
	-02.	mais	là	rendévous	-a	pou	mars
		mais	là.ADV	rendez-vous	ART.DEF	pour.PREP	mars
		CONJ	ADV	N	DET	ADP	N
		<i>mais là le rendez-vous est prévu pour mars hein ↗ /</i>					
		but here the appointment is for March, right?					
	090.						
	-03.	i	pa	gen	rendévous	avant ↘ /	
		3SG	pas.NEG	avoir	rendez-vous	avant.PREP	
		PRN	PRT	V	N	ADP	
		<i>il n'y a pas de rendez-vous avant /</i> There are no appointments before					

Fig. 2: Visualization of line 90 of a plurilingual corpus with multilayered glosses

This extract exemplifies two major adaptations for heterogeneous corpora that we have introduced, concerning 1) the language of the utterance, which is “multiple” in most cases (it appears in yellow in our transcripts), and 2) the allocation of the multilingual attribute to various morphosyntactic segments (it appears in blue here with the various possibilities noted as if they were on a musical staff). Some segments are attributable to several languages and lexical items, such as “rendez-vous” in French or “randévous” in Creole; these are quite common examples though not the most interesting ones. We decided to make these interpretative possibilities visible in the transcription in any case, meaning that at the representational level of transcription (Bucholtz, 2000) we make various interpretative possibilities visible (in terms of language sources) without choosing between them.

In order to visualize extract as shown, we first need to annotate the corpus in XML, as shown in (2) for only the first line of the interaction:

(2)

```
<speech turn text lang=mul> <segment lang=fr> bon écoutez </segment> <segment lang=gcf> nou ka roupran
</segment> <segment lang=mul> <alternative transcript lang= gcf> randévous </alternative transcript>
```

⁹ A PERL concordancer allows for complex queries on the linguistic data; queries on the associated metadata allow for combining quantitative results with the speakers’ linguistic background and biography or the specificities of the sociolinguistic multilingual setting, for example.

¹⁰ Line 1: transcript with French in roman and Creole in bold; line 2: interlinear translation morpheme by morpheme with morphematic gloss derived from the Leipzig Glossing Rules; line 3: part of speech (POSTAGS); line 4 and 5: translation (both in French and English here).

ordinary plurilingual speakers in various everyday life situations around the world, such as Spanish dialect contact and Quichua in Colombia (Sánchez Moreano, 2015), Casamance Creole, French, and Wolof in Senegal (Nunez and Léglise, 2017), and Arabic dialects in contact with various languages (Manfredi and Istanbulu, 2019). Those recordings involve all sorts of cross-cultural contacts, multilingual speech turns, and multi-participant interactions in which different styles may also occur. The finalized corpora are associated with a relational database storing metadata on speakers, their language biography, and linguistic repertoires, and on the multilingual setting both in synchrony and diachrony (including language ideologies and power relations on the micro- and macro-level), enabling the analysis of remarkable phenomena in the light of various linguistic and sociolinguistic variables (see Léglise and Alby, 2016: 374–375). All this information is available as secondary data for the analysis. The next section shows how this kind of transcript makes it possible to interpret the data from several perspectives, and in particular to look at simultaneities in heterogeneous practices.

4. Bivalency and simultaneities in heterogeneous and multilingual practices

A particularly important point regarding heterogeneous practices is linked to the question of simultaneity in plurilingual communication. Here again a number of terms have been coined to refer to the fact that – as we would say – some elements could be attributable to at least two languages or varieties with various consequences: these homophonous diamorphs (Muysken, 1990) or cases of bivalency (Woolard, 1998) could lead to interlingual identities (Haugen, 1972), interlects (Prudent, 1981), code gliding (Baggioni, 1992), or floating zones (Ledegen, 2012). Simultaneity refers to “the use by a bilingual of words or segments that could ‘belong’ equally, descriptively and even prescriptively, to both codes” (Woolard, 1998: 7). Previous studies have focused for example on close Romance languages in contact as in Spain (see Woolard (1998) for Catalan and Castilian and Alvarez-Cáccamo (1990) for Galician and Castilian).

In my view, the term bivalency perhaps shows most clearly the “simultaneous membership of an element in more than one linguistic system” (Woolard, 1998: 6). Woolard gives a good example of Catalan and Castilian, where at the beginning of a comedy sketch, in “El saben aquel” (Do you know the one) the initial pronoun is Catalan, the verb both Catalan and Castilian, and the demonstrative pronoun Castilian, enabling the speaker to play with his linguistic resources:

5)	<i>el</i>	<i>saben</i>	<i>aquel</i>
	3SG	know.3PL	3SG
	Catalan	Catalan or castillan	Castillan
	'do you know the one'		

“We may discern in single utterances the simultaneous and equally concrete (or equally ephemeral) presence of more than one value of a paradigmatic contrast,” she writes (ibid.,: 5). “A bilingual practice can dismantle (but does not simply neutralize) binary distinctions, in ‘an undoing yet preserving of all opposition’ that ‘keeps alive an unresolved contradiction’ (Spivak in Derrida, 1974: xx 1)” (ibid., 6). The speaker is “drawing on similarities to inhibit definition of the variety they are using.” For Woolard, bivalency is thus a linguistic resource that could be “strategically marshaled and rhetorically manipulated” by speakers.

