Embedding in Shawi narrations: A quantitative analysis of embedding in a post-colonial Amazonian indigenous society

LUIS MIGUEL ROJAS-BERSCIA[a] ,
TOMAS LEHECKA[b] , SIMON A. CLAASSEN[c] ,
A. A. K. PEUTE[c] , MOISÉS PINEDO ESCOBEDO[d] ,
SEGUNDO PINEDO ESCOBEDO[d] ,
ABIMAEL HUIÑAPI TANGOA[e] AND
ELIO YUMI PIZANGO[f]

[a]*Max Planck Institute for Psycholinguistics,
Radboud Universiteit Nijmegen, The Netherlands, and
University of Queensland, Australia*
[b]*Max Planck Institute for Psycholinguistics, The Netherlands and
University of Basel, Switzerland*
[c]*Radboud Universiteit Nijmegen and
Max Planck Institute for Psycholinguistics, The Netherlands*
[d]*Comunidad Nativa Santa María de Cahuapanas, Peru*
[e]*Comunidad Nativa Pueblo Chayahuita, Peru*
[f]*Comunidad Nativa Santa Rosa, Balsapuerto, Peru*

ABSTRACT

In this article, we provide the first quantitative account of the frequent use of embedding in Shawi, a Kawapanan language spoken in Peruvian Northwestern Amazonia. We collected a corpus of ninety-two Frog Stories (Mayer 1969) from three different field sites in 2015 and 2016. Using the glossed corpus as our data, we conducted a generalised mixed model analysis, where we predicted the use of embedding with several macrosocial variables, such as gender, age, and education level. We show that bilingualism (Amazonian Spanish-Shawi) and education, mostly restricted by complex gender differences in Shawi communities, play a significant role in the establishment of linguistic preferences in narration. Moreover, we argue that the use of embedding reflects the impact of the mestizo[1] society from the nineteenth century until today in Santa Maria de Cahuapanas, reshaping not only Shawi demographics but also linguistic practices. (Post-colonial societies, Amazonian linguistics, Kawapanan, Shawi, embedding, language variation and change, contact linguistics)*

427

INTRODUCTION

South America is a crucible of linguistic diversity. A contemporary survey suggests the existence of at least 118 genetic units (forty-eight groupings + seventy isolates) (O'Connor & Muysken 2014). What were/are the social dynamics behind the great language diversification in South America? Contemporary sociolinguistic studies on language variation and change in South American small indigenous communities could shed light on this issue, by informing us on how this great diversification could have come to be (Rojas-Berscia 2019b). In this article, we look for the first time at one aspect of contemporary syntactic variation in Shawi, a Kawapanan language spoken in Peruvian northwestern Amazonia.

The syntax of indigenous Amazonian languages has been in the spotlight for at least a decade and a half, since the publication in *Current Anthropology* of Everett's work on the lack of embedding in Pirahã[2] (Everett 2005). Researchers from diverse subdisciplines within linguistics have subsequently conducted surveys and experiments on different languages of the world to follow up on these claims, showing that (complex) embedding is culture-fostered, and therefore not necessarily an innate aspect of language.[3] Although embedding does exist in the languages studied, the degree to which it is used seems to be related to more formal registers (Laury & Ono 2010), formal schooling (Dąbrowska 1997), language-dependent patterns of grammaticalisation (Mithun 2010), and language contact (Sakel & Stapert 2010).

In this article, we adopt a sociolinguistic perspective with regard to embedding, and present a quantitative analysis of embedding in a Shawi corpus. Shawi displays different formal strategies for embedding. However, while their existence is not under discussion, they seem not to be used by all speakers of the language with the same frequency. One of our very first observations in the field when collecting data was that Shawi men overall seem to prefer the use of embedded constructions, in opposition to chains of non-embedded clauses, when narrating a story. We decided to study this phenomenon from a quantitative perspective as a starting point. We then analysed the distribution of embedding strategies in the corpus and the social motivations behind the frequent use of these in Shawi narratives. For this, we collected a corpus of ninety-two Frog Stories (Mayer 1969) in the Shawi town of Santa María de Cahuapanas, and some communities around the village of Balsapuerto and along the Sillay River, using the Max Planck Institute Shawi Fieldkit, a tablet app specially designed for mass data collection in the field with ELAN-ready files as an output.

In narratives, speakers have the option to bind sentences together into a coherent story using two different strategies:

    a. Chain of non-embedded sentences (coherence solely through temporal/linear order in the narrative)
    b. Embedding (coherence through syntactic relationships)

Below we present two examples. (1) displays a chain of non-embedded sentences, and (2) displays the forms to be analysed in use: instead of chains of non-embedded

sentences, embedding strategies are deployed. Embedders are in bold; glossing abbreviations are listed in the appendix.

(1) (FS_Sillay_CH_160329)
Ni'ni'ra   wa'washa   ni'-sa-r-in-ø.   Tururu'   wens-a-r-in.
dog   child   see-PROG-N.FUT-3MIN.A-3MIN.O   frog   sit-PROG-N.FUT-3MIN.S

Wa'washa   ni'nira-ri   wens-apu-r-in.
child   dog-ERG   sit-together-N.FUT-3MIN.S

'The dog is observing the child. The frog is sitting. The dog is sitting next to the child.'

(2) (FS_Sillay_JTC_162803)
Wa'washa-wa   ni'nira-re'   pipa-a-**watu**-n   nu'-te-sa-r-in
child-OFF   dog-COM   lay.down-PROG-SEQ-3MIN.S   look-VM-PROG-N.FUT-3MIN.S

ni'ni-ra'wa-re'-ø
dog-DIM-COM-3MIN

'The child, after laying down with the dog, was looking (at something) with it.'

