

SocioBaGS

Macro- and micro-variation in Bantu grammatical gender systems and their sociolinguistic correlates

Funded by the *Agence Nationale de la Recherche* (ANR) and the *Deutsche Forschungsgemeinschaft* (DFG), as part of the FRAL funding scheme

Principal investigators and project coordinators:

Francesca DI GARBO & Annemarie VERKERK

Core members of the project and international consortium:

Antti LAINE, Anne-Maria FEHN, Michael KARANI, and Sara PACCHIAROTTI

Funding period: 2024-2027

Summary

Variation and change are inherent to language. While this is an uncontroversial statement in linguistics, the mechanisms that steer language variation and change are debated. Our project addresses this debate by investigating variation and change in the nominal classification systems of the Bantu languages, a.k.a. grammatical gender systems. Bantu is one of the world's largest language families. At each of the borders of the Bantu spread zone – northwestern, northeastern, and southern – Bantu languages are in contact with other language families that we investigate. These encompass distantly related branches of the Niger-Congo family, such as Ubangi, and genealogically unrelated groupings such as Cushitic, Nilotic, Central Sudanic, Khoe-Kwadi, Kx'a, and Tuu. Within the Bantu spread zone, inter-Bantu contact can be intense and sustained. We investigate how Bantu gender systems vary and change in languages spoken at the border with non-Bantu and/or other Bantu languages. By comparing changes that occur in Bantu languages in contact with non-Bantu and Bantu languages, we study how processes of language convergence differ depending on whether languages in contact share ancestry. In doing so, we examine the sociohistorical correlates of language change across different language contact ecologies and shed new light on the language and population history of sub-Saharan Africa. Language contact is well-researched, but the outcomes of contact are still hard to predict, and there is a lack of research on comparing contact effects in unrelated and related languages. Our research consortium will tackle these issues with new methods and bring fresh answers to it. We investigate contact-induced change in Bantu gender systems through large-scale comparative analyses and field-based studies conducted in several regions of the Bantu-speaking world. The two teams consist of researchers with leading expertise in Bantu historical-comparative and contact linguistics, the language and population history of the Bantu people and Africa, language typology, sociolinguistics, and quantitative methods in the language sciences.

1. Context, positioning and objectives of the project proposal

1.1. Research objectives and hypotheses

Variation and change are inherent to language. While this is an uncontroversial statement in linguistics, the mechanisms that steer language variation and change are debated. Our project addresses this fundamental debate in the language sciences by focusing on **how languages change through contact with neighbouring communities**. More specifically, we investigate patterns of language variation and change in the nominal classification systems of the Bantu languages.

Bantu is one of the world's largest language families. At each of the borders of the Bantu spread zone – northwestern, northeastern, and southern – Bantu languages are in contact with other language families. These encompass distantly related branches of the Niger-Congo family, such as Ubangi, as well as genealogically unrelated groupings such as Cushitic, Nilotic, Central Sudanic, Khoe-Kwadi, Kx'a, and Tuu. Within the Bantu spread zone, inter-Bantu contact can be intense and sustained.

The Bantu languages have rich nominal classification systems that are coded through complex patterns of grammatical agreement and are known in the literature as **grammatical gender systems** (Corbett 1991, Van de Velde 2019). Gender systems are highly stable (Nichols 2003), yet particularly prone to erode as a result of language contact (Trudgill 1999, Di Garbo 2020). This tendency has also been confirmed for selected areas of the Bantu-speaking world (Di Garbo and Verkerk 2022; Verkerk and Di Garbo 2022 on northwestern Bantu languages).

In this project, we investigate how Bantu gender systems vary and change across the whole of the Bantu-speaking world and aim to explain this variation in relation to varying language contact dynamics, both within-family and with non-Bantu languages. Our general objectives are:

- (1) to study how **Bantu gender systems vary**,
- (2) to focus on typological variation in **language contact ecologies**, by systematically comparing contact scenarios featuring **Bantu–non-Bantu interactions** with contact scenarios involving only **Bantu languages**,
- (3) to use this wealth of synchronic data on ongoing variation to draw **diachronic inferences** about trajectories of change across contact scenarios, using both quantitative and qualitative methods.
- (4) to consider how this speaks to the nature of **contact-induced change more generally**.

To be able to model the effect of language ecology on Bantu gender systems, we study three different contact scenarios, namely:

- A. **Bantu languages in contact with non-Bantu neighbouring languages that typically do not have Bantu-like gender systems (henceforth, Bantu <-> non-Bantu)**. We hypothesise that this scenario may lead to either the restructuring or the erosion of grammatical gender (as already partially confirmed for northwestern Bantu by Di Garbo and Verkerk 2022; Verkerk and Di Garbo 2022).
- B. **Bantu languages spoken by relatively close-knit communities in contact with one or several Bantu lingua francas that typically have highly eroded gender systems** (see Mufwene 2003) **(henceforth, Bantu <-> BantuLinguaFranca)**. For this scenario, we predict that languages may either converge with the eroded gender systems attested in the contact lingua franca or rather retain more conservative features.
- C. **Bantu languages spoken by relatively close-knit communities that influence one another through a simple diffusion model (henceforth, Bantu <-> Bantu)**. Under this scenario, we hypothesise that similarities in the structural make-up of gender systems are maintained and reinforced through sustained interactions between neighbouring language communities.

These contact scenarios are necessarily simplified for the ease of presentation here and are not mutually exclusive. We assume that the gender systems of many Bantu languages may be influenced by more than one of these scenarios at once, as well as by yet other ones that are not captured here. In addition, while the three scenarios mentioned above provide a rather schematic representation of contact settings, which is essentially based on the genealogical affiliation (Bantu vs. non-Bantu) or the status (language of wider communication vs. ‘smaller’ languages) of the contact languages in question, in this project, we will work with fine-grained models of contact scenarios that attempt at targeting such things as the intensity of contact and the attitudes that speakers have towards given contact settings.

Our research design is based on a **multipronged approach** (a detailed description is provided in section 1.3), which combines **large-scale comparative analyses**, focusing on the whole of the Bantu family and using **state of the art methods in quantitative sociolinguistic typology**, with **smaller-scale studies** based on **fieldwork** in specific regions of the Bantu-speaking world. A research design of this kind requires extensive expertise in a diverse range of methods of data collection and analysis. Because of its grounding in **data and method triangulation**, this approach has the advantage of likely leading to more robust results than a single-method approach.

The research consortium consists of researchers with leading expertise in Bantu historical-comparative and contact linguistics, language typology, sociolinguistics, and quantitative methods, as well as the language and population history of the Bantu people and Africa. **The expertise of the individual members of the consortium consolidates our objectives**, as well as the methods and tools we aim at using in order to reach them. This in itself mitigates any potential risks related to using a range of different data and methods of analysis.

By systematically **comparing contact between unrelated languages with contact between related languages**, and by bringing **speech communities and their language ecologies to the fore**, we anticipate that our approach will take current research on the impact of population contact on processes of language change to a new level. We also estimate that the results that will be brought about by this project will have **a long-lasting impact in the field of Bantu linguistics**. Finally, while our focus is on one specific grammatical domain, and one family of related languages, we are confident that the methodologies implemented and tested in this project can be applied to **other grammatical domains as well as to language families and areas of the world beyond Bantu and sub-Saharan Africa**.

1.2. Positioning relative to the state of the art

1.2.1. Sociolinguistic typology

This project is a work in **sociolinguistic typology**, which attempts to explain typological distributions through the analysis of language contact scenarios and the sociohistorical ecology of language communities at large. Sociolinguistic typology has become well-known through the work of Trudgill (2011), who proposes direct links between the social make-up of language communities and the distribution of types of language structure. Among the key points of Trudgill’s manifesto is the idea that close-knit, isolated communities favour complex language structures while communities with looser social network structures and intense interpopulation contact favour their loss.

Much of the existing work in sociolinguistic typology is of a distinct **quantitative nature** (Lupyan and Dale 2010, Sinnemäki & Di Garbo 2018), often drawing upon world-wide language samples with little eye for either the diachronic processes that may account for the attested linguistic patterns or the relevant socio-historical scenarios. A careful examination of the **sociolinguistic history** of the sampled language communities, united with in-depth analyses of **attested pathways of change** and their distribution, are of the utmost importance when testing hypotheses about contact-induced change. Yet, it is only in recent years that such issues of variable design have become topical in the field. The scientific coordinators of our research group have played a key role in bringing these issues to the fore in several of their recent publications (see Verkerk and Di Garbo 2022; Di Garbo et al. 2021; Di Garbo and Napoleão de Souza 2023). The study by Verkerk and Di Garbo (2022), in particular, is

pioneer in applying phylogenetic comparative methods¹ to the study of contact effects in a typological perspective.

In this project, we build upon these recent contributions to develop a research program that is strongly geared towards methodology. We focus on one specific region of the world, **sub-Saharan Africa**, and one language family, **Bantu**. Working with controlled geographical and genealogical settings allows us to explore the depths of the sociolinguistic history of Bantu speaker communities. At the same time, focusing on the languages of the Bantu people and neighbouring language communities enables us to build a large data set with a fine-tuned and multilayered inventory of sociohistorical and linguistic variables, building on history and sociology in an explicit **interdisciplinary fashion**. All these components are essential for the purpose of large-scale comparative research. They also provide the most appropriate ground for digging further into qualitative investigations of specific areas and contact settings.