Based on our observation of extensive corpora in the Clapoty project, I propose the idea that the presence of features that may belong to two or more language varieties can sometimes be particularly frequent, especially in endolingual settings, i.e. among plurilinguals sharing the same languages (De Pietro, 1988) or linguistic repertoires. We compared contact between unrelated languages in a multilingual environment such as Casamance in Senegal (Nunez, 2015) and contact in postcolonial settings where a Creole and its lexifier are still in contact, such as French and Creole on the island of Réunion (Ledegen, 2012) and in French Guiana (Léglise, 2013), and also contact in multilingual settings involving related languages and varieties, such as various English-based Creoles varieties in French Guiana and Suriname (Migge and Léglise, 2013) and Spanish varieties in Colombia (Sánchez Moreano, 2015). Ledegen (2012) has documented that bivalent predicates make

up 16 per cent of all predicates in the decreolization context of Réunion, whereas in a multilingual context such as Casamance only 4 per cent of the linguistic features are bivalent across Portuguese-based Casamancese Creole, Wolof, and French (Nunez, 2015). Although 4 per cent is low, it is far from negligible, and this dimension is generally neglected in plurilingual corpora. Although 16 per cent is somewhat higher, this bivalency has not previously been operationalized in plurilingual corpora. Moreover, in some of our corpora including several English-based Creoles in French Guiana, up to 80 per cent of the features can be bivalent. We will concentrate on this situation here and show some of its consequences for linguistic analysis.

From a linguist's perspective, the multilingual ecology of western French Guiana¹² includes, among a dozen assorted languages, at least two English-based Creoles: one is Sranan Tongo, a language that developed among the Afro-Surinamese Creole population, especially in the coastal region, and is now a vehicular language in Suriname and also in western French Guiana; the other is Eastern Maroon Creole, a general designation for the languages spoken by the Maroon communities that emerged between 1690 and 1800 as a form of resistance to slavery in Suriname. Each community is headed by a paramount chief, consists of lineages and matrilineal clans, and traditionally identifies itself through an ethnic denomination similar to its language name variety (Aluku, Ndyuka, or Pamaka in French Guiana). Maroons may also refer to their language as *businenge tongo*, a concept close to the label Eastern Maroon Creole used by linguists. Since the 1980s the civil war in Suriname, massive displacement, and urbanization have profoundly altered lifestyles and language practices. The following extracts were recorded among Maroons in western French Guiana during long-term ethnographic fieldwork (Migge and Léglise, 2013: ch. 8).

In some of these corpora, most turns are composed of bivalent or shared elements attributable to Sranan and/or Eastern Maroon Creole. For example, of the nine elements in the extract (6), six are shared by Sranan Tongo (bold underlined) and Eastern Maroon Creole (EMC, in italics), one is clearly associated with Sranan Tongo (*no*), and two are associated with EMC (*leli* and *puu*).

- 6) **fu** **a** **no** **go** *leli* **fu** *puu* **a** **moni**
fu *a* *go* *leli* *fu* *a* *moni*
 FOR.COMP 3SG.SUBJ NEG go learn FOR.COMP pull ART.DEF.SG money
 'so that she does not learn to take money and waste it all'

Transcribing and annotating such corpora poses some irresolvable questions. A classic TEI perspective on codeswitching would try to find the matrix language of the following extract and treat it either (7) as Eastern Maroon Creole talk with some Sranan Tongo insertions (bold underlined) or (8) as Sranan Tongo talk with some Eastern Maroon Creole insertions (in italics). The POS annotation and translation are given in (9) with no indication of languages.

- 7) *i* **ne e ferstan** *san mi e du nou / mi* **kan rei tu trip, i ferstan tok / mi kan lei tu trip / mi e lei**
den man fu mi
 EMC talk + Sranan insertions

- 8) **i ne e ferstan san mi e du nou. Mi kan rei tu trip, i ferstan tok, mi kan lei tu trip / mi e lei**
den man fu mi
 Sranan talk + EMC insertions

- 9)  *ne e ferstan san mi e du nou*
 2SG NEG IMPF understand what 1SG IMPF do now
 'you don't understand what I'm doing now'

¹² I provide very basic information here for the understanding of this complex ecology. For a more detailed presentation, see Léglise and Migge (2006).

mi kan rei tu trip i ferstan tok
 1SG can drive two trips 2SG understand right

'I can drive two trips, you understand right'

mi kan lei tu trip mi e lei den man fu mi
 1SG can drive two trips 1SG IMPF drive DET man POS 1SG
 'I can drive two trips, I drive my guys'

Of course, defining the language of an utterance or a talk is highly ideological, and this definition can change depending on the people involved and the circumstances. In this last case, a linguist could argue that, as almost all the elements could be Sranan, the matrix language of the utterance should be Sranan. The speaker could argue instead that because of his ethnicity, being a young Maroon adult, the language of the utterance is the variety he speaks and belongs to (i.e. EMC or a specific variety such as Ndyuka). When working on this example, I discussed it at length with my colleague Bettina Migge, a specialist in EMC. We could see both perspectives. Based on the ethnicity of the speaker, she argued in favor of (7) as the right transcript to adopt, whereas I was looking for formal distinctions as solid proofs and did not find any, so I argued that (8) was also a possibility. This made me realize that it was not possible to decide between the two interpretations from a honest perspective. So, instead of trying to see 'language identification' as a problem, I propose to see it as a chance, an opportunity to reveal various possible interpretations.

Once we recognize the impossibility of choosing between two interpretations, that is to say two possible attributions of language labels to linguistic forms – on a large scale, not only for some isolated features – we can see the point of the annotation standards proposed in section 3 and the double transcription as exemplified in (10):

10) *i ne ferstan san mi E du nou*
i E san mi E du nou
 2SG NEG IMPF understand what 1SG IMPF do now
 'you don't understand what I'm doing now'

mi kan rei tu trip i ferstan tok
mi tu i
 1SG can drive two trips 2SG understand right
 'I can drive two trips, you understand right'

mi kan tu trip mi e den man fu mi
mi lei tu mi e lei den man fu mi
 1SG can drive two trips 1SG IMPF drive DET man POS 1SG
 'I can drive two trips, I drive my guys'

This presentation not only makes it possible to see various perspectives and interpretations, it invites the reader to consider those various interpretations, either sequentially (as EMC talk + Sranan insertions or Sranan talk + EMC insertions) or simultaneously because we could also argue that (10) is neither EMC or Sranan but EMC and Sranan at the same time. Bivalent elements make us see simultaneities in multilingual practices. Bivalent elements transcribed on a musical staff are a good example of Bakhtin's polyphony: a speech is full of multiple voices or speaking positions (Holquist, 1990; Woolard, 1998).