Ni'ni'-rawa   putiria-puchin   ninin-pa   mu'ten   pu'mu-r-in-ø,
dog-DIM   bottle-COMP   hole-ALL   head;3.POSS put-N.FUT-3MIN.A-3MIN.O

wa'washa-'wa-nta   sunpiriu-(p)uchin   ninin-ø-**su'**   ma'-sa-r-in-ø.
child-DIM-ADDIT   hat-COMP   hole-3MIN.COP-NMLZ   grab-PROG.NFUT-3MIN.A-3MIN.O

'The dog pushed his head through the hole of a bottle, and the child is grabbing what is the hole of something like a hat.'

Our initial results suggest that the frequency of embedding is primarily predicted by the speaker's level of education, where consultants with a higher education display a preference for the use of embedding in narrations. In addition, we also observe a significant gender difference (men using embedding more than women), as well as a systematic regional variation, possibly related to the impact of migrations of Spanish speakers to the Cahuapanas region since the nineteenth century. We hypothesise that the gender effect is not causally related to the preference for embedding (or the evidence of gendered speech), but is indirectly related to it, as part of a whole conglomerate of variables that account for this phenomenon in particular.

The article is structured as follows: first, we provide a short ethnographic description of the Shawi in order to lay a foundation for the social variables used in our analysis. We then focus on the use of the four most common embedding morphemes: (i) the general nominaliser -su' (Rojas-Berscia 2019a), (ii) the sequential -watu, (iii) the simultaneous -se', and (iv) the causal -tun. This is followed by a

section on the methods used to collect the data. We then present our findings based on a multivariate analysis of our corpus. Finally, we close the article with a discussion of our results and the historically ongoing transformations in the linguistic practices of Shawi speakers in general.

THE SHAWI OF NORTHWESTERN AMAZONIA

Shawi or Chayahuita (ISO 639-3: cbt) is a Kawapanan language spoken by the swidden agriculturalist Shawi in Peru. The Shawi inhabit the areas alongside the Sillay, Cahuapanas, and Paranapura rivers, and various adjacent streams, in the provinces of Alto Amazonas and Datem del Marañón, in the region of Loreto, as well as the areas alongside the Shanusi River in the province of Lamas, in the region of San Martín (see Figure 1). By and large, Shawi is considered a vital language. Today, it is spoken by ca. 21,000 people, according to a census carried out in 2007 by the Peruvian National Statistics Institute (Instituto Nacional de Estadística e Informática 2009). The language is still transmitted to children as their mother tongue and is used in household situations as well as in everyday interactions. Today, Shawi is said to have three dialects—Paranapura, Cahuapanas, and Chayahuita—which basically refer to the three main Shawi settlements: Balsapuerto, Santa María de Cahuapanas, and the villages along the Sillay River, respectively (Rojas-Berscia 2019b). According to Barraza de García (2005), Shawi dialects display little to no morphosyntactic variation, although some phonological variation can be found. Rojas-Berscia (2019b) showed that there is indeed phonological variation in Shawi. However, more than just a diatopic trend, it was socially motivated.

Prior to the arrival of the Spaniards in the seventeenth century and the Jesuit missionary enterprise, the Shawi were neither culturally nor linguistically unified. Several vernaculars, such as Cahuapanas, Paranapura, Chayahuita, and Concho, were probably spoken in the mountains of Chayabitas, along the Andean foothills of the modern Shawi area. Only after the establishment of the first Jesuit missions, known as *reducciones*, such as Nuestra Señora de la Presentación de Chayabitas, did a new unified standard language come to be used for political and religious purposes. This is probably one of the reasons why the so-called *lengua general* (lit. 'general language'), Quechua, never attained the status of lingua franca in the Shawi-speaking area as it did in the rest of the region and was relegated to the trading sphere.

After the expulsion of the Jesuits in 1767, some *reducciones*[4] were abandoned (Fuentes 1988; Ochoa-Gilonne 2007). Others changed location but retained their social organisation. It was not until the boom of rubber, barbasco, and cow tree[5] (ca. 1870s–1940s), that complex dynamics of trade, intermarriage, and population displacement reshaped the demographics of the Shawi area (Fuentes 1988). The small Shawi villages began to lose political relevance, while larger cities, such as Moyobamba and Iquitos, which became the capital of the region of
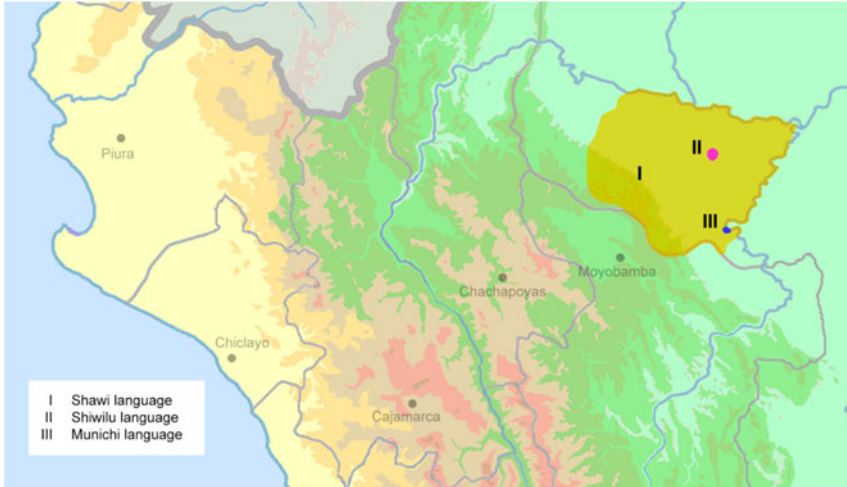
FIGURE 1. Map of the Shawi language area (adapted from https://commons.wikimedia.org/wiki/File:Regiones_naturales_del_Perú.png).

Loreto in 1897, became the new centres for political affairs and trade. The Shawi area, however, remained an important crossroads for the extraction of barbasco and cow tree. Shawi labour was also of special importance during the Amazon rubber boom, making many men leave their native villages for cities such as Iquitos.