1.2.2 Grammatical gender

We investigate sociohistorical correlates of variation and change in nominal morphosyntax, focusing on diachronic change in **nominal classification systems**. These divide the nominal lexicon of a language into classes based on shared semantic, morphosyntactic and/or phonological properties (Seifart 2010). The Bantu languages have nominal classification systems that are coded through complex patterns of grammatical agreement and are known in the literature as **grammatical gender systems** (Corbett 1991, Van de Velde 2019).

As mentioned above, gender systems are highly stable (Nichols 2003), yet prone to change in situations of language contact (Trudgill 1999, Di Garbo 2020). This is because the patterns of inflections whereby gender distinctions are coded on agreement targets (modifiers, predicates, and pronouns) are hard to acquire non-natively, and thus likely to undergo erosion or reanalysis through second language learning. Gender systems may be reanalyzed to encode animacy, such that gender marking becomes polarized around the coding of the distinction between living and non-living entities (Igartua & Santazilia 2018; Seifart 2018; Güldemann 2021). In our recent work on the northern Bantu borderlands, we show that animacy-based gender systems are likely to emerge and spread across (related and unrelated) languages in contact (Güldemann et al. 2021, Verkerk and Di Garbo 2022) where compositional meaning structures typically abound (Kempe & Brooks 2018). Other ways in which gender systems change are exemplified below, with reference to Bantu.

1.2.3 Bantu gender systems and their sociogeographic correlates

The Bantu language family (also known as Narrow Bantu) is a large (550+ languages) subfamily of Niger-Congo, the biggest language family of Africa (Bendor-Samuel & Hartell 1989). Bantu languages are famous for their remarkably rich gender systems. These systems typically consist of **more than five gender distinctions** and a combination of **semantic and formal assignment criteria**, which are not based on sex (Van de Velde 2019). Gender distinctions are marked through prefixes, which also code for number. Adnominal modifiers, pronouns, and verbs typically inflect in agreement with the gender of nouns, which are mostly also overtly marked for gender.

While this description is typical for most Bantu gender systems and fits Proto-Bantu reconstruction (Meeussen 1967; Van de Velde 2019), Bantu gender systems also vary, sometimes in very conspicuous ways, and on different levels. Here we focus on pan-Bantu processes of variation and change on three levels: (1) **number of gender distinctions**, (2) **types of agreement patterns**; and (3) patterns of **gender assignment**. These domains of analysis are associated with renowned patterns of worldwide crosslinguistic variation in the gender domain (Corbett 1991; Audring 2017).

With respect to the number of gender distinctions, Bantu languages with conservative gender systems tend to exhibit seven or more class distinctions (Maho 1999: 54). Both Maho (1999) and Di Garbo & Verkerk (2022) find that languages with highly eroded systems, featuring only two genders (the animate and the inanimate) or no gender at all cluster around the northern Bantu borderlands.

¹ These are a set of methods that model language change on the branches of a phylogenetic tree, see Verkerk (2014).

Similarly, the loss of the locative, diminutive and augmentative genders has been reported to be a relatively common feature among the southern Bantu languages (Güldemann 1999, Gibson et al. 2017), but not elsewhere.

With respect to types of agreement patterns, it has been shown that several Bantu languages display instances of semantic agreement whereby, irrespective of their lexically-specified gender, nouns denoting humans and/or animate entities take agreement in gender 1/2, which is the default lexical gender for most human nouns (see Van de Velde 2019 for other types of semantic agreement). Di Garbo & Verkerk (2022) find that animacy-based agreement is wide-spread in the northwestern² Bantu area (attested in 51 out of the 179 sampled languages), which is the most linguistically diverse area of the Bantu-speaking world (Nurse and Phillipson 2003: 165). These findings contrast with earlier accounts, which conceived of animacy-based agreement as attested only in eastern coastal Bantu languages (Contini-Morava 2008, see earlier documentation by Wald 1975).

Finally, with respect to gender assignment, studies focusing on individual Bantu languages, from the perspective of cognitive semantics (e.g., Contini-Morava 1994) or language acquisition (e.g. Demuth 2000) abound, and attempts at reconstructing the semantics of gender distinctions in Proto-Bantu have also been made (Denny & Creider 1996). However, to the best of our knowledge, variation and change in the gender assignment principles of Bantu languages in contact with one another have not been devoted much attention in the literature.

Language contact indeed plays an important role in explaining variation and change of Bantu gender systems. Verkerk and Di Garbo (2022) find that in northwestern Bantu languages, radical animacy-based restructuring of gender systems is most likely to occur in languages bordering with Ubangi and Central Sudanic and in Bantu languages of wider communication. Güldemann et al. (2021) show that many of these neighbouring non-Bantu languages are also characterised by animacy-based gender systems, which supports the idea that patterns of restructuring spread and cluster areally. The loss of the locative, diminutive, and augmentative genders in southern Bantu has been related to the emergence of suffixes of the same function, which grammaticalised because of Khoisan influence (Güldemann 1999). With respect to gender assignment, the impact of sociolinguistic factors on variation and change in the gender assignment of loanwords has been studied based on individual languages and contact scenarios (Mous 2001; Gunnink et al. 2015). Despite this scattered evidence, there is still a lack of comprehensive comparative and small-scale studies addressing contact-induced change in the gender domain in Bantu languages. **Any new findings** on why some processes of change are only attested in some Bantu ecologies, but not others, additionally have the potential of **shedding light on Bantu linguistic and population history and on sociohistorical correlates of language change**.

Table 1 provides a list of earlier work conducted independently and/or collaboratively by selected members of the consortium, which addresses themes that are relevant to the project, from sociolinguistic typology to comparative Bantu grammar.

Table 1: List of earlier work already carried out by selected members of the consortium on the theme of the project (see bibliography for the full reference entries).

1	Di Garbo and Verkerk (2022). <i>A typology of northwestern Bantu gender systems</i>
2	Verkerk and Di Garbo (2022). <i>Sociogeographic correlates of typological variation in northwestern Bantu gender systems</i>
3	Di Garbo (2020). <i>The complexity of grammatical gender and language ecology</i> .
4	Aunio et al. (2019). <i>The Mara Languages JE40</i> .
5	Bernander & Laine (2020). <i>The formation of existential constructions in Western Serengeti – a micro-comparative exploration of variation and change</i>
6	Fehn (2017). <i>Nominal gender marking and case in Ts'ixa</i> .
7	Fehn and Phiri (2017). <i>Nominal Marking in Northern Tshwa (Kalahari Khoe)</i> .
8	Karani (2018). <i>Syntactic categories and argument structure in Parakuyo-Maasai</i>
9	Pacchiarotti and Bostoen (2021). <i>The Evolution of the Ngwi noun class system (West-Coastal Bantu, B861, DRC)</i>
10	Pacchiarotti et al. (2019). <i>Untangling the West-Coastal Bantu Mess: Identification, Geography and Phylogeny of the Bantu B50-80 Languages</i> .

² Here and in the rest of this project proposal, we use the labels northwestern, eastern, and southern Bantu in a geographical sense, without referring to any specific genealogical classification.

1.3. Methodology and risk management

1.3.1 Methodology

The SocioBaGS project relies on the combination of two main methodological approaches. On the one hand, we conduct large-scale comparative studies of the gender systems and sociolinguistic profiles of 150 to 200 Bantu languages sampled from all regions of the Bantu-speaking world. On the other hand, we run a collection of regional field-based studies which zoom in on specific communities, each targeting one of the three scenarios outlined in section 1.1: **Bantu <-> non-Bantu**; **Bantu <-> Bantu-Lingua-Franca**; and **Bantu <-> Bantu contact**.³

For the **linguistic data collection**, the same questionnaire will be used both in the large-scale and regional studies. This tool has been developed by the Scientific Coordinators in their earlier joint work and builds upon established research on the typology of gender systems (Di Garbo and Verkerk 2022; Corbett 1991).

For the **sociolinguistic data collection**, we will use adapted versions of a newly developed tool for comparative sociolinguistic research. This is the **GramAdapt sociolinguistic questionnaire**, which has been designed and implemented at the University of Helsinki in the context of the ERC-funded project *Linguistic Adaptation*, of which Di Garbo is a member. The questionnaire was developed to enable comparative qualitative and quantitative investigations of interactions between pairs of language communities in contact. Importantly, it is meant to be answered by experts of individual language communities and/or through the use of published sources, rather than by community members (Di Garbo et al. 2021; Kashima et al. under revision). It consists of two parts: the **Overview Questionnaire**, which considers macro-level aspects of the sociolinguistic settings of language communities (e.g. demography, political complexity, subsistence practices; language policies and ideologies) and the **Domains Questionnaire**, which looks at interactions between pairs of language communities in contact across six social domains: family and kin; exchange practices (including marriage); daily interactions beyond the family; labour; knowledge exchange (education, religion). In the context of the SocioBaGS project, sociolinguistic data for the large-scale comparative studies will be collected based on an adapted version of the GramAdapt Overview Questionnaire. The sociolinguistic data for the regional studies on Bantu <-> non-Bantu and Bantu <-> Bantu interactions will be based on an adapted version of the GramAdapt Domains Questionnaire. The adjustments we plan to make to the current version of the two Questionnaires mostly concern reducing the number of questions (currently, the GramAdapt Questionnaire consists of 250+ questions).