Based on previous work on social identities among Maroons, we know that although Maroons identify with their traditional language and ethnicity and tend to reject Sranan as the language of the Afro-Surinamese Creole population, urban Maroon adults frequently use Sranan, which is also indexical of the urban environment. Migge (2007) showed that young adults generally employ codeswitching between EMC and Sranan to construct themselves as sophisticated and urbanized Maroons, and employ code-mixing to assert membership in the social group of young men whose salient characteristic is modern urban sophistication. But this switch between linguistic features associated with one or the other language is not the only interesting feature. Bivalent elements, as exemplified in (10), are, in my view, the best way to display this dual membership – not entirely Eastern Maroon Creole nor entirely Sranan, not entirely traditional Maroon nor an urban member of the Afro-Surinamese population but both at the same time.

We see clearly here how languages may overlap in multilingual practices that are heterogeneous, just as cultural and identity constructs do, making it irrelevant to draw arbitrary lines and boundaries between linguistic resources in a single and simple transcript based on the principle of univocity. This ought to have an impact on the way linguists consider the attribution of language labels to linguistic forms in multilingual practices: while labels are necessary in documentary or corpus linguistics in order to clearly identify forms and enable subsequent searches, they should be as open-minded as possible and should include the mention of various possible perspectives and interpretations, allowing the reader to construct her own¹³. This also offers the possibility of studying simultaneities and play with language boundaries during interactions, as we shall see in the next section.

5. Performing fuzzy language boundaries or marking them within heterogeneous practices

I would like to show in this section how the methodology for annotating plurilingual corpora, revealing both heterogeneity and simultaneities, enables sociolinguists to document how language boundaries may be constructed or deconstructed in face-to-face interaction. The use of specific linguistic resources in multilingual practices may occur unnoticed in most of the cases or may be very meaningful – as in the case of reformulations.

As already exemplified in extracts (6)-(10), among people sharing the same languages or close language varieties within their linguistic repertoires, for example among peer groups, we see the tendency to use common, shared elements in very long extracts. Here is another extract with EMC and Sranan where the vast majority of the linguistic features are bivalent (EMC and Sranan) and where some insertions in English are used (such as *freedom*). This is a common way of speaking for young adults living in urban centers in western French Guiana and Suriname:

¹³ As data can be analyzed through different lenses, this kind of double annotation could enable a structural linguist to analyze the data through the lens of koineization or convergence among languages in contact – or to choose a univocal matrix language based on her criteria and interpretation. Showing bivalency in the transcript does not reveal a theoretical paradigm but allows various paradigms to read the data.

003. -01.	i	wani	go	na ¹	dape ²	a	didon														
	i	wani	go	na	dape	a	didon														
	2.SG.SBJ	want.MOD	go	at.PREP.	there	3.SG.SBJ	ly.down														
	PRN	PRT	V	ADP	ADV	PRN	V														
	<i>Do you want to go there where he is buried?</i>																				
003. -02.	san ³	a	man	dede	dape?																
	san	a	man	dede	dape?																
	INT	ART.DEF.SG	man	die	there																
	PRN	[DET	N]GNV	V	ADV																
	<i>Why did the guy die there?</i>																				
003. -03.	Na	a	freedom	a	man	e	suku	A	bun	san	a	man	si,								
	Na	a	freedom	a	man	e	suku	A	bun	san	a	man	si,								
	FOC	ART.DEF.SG	freedom	ART.DEF.SG	man	AS.IPFV	search	3.SG.SBJ	good	what.INT	ART.DEF.SG	man	see								
	PRT	[DET	N]GNV	[DET	N]GNV	PRT	V	PRN	V	PRN	[DET	N]GNV	V								
	<i>it's freedom that the guy was looking for. It's good what the guy saw/experienced,</i>																				
003. -04.	blakaman	ná	a	freedom	ete.	Luku	fa	u	de	dape ⁴	u	ná	a	freedom	ete	eee					
	blakaman	ná	a	freedom	ete	Luku	fa	u	de	dape	u	ná	a	freedom	ete	eee					
	[N]GNV	PRT	V	[N]GNV	ADV	V	PRN	PRN	V	ADV	PRN	PRT	V	[N]GNV	ADV						
	<i>black people don't have freedom yet. Look at us there, we are not free/we don't have freedom yet, yes, yes, yes, yes</i>																				

Fig.: 3 Visualization of the “erasure of language boundaries” between EMC and Sranan

I propose to refer to moments in the interaction when it is not possible to distinguish between the languages in play – figured in blue in Fig. 3 – as an “erasure of language boundaries.” By indiscriminately using language resources which may belong to one or the other language, although specific marks or terms could be used, the speakers produce indistinction and then perform blurred boundaries between languages. Of course, it might be argued that EMC and Sranan share a lot of lexical and syntactic features because they are both the product of slave trade history, and that the use of bivalent elements here could be accidental. But the reason why sharing common words is not enough to explain so much overlap between the linguistic resources used during this recording is firstly that alternative wordings (or marked elements) sounding more “traditional/Maroon” or more “urban/Sranan” are possible, and secondly that speakers do have a term to refer to this kind of heterogeneous practice, namely *Takitaki*.

This term, widely circulating at the time of the recordings, is particularly useful because it avoids reference to ethnic identifications (Pamaka, Ndyuka, Aluku) or to language boundaries (EMC vs. Sranan). *Takitaki* meaning originally “to chat” has long been depreciative and served as a hetero-designation, but in the 2000s and 2010s it has been taken up by Maroons in a re-appropriation move and appears to be very useful for hiding ethnolinguistic differences (Léglise and Migge, 2006). By referring to their language as *Takitaki*, speakers avoid referring to EMC or Sranan. Hence, the process of erasing language boundaries that produces indistinction and fuzziness in speech is performed by speakers through the use of bivalent or unmarked elements, shared by the two languages or varieties, and at the same time by the name *Takitaki*, a term that is underdefined, its own meaning being vague.