These dynamics created a constant demographic exchange between the highlands and the lowlands, making many mestizos from the highlands of Moyobamba move to the lowlands and settle among the Shawi in the late nineteenth century and the first decades of the twentieth:

> The migration waves from the highlands (inhabitants of Moyobamba) gave rise progressively to a white-mestizo society, as a separate and dominant sphere, superimposed over the indigenous sphere, which fell back on itself and was made up of the ancient villages-reducciones of Mainas. This ethnic separation—which persists until today—carries on despite the transit of people between villages, fundamentally through biological miscegenation. That is how we explain the presence of mestizo last names in Shawi communities, preserving, however, an indigenous identity.[6] (Fuentes 1988:31)

Of the three regions included in our analysis, the sharp division between the mestizo and the Shawi society is still felt in Balsapuerto, where Shawi women and children are monolingual,[7] and Spanish last names are rarely found among the Shawi. Santa María de Cahuapanas is the opposite. There, the mestizo and the Shawi societies are nowadays almost inseparable. Many mestizo-Shawi couples have been formed since the rubber boom, and since then many children were born. Not only did they inherit the last names of their mestizo parents, but also their vernaculars. Santa María and its satellite towns are known until today for their high occurrence

of Spanish last names, as well as for their increasing number of Spanish speakers. Most Shawi children in Santa María have a good command of both Shawi and Spanish. They could be considered compound bilinguals. The communities along the margins of the Cahuapanas River are in a rapid process of westernisation, not only perceivable through language practices and last names, but also through clothing and house apparel: Shawi women no longer wear traditional clothes and the traditional Shawi plates and bowls have been replaced by cheap plastic ones.[8] The situation in Sillay is probably the only one that resembles the traditional Shawi village dynamism prior to the advent of western migrants. Sillay preserves strong Shawi traditions, and there are barely any mestizos living in the region. Sillay is, to a large extent, monolingual.

Westernisation and, therefore, Hispanicisation have been somewhat boosted by the arrival of Catholic and protestant missionaries, the start of education, and, subsequently, intercultural bilingual education. Today, the villages undergoing a rapid westernisation are those along the margins of the Paranapura and Cachiyacu rivers, such as Balsapuerto. However, the phenomenon can be observed throughout the whole Shawi area in different degrees: Shawi central villages, such as Santa María de Cahuapanas, have more schools, churches, and so on, while satellite villages, like those around Balsapuerto or along the Sillay River, sometimes only have one primary school. A similar cultural impact to that of Santa María de Cahuapanas is yet to be seen. In the meantime, high rates of monolingualism among women and children are still the rule in these regions.

Intercultural bilingual education began sometime in the late seventies with the incursion of the Summer Institute of Linguistics (SIL) missionaries. They were the first who created a phonologically based Shawi alphabet, translated the Bible into Shawi, and created the first teaching materials for the Shawi in Shawi (Yris Barraza de García, Nila Vigil Oliveros, and Virginia López p.c.). The Shawi still remember Helen Hart and George Hart's efforts in the villages of Santa María de Cahuapanas, Palmiche (Sillay), and in the surroundings of Balsapuerto in getting as many consultants as possible for the translation and crafting of the materials. After the departure of the SIL missionaries, these efforts did not stop, but were followed by experts from the CAAAP (Centro Amazónico de Antropología y Aplicación Práctica) in the eighties and early nineties (Nila Vigil Oliveros and Virginia López p.c.). Some of the materials created in this period are still preserved by some middle-aged Shawi, who proudly show to newcomers that their primary education was conducted in their original language. Even in villages such as Varadero in the Paranapura, where Shawi and Shiwilu have long been in retreat, Shawi was also taught as a second language. These efforts were consequently followed by the FORMABIAP (Programa de Formación de Maestros Bilingües de la Amazonía Peruana, lit. 'Programme for the Formation of Bilingual Teachers of Peruvian Amazonia'), who provided the first organised programme of higher education for people in the Amazon interested in teaching in their own native language (Yris Barraza de García p.c.). This programme is still running nowadays and many Shawi have benefitted from such an education. For the past

decade, the Ministry of Education has boosted the already local bilingual education programmes, launching a national intercultural bilingual education agenda, which has so far benefitted most of the Shawi speaking area. Most primary schools nowadays offer a bilingual curriculum. Some local teachers in Santa María de Cahuapanas told us that each day has its own language. Some days the teaching is carried out in Shawi. Some other days, the teaching is carried out in Spanish. They want children to be fluent in both languages. In villages like Santa María de Cahuapanas, this is not far from realistic, given the already existent compound bilingualism. In the communities around Balsapuerto, or Balsapuerto itself, this is more difficult. Children do not learn Spanish at home, and in most cases, only male children have the chance of getting in contact with Spanish-speaking foreigners. As regards Sillay, its situation deserves a little more ethnographic work to have a final say, given the difficulty of accessing the communities and the taboos associated to talking to foreigners about local practices.

EMBEDDING IN SHAWI

Before delving into embedding proper, some useful features of Shawi grammar are introduced for a better understanding of the examples.

Grammatically speaking, Shawi is predominantly a predicate-final language, that is, it displays an AOV/SV order in the clausal domain and most derivational as well as inflectional processes are done by means of suffixation, although some prefixes exist (see (3) below). It is also an agglutinative and synthetic language, with a simple phoneme inventory: nine consonants, two glides, and four vowels (a <a>, i <i>, o <u>, and ə <e>). The typology of the language resembles that of Andean languages, particularly Quechua and Aymara (Valenzuela 2015; van Gijn & Muysken 2020) although the fact of having a centroid such as /ə/, deploying Arawak-like valency-changing operators, and displaying a particular ergative-marking system (q.v. Rojas-Berscia & Bourdeau 2018) brings Shawi typologically closer to lowland Amazonian languages.

(3) (FS_Cahuapanas_MIP_160323)
   Wa'washa-re'e tuwayu-ri ichi-pa'ya-n-in,   anu-t-a-r-in.
   child-DIM   owl-ERG   SOC.CAUS-be.scared-N.FUT-3 fall-VM-PROG-N.FUT-3
   'The owl frightened the child and he (the child) is falling.'