One of the methodological strengths of the SocioBaGS project is thus the variety of data types that we collect. However, varied data types also require **varied methods** of analysis. The composition of our respective teams (described in detail in section 1.3.2) is meant to address exactly this need. The following approaches are employed in the project to address the peculiarities of the data that we collect:

- Qualitative and quantitative methods for **analysing comparative data** on language structures and sociolinguistic profiles. This type of expertise will be essential for WP 1 and 2.
- Qualitative methods for **analysing corpus data collected in the field**. This type of expertise is mostly channelled into the field-based studies that we propose for WP 3. Data collection will mostly be based on elicitation, but the data will be cross-checked with corpus data based on narrative that exists for some of the target languages.
- Quantitative methods for **modelling the interaction between types of gender systems and sociolinguistic variables**. This type of expertise is essential for analysing the typological data resulting from WP 1 and 2 and will be mostly applied in WP4.

³ A similar approach, based on a combination of macro- and micro-level investigations of Bantu morphosyntax, has been adopted by the recently completed project ‘Morphosyntactic Variation in Bantu: Typology, contact and change’. The project investigated a wealth of morphosyntactic features, including grammatical gender, and their distribution across the whole of the Bantu family. See, for instance, Marten et al. (2022) for a comparative overview, based on phylogenetic comparative methods, and Gibson and Marten (2019) for a study of morphosyntactic change in one particular language, the Bantu language Rangi.

- Qualitative and quantitative methods for **historical-comparative reconstruction**. The wealth of synchronic data that we collect will enable us to draw inferences on possible trajectories of change in Bantu gender systems. While, in our project, we do not have a study that is solely devoted to historical-comparative reconstruction, considerations pertaining to diachrony will be instrumental to the studies we run in WP3. In addition, the quantitative methods we employ in WP4 will always rely on phylogeny (that is, on information pertaining to the genealogy of Bantu languages), through the use of phylogenetic comparative methods.

Combining accurate variable design, high-quality and fine-grained data, and a variety of methods to tackle these data is key to investigate sociolinguistic correlates of language structures and to empirically verify any claimed relation between languages and their socio-historical environment (Kaius and Di Garbo 2018; Verkerk & Di Garbo 2022; Di Garbo et al. 2023). The SocioBaGS' research consortium will put a strong focus on best practices of variable design and data analysis. This priority is reflected by each of the work packages our research programme relies upon, which we describe in the next section.

1.3.2 Scientific programme

Our research program consists of **four Work Packages (WPs)**, articulated into several smaller studies. We conduct extensive studies of Bantu gender restructuring across the whole family, in a comparative perspective (WP1) and focusing on fieldwork data from selected Bantu-speaking regions (WP3). We also study what type of gender system, if any, is attested in families that have been in contact with Bantu and in Bantu languages of wider communication (WP2). These three WPs provide us with a baseline to assess the nature and direction of contact influence through a **variety of data types**, from large-scale typological and sociolinguistic data to fieldwork data that zoom in on individual language communities. This **multilayered study design** culminates in WP4 where we bring together all the evidence from WP1, WP2, and WP3 and run a pan-Bantu study of sociolinguistic and sociohistorical correlates of variation and change in Bantu gender systems using state-of-the art statistical methodologies (Study 4.1 and 4.2). A theoretical and methodological review paper concludes the project (Study 4.3).

Figure 1 shows the location of the Bantu language family in sub-Saharan Africa in the context of neighbouring non-Bantu groupings and is thus an illustration of our object of study. The orange-coloured dots represent the Bantu languages that the two Scientific Coordinators, Francesca Di Garbo and Annemarie Verkerk, studied in their earlier work (Di Garbo and Verkerk 2022; Verkerk and Di Garbo 2022). The black-coloured dots represent the languages that the SocioBaGS project will target through a combination of large-scale comparative studies and regional studies of individual communities. The text boxes in light pink point at the non-Bantu language groupings that are located on the northwestern, northeastern, and southern fringe of the Bantu-speaking world. These are the three Bantu-non-Bantu borderlands that we target in our regional studies.

1.3.2.1 WP1: Comparative Bantu gender systems and sociolinguistics

Study 1.1 is a **comparative study of systems of gender marking in the Bantu language family**. It extends the work on northwestern Bantu languages by Di Garbo and Verkerk (2022) to southern and eastern Bantu. We investigate the extent of the distribution of animacy-based agreement beyond northwestern Bantu, and study what other processes of restructuring may trigger gender erosion and loss beyond animacy-based agreement. The data will be collected through reference grammars and expert consultation, using an updated version of the coding sheet for Bantu gender systems, which Di Garbo and Verkerk (2022) built for their work on northwestern Bantu. The data will be analysed qualitatively as well as with the support of quantitative methodologies, including phylogenetic comparative methods. These methods will allow us to develop an understanding of the distribution of restructured gender systems, types of restructuring and any genealogical biases therein. **Study 1.2** is a **comprehensive sociolinguistic typological survey of the sampled Bantu languages**, which brings together the language sample built for Study 1.1 and the northwestern Bantu sample used by Verkerk and Di Garbo (2022). In Verkerk and Di Garbo (2022), we used a combination of demographic and

geographical variables to test the impact of contact on the degree of animacy-based restructuring in northwestern Bantu gender marking. In this study, we will collect extensive and fine-grained sociolinguistic information on all sampled Bantu languages using existing sociolinguistic and ethnographic literature and based on a revised version of the GramAdapt Overview Questionnaire (for variable design, Di Garbo et al. 2021, Kashima et al under revision).

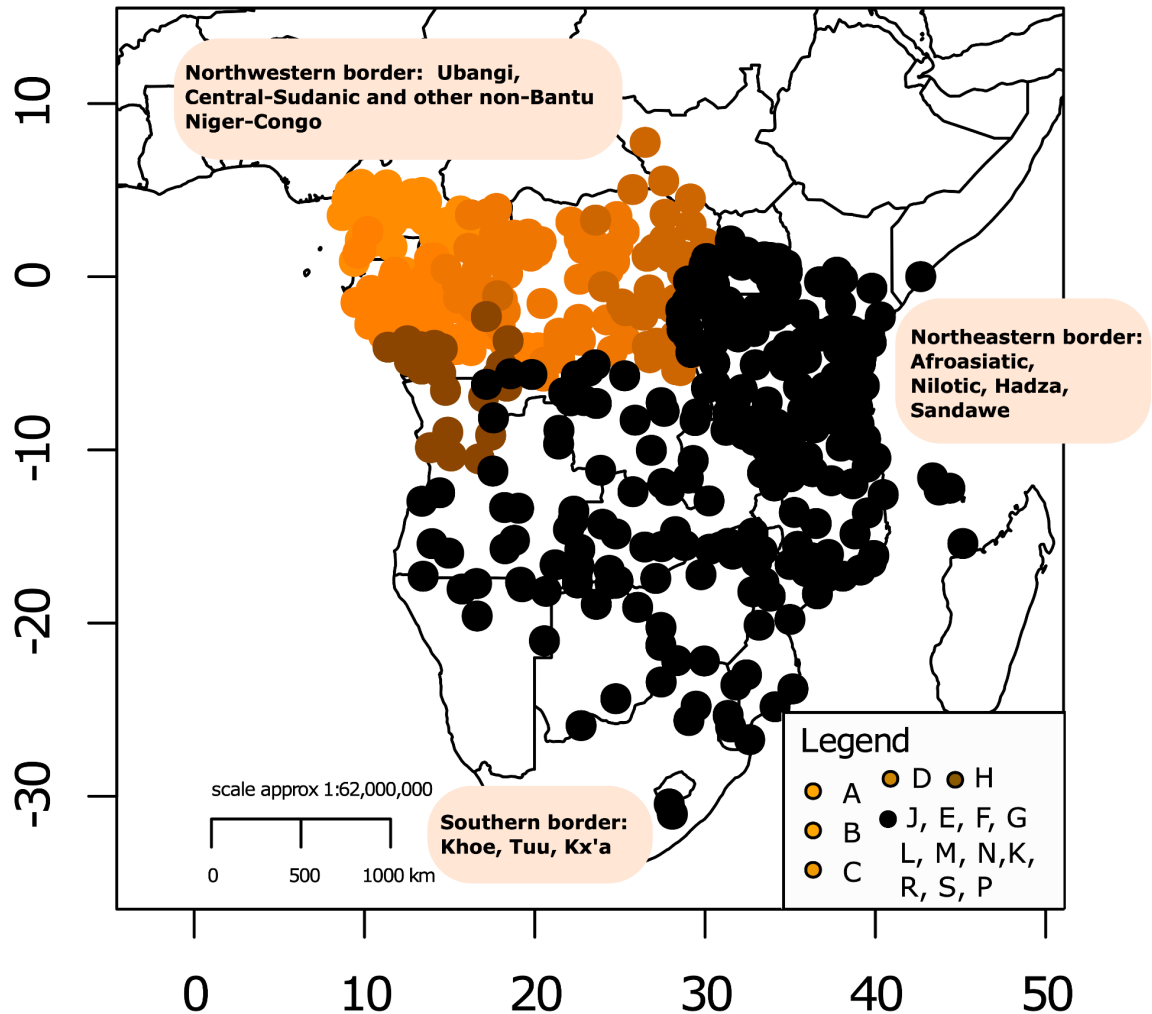


Figure 1: The area under study. The dots coloured in different shades of orange represent languages in the scope of earlier work by Di Garbo and Verkerk (2022) and Verkerk and Di Garbo (2022). The black dots stand for the languages we target in the SocioBaGS project. These languages belong to different geographical subgroupings, known as the Guthrie zones, which are labelled with alphabetical letters (provided in the legend). The map is a revised version of Figure 1 in Di Garbo and Verkerk (2022: 1181).