The use of bivalent elements by plurilingual social actors might be seen as a performance, in the sense that language practices construct social meaning and produce social effects (Boutet, Fiala, and Simonin-Grumbach, 1976). One of these effects could be linked here to processes of linguistic and cultural homogenization accomplished through the creation of a pan-Maroon identity, where Maroons avoid making reference to their ethnolinguistic differences (Léglise and Migge, 2006). Another social effect is linked to processes of differentiation in the urban context as opposed to the traditional monolingual context (Migge, 2007). These macro-processes of homogenization and

differentiation¹⁴ are included under the umbrella of the term *Takitaki*. Concretely, these processes are also able, by the use of linguistic resources at a micro level, to blur language boundaries or display them in a way that might lead to the emergence of new practices.

In some cases, the polylinguality is more obvious and there is a wider use of linguistic features attributable to more languages. In the following extract, which could be described as code-mixing in Auer (1999) typology or polylinguaging in Jørgensen et al. (2011) view, the interlocutors are using five different linguistic resources (being called languages or varieties): EMC (in its Ndyuka variety, in italics), Sranan Tongo (bold underline), undetermined elements (which can be both EMC and Sranan), undetermined French or French Guianese Creole (double underline), and Dutch (bold). By using all these different linguistic resources, through their heterogeneous language practices the interlocutors are performing being urban and modern, affiliating with all the languages that are present in the public sphere in urban centers such as Saint-Laurent-du-Maroni in French Guiana or Paramaribo in Suriname. Both interlocutors are languaging here, unmarking boundaries between EMC and Sranan and even between French and French Guianese Creole. We can categorize this extract as a good example of convergence and similarity in the way social actors use all kinds of linguistic resources from their repertoire within their heterogeneous language practices.

- 12) 1. B Da *i de anga congé* nounou?
'So you're on vacation right now?'
2. E Aii, mi de **nanga congé** nou te **lek' tra** mun, bigin fu **tra** mun. Le sept, da mi bigin baka.
'Yes, I am on vacation now until like next month, beginning of next month. The seventh, I'll start working again.'
3. B *Soutu wooko* i e du?
'What kind of work do you do?'
4. E **Sortu wroko** mi e du? Mi e du wan sers.
'What kind of work I am doing? I am working as a security guard.'
5. B Mh?
'Mh?'
6. E Wan sers mi e du, wan **tra wroko leki** soudati, fu la mairie / lameri gi la mairie / lameri. Ma mi **hoop** taki **nanga kontrakt**, den man *ná* e gi wan langa **kontrakt**, **siksi** mun, ef i e **wroko** bun, den man gi i **siksi** mun baka, te **nanga** tu *yali*, a *kaba*, den man **stop** en.
'I am working as a *security guard*, another type of work like a soldier, for the town hall. But I **hoop** that with a **contract**, those people *do not* make long **contracts**, **six** months, if you **work** well, the people give you **six** months again, **until** about two years, it's **over**, they **stop** it.'

However, this last extract is also interesting as a case of non-alignment to the language introduced by the interlocutor. I use here the concepts of convergence and alignment in the sequentiality of speech turns (Auer, 1995). What happens on line 4 is that E does not align with B's language choice: he reformulates in Sranan (**sortu wroko**) what she says in EMC (*soutu wooko* "kind of work"). As pointed out by Auer (1995, 131–132), this kind of alternation may have indexicality and

¹⁴ Social and linguistic homogenization and differentiation are linked to language ideology, ideology being a good sorting device to understand the linguistic and social processes taking place (Migge and Léglise: 2013, chap 9).

perform both participant- and discourse-related functions. The exact sequentiality is provided in the following transcript (13), which annotates the alternations:

- | | | | | |
|-----|------|--------------------|---|-----------------------------|
| 13) | 3. B | <i>Soutu wooko</i> | i e du? | |
| | | ‘What kind of work | do you do?’ | |
| | | EMC | Undetermined | <i>EMC+undetermined</i> |
| | 4. E | <u>Sortu wroko</u> | mi e du? | |
| | | ‘What kind of work | I am doing?’ | |
| | | Sranan tongo |  | <u>Sranan</u> +undetermined |

E moves away from the language choice operated on the previous line by B; he dis-aligns. While she used an EMC form and an undetermined form (between Sranan and EMC), he reformulates with a Sranan form and an undetermined form. However, we know that reformulations initiated by others (or hetero-repetitions) (Ursi et al., 2018) can be particularly significant in the interaction, especially when used as a repair (Gafaranga, 2012; Kitinger, 2013). This dis-alignment can be interpreted as a remedy, in a movement of linguistic differentiation that clearly points out that the wording or way of speaking used by the other is not that of the speaker. This differentiation can be meaningful, in the sense that the use of both EMC and Sranan is indexical of specific affiliations, referring the other (B) to an identity (female, with a village way of speaking) that is not shared by E, who projects a masculine and modern identity with features from Sranan Tongo urban language. The features used by B are associated with a traditional Maroon ethnic language (EMC) linked to rural zones and traditional ways of speaking, which is appropriate for women’s talk but not for the assertion of a modern form of masculinity. We can imagine it is linked to E and B’s gender but also to the subject of the conversation: E is talking about the kind of masculine work he is doing (as a kind of soldier for the town hall).

This extract is a good example of heterogeneous language practices in which social actors can converge in the way they mix linguistic features and unmark language boundaries, performing a common modern identity but at the same time can, at each point, show disaffiliation from the wording choice of the interlocutor and perform some other kind of identity (masculinity, for example) and clearly mark language or dialectal boundaries, in this case excluding the other as “a woman” and “traditional.”