The notion of embedding we deploy in this article is that of *grammatical embedding*, also known as subordination or hypotaxis, and which refers 'to all types of clauses occurring as subordinate parts of their superordinate clauses (which may be either main or subordinate)' (Karlsson 2010:108). As such, we can distinguish between morphological and syntactic embedding, depending on whether the embedding operators or markers are morphological or syntactic.

Grammatical embedding is present only when there is a verbal form (finite verb, infinitive, a participle, or a nominalised verb[9]). This entails that grammatical embedding comes in degrees. In some cases, we have *full-clausal embedding*, where the verb is finite. In other cases, we have *reduced embedding*, where the verb form is in either a participial or infinitival form. Unlike reduced embedding, which is easier to diagnose, full-clausal embedding can sometimes be confused with parataxis. In languages such as German or Dutch, word order and verbal mood (subjunctive) are direct indicators of full-clausal embedding. In other languages, it is trickier since there may be no clear signals of embedding. In such situations, a semantic criterion can be applied, which consists in sorting out what the sentence as a whole asserts and what the resultant meaning is if the whole sentence is placed under a higher operator such as negation. For example, a sentence like *After Doris left, he watched TV* is a case of embedding, not of parataxis, since its negation is *After Doris left, he **did not** watch TV* (Seuren 2021).

Our discussion of embedding deals with five markers, all of which contain a verbal form and, therefore, instantiate embedding.

1. *-su'* in relative clauses, which contains a finite verb form (see weak nominalization in Rojas-Berscia 2019a), as well as in purposive constructions, (*-ka*)…*su'*
2. *-watu* as a sequential marker that occurs with a reduced verb form (tenseless)
3. *-se'* as a simultaneous marker that also occurs with a reduced verb form (tenseless)
4. *-tun,* meaning 'because', that goes with a full finite verb form
5. *nitun*, also meaning 'because', that also goes with a full finite verb form

These markers are henceforth referred to as embedders. As seen, embedding in Shawi is mainly achieved through suffixation. In what follows, we illustrate the use of these markers by examples from our data.

The general nominaliser *-su'* (in bold in the examples) is mainly used in relativisation as in (4), as well as in purposive constructions as in (5).

(4) (FS_Balsapuerto_AYP_160516)
Kankan-pa' anu-te-r-in-**su'**   ni'-sa-r-in-ø.
wasp-ALL   fall-VM-N.FUT-3MIN.S-NMLZ see-PROG-N.FUT-3MIN.A-3MIN.O
'He saw the one who fell next to the wasps.'

(5) (FS_Cahuapanas_PIA_160323)
Wa'washa ni'nira-re-na   ma'su'-t-a-r-in   tururu'
child   dog-COM-3AUG   something-VM-PROG-N.FUT-3MIN.S frog

maka-**ka**-ø-ø-**su'**-mare'.
grab-PURP-3MIN.A-3MIN.O-NMLZ-BEN
'The child did something with the dog, so that he [could] grab the frog.'

The sequential marker *-watu/-atu* (in bold in the examples) is used to describe an event which happened immediately before the event described by the main verb as in (6). However, in many cases, it also portrays simultaneous action as in (7).

(6) (FS_Balsapuerto_MH_160517)
Yu   amite-mu'tuka-**watu**-n   wa'washa   pa'-sa-r-in.[10]
deer put-head-SEQ-3MIN.S   child   go-PROG-N.FUT-3MIN.S
'After putting the child on his head, the deer goes.'

(7) (FS_Balsapuerto_ER_160516)
Tururu'   se'ke-**atu**-n   wense-ra-pi.
frog   grab.hand-SEQ-3MIN.S   sit-PROG-N.FUT;3AUG.S
'The frogs are sitting while grabbing their hands.'

The simultaneous marker *-se'* (in bold in the example) describes an event that occurs at the same time as the main event described by the main verb, given in (8). This marker is progressively being replaced by the more frequent suffix *-watu*, which can also convey the same meaning.

(8) (FS_Cahuapanas_LChS_160323)
Apura-watu-n-ø   pa'-sa-**se'**-ø   kankan
do.this-SEQ-3MIN.A-3MIN.O   go-PROG-SIM-3MIN.S   wasp

kenan-kun-n-in-ø   wa'washa.
find-ANDAT-N.FUT-3MIN.A-3MIN.O   child
'After doing this, while walking, the child found the wasps.'

The causal marker *-tun* (in bold in the examples) describes an event that triggers the occurrence of the event described by the main verb as in (9). In some Southern Shawi varieties, it occurs as an independent lexeme, *nitun*, fossilised with the verb 'to be', given in (10).

(9) (FS_Cahuapanas_CPE_160323)
Ina   isu' ni'ni' nu'wi-t-a-r-in-**tun**   wa'an-ne-n-nta'
3MIN this dog   bark-VM-PROG-N.FUT-3MIN.S-because owner-ALIEN-3MIN.POSS-ADDIT

wa'an-ne-n-ri   a'pani-sa-r-in-ø.
owner-ALIEN-3MIN.POSS-ERG   answer-PROG-N.FUT-3MIN.A-3MIN.O
'The (dog) owner was scolding (the dog), because the dog was barking.'

(10) (Hart 1988:205)
Aruse nani   kaya-r-in   **nitun**,   sena-r-awe-ø.
rice   already be.ripe-N.FUT-3MIN.S because harvest-PROG-N.FUT-1AUG.EXCL.A-3MIN.O
'Since the rice is mature, we harvest it.'

METHODS

Stimuli and procedure

We worked with stimulus known as the Frog Story (Mayer 1969), a set of pictures depicting the adventures of a child and his dog looking for a rogue frog. Given the relative cultural neutrality of the settings depicted by the story, it is commonly used as a set of stimuli for elicitation of narratives. In this case, we implemented the story in the Max Planck Institute Shawi Fieldkit (Withers 2015), an Android application developed at the Max Planck Institute for Psycholinguistics specifically designed for the implementation of diverse stimuli sets. The application is used both for elicitation and data collection, resulting in files immediately compatible with ELAN software (2018), which shortens the processing time of the raw data after collection. We used the Frog Story stimulus since its use results in the same narration with a similar number of narratological units told by ninety-two participants. Although these elicited narrations cannot be said to be entirely naturalistic, their nature allows us to carry out a more controlled quantificational study.