The type of information we will gather for each of the sampled languages includes, for instance, whether it is in contact with any other Bantu lingua franca, other Bantu and non-Bantu languages.

The two studies which are planned for this WP will be coordinated by the German side of our research consortium. The PhD student hired in Saarbrücken will be responsible for collecting and analysing the linguistic and sociolinguistic comparative data. S/he will work under the supervision of Verkerk and Di Garbo and will also benefit from assistance from two student assistants. WP1 will span the entire duration of the project (36 months). Its main outcomes will be the PhD student's **doctoral dissertation** and **three or more research articles** co-authored primarily by the PhD student and possibly the two scientific coordinators and the postdoc. The main risks of this WP are related to keeping track of the doctoral student's research progress, while ensuring their well-being and

successful integration in the larger team. Regular supervision meetings, student assistant support, as well as the fact that the doctoral student's tasks will be also part of the workflow shared by all members of the core team will help mitigate these risks and grant a successful and timely completion of the doctoral dissertation.

1.3.2.2. WP2: Comparative studies of gender systems in the target contact languages

Study 2.1 is a comprehensive **overview of nominal classification in non-Bantu contact languages**. Here we gather data on the nominal classification systems attested in non-Bantu language families that have been in contact with Bantu, elaborating on such studies as Güldeman (2000, 2021). The language families in focus include Ubangi (Niger-Congo), Cushitic (Afro-Asiatic), Nilotic, Central Sudanic, Khoe-Kwadi, Kx'a, and Tuu⁴. In order to shed light on the diachronic scenarios which may have given rise to the attested typological distributions, reconstructions of proto-languages will be produced, whenever possible and feasible. Study 2.1 will be **coordinated by the French side of the consortium**, and in close collaboration with all members of the advisory board. Data will be collected through a task-based workshop, organised at the beginning of the second year of funding (January 2025). Our collaborators will be invited to Aix-en-Provence for a three-day workshop held at the *Laboratoire Parole et Langage*, during which they will contribute data on their languages of expertise and/or present their ongoing research on topics that are relevant to the project agenda. A questionnaire developed by Di Garbo, Laine, Verkerk and the PhD student during the first year of funding will be used as the basis for data collection and the coding sheet for the resulting database. This questionnaire will largely overlap with the one used for Study 1.1, but it will also include a few more questions, targeting nominal classification systems beyond grammatical gender, which may exist in the non-Bantu contact languages. In addition to the shared data collection tasks, presentations will be held throughout the duration of the workshop to enhance knowledge exchange and spark collaboration between all participants. This joint data collection effort will culminate in a special issue on nominal classification strategies in non-Bantu languages in contact with Bantu languages, to be submitted to an open access journal in the field of African linguistics, such as *Linguistique et Langues Africaines*, and co-edited by Di Garbo, Laine, Verkerk and the PhD student. The workshop will be publicly advertised and open to contributions from outside the consortium. However, only the travel and accommodation expenses of the members of the consortium will be covered by the project (for details, see budget).

Study 2.2 is an **analysis of the gender systems of Bantu lingua francas**. This study focuses on the typological make-up of the gender systems of a sample of Bantu lingua francas, pidgins and creoles, based on descriptive resources. So far, we observe that Bantu lingua francas, pidgins, and creoles do not constitute a uniform typological profile in terms of the structural features of their gender systems. Attested variation may be explained as a result of distinct processes of language change embedded in specific sociolinguistic scenarios (see also Blasi et al. 2017, who argue against the concept of a 'creole type'). Study 2.2 will be **coordinated by the German side** of the consortium. Comparative data on the gender systems attested in a sample of about 10 Bantu lingua francas will be collected using the same coding sheet developed for WP1, based on reference grammars and expert consultation. The languages of the sample will be selected in order to represent lingua francas from all three different macro-regions of the Bantu-speaking world (northwestern, eastern, and southern). The study will result in the first ever **typological investigation of gender systems in a sample of genealogically related languages of wider communication**. In addition to providing a baseline for better understanding contact influences of Bantu languages of wider communication on other Bantu languages, this type of data provides a unique perspective to the study of language change in contact situations. By comparing languages which share a history of creolization, but also their ancestry, we will be in a position of better understanding **whether the gender systems of these languages have changed in similar ways and, if not, what may have prompted these different pathways of change**.

⁴ The three latter language families are often lumped together under the umbrella label 'Khoisan' (Güldemann 2014).

This study will also benefit from collaboration with S. Kriegel, who is a leading expert on grammaticalization phenomena in creole languages and has extensively worked on the notion of convergence in contact-induced change. The results of Study 2.2 will be published in the form of a research article co-authored by Verkerk, Di Garbo and Kriegel.

One potential risk related to this WP lies in the comparability of the crosslinguistic data resulting from the two studies. Using the same questionnaire as a data collection tool (with a few additional questions for Study 2.1) and coordinating the work of the external collaborators through a task-based workshop (in the case of Study 2.1), will help mitigate this risk.

1.3.2.3. WP3: Regional studies of variation and change in Bantu gender systems

This WP contains a collection of **four case studies** zooming in on **specific regions of the Bantu-speaking world** where Bantu languages are spoken at the border with non-Bantu language families. All three fringes of the Bantu spread zone (northwestern, northeastern and southern) will be addressed in this WP. The four studies will differ not only with respect to the contact scenarios in focus, but also in terms of their breadth, scope, and implementation.

Study 3.1 and 3.2, will be Laine's main responsibility. Their focus will be on two different contact scenarios from the northeastern region of the Bantu spread zone, and their implementation will entail fieldwork specifically conducted for the purpose of this project. Study 3.2 will be run in collaboration with Michael Karani from the University of Dar es Salaam. Study 3.3, and 3.4 will focus on the northwestern and southern fringe of the Bantu-speaking world, respectively. These studies will be run in close collaboration with two members of our advisory board who have extensive expertise in western and southern Bantu languages, and their respective contact ecologies. Sara Pacchiarotti will be our collaborator for the northwestern Bantu study, while Anne-Maria Fehn will be involved in the southern Bantu study. In order to coordinate the implementation of these collaborative studies, we will invite our collaborators to Aix-en-Provence and Saarbrücken for 8 to 10 weeks research visits. Michael Karani's research visit will be based in Aix-en-Provence, while Sara Pacchiarotti and Anne-Maria Fehn's visits will be based in Saarbrücken. The design and objectives of each case study are described in the remainder of this section.

Study 3.1 focuses on the **gender agreement systems of the Mara languages of Tanzania and the effects of Bantu <-> Bantu contact on their development** (scenario C as outlined in Section 1.1). The Mara Region is located in northern Tanzania and is home to three distinct groups of Great Lakes Bantu languages – North Mara, South Mara and Suguti – numbering up to 20 language varieties altogether (cf. Aunio et al. 2019). The languages of the Mara region are illustrated in Figure 2. While the region is among the earliest sites where Bantu speakers entered East Africa, the picture of subsequent migrations and the genealogical relationships within Great Lakes Bantu are not well understood (Nurse 1999). It seems that North Mara, South Mara and Suguti speaking peoples entered the region from different directions, and the North and South Mara branches, in particular, offer a fertile ground for historical studies of contact and convergence between these groups of related languages.⁵ Their study also has implications for the genealogy of the Great Lakes Bantu branch more generally. The gender systems of the Mara languages feature some reorganisation of the diminutives and locatives especially, but the isoglosses do not neatly match proposed genealogical divisions. The study will investigate the gender systems of the Mara languages with the primary aim of reconstructing the systems of North Mara and South Mara and situate these within the wider Great Lakes Bantu family. Any retention and innovation in this domain of grammar will then be explained in terms of different possible socio-historical scenarios. In addition to these comparative analyses, the study will also target any potential influence of Swahili, the lingua franca of the area, on Mara gender systems. Two potentially relevant topics are the spreading of animacy-based agreement, which is pervasive in Swahili, but thus far unattested in the Mara languages (Laine 2023), and the choice of the diminutive and augmentative

⁵ The Bantu languages of the Mara region have also been in contact with early Cushitic and Nilotic-speaking groups (Ehret 1999; Roth and Gibson 2019). The impact of these historical contacts on structural features of the Mara languages falls outside the scope of this project.

classes, which, at least in the Western Serengeti languages, seems to be increasingly aligning with the patterns that are most prevalent in Swahili. Fieldwork exploring not only the gender systems, but also current and historical sociolinguistic settings, will be conducted on nine different language communities located in the Mara region, where Laine has extensive fieldwork experience and existing contacts from his previous work. A fieldwork period of around three months will take place in the autumn of 2024, with a shorter follow-up (less than one month) visit in late 2025.