6. Conclusion

Trying to document multilingual practices through the use of a precise annotation tool reveals the intrinsic heterogeneity of these practices. This non-aprioristic method of annotation enables linguists to explore heterogeneous corpora made up of various linguistic resources attributable to various languages, dialects, registers, and styles – sometimes sequentially and sometimes simultaneously. It enables sociolinguists to show how language boundaries are constructed and deconstructed in face-to-face interaction. Just as the speakers’ discourses about their language practices are not fixed but draw fluctuating boundaries that are constantly being built and deconstructed (Canut, 2001; Greco and Auzanneau, 2018), in the same way multilingual language practices at a micro-level draw, construct, or deconstruct language or language variety boundaries. Speakers either erase them or reintroduce them, sometimes deliberately.

If for linguists, in retrospect, knowing in which language one expresses oneself is a fundamental question, for the social actors who express themselves, at the time of production, this question probably does not make sense in many situations: they just express themselves and communicate through the activity of language among other means of communication as gesture, posture and other non-verbal modes. Social actors, as “language beings” or “language users,” may draw on similarities across languages and varieties. In multilingual exolingual settings that we did not mention in this paper, bivalent elements are commonly used to build comprehension (see Légise (2017) for multilingual interactions at the hospital between patients and doctors who do not share the same linguistic repertoires).

In multilingual endolingual settings, we can document that bivalent elements may be more or less used: in our corpora they make up between 4 and 80 percent of the linguistic features depending on the characteristics of the language contact situation. Particularly in endolingual settings involving closed languages, unmarked or bivalent elements are linguistic resources that may be used to blur (language and social) boundaries and perform various identities emerging against a background where marked elements from traditional languages and affiliation are linked. They illustrate Bakhtin's polyphony and the potential multiple voices or speaking positions (Holquist, 1990; Woolard 1998). We could argue that bivalent elements, as any other linguistic resource enable social actors to "sociolinguistically claim opposed but nested identities" (Kroskrity: 2000: 340) but, as specific resources, neither from a language or from the other, they also enable to erase previous affiliations and iconicity and allow to create new meanings.

At the same time, in the same multilingual interaction, social actors may on the one hand draw on linguistic similarities and project common identities but on the other hand draw also on a single linguistic feature as a difference, not to converge and align but disalign and construct language boundaries. This can happen any time, at every single moment of the interaction. As we saw, both bivalent and multilingual resources may be used during social interaction as acts of identity (Bucholtz and Hall, 2010), specifically in processes of differentiation.

Documenting multilingual heterogeneous practices and the role of bivalent elements allows to follow micro-interactional moves of convergence and divergence. If they are of interest for interactional sociolinguistics, they might also be of interest for documentary and contact linguistics. It is indeed vital to document heterogeneous practices and bivalent elements in interaction in order to understand ongoing social and linguistic processes of homogenization and divergence at play in koineization and the emergence of new language practices that might become identified and claimed varieties someday.

References

- Adamou, Evangelia. 2010. Bilingual speech and language ecology in Greek Thrace: Romani and Pomak in contact with Turkish. *Language in Society* 39 (2): 147–71.
- Alby, Sophie, and Bettina Migge. 2007. Alternances codiques en Guyane française: les cas du kali'na et du nenge. In Isabelle Léglise and Bettina Migge (eds.), *Pratiques et représentations linguistiques en Guyane: regards croisés*, , 49–72. Paris: IRD Editions.
- Alvarez-Cáccamo, Celso. 1990. Rethinking conversational code-switching: codes, speech varieties, and contextualization. *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society*, February 16-19, 1990, 3–16. Berkeley: Berkeley Linguistics Society.
- Auer, Peter. 1995. The pragmatics of code-switching: a sequential approach. In Lesley Milroy and Pieter Muysken (eds.) *One Speaker, Two Languages: Cross-Disciplinary Perspectives on Code-Switching*, 115–35. Cambridge: Cambridge University Press.
- , ed. 1998. *Code-switching in Conversation. Language, Interaction and Identity*. London and New York: Routledge.
- . 1999. From codeswitching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. *The International Journal of Bilingualism* 4 (3): 309–32.
- . 2005. A postscript: code-switching and social identity. *Journal of Pragmatics* 37: 403–10.
- . Forthcoming. 'Translanguaging' or 'doing languages'? Multilingual practices and the notion of 'codes'. In Jeff MacSwan (ed.) *Language(s): Multilingualism and Its Consequences*. Bristol: Multilingual Matters.
- Bagga-Gupta, Sangeeta, and Giulia Messina Dahlberg. 2018. Meaning-making or heterogeneity in the areas of language and identity? The case of translanguaging and nyanlända (newly-arrived) across time and space. *International Journal of Multilingualism* 15 (4): 383–411. <https://doi.org/10.1080/14790718.2018.1468446>.