Consultants were presented with the set of stimuli in the presence of the experimenter, which was either the first author of this article, who is a fluent speaker of Shawi, or one of the Shawi native assistants, who were trained for data collection prior to the arrival at the field sites.[11] In the latter case, this allowed us to reduce the observer paradox and get data from otherwise relatively unreachable groups for foreigners, such as women. Consultants were told that they would observe a set of pictures depicting a story, involving a child, a dog, and a frog, and that they should express in Shawi as naturally as possible what they saw on the pictures as if they were telling a story. On average, the experiment lasted four minutes. On a number of occasions, consultants refused to do the task, claiming that they did not know what to do or that they simply did not want to. In these cases, they were not included in the corpus for this study.

Consultants

Frog stories were collected from ninety-two Shawi consultants from three locations: Balsapuerto, Sillay, and Cahuapanas. In addition, we also collected some data from the High Paranapura, Chayahuita, and Shanusi, but these were not taken into account for the present study, as the number of consultants from these locations was too low for reliable statistical analysis. All consultants were paid the Peruvian equivalent of 3–5 € (euros) for participation in the task.

Collection of background data

We systematised data collection as much as possible, given the difficulties of such an enterprise in the field. We collected data from consultants from different age groups, ranging from young adults (min. fifteen years old) to elders (max.

seventy-four-years old). The native Shawi intuitions of age groups differ from those in most modern western cultures today. For example, what westerners would consider an adult (e.g. a forty-five-year-old) is considered an elder in Shawi communities. In contrast, a sixteen-year-old is considered an adult among the Shawi, but not necessarily so by westerners.

We included the following background variables in our analysis: age, gender (male/female), location (Balsapuerto, Cahuapanas, Sillay), education (primary/middle/higher), and occupation (farmer, hunter, teacher, etc.) With regard to education, most Shawi know at least how to write their names and do simple math. Consultants, whom we classified as having a primary level of education, had either only primary education (one to six years of school) or no education at all. Consultants who have some high school education or who finished high school were classified as having a middle level of education. Finally, consultants who went to university or a technical school after completing high school and completed or are in the process of completing a degree were classified as having higher level of education.

In some cases, consultants were initially reluctant to share their background information with us, especially regarding their level of education if they had not received any. This power-relation bias was in most cases overcome through a good command of Shawi, the help of our Shawi assistants, and day to day interaction. Consultants who did not know or did not want to share their age with us were not included in the data.

RESULTS

Descriptive statistics of the corpus

The corpus consists of frog stories told by ninety-two consultants from three different regions: Balsapuerto, Cahuapanas, and Sillay. Table 1 summarises the data in terms of the number subjects from each region.

As Table 1 shows, the speaker samples from the three regions are of comparable size. Moreover, the number of female and male subjects, as well as their mean age, are relatively similar across the three populations. Thus, in the context of field-linguistic studies in general, our corpus covers a relatively large and balanced population, enabling a more detailed quantitative analysis of linguistic features in the data.

There is considerable variation regarding the length of the individual stories and, thereby, their level of detail. During the transcription and glossing of the stories, we divided each story into narratologically identifiable utterances. On average, the stories consist of twenty-eight utterances (both mean and median), which, quite logically, is similar to the number of pictures in the frog story storybook (twenty-four). However, the shortest story in the corpus consists of just sixteen utterances and the longest of thirty-five utterances. In addition, the subjects vary noticeably

TABLE 1. *Summary of data.*

Region	Total # of subjects	# of females	# of males	Mean age
Balsapuerto	27	10	17	33.7
Cahuapanas	35	15	20	37.9
Sillay	30	15	15	37.9
Total	*92*	*40*	*52*	*36.7*

concerning the average length of their utterances. The mean utterance length per subject ranges from 1.6 to 7.7 words (the longest utterance in the corpus is nineteen words). The mean utterance length in the corpus as a whole is 3.7 words, while the median is 3.5 words per utterance. In total, the corpus consists of 2,569 utterances, and these contain 9,348 words. As Shawi is a highly synthetic language, the number of morphemes in the corpus is more than double the number of words: 23,319.

Embedding in the corpus

Of the 2,569 utterances in the corpus, 446 (or 17%) contain some type of embedding. The total number of tokens of embedders (i.e. suffixes with embedding function) in the corpus is 526. In other words, some of the 446 utterances with embedding contain several instances of embedders. However, the use of embedding is clearly not uniformly distributed in the corpus. Figure 2 shows the number of tokens of embedders in each story (i.e. per consultant), ordered from most to least instances of embedding.

It is evident from Figure 2 that the number of embedders per story follows an exponential distribution. Of the ninety-two subjects, fifteen do not use any form of embedding in the story, and another fifteen use it only once. On the other hand, the twenty-one subjects who use embedding the most account for 67% of all tokens of embedders in the corpus. Therefore, it is crucial to ask what makes some subjects more likely to use embedding than others, and whether this difference can be explained by macro social variables in our population. We describe the results of such an analysis using generalised linear mixed models in *Social variables as predictors of embedding* below.

In addition to acknowledging the between-subject variation, it is also important to distinguish between the different expressions of embedding in the corpus, as the individual embedders vary in frequency. This variation is clearly visible from Table 2.