Study 3.2 focuses on **Temi (also known by the exonym Sonjo), the southernmost language in the Central Kenya Bantu group, spoken in the Ngorongoro Region of Tanzania**, to the west of Lake Natron. The Temi, whose language is largely undocumented (a short sketch by Nurse & Rottland 1991/1992 being the only published descriptive resource we are aware of), are entirely surrounded by the traditionally pastoralist Masai people, who speak an unrelated East Nilotic language known as Masai (alternative names: Maa, Maasai; here we follow the the Glottolog label). The Temi are farmers co-dependent with the Masai, and they are in intense language contact with them. Due to the relative sizes of the populations and the social dynamics between the groups, the Temi are likely to speak Masai, but it is unlikely that many Masai would know Temi. The approximate location of Temi and Masai-speaking communities is shown in Figure 3.

To the best of our knowledge, very little is presently known about the dynamics of contact between the Temi and Masai or their possible effect on language structures.⁶ Our study tackles the sociolinguistic situation and the contact dynamics between the Temi and the Masai, as well as the potential influence of the sex-based gender system of Masai (Payne 1998) on the Temi gender system. As such, the study explores one instance of Bantu <-> non-Bantu contact as posited in Section 1.1 (scenario A). The study will be run in collaboration with Michael Karani from the University of Dar es Salaam. Karani is a Masai speaker and is familiar with the region; contacts with the Temi will therefore be relatively easy to establish, and access to the area will be facilitated by his expertise.

The **Temi language and people are of great interest to the history of migration and language contact in the African Great Lakes and Rift Valley regions**. While Temi seems to be genealogically linked to the Kenyan Bantu languages to the north, the South Mara peoples have oral histories that trace their ancestry to the east and southeast, that is, to areas across the modern-day Serengeti National Park (Shetler 2003). In particular, the Ikoma, Ishenyi, Nata and Ngoreme consider the Temi their relatives and claim that the respective groups used to live together. The linguistic links between the Mara languages and Temi remain uninvestigated. A larger question yet, which the data resulting from this study may contribute to tackle, is the exact details of ancient links between some eastern Great Lakes Bantu languages and the wider Central Kenya languages. Some similarities have been noted between them that seem to question some of the bases upon which the Great Lakes genetic unit has been built (Nurse 1999). Finally, due to the scarcity of existing descriptive work on Temi, all corpus data and synchronic grammatical descriptions generated as by-products of this project will be valuable contributions to the basic documentation of eastern Bantu languages and the world's languages more generally.

⁶ Nurse and Rottland (1991/1992: 1999) mention that lexical influence of Masai on Temi is “surprisingly light”. On the other hand, they report several loanwords from neighbouring southern Cushitic languages.



Figure 2: The Mara languages. The languages targeted in Study 3.1 are: Simbiti, Kuria, Kabwa (North Mara); Ikizu, Ikoma, Ishenye, Zanki (South Mara); Jita, Kwaya (Suguti).



Figure 3: Locations of the Temi and Masai-speaking communities. The Masai speaking area is very vast, extending all the way to the Serengeti in the west and up into Kenya in the North. Thus, this map provides only an approximate representation of the location of Masai-speaking communities.

Study 3.3 focuses on the northwestern border of the Bantu area, in the south of the Central African Republic and the north of Congo-Kinshasa, where Bantu languages are spoken alongside Ubangi (Niger-Congo) and Central Sudanic languages, and language shift is pervasive and multidirectional (Bursens 1958; McMaster 2005; Boyeldieu 1990; Fultz and Morgan 2017). Here, Di Garbo and Verkerk (2022) find a mix of gender systems, including languages with heavily restructured gender systems (Kako, Pande, Mbati, and Polri in the west, and Bodo, Homa, Kari in the east) as well as languages with a more or less intact gender system which also features animacy-based agreement, such as Ngombe. The Ubangi languages show a variety of gender systems, including pronominal gender systems (Zande, Mbayic), Bantu-type gender systems (in Mbayic: Mba, Ndunga, 'Dongo), attrited Bantu-type gender systems (in Mbayic: Ma), and even possessive classifiers ('Dongoko) (Pasch 2020: 234-235). Finally, many other Ubangi languages do not have grammatical gender, but they show an animate/inanimate contrast elsewhere in their grammar that is important for studying animacy-based restructuring of gender systems in this area (Güldemann et al. 2021). The focus of this study is the influence of Bantu-Ubangi contact on the gender system(s) of one or more relevant Bantu languages. Possible candidates are for example Pande in the west of this area; Pande has a completely restructured gender system, with exclusively animacy-based gender agreement solely on verbs (Richardson 1957, Di Garbo and Verkerk 2022). It is surrounded to the north by Ubangi languages without gender systems (Gundi, Banda-Yangere, Southwest Gbaya), as shown in Figure 4. Investigating Pande and its Ubangi neighbours more closely from a sociolinguistic perspective would support or disprove the contact hypothesis. Another option is the area including and surrounding the Ituri forest in eastern Congo-Kinshasa, where not only Bantu restructured gender systems are found, but also Zande and Mbayic, which all have unusual gender systems from an Ubangi perspective (Pasch 2020). Additionally, Central Sudanic languages are widely spoken in this area, which also typically lack grammatical gender systems (Dimmendaal 2000).

This study will be spearheaded by Sara Pacchiarotti, who has a long-term research interest in the area (Pacchiarotti et al. 2019, Pacchiarotti & Bostoen 2021, 2022) and aims to conduct independently funded fieldwork in the area during the SocioBaGS project. Verkerk will help to integrate comparative data from Di Garbo and Verkerk (2022) and the findings from Study 2.1 on the non-Bantu languages spoken along the northwestern Bantu border with new and fine-grained data on Bantu and Ubangi languages in contact generated by Pacchiarotti. The focus will be on less-studied languages in line with Pacchiarotti's aims for funding that is currently under review. Hence this project will also make a direct contribution to the description of gender systems in poorly described languages of the northern Bantu borderlands, which face threat of extinction.



Figure 4: Approximative locations of Pande and neighbouring non-Bantu languages

Study 3.4, in turn, targets the southern border of the Bantu area in Namibia and Botswana, where Bantu languages are spoken in close contact with Khoe–Kwadi, Kx’a, and Tuu (so-called ‘Khoisan’ languages). The dynamics between Khoisan and Bantu populations have been well described (Gunnink et al 2015, Pakendorf et al 2017, Sands and Gunnink 2019), however, detailed studies on contact-induced change of nominal classification systems are lacking. Grammatical gender in Khoe-Kwadi languages is mostly sex-based. It is marked by cumulative morphemes which also include information on person, number and sometimes case. While not all Khoe-Kwadi languages regularly mark nouns for grammatical gender, the gender system is still visible in the pronouns. Known from previous studies is Khoisan influence in the loss of diminutive and locative noun classes in southern Bantu (Güldemann 1999); other divergence in southern Bantu gender systems (e.g. those examined by Gowlett and Dowling 2015) have not been systematically investigated from a contact perspective. Here, we target the Okavango River and Delta areas, where intensive contact between Bantu and Khoe-Kwadi speakers is ongoing. The specific contact scenarios involve contact between Kxoe (a.k.a Khwe, here we follow the Glottocode label), Mbukushu (Bantu) and Yeyi (Bantu), between ||Ani and Yeyi, and between Ts’ixa and Yeyi (approximative locations shown in Figure 5). All Kalahari Khoe languages of the Okavango Delta mark about 75% of their nouns for grammatical gender and it may be assumed that the gender-number markers function as specific articles in those languages. The study will be run in

collaboration with Anne-Marie Fehn who has a long-term research interest in Bantu <-> non-Bantu contact in southern Africa (Fehn 2019, 2020a, 2020b, Fehn et al. 2022) and is currently leading fieldwork-based research on the Okavango River and Delta areas. Similarly to Study 3.2, Verkerk will help to integrate comparative data from Study 1.2 on southern Bantu gender systems and Study 2.1 on the non-Bantu languages spoken along the southern Bantu border with new and fine-grained data on the Okavango River languages generated by Fehn.