- Baggioni, Daniel. 1992. Langues mixtes, discours métisses et conflits d'identités. In Jean-Luc Alber, Claudine Bavoux, and Michel Watin (eds.) *Métissages, Tome 2, Linguistique et anthropologie*, 185–96. Paris: L'Harmattan and Université de la Réunion.
- Bakhtin Mikhail Mikhaïlovich / Voloshinov, V. N. 1973. *Marxism and the Philosophy of Language*. New York: Seminar Press.
- Blanche-Benveniste, Claire. 1993. Une description linguistique du français parlé. *Gré des langues* 5: 8–29.
- Blommaert, Jan, and Ad Backus. 2011. Repertoires revisited: 'Knowing language' in superdiversity. *Working Papers in Urban Language and Literacies* 67: 1–26.
- Blommaert, Jan, James Collins, and Steff Slembrouck. 2005. Spaces of multilingualism. *Language and Communication* 25: 197–216.
- Botley, Simon, Anthony Mark McEnery, and Andrew Wilson. 2000. *Multilingual Corpora in Teaching and Research*. Amsterdam: Rodopi
- Boutet, Josiane. 1995. Une linguistique de l'activité. In Janine Bouscaren, Jean-Jacques Franckel, and Stéphane Robert (eds.) *Langues et langage. Problèmes et raisonnements en linguistique. Mélanges offerts à A. Culioli*, 393–403. Paris: PUF.
- Boutet, Josiane, Pierre Fiala, and Jenny Simonin-Grumbach. 1976. Sociolinguistique ou sociologie du langage? *Critique* 344: 68–85.
- Bucholtz, Mary. 2000. The politics of transcription. *Journal of Pragmatics* 32 (10): 1439–65. [https://doi.org/10.1016/S0378-2166\(99\)00094-6](https://doi.org/10.1016/S0378-2166(99)00094-6).
- Bucholtz, Mary, and Kira Hall. 2005. Identity and Interaction : A sociocultural linguistic approach. *Discourse Studies* 7 (4-5): 585--614.
- Canagarajah, Suresh. 2012. *Translingual Practice: Global Englishes and Cosmopolitan Relations*. London: Routledge.
- . 2017. *Translingual Practices and Neoliberal Policies: Attitudes and Strategies of African Skilled Migrants in Anglophone Workplaces*. SpringerBriefs in Linguistics. New York: Springer International Publishing.
- Canut, Cécile. 2001. A la frontière des langues. Figures de la démarcation. *Cahiers d'Etudes Africaines* 163–164: 443–63.
- Carlo, Pierpaolo Di. 2018. Towards an understanding of African endogenous multilingualism: Ethnography, language ideologies, and the supernatural. *International Journal of the Sociology of Language* 2018 (254): 139–63. <https://doi.org/10.1515/ijsl-2018-0037>.
- Creese, Angela, and Adrian Blackledge. 2010. Translanguaging in the bilingual classroom: A pedagogy for learning and teaching? *The Modern Language Journal* 94 (1): 103–115.
- . 2011. Separate and flexible bilingualism in complementary schools: Multiple language practices in interrelationship. *Journal of Pragmatics, Multilingual Structures and Agencies*, 43 (5): 1196–1208. <https://doi.org/10.1016/j.pragma.2010.10.006>.
- De Pietro, Jean-François. 1988. Vers une typologie des situations de contacts linguistiques. *Langage et Société* 43: 65–89.
- Epps, Patience. 2018. Contrasting linguistic ecologies: Indigenous and colonially mediated language contact in Northwest Amazonia. *Language and Communication, Indigenous Multilingualisms*, 62 (September): 156–69. <https://doi.org/10.1016/j.langcom.2018.04.010>.
- Fishman, Joshua A. 1965. Who speaks what language to whom and when? *La Linguistique* 2: 67–88.
- Friedman, Victor A., and Brian D. Joseph. 2017. Reassessing Sprachbunds: A View from the Balkans. In Raymond Hickey (ed) *The Cambridge Handbook of Areal Linguistics*, 55–87. Cambridge

- Handbooks in Language and Linguistics. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781107279872.005>.
- Gafaranga, Joseph. 2012. Language alternation and conversational repair in bilingual conversation. *International Journal of Bilingualism* 16 (4): 501–27. <https://doi.org/10.1177/1367006911429520>.
- Gal Susan and Irvine Judith. 1995. The Boundaries of Languages and Disciplines: How Ideologies Construct Difference. *Social Research* 82 (4), pp. 967–1001.
- García, Ofelia. 2009. Education, multilingualism and translanguaging. In Ajit Mohanty, Minati Panda, Robert Phillipson, and Tove Skutnabb-Kangas (eds.) *Multilingual Education for Social Justice: Globalising the Local*, , 140–58. New Delhi: Orient Blackwan.
- García, Ofelia, and L. Wei. 2013. *Translanguaging: Language, Bilingualism and Education*. 2014 edition. London: Palgrave Pivot.
- Gorter, Durk, Heiko F. Marten, and Luk Van Mensel. 2011. *Minority Languages in the Linguistic Landscape*. Basingstoke: Springer.
- Greco, Luca, and Michelle Auzanneau. 2018. *Dessiner les frontières*. Lyon: ENS Editions.
- Grenoble, Lenore A. 2007. The importance and challenges of documenting pragmatics. In Peter K. Austin (ed) *Language Documentation and Description*, , 4:145–62. London: SOAS.
- Grenoble, Lenore A., and Louanna Furbee-Losee, eds. 2010. *Language Documentation: Practice and Values*. Amsterdam: John Benjamins Publishing Company.
- Gumperz, John J. 1958. Dialect differences and social stratification in a North Indian village. *American Anthropologist* 60 (4): 668–82.
- . 1982. *Language and social identity*. Cambridge: Cambridge University Press.
- Hall, Kira, and Chad Nilep. 2015. Code-switching, identity, and globalization. In Deborah Tannen, Heidi E. Hamilton, and Deborah Schiffrin (eds.) *The Handbook of Discourse Analysis*, 597–619. Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Haugen, Einar Ingvald. 1972. *The Ecology of Language*. Stanford, CA: Stanford University Press.
- Heine, Bernd, and Tania Kuteva. 2005. *Language Contact and Grammatical Change*. Cambridge: Cambridge University Press.
- Heugh, Kathleen, and Christopher Stroud. 2019. Diversities, affinities and diasporas: a southern lens and methodology for understanding multilingualisms. *Current Issues in Language Planning* 20 (1): 1–15. <https://doi.org/10.1080/14664208.2018.1507543>.
- Holquist, Michael. 1990. *Dialogism: Bakhtin and His World*. London: Routledge.
- Hymes Dell 1967, Why Linguistics Needs the Sociologist, *Social Research* 34 (4), pp. 632–647.
- Irvine, Judith T., and Susan Gal. 2000. Language ideology and linguistic differentiation. In Paul V. Kroskrity (ed) *Regimes of Language. Ideology, Politics and Identity*, , 35–83. Santa Fe/Oxford: School of American Research Press/ James Currey.
- Jacquemet, Marco. 2005. Transidiomatic practices: Language and power in the age of globalization. *Language and Communication* 25 (3): 257–77. <https://doi.org/10.1016/j.langcom.2005.05.001>.
- Jake, Janice L., Carol Myers-Scotton, and Steven Gross. 2005. A response to MacSwan (2005): Keeping the matrix language. *Bilingualism: Language and Cognition* 8 (03): 271–76. <https://doi.org/10.1017/S1366728905002300>.
- Jørgensen, Jens Normann, and Kaspers Juffermans. 2011. *Languaging*. <http://toolkit.socsci.uva.nl/docs/languaging.html>.