The sequential -*watu* is by far the most frequent syntactic embedder in the corpus, accounting for 300 of the 526 tokens, or 57%. Hence, using the sequential is clearly the dominating embedding strategy in our corpus. However, it is possible that this pervasiveness of the sequential suffix is, at least partly, due to the nature of frog stories as linguistic data, where the stories describe a series of events one after
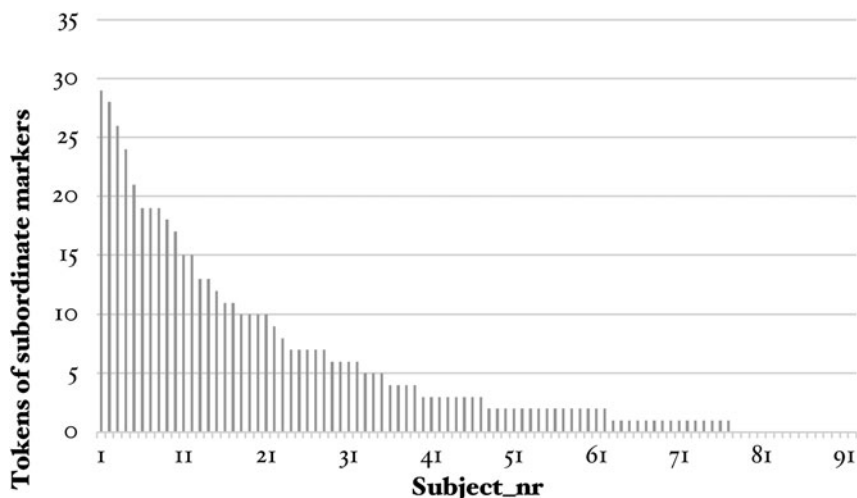
FIGURE 2. Number of tokens of subordinate markers per consultant.

the other, that is, sequentially. In other words, our corpus has certain specific characteristics (such as the sequential nature of narratives) which could promote the prevalence of certain linguistic features (such as the sequential suffix) over others. Therefore, these results do not necessarily always reflect performance in the language as a whole.

The second most frequent embedder is the nominaliser -*su'*, which accounts for 23% of the tokens of all embedders in the corpus. The causal -*tun* accounts for 13%. The purposive and the simultaneous embedders are very infrequent and account for just 4% and 3% of the total number of embedder tokens, respectively.

Social variables as predictors of embedding

In order to analyse the role of the social variables in the use of embedding in the corpus, we employed a generalised linear mixed effects model analysis. Our model predicted the use of embedding based on the social characteristics of a consultant. However, given the considerable variation in utterance lengths between the consultants in our corpus (see above), and given that linguistic elements are logically more likely to occur in longer than in shorter utterances, it is important that our analysis take this variation in utterance length into account. In addition, a sociolinguistic model needs to include the effect of any potential idiosyncratic variation between the subjects, which, again, can be partly dependent on how long the utterances are. Generalised linear mixed models allow us to control for both the effects of differing utterance lengths, as well as for idiosyncratic linguistic behaviour. Due to the individual embedders having relatively few tokens in the corpus (with,

TABLE 2. *Frequency of individual syntactic embedders in the corpus.*

Syntactic embedders	Tokens in corpus
SEQ	300 (57,03%)
NMLZ	122 (23,19%)
because	68 (12,93%)
PURP	20 (3,8%)
SIM	16 (3,04%)
Total	526

perhaps, the exception of the sequential; see above), we modelled the usage of embedders in general (without differentiating between the different expressions).

We used the *lme4* package (Bates, Mächler, Bolker, & Walker 2015) in R (R Core Team 2017) to run the models. We defined *the number of tokens of embedders in an utterance* as our dependent variable. These counts range from zero to five and follow a Poisson distribution, with most utterances containing zero instances of embedding (see above). We checked that the data does not show signs of overdispersion. Hence, we specified 'poisson' as the family for our generalised linear mixed model and 'log' as our model link function. As fixed effects, we entered the following social variables: *gender*, *age*, *region*, *education*, and *occupation* (without interaction term). As random effects, we had an intercept for *utterance length* (operationalised as the number of morphemes in an utterance[12]) as well as an intercept for *subjects*, with the latter including a random slope for the effects of utterance length. We checked that the effects do not demonstrate signs of collinearity. *P*-values for the effects of individual variables were obtained by likelihood-ratio tests where the best-fit model with the effect in question was compared against a model without that specific effect.[13]

After model comparison, the model with the best fit[14] included both of the random effects entered in the analysis—that is, utterance length, expressed in R-notation as (1|nr_of_morphemes), and by-subject variation including random slopes for effects of utterance length (1 + nr_of_morphemes|subject_id)—as well as the fixed effects of education, region, and gender. In other words, we found no significant effect of either age or occupation on the frequency of embedding. The results for the fixed and the random effects in the model with the best fit are presented in Table 3.

As expected, utterance length had the strongest effect on the number of embeddings in an utterance: the longer the utterance is, the more likely it contains a form of embedding ($\chi^2(1) = 146.1$, $p < 0.0001$). Furthermore, the idiosyncratic by-subject variation also has a significant effect on the use of embedding ($\chi^2(3) = 28.5$, $p < 0.0001$) and allowing for by-subject random slopes for the effect of utterance length further increases the explanatory power of the idiosyncratic variation ($\chi^2(2) = 10.4$, $p = 0.006$). In other words, some subjects use a form

TABLE 3. *Results of the mixed model (Significance codes: 0.05 *, 0.01 **, 0.001 ***).*

Random effects					
Groups	Name	Variance	Std. dev.		
Subject_id	Intercept	0.921391	0.95989		
	Nr_of_morphemes	0.001515	0.03892		
Nr_of_morphemes	Intercept	2.072556	1.43964		
FIXED EFFECTS					
	ESTIMATE	STD. ERROR	Z-VALUE	P-VALUE	
(Intercept)	-1.80105	0.31337	-5.747	9.07e-09 ***	
GenderWoman	-0.33304	0.14622	-2.278	0.022746 *	
RegionCahuapanas	0.39287	0.16595	2.367	0.017916 *	
RegionSillay	0.03029	0.19404	0.156	0.875943	
Education2	0.37398	0.16675	2.243	0.024915 *	
Education3	0.72892	0.18972	3.842	0.000122 ***	

of embedding more frequently than others, and this idiosyncratic linguistic behaviour manifests itself more or less clearly depending on the length of the utterances.