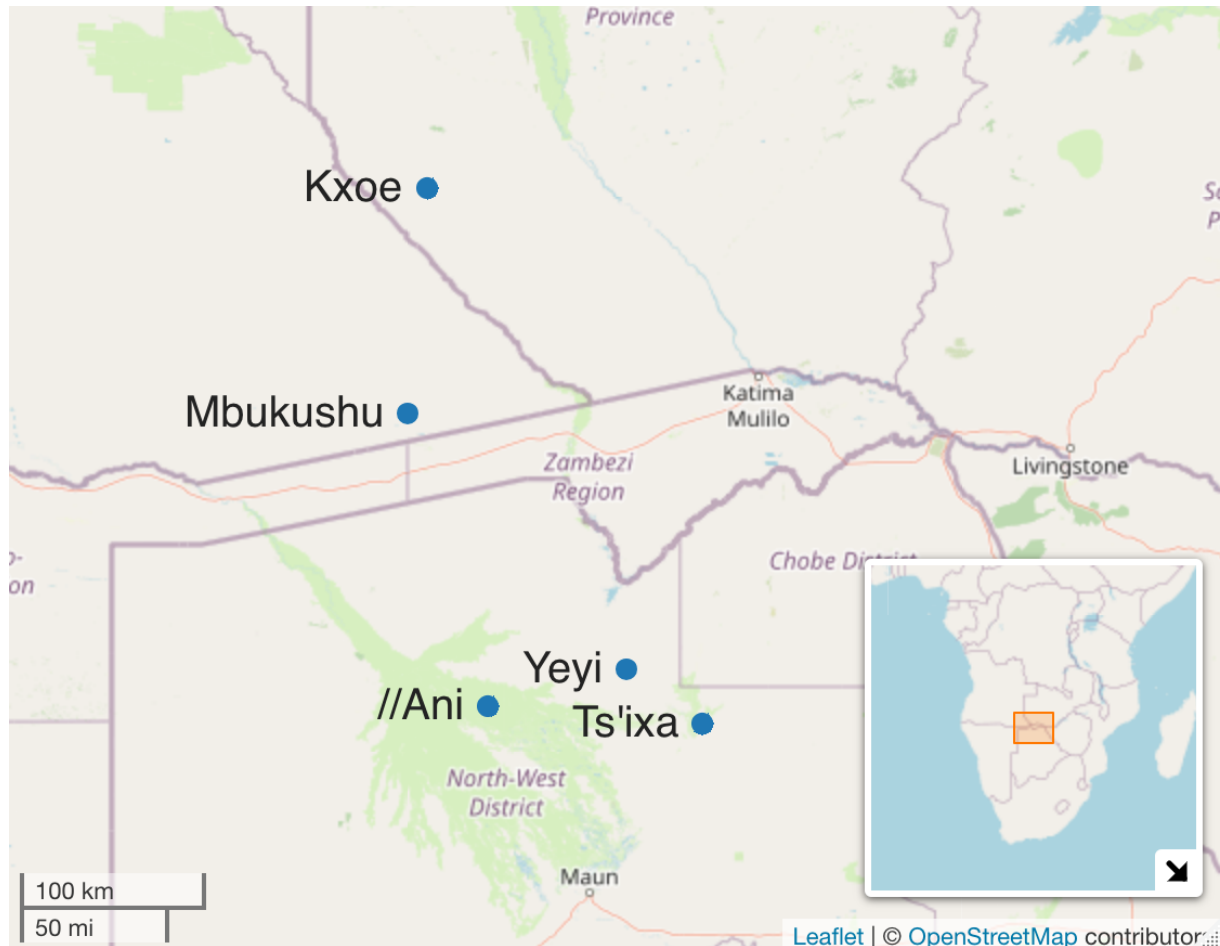


Figure 5: Approximative locations of the languages of the Okavango River and Delta areas, featured in Study 3.4

1.3.2.4 WP 4: Bringing all evidence together – A pan-Bantu sociolinguistic typology of gender systems

WP4 consists of three interconnected studies. Study 4.1 aims at reconciling the **macro- and micro-level approaches to contact settings, linguistic structures, and their mutual interactions that we develop throughout the project**. The focus of the study is twofold. On the one hand, it aims at modelling how change 'spreads' in different language ecologies, for example through the influence of lingua francas, and/or through contact with genealogically unrelated languages. On the other hand, it attempts at 'scaling up' the findings from regional studies so that they can be compared and mapped onto the typological data. Concretely, this is made possible by the fact that the **sociolinguistic data** of the comparative study and those of the regional studies are of a **different, yet interrelated nature**. The comparative sociolinguistic data, collected through an adapted version of the GramAdapt Questionnaire (Di Garbo et al. 2021; Kashima et al. under revision) lend themselves to **classify languages in terms of broad sociolinguistic profiles**, related to such macro-level variables as demography, political complexity, subsistence type, and literacy. Conversely, the sociolinguistic data collected through the regional studies conducted in WP3 zoom in onto interactions between communities in contact, based on an adapted version of the GramAdapt Domains Questionnaire (Di

Garbo et al. 2021 ; Kashima et al. under revision). These data consider such things as whether people from these communities intermarry, live close to each other, practice trade or ceremonial exchange, work, or exchange knowledge together, and how often, for how long, and with which attitudes they have been doing this for. They thus lend themselves to **measures of contact intensity and density**. Finally, given that for each domain we ask whether children are present and involved, and which languages are preferably used in each domain, these data can be also used to characterise **types of bi-/multilingual communities** (adult vs. child bi-/multilingualism; symmetrical vs. asymmetrical bi-multilingualism). Study 4.1 will thus seek to establish whether there is a relation between sociolinguistic profiles of languages at large and degrees and types of contact at the micro-level, and how this may in turn relate to the type of contact effects that we observe at the linguistic level. The study will be a major theoretical and methodological contribution to (typological) research on language contact and language change. It will address current desiderata in sociolinguistic typology (Di Garbo et al 2021, Shcherbakova et al. 2022), by bringing together, probably for the first time, macro- and micro-perspectives on the modelling of sociolinguistic correlates of variation and change in one large family of closely related languages and one domain of grammar,

Study 4.2 builds on the results of WP 4.1 to derive quantitative measures of language contact. Based on these measures, and using Bayesian generalised mixed models, we run statistical analyses where we test different hypotheses about the **distribution of attested types of gender systems and their sociolinguistic correlates**. The analyses will target all aspects of gender systems that we do research on in WP 1, 2 and 3, such as the number of gender distinctions (e.g.: *are bipartite gender systems only attested in the northern Bantu borderlands' languages? If so, why?*), and types of agreement patterns (e.g.: *what is the southernmost fringe of animacy-based agreement and what contact dynamics explain it?*). The study design is modelled on Verkerk and Di Garbo (2022) but takes scope on the whole of the Bantu family, thus providing a major contribution to comparative Bantu linguistics, sociolinguistics typology and, more generally, quantitative methods in the language sciences.

Study 4.1 and 4.2 will be a collaboration by Di Garbo, Verkerk, Laine, the PhD Student, and the Research Engineer who will be hired during the last year of funding. However, the Research Engineer will have a leading role in the completion of Study 4.1 as s/he will be responsible for setting up the **statistical tools and methods** that will enable the team to **capitalise on the different types of data** obtained earlier on during the project. The Research Engineer's expertise in data science and quantitative methods will thus be crucial to the set-up and implementation of the study and will also ensure its feasibility.

Study 4.3 is a **review article focusing on the impact of the project for sociolinguistic typology, and its long-term repercussions beyond Bantu and beyond grammatical gender**. This article, drawing heavily on WP3 but also on WP1 and 2, will provide the first-ever comparative review of contact-induced change between related and unrelated languages, the differences between which are still poorly understood in the field of areal and contact linguistics (for a recent discussion see McColl Millar 2016). In addition, the study will assess benefits and drawbacks of shifting the focus of comparative language contact research **from individual users** (as is common in research on bilingualism and second language acquisition) **to communities of users**. The study will make a substantial contribution to theory and methods of research in the language sciences. It will be conducted by Di Garbo and Verkerk and will also benefit from collaboration with S. Kriegel, who, as mentioned earlier on, is a renowned expert on models of language change in language contact situations.

WP4 hinges upon the outcomes of the three other WPs within the larger project. Thus, its completion naturally entails a certain degree of unpredictability and the possibility of having to adjust its goals and outcomes to the progress of the larger project. What mitigates these risks is the fact that the WP comes at the very end of the larger project. Thus, while the goals and expected outcomes of the three studies may be subject to adjustments, the implementation of these studies does not interfere with the implementation of any of the other work packages within the project.

1.3.3 Summary and risks

To summarise, the SocioBaGS project will consist of four WPs, each articulated in several smaller studies. As detailed in the previous sections, these are:

1. Comparative Bantu gender systems and sociolinguistics (coordinated by the German team)
2. Comparative studies of gender systems in the target contact languages (one block of studies coordinated by the French team and another one by the German team)
3. Regional studies of variation and change in Bantu gender systems (one block of studies coordinated by the French team and another one by the German team)
4. Bringing all the evidence together: A pan-Bantu sociolinguistic typology of gender systems (coordinated by the French team)

Even though different WPs and different studies therein are captained by one or the other team, the members of the consortium will collaborate across teams all throughout the duration of the project. Each team will have its own internal weekly meetings, while the core members of the consortium will meet online once every month (or every two weeks, depending on the unfolding of the project activities). Finally, the task-based workshop organised during the second year of funding will also be an opportunity to meet with all the external collaborators and any other researcher interested in the project. The workflow of the project is summarised in the Gantt chart provided in Figure 6.

One risk component in the organisation of the workflow concerns the collaborative studies we plan to run with three of the external collaborators (Kharani, Pacchiarotti, and Fehn). These researchers will not be salaried by the SocioBaGS project, but their visits to Aix-en-Provence and Saarbrücken are budgeted in the project. The risks implied by their status as external collaborators are mitigated by the fact that all three researchers currently benefit from either permanent (Karani and Fehn) or medium-to-long term academic positions (Pacchiarotti). If indeed, one or more of the external collaborators should not be available for whatever reasons by the time the project starts, we would approach other researchers to conduct field-based studies in the respective areas of interest. There are several options for this (which we will not elaborate on here, without consulting the people involved first), including Africanists in France, Germany, and internationally. These options also include some of the other contributors to the SocioBaGS project and their colleagues.