- Jørgensen, Jens Normann, Martha Sif Karrebaek, Lian Malai Madsen, and Janus Spindler Møller. 2011. Polylinguaging in superdiversity. *Diversities* 13 (2). www.unesco.org/shs/diversities/vol13/issue2/art2.
- Kitzinger, Celia. 2013. Repair. In Jack Sidnell and T. Stivers (eds.) *The Handbook of Conversation Analysis*, 229–56. Malden: Wiley-Blackwell.
- Kroskrity, Paul, V. 2000. Language Ideologies in the Expression and Representation of Arizona Tewa Identity. In Paul V. Kroskrity (ed) *Regimes of Language: Ideologies, Politics, and Identities*. Sante Fe/Oxford: James Currey.
- Ledegen, Gudrun. 2012. Prédicats ‘flottants’ entre le créole acrolectal et le français à la Réunion: exploration d’une zone ambiguë. In Claudine Chamoreau and Laurence Goury (eds.) *Changement linguistique et langues en contact*, 251–70. Paris: CNRS Editions.
- Léglise, Isabelle. 2013. Multilinguisme, variation, contact. Des pratiques langagières sur le terrain à l’analyse de corpus hétérogènes. Habilitation, INALCO.
- . 2017. Multilinguisme et hétérogénéité des pratiques langagières. Nouveaux chantiers et enjeux du Global South. Multilingualism and heterogeneity of language practices. New focus areas and issues of the global South. *Langage et société*, 160–161: 251–66.
- . 2018. Pratiques langagières plurilingues et frontières de langues. In Luca Greco and Michelle Auzanneau (eds.) *Dessiner les frontières*, 143–69. Lyon: ENS Editions.
- Léglise, Isabelle, and Sophie Alby. 2006. Minorization and the process of (de)minoritization: The case of Kali’na in French Guiana. *International Journal of the Sociology of Language* 182: 67–86.
- . 2016. Plurilingual corpora and polylinguaging: Where corpus linguistics meets contact linguistics. *Sociolinguistic Studies* 10 (3): 357–81.
- Léglise, Isabelle, and Claudine Chamoreau, eds. 2013. *The Interplay of Variation and Change in Contact Settings*. Amsterdam: John Benjamins.
- Léglise, Isabelle, and Bettina Migge. 2005. Pour une étude des contacts de langues en synchronie: quelques exemples tirés du cas guyanais. *Trace (CEMCA, Mexico)* 47: 113–31.
- . 2006. Language naming practices, ideologies and linguistic practices: Toward a comprehensive description of language varieties. *Language in Society* 35: 313–39.
- . 2019. Language and identity construction on the French Guiana-Suriname border. *International Journal of Multilingualism*, July, 1–15. <https://doi.org/10.1080/14790718.2019.1633332>.
- Lüdi, Georges. 1994. Dénomination médiante et bricolage lexical en situation exolingue. *Acquisition et interaction en langue étrangère*, 3: 115–46.
- Lüpke, Friederike. 2010. Multilingualism and language contact in West Africa: Towards a holistic perspective. *Journal of Language Contact* 3 (1): 1–12. <https://doi.org/10.1163/19552629-90000002>.
- . 2016. Uncovering small-scale multilingualism. *Critical Multilingualism Studies* 4 (2): 35–74.
- Lüpke, Friederike, and Rachel Watson. 2020. Language contact in West Africa. In Evangelia Adamou and Yaron Matras (eds.) *Language Contact in West Africa*, 528–49. Abingdon / New York: Routledge. <https://doi.org/10.4324/9781351109154-31>.
- Makoni, Sinfree  and Alastair Pennycook. 2007. *Disinventing and Reconstituting Languages*. Bristol: Multilingual Matters.

- Manfredi, Stefano, and Suat Istanbulu. 2019. A multilayered and data-driven method for exploring Arabic in multilingual settings. In Werner Arnold and Maciej Klimiuk (eds.) *Arabic Dialectology: Methodology and Field Research*, . Wiesbaden: Harrassowitz.
- Manfredi, Stefano, Marie-Claude Simeone-Senelle, and Mauro Tosco. 2015. Language contact, borrowing and code-switching. In Amina Mettouchi, Martine Vanhove, and Dominique Caubet (eds.) *Corpus-Based Studies of Lesser-Described Languages: The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*, , 283–308. Amsterdam: Benjamins.
- Matras, Yaron, and Peter Bakker. 2003. *The Mixed Language Debate: Theoretical and Empirical Advances*. Berlin: Walter de Gruyter
- Mettouchi, Amina, Martine Vanhove and Dominique Caubet. eds. 2015. *Corpus-Based Studies of Lesser-Described Languages: The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Migge, Bettina. 2007. Codeswitching and social identities in the Eastern Maroon community of Suriname and French Guiana. *Journal of Sociolinguistics* 11 (1): 53–72.
- . 2020. Researching endangered languages: Critical reflections on field and documentary linguistics. In *Colonial and Decolonial Linguistics – Knowledges and Epistemes*, edited by Anne Storch, Ana Deumert, and Nick Shepherd. Oxford: Oxford University Press, 157–75.
- Migge, Bettina, and Isabelle Léglise. 2011. On the emergence of new language varieties: The case of the Eastern Maroon Creole in French Guiana. In *Variation in the Caribbean: From Creole Continua to Individual Agency*, edited by Lars Hinrichs and Joseph Farquharson, 181–99. Amsterdam: John Benjamins.
- . 2013. *Exploring Language in a Multilingual Context: Variation, Interaction and Ideology in Language Documentation*. Cambridge: Cambridge University Press.
- Mondada, Lorenza. 2002. Pour une approche interactionnelle de la catégorisation des ressources linguistiques par les locuteurs. In Véronique Castellotti and Didier de Robillard (eds.) *France pays de contacts de langues*, . 23–35. Louvain-la-neuve: Cahiers de l'Institut Linguistique de Louvain
- . 2012. L'organisation émergente des ressources multimodales dans l'interaction en lingua franca: entre progressivité et intersubjectivité. *Bulletin Suisse de Linguistique Appliquée* 95: 97–121.
- . 2018. Transcription in linguistics. In Lia Litosseliti (ed.) *Research Methods in Linguistics*, , 2nd ed., 85–114. London/New York: Bloomsbury Academic.
- Muysken, Pieter. 1990. Concepts, Methodology and Data in Language Contact Research; In *Papers for the Workshop on Concept, Methodology and Data. Network on Code-Switching and Language Contact*. 15–30. Strasbourg: European Science Foundation.
- Myers-Scotton, Carol. 1993. *Duelling languages: Grammatical structures in codeswitching*. Oxford: Clarendon Press.
- Nilep, Chad. 2016. 'Code Switching' in Sociocultural Linguistics. *Colorado Research in Linguistics* 16. DOI: 10.31235/osf.io/jfecn
- Nunez, Joseph Jean François. 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 dissertation. Sorbonne Paris Cité. <http://www.theses.fr/2015USPCF016>.
- Nunez, Joseph Jean François, and Isabelle Léglise. 2017. Ce que les pratiques langagières plurilingues au Sénégal disent à la linguistique de contact. In Michelle Auzanneau, Margaret Bento, and Malory Leclere (eds.) *Espaces, mobilités et éducation plurilingues: éclairages d'Afrique ou d'ailleurs*, , 99–119. Paris: Editions des archives contemporaines