As far as the fixed effects go, the results of the model show that Education3— that is, higher education—has a significant effect on the use of embedding. A likelihood-ratio test confirms that the education variable adds significantly to the explanatory power of the model ($\chi^2(2) = 11.7$, $p = 0.003$). The categories used in the coding of the education variable were '1', '2', and '3', designating lowest to highest levels of education. Thus, the results from our model show that the higher the level of education, the more likely a subject is to use embedding (as demonstrated by the positive number in the Estimate column). For the sake of illustration, the positive correlation between the frequency of embedding and the level of education can be visualised in a simple plot, as in Figure 3.

As Figure 3 demonstrates, the mean number of embedders per utterance is ca. 0.1 for the subjects with the lowest level of education, 0.2 for the subjects with the middle level of education, and over 0.5 for the subjects with the highest level of education. Put differently, stories told by subjects with higher education contain over five times more embedding than stories by subjects with no or only primary education. However, it is important to emphasise that the relationship between the frequency of embedding and the level of education is, in reality, not as straightforward as Figure 3 suggests (the figure is used here only as a means of simplified illustration and comparison). This is because education also correlates positively with utterance length, which, as we demonstrated above, is the single most significant factor explaining the frequency of embedders. It is, therefore, expected that education and frequency of embedding show signs of correlation when examined independently of all other variables. The benefit of the mixed model analysis, then, is showing that education remains a significant variable above and beyond
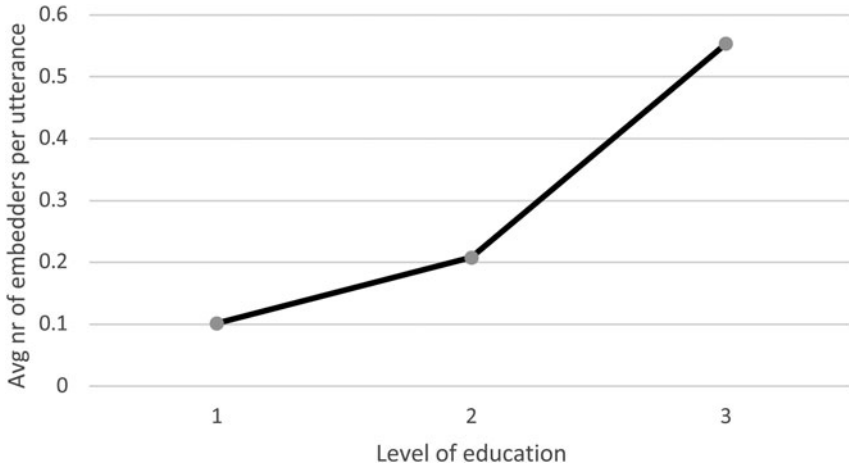
FIGURE 3. Average number of syntactic embedders per utterance according to subjects' level of education.

the main effect of utterance length (and the effects of other variables as well). We discuss the potential explanations for the effect of education on the usage patterns of embedding in the next section.

Looking once more back at our mixed effects model, the results show that gender and region also play a role in explaining the frequency of embedding in the corpus. More precisely, the results indicate that subjects in Cahuapanas use embedders more frequently than those in either Balsapuerto or Sillay do (as demonstrated by the positive Estimate value for RegionCahuapanas and the p-value 0.0179). Overall, the variable of region has a significant effect on the explanatory power of the model ($\chi^2(2) = 8.05, p = 0.018$). Likewise, our analysis demonstrates a significant effect of gender ($\chi^2(1) = 4.61, p = 0.032$). In particular, the model indicates that women use embedders less frequently than men do (who are grouped in the model under the intercept).

To summarise our findings, our analysis demonstrates that the use of embedding varies systematically with the social variables of education, region, and gender. We found no significant effect of either age or occupation. We discuss the potential causes and implications of these results in the following section.

DISCUSSION

Shawi-Spanish bilingualism in Cahuapanas

The Shawi have been in contact with the West for the past three centuries, constantly reshaping their identity without losing a sense of cultural and linguistic unity (Fuentes 1988:13). This perpetual influx of foreign groups into the Shawi

communities is not only perceivable demographically, but is evident at the linguistic level as well.

In THE SHAWI OF NORTHWESTERN AMAZONIA above, we discussed the relevance of miscegenation in the region of Cahuapanas. Mestizos from the highlands settled in the villages along the margins of the Cahuapanas River, possibly motivated by the abundance of fish poison or barbasco. The arrival of these newcomers during the last decades of the nineteenth century and the first decades of the twentieth century did not remain unnoticed. They married Shawi women, built their own houses, and brought their own customs. Even nowadays, this seems to be a common phenomenon. Adventurers from the highlands of the region of Amazonas arrive in Santa María de Cahuapanas looking for new opportunities. Today, most villages along the Cahuapanas River are bilingual. Children speak the Shawi vernacular of the region along with Amazonian Spanish.[15] As argued earlier, they fit the category of compound bilinguals.

The bicultural and, therefore, bilingual nature of Santa María de Cahuapanas, despite its great distance from major cities, makes it accessible both to Shawi and non-Shawi newcomers. Most Shawi from Santa María de Cahuapanas are brought up in Shawi and Spanish. They are acquainted with the oral transmission practices of the elders, but also Western bookkeeping, Bible translation, and writing are not foreign to them. Many of them send their children to school to be literate in Shawi and Spanish. Remarkably, some of them claim they speak better Shawi than their *hermanos*, lit. 'brothers', from Balsapuerto. For them, the *Balsachos*[16] sometimes speak 'broken Shawi'. It appears that the influence of Spanish speaking mestizos in the region introduced the notion of 'standard', creating an ideology around the 'pureness' of the Cahuapanas dialect. For Cahuapanas speakers, their dialect is closer to the 'true' Shawi, *el verdadero*, as they say.