2. Impact and consequences of the project

Our teams bring together researchers with leading expertise in Bantu historical-comparative and contact linguistics, the language history of Africa, language description, language typology, sociolinguistics, and quantitative methodologies for comparative linguistic research. As such, they are in a **unique position to successfully undertake the tasks entailed by the project** and to make a long-lasting contribution to the study of linguistic diversity and its interaction with the sociohistorical environment. **Eastern Africa expertise will be a distinctive feature of the French team** and will overlap with ongoing research on the sound systems of the languages of the Great Tanzanian Rift Valley at the *Laboratoire Parole et Langage*, and in collaboration with Michael Karani. In addition to the regional studies on northwestern and southern Bantu hosted in Saarbrücken, the focus of the **German team will be on large-scale comparisons of language structure and sociolinguistic profiles**. These two foci will converge towards the final objective of producing a comprehensive model of variation and change in language contact ecologies. In this framework, macro- and micro- perspectives on Bantu gender systems are brought together through state-of-the art statistical methods modelling the rise and spread of variation and change at the individual and community-level.

By pursuing this research agenda, the SocioBaGS project will produce a **wealth of data on Bantu grammar, sociolinguistics, and the typological profile of non-Bantu neighbour languages**. The methods we implement combine **large-scale comparative data** with **fieldwork data** to zoom in on ongoing processes of language variation and change. Comparing evidence across methods (comparative vs. fieldwork) and data types (Bantu-internal and Bantu-external contact) allows us to test hypotheses about scenarios of contact-induced change in unprecedented ways and will refine the standards of research in sociolinguistic typology. Finally, the **large body of knowledge on comparative Bantu sociolinguistics** resulting from the project may be used for a variety of purposes, typological but also regarding multilingualism, language shift and maintenance, and language documentation. Our project addresses an oft-repeated **call for ecologically grounded approaches to the study of linguistic diversity** which has been raised in recent years in connection with **all aspects of large-scale comparative linguistic research**. This includes dense instead of sparse sampling (Bickel 2017; Verkerk and Di Garbo 2022), open source and detailed data on sociolinguistic variables (Di Garbo et al. 2021), and rooting contact-based explanations in historical and social contexts (Thomason 2010). The SocioBaGS consortium is committed to significantly advance this debate by developing new and bold ways of tackling it while building upon earlier research initiated by the two Scientific Coordinators and upon the expertise of the other members of the two teams.

Our dissemination strategy involves presenting both preliminary work and/or study designs as well as fully analysed results at workshops, conferences, and in articles. We aim to publish in peer-reviewed high-quality specialist journals (e.g., *Language, Linguistic Typology, Journal of Language Contact, Journal of African Languages and Linguistics, Linguistics, Language Dynamics and Change, Diachronica, Journal of Historical Sociolinguistics*). We strongly believe in **making data (and code) accessible** alongside publications. We have done so in the past and commit to this practice for the current project. We also believe in **open access**, for which we have partly budgeted. Conferences whose attendance we prioritise are the biannual Bantu conferences, the World Congress of African Linguistics, the biannual meeting of the Association of Linguistic Typology, the annual meeting of the Societas Linguistica Europaea, the International Conferences for Historical Linguistics, and the International Conference on Sociolinguistics. Budgeting has been provided for attending these; **online conferences** and the possibility of linking team meetings to conference travel will be preferred in light of **sustainability**. In addition to participating in scientific conferences and other types of scholarly events, we also plan to regularly engage in dissemination activities for the **wider public** (cf. budget).

Finally, two of the studies in WP3 rely on **fieldwork directly funded by the project**. Field trips are scheduled to take place starting from the second half of the first year of the project. In the event that fieldwork is not possible for unforeseen reasons going beyond our control, consultation with language

informants will take place online. The budget planned for the field trips may be used (pending negotiations with the funders) to compensate research assistants from the local communities. Participation in research by language consultants will be voluntary and based on informed consent. Their involvement will always be acknowledged in our research outputs (anonymously if they wish). Compensation will be provided to the language consultants in accordance with local recommendations (cf. budget).

3. Bibliography

- Audring, J. 2017. Calibrating complexity: How complex is a gender system? *Language Sciences* 60. 53–68.
- Aunio, L., H. Robinson, T. Roth, O. Stegen & J. B. Walker. 2019. The Mara Languages JE40. In M. van de Velde, K. Bostoen, D. Nurse & G. Philippson (eds.), *The Bantu Languages* (2nd edn), 501–532. London & New York: Routledge.
- Bendor-Samuel, J. & R. L. Hartell (eds.) 1989. *The Niger-Congo languages: A classification and description of Africa's largest language family*. Lanham, Maryland: University Press of America.
- Bentz, C., Dediu, D., Verkerk, A. and Jäger, G. (2018) The evolution of language families is shaped by the environment beyond neutral drift. *Nature Human Behavior* 2: 816–821. [10.1038/s41562-018-0457-6](https://doi.org/10.1038/s41562-018-0457-6)
- Bernander, R. & A. Laine. 2020. The formation of existential constructions in Western Serengeti – a micro-comparative exploration of variation and change. *Africana Linguistica* 26. 59–102.
- Bickel, B. 2017. Areas and universals. In R. Hickey (ed.), *The Cambridge handbook of areal linguistics*, 40–55. Cambridge: Cambridge University Press.
- Blasi, D. E., S. M. Michaelis & M. Haspelmath. 2017. Grammars are robustly transmitted even during the emergence of creole languages. *Nature Human Behaviour* 1. 723–729.
- Boyeldieu, P. 1990. L'extension oubanguienne des langues « sara ». *Journal des Africanistes* 60(1). 77–106.
- Burssens, H. 1958. *Les Peuplades de l'entre Congo-Ubangi (Ngbandi, Ngbaka, Mbanza, Ngombe et Gens d'eau)*. Tervuren.
- Contini-Morava, E. 2008. Human relationship terms, discourse prominence, and asymmetrical animacy in Swahili. *Journal of African Languages and Linguistics* 29(2). 127–171.
- Contini-Morava, E. 1994. *Noun classification in Swahili*. Report available at <http://www2.iath.virginia.edu/swahili/oldversion/oldswahili.html>.
- Corbett, G. G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Demuth, K. 2000. Bantu noun class systems: Loanword and acquisition evidence of semantic productivity. In G. Senft (ed.), *Systems of noun classification*, 270–292. Cambridge: Cambridge University Press.
- Di Garbo, F. 2020. The complexity of grammatical gender and language ecology. In P. Arkadiev & F. Gardani (eds.), *The complexities of morphology*, 193–229. Oxford: Oxford University Press.
- Di Garbo, F. & R. Napoleão de Souza. 2023. A sampling technique for worldwide comparisons of language contact scenarios. *Linguistic Typology*. <https://doi.org/10.1515/lingty-2022-0005>.
- Di Garbo, F. & A. Verkerk. 2022. A typology of northwestern Bantu gender systems. *Linguistics* 60(4). 1169–1239. <https://doi.org/10.1515/ling-2020-0217>.
- Di Garbo, F., E. Kashima, R. Napoleão de Souza & K. Sinnemäki. 2021. Concepts and methods for integrating language typology and sociolinguistics. In S. Ballarè & G. Inglese (eds.), *Atti del workshop SLI sociolinguistica e tipologia: Verso un approccio integrato allo studio della variazione*, 143–176. Milano: Officina 21.
- Dimmendaal, G. 2000. Number marking and noun categorization in Nilo-Saharan languages. *Anthropological linguistics*, 214–261.
- Denny, J. P. & C. A. Creider. 1986. The semantics of noun classes in proto-Bantu. In C. G. Craig (ed.), *Noun classes and categorization*, 217–239. Amsterdam: John Benjamins.
- Ehret, C. 1999. Subclassifying Bantu: the evidence of stem morpheme innovations. In J.-M. Hombert & L. M. Hyman (eds.), *Bantu historical linguistics: theoretical and empirical perspectives*, 43–147. Stanford: CSLI (Center for the Study of Language and Information)
- Fehn, A.-M. 2020a. Click loss in Khoe-Kwadi. In B. Sands (ed.), *Click Consonants*, 291–335. Leiden: Brill. DOI: https://doi.org/10.1163/9789004424357_010
- Fehn, A.-M. 2020b. Click replacement and loss in Ju. In B. Sands (ed.), *Click Consonants*, 336–355. Leiden: Brill. DOI: https://doi.org/10.1163/9789004424357_011