- Ochs, Elinor. 1979. Transcription as theory. In Elinor Ochs and Bambi B. Schieffelin (eds.) *Developmental Pragmatics*, , 43–72. New York: Academic Press.
- O'Rourke, Bernadette. 2005. Attitudes towards minority languages: An investigation of young people's attitudes towards Irish and Galician. PhD Thesis. Dublin City University
- O'Shannessy, Carmel. 2006. Language contact and children's bilingual acquisition: learning a mixed language and Warlpiri in northern Australia. PhD Dissertation. Sydney: University of Sydney.
- Otheguy, Ricardo, Ofelia García, and Wallis Reid. 2015. Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Applied Linguistics Review* 6 (3): 281–307. <https://doi.org/10.1515/applirev-2015-0014>.
- Pietikainen, Sari, and Helen Kelly-Holmes. 2013. *Multilingualism and the Periphery*. Oxford: Oxford University Press.
- Pike, Kenneth L. 1964. Towards a theory of the structure of human behavior. In Dell Hymes (ed.) *Language in culture and society*, , 54–62. New York: Harper & Row.
- Poplack, Shana. 1981. The syntactic structure and social function of code-switching. In Richard P. Duran (ed.) *Latino Language and Communicative Behavior*, , 169–84. Norwood, NJ: Ablex.
- Prudent, Lambert-Félix. 1981. Diglossie et interlecte. *Langages* 61: 13–38.
- Rampton, Ben. 1998. Language crossing and the redefinition of reality. In Peter Auer (ed.) *Code-Switching in Conversation: Language, Interaction and Identity*, , 290–317. London/New York: Routledge.
- . 2005. *Crossing: Language and Ethnicity among Adolescents*. 2nd ed. Manchester, UK /Northampton, MA: St. Jerome Publishing.
- Sánchez Moreano, Santiago. 2015. Conséquences linguistiques et identitaires du contact linguistique et dialectal à Cali (Colombie): le cas des variations de l'ordre des constituants. PhD dissertation, Paris: Université Paris Sorbonne Cité, Paris Diderot. <https://hal.archives-ouvertes.fr/tel-01314442>.
- Stell, Gerald. 2015. Towards an integrated approach to structural and conversational aspects of code-switching through macrosociolinguistic predictors. In Kofi Yakpo and Gerald Stell (eds.) *Code-Switching Between Structural and Sociolinguistic Perspectives*, , 117–38. Berlin and New York: Walter de Gruyter.
- Stroud, Christopher. 1998. Perspectives on cultural variability of discourse and some implications for code-switching. In Peter Auer (ed.) *Code-Switching in Conversation: Language, Interaction and Identity*, , 321–48. London: Routledge.
- Thomason, Sarah G. 2001. Contact-induced typological change. In Martin Haspelmath, Ekkerard Koenig, Wulf Oesterreicher, and Wolfgang Raible (eds.) *Language Typology and Language Universals, Sprachtypologie und sprachliche Universalien*,. Vol. 2. Berlin and New York: Walter de Gruyter.
- Ursi, Biagio, Florence Oloff, Lorenza Mondada, and Véronique Traverso. 2018. Diversité des répétitions et des reformulations dans les interactions orales: défis analytiques et conception d'un outil de détection automatique. *Langages* 212 (4): 87–104.
- Vaillant, Pascal, and Isabelle Léglise. 2014. A la croisée des langues: Annotation et fouille de corpus plurilingues. *Revue des Nouvelles Technologies de l'Information RNTI-SHS-2*: 81–100.
- Vaughan, Jill. 2018. 'We talk in saltwater words': Dimensionalisation of dialectal variation in multilingual Arnhem Land. *Language and Communication, Indigenous Multilingualisms*, 62 (September): 119–32. <https://doi.org/10.1016/j.langcom.2017.10.002>.
- Winford, Donald. 2003. *An Introduction to Contact Linguistics*. Oxford: Blackwell.

Woolard, Kathryn A. 1998. Simultaneity and bivalency as strategies in bilingualism. *Journal of Linguistic Anthropology* 8 (1): 3–29. <https://doi.org/10.1525/jlin.1998.8.1.3>.