- Fehn, A.-M. 2017. Nominal gender marking and case in Ts'ixa. In A. Fehn (ed.), *Khoisan Languages and Linguistics. Proceedings of the 4th International Symposium, Riezlern 2011*, 121–156. Cologne: Rüdiger Köppe.
- Fehn, A.-M. 2019. Kuvale: A Bantu language of southwestern Angola. *Journal of African Languages and Linguistics* 40 (2). 235–270.
- Fehn, A.-M., B. Amorim & J. Rocha. 2022. The linguistic and genetic landscape of southern Africa. *Journal of Anthropological Sciences* 100.–265.
- Fehn, A. & A. Phir. 2017. Nominal Marking in Northern Tshwa (Kalahari Khoe). In A. Andrason & T. Biberauer (eds.), *Iets kleins/'n Klein ietsie: a festschrift to honour Johan Oosthuizen on the occasion of his retirement* (Stellenbosch Papers in Linguistics (Plus) 48), 105–122.
- Fultz, J. & D. Morgan. 2017. *Enquête dialectale de l'Ubangi et de la Mongala, région de l'Equateur, République du Zaïre*. Dallas: SIL International.
- Gibson, H. & L. Marten. Probing the interaction of language contact and internal innovation: four case studies of morphosyntactic change in Rangi. *Studies in African Linguistics* 48(1). 63–92.
- Gibson, H., R. Guérois & L. Marten. 2017. Patterns and developments in the marking of diminutives in Bantu. *Nordic Journal of African Studies* 26 (4). 344–383.
- Gowlett, D. & T. Dowling. 2015. Incipient merger of Cls 11 and 5 in Xhosa? *South African Journal of African Languages* 35. 67–81.
- Gunnink, H, B. Sands, B. Pakendorf & K. Bostoen. 2015. Prehistoric language contact in the Kavango-Zambezi transfrontier area: Khoisan influence on southwestern Bantu languages. *Journal of African Languages and Linguistics* 36(2). 193–232.
- Güldemann, T. 1999. Head-initial meets head-final: nominal suffixes in eastern and southern Bantu from a historical perspective. *Studies in African Linguistics* 28,1. 49–91.
- Güldemann, T. 2000. Noun categorization in Non-Khoe lineages of Khoisan. *Afrikanistische Arbeitspapiere* 63: 5–33.
- Güldemann, T. 2014. 'Khoisan' linguistic classification today. In T. Güldemann & A.-M. Fehn (eds.), *Beyond 'Khoisan': Historical relations in the Kalahari basin*, 1–40. Amsterdam: John Benjamins.
- Güldemann, T. & I. Fiedler. 2019. Niger-Congo “noun classes” conflate gender with deriflection. In F. Di Garbo, B. Olsson & B. Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 95–145. Berlin: Language Science Press.
- Güldemann, T. 2021. From asymmetric grammatical treatment of nouns to gender: Animacy-based noun classification in Central Africa and its typological significance. Paper presented at *African Linguistics Research Coll.* Berlin 27/4/2021
- Güldemann, T, F. Di Garbo & A. Verkerk. 2021. Contact-induced change in gender systems of Northern Bantu languages. Paper presented at *WOCAL 10*, Leiden.
- Igartua, I. & E. Santazilia. 2018. How animacy and natural gender constrain morphological complexity: Evidence from diachrony. *Open Linguistics* 4 (1). 438–52.
- Karani, M. 2018. *Syntactic categories and argument structure in Parakuyo-Maasai* (Doctoral dissertation, Stellenbosch: Stellenbosch University).
- Kashima, E. F. Di Garbo, O. Khanina, & R. Singer. Under revision. The design principles of a sociolinguistic-typological questionnaire for language contact research.
- Kempe, V. & P. J. Brooks. 2018. Linking adult second language learning and diachronic change: A cautionary note. *Frontiers in Psychology* 9 (April). 621–625.
- Laine, A. 2023 [accepted]. *Microvariation in the Western Serengeti languages: Comparative Morphosyntax of Ikoma, Ishenyi, Nata and Ngoreme*. University of Helsinki PhD dissertation.
- Laine, A, R. Bernander & H. Gibson. Forthcoming. Negation in Ngoreme. In M. Miestamo & L. Veselinova (eds.), *Negation in the Languages of the World*. Berlin: Language Science Press.
- Lupyan, G. & R. Dale. 2010. Language structure is partly determined by social structure. *PLoS ONE* 5 (1). e8559.
- Maho, J. 1999. *A comparative study of Bantu noun classes*. Göteborg: Acta Universitatis Gothoburgensis.

- Marten, L., P. Edelsten, H. Gibson, R. Grollemund, R. Guérois & T. Rama. 2022. Inheritance and convergence in Bantu lexicon and morphosyntax: A phylogenetic study. Paper presented at *The International Conference of Historical Linguistics*. Oxford University.
- McColl Millar, R. *Contact: The interaction of closely related linguistic varieties and the history of English*. Edinburgh: Edinburgh University Press.
- McMaster, M. A. 2005. Language shift and its reflection in African archaeology: Cord rouletting in the Uele and interlacustrine regions. *Azania* 40(43–72).
- Mous, M. (2001). Paralexification in language intertwining. In N. Smith & T. Veenstra (eds.), *Creolization and contact*, 113–123. Amsterdam: John Benjamins.
- Mufwene, S. S. 2003. Contact languages in the Bantu area. In G. Philippson & D. Nurse (eds.), *The Bantu languages*, 195–208. London: Routledge.
- Nichols, J. 2003. Diversity and stability in language. In R. D. Janda & B. D. Joseph (eds.), *Handbook of historical linguistics*, 283–310. London: Blackwell.
- Nurse, D. 1988. The diachronic background to the language communities of Southwestern Tanzania. *Sprache und Geschichte in Afrika* 9: 15–115.
- Nurse, D. 1999. Towards a historical classification of East African Bantu languages. In J.-M. Hombert & L. M. Hyman (eds.), *Bantu Historical Linguistics*, 1–42. Stanford: CSLI Publications.
- Nurse, D. & G. Philippson. 2003. Towards a historical classification of the Bantu languages. In D. Nurse & G. Philippson (eds.), *The Bantu languages*, 164–181. London: Routledge.
- Nurse, D. and F. Rottland. 1991/1992. Sonjo: description, classification, history. *Sprache und Geschichte in Afrika (SUGIA)* 12/13. 171–289.
- Pacchiarotti, S., N. Chousou-Polydouri & K. Bostoen. 2019. Untangling the West-Coastal Bantu Mess: Identification, Geography and Phylogeny of the Bantu B50-80 Languages. *Africana Linguistica* 25. 1–74.
- Pacchiarotti, S. & K. Bostoen. 2021. The Evolution of the Ngwi noun class system (West-Coastal Bantu, B861, DRC). *Language in Africa* 2 (3). 11–67.
- Pacchiarotti, S. & K. Bostoen. 2022. Erratic velars in West-Coastal Bantu: Explaining irregular sound change in Central Africa. *Journal of Historical Linguistics* 12 (3). 381–445.
- Pakendorf, B., H. Gunnink, B. Sands & K. Bostoen. 2017. Prehistoric Bantu-Khoisan language contact: A cross-disciplinary approach. *Language Dynamics and Change* 7(1). 1–46.
- Pasch, Helma. 2020. Ubangi. In R. Vossen & G. J. Dimmendaal (eds.), *The Oxford handbook of African languages*, 231–238. Oxford: Oxford University Press.
- Payne, D. L. 1998. Maasai gender in typological perspective. *Studies in African Linguistics* 27. 159–175.
- Richardson, I. 1957. *Linguistic survey of the Northern Bantu Borderland*. London: Oxford University Press.
- Roth, T. & H. Gibson. 2019. Differential diagnoses of language change in Ngoreme, a Bantu language of Tanzania. *Africana Linguistica* 15. 281–319.
- Sands, B. & H. Gunnink. 2019. Clicks on the fringes of the Kalahari Basin Area. In E. Clem, P. Jenks & H. Sande (eds.), *Theory and description in African linguistics: Selected papers from the 47th Annual Conference on African Linguistics*, 703–724. Berlin: Language Science Press.
- Seifart, F. 2010. Nominal classification. *Language and Linguistics Compass* 4(8). 719–36.
- Seifart, F. 2018. The semantic reduction of the noun universe and the diachrony of nominal classification. In W. B. McGregor & S. Wichmann (eds.), *The diachrony of classification systems*, 9–32. Amsterdam: John Benjamins.
- Shcherbakova O., S. M. Michaelis, H. Haynie, S. Greenhill, D. Blasi, R. Gray, V. Gast & H. Skirgård. *Grammatical complexity is only weakly influenced by the sociolinguistic environment*. Paper presented at the 14th Conference of the Association for Linguistic Typology. University of Austin: Texas.
- Shirtz, S., Talamo, L., Verkerk, A. 2021. The evolutionary dynamics of negative existentials in Indo-European. *Frontiers in Communication* 6: 661862. [10.3389/fcomm.2021.661862](https://doi.org/10.3389/fcomm.2021.661862).
- Sinnemäki, K. & F. Di Garbo. 2018. Language structures may adapt to the sociolinguistic environment, but it matters what and how you count: A typological study of verbal and nominal complexity. *Frontiers in Psychology* 9 (August). 89–123.

- Thomason, S. 2010. Contact Explanations in Linguistics. In R. Hickey (ed.), *The Handbook of Language Contact*, 31–47. Chichester: Wiley-Blackwell.
- Trudgill, P. 1999. Language contact and the function of linguistic gender. *Poznań Studies in Contemporary Linguistics* 35. 133–52.
- Trudgill, P. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.
- Van de Velde, M. 2019. Nominal morphology and syntax. In M. Van de Velde, K. Bostoen, D. Nurse & G. Philippson (eds.), *The Bantu languages* (2nd edn), 237–269. London: Routledge.
- Verkerk, A. 2014. *The evolutionary dynamics of motion event encoding*. Radboud University Nijmegen PhD dissertation.
- Verkerk, A. (2019) Detecting non-tree-like signal using multiple tree topologies. *Journal of Historical Linguistics* 9(1): 9–69. [10.1075/jhl.17009.ver](https://doi.org/10.1075/jhl.17009.ver)
- Verkerk, A. & Di Garbo, F. 2022. Sociogeographic correlates of typological variation in northwestern Bantu gender systems. *Language Dynamics and Change* 12(2). 155–223. doi: <https://doi.org/10.1163/22105832-bja10017>
- Wald, B. V. 1975. Animate concord in Northeast coastal Bantu: Its linguistic and social implications as a case of grammatical convergence. *Studies in African Linguistics* 6. 267–314.