When transcribing Cahuapanas Frog Stories with the help of our consultants from Chayahuita (Sillay) and Balsapuerto, we also encountered some difficulties. They were constantly complaining about the complexity of the texts. They said they were saying 'unnecessary things', *no dicen lo necesario.* The difficulty did not lie on the lexical and phonological differences, which certainly exist, but on the pervasive use of very long sentences that contained embedders.

These differences, possibly a result from the constant influx of Andean and Amazonian Spanish speakers since the nineteenth century, could be the reason behind the probably more complex narrations found in Cahuapanas Shawi, displayed in the form of a more frequent occurrence of embedded constructions. Interestingly, this did not lead to any sort of linguistic simplification in Cahuapanas Shawi, but complexification in terms of discourse strategies (Trudgill 2011:62), possibly due to child language acquisition in a community where both Amazonian Spanish and Cahuapanas Shawi are in an equal relationship. As embedded constructions were not borrowed into Cahuapanas Shawi, we could argue for this to be the case of a special type of system-preserving change (Aikhenvald 2006:20), where contact with Spanish would have triggered no creation of new categories

or constructions, but an enhancement an existing feature (q.v. Aikhenvald 2006:22), that is, the use of embedders. As is the case of Basque and Israeli Hebrew, where pre-existing analytic tendencies were strengthened by contact with Indo-European languages, we could argue that contact with a variety of Spanish enhanced the occurrence of embedded constructions. This would go hand in hand with language contact studies on the increasing pervasiveness or emergence of embedded constructions in discourse (Sakel & Stapert 2010).

We argue, then, that, in the case of Cahuapanas Shawi, the increasing pervasiveness of usage of embedded constructions seems to have been triggered by Spanish-Shawi contact since the nineteenth century, and later boosted by the incursion and success of western schooling programmes in the area in the second half of the twentieth century, as described in THE SHAWI OF NORTHWESTERN AMAZONIA above. Therefore, we claim that embedding in Cahuapanas is a grammatical *sociolinguistic indicator*, or a first order index (Silverstein 2003), which puts in evidence the different types of social dynamics that the region of Cahuapanas underwent from the nineteenth century onwards as opposed to the rest of the Shawi speaking area.

Embedding and education

Education remains a significant macro social variable above and beyond the main effect of utterance length. More and more, speakers from all Shawi areas have access to education beyond high school. Some speakers had the chance of leaving their communities and attending higher education in order to become teachers in the local intercultural bilingual schools of their respective communities. However, something to be taken into account is that higher education in Peru is carried out only in Spanish. As such, students are exposed to academic Spanish in classrooms and readings. They have to learn to give presentations about complex topics, write essays, and write their bachelor's theses in academic Spanish only. This means that, by the end of their university studies, they should be acquainted with the formal variety of Spanish, in which the use of embedding abounds. We surmise this not to be coincidental.

In this article, we assumed a sociolinguistic and quantitative point of view, taking the presence of embedding as our focus of study. We suggest that the variation found in the number of embedders in Shawi speakers, as with other morphosyntactic variables in other sociolinguistic studies, is used as an intensifier, that is, 'the degree of investment a speaker is making for [a] proposition,… [i]n this case, variation imbues a referential sign with additional indexical meaning' (Eckert & Labov 2017:469). Thus, speakers using more embedders are not just copying a strategy used in narrations in formal Spanish (i.e. there is no structural borrowing),[17] but are introducing themselves to the interlocutors as educated beings. From the perspective of language contact, following Matras, Shawi educated speakers, who by default possess a non-polyglossic repertoire, would be enhancing their use of embedders 'using similar contextual inferences and meaning abstraction in

the replica language [Shawi here] as are applied in the model language [Spanish]' (Matras 2011:154), with a particular communicative goal (Matras 2009, 2011:143). Moreover, the fact that educated speakers use more embedders is not just a mere number correlation. Shawi speakers in the area often refer to educated speakers such as Josué, the leader of the Palmiche community of Sillay, or Elio Yumi, an assistant of ours and Balsapuerto speaker who attended university, as examples of how to talk. Educated speakers are commonly referred to as the 'best' speakers of Shawi in their communities or the ones who *hablan bonito* 'speak nicely'. These findings go hand in hand with those by Dąbrowska (1997). We therefore suggest that the use of embedded constructions in Shawi is a linguistic *social marker* (Labov 1972)*,* or second order index.

Unfortunately, there are no studies on the Spanish of bilingual Shawi speakers at the university level. However, Virginia López, a doctoral candidate at the Universidad Nacional Mayor de San Marcos in Peru reports the overall abundance of parataxis and lack of subordination (in the form of *because*-clauses) in the written Spanish of Shawi teachers (Virginia López p.c.). Could it be the case that Cahuapanas Shawi, and not Spanish, is being reinscribed as a part of a new 'educated' register? This will be a topic worth looking at in the future, given the absence of evaluations from our consultants from Balsapuerto and Sillay as regards the 'pureness' or 'properness' of the Cahuapanas variety.

'Gendered speech'

Is embedding also a feature of gendered speech? As already mentioned in THE SHAWI OF NORTHWESTERN AMAZONIA above, Shawi communities have defined gender roles, with great inequality between the genders (Dradi 1987). Shawi women are commonly constrained to the household sphere. They only go where their parents/husbands go. They attend primary schools to learn how to read and write, sometimes helped by nuns or protestant missionaries who help them convince their reluctant parents of the importance of schooling for women. Most of the time, they remain monolingual. In contrast, Shawi men are supported to pursue becoming good merchants, hunters, and navigators. They are always sent to school and become bilingual from a very young age. Young boys are commonly taken along with their fathers on hunting or trading excursions. During the latter, they have their first contact with Amazonian Spanish.[18] We interpret the results of our variation analysis signalling that these social differences between women and men are indirectly manifested in the linguistic practices of the Shawi speakers. As such, this is not a case of gendered speech. Men, having by default more access to bilingualism and education, would tend to be more used to Spanish linguistic practices. Therefore, when narrating a story, it is more common for them to use embedding. The default difficulty or prohibition of access to schooling and bilingualism for women would explain the less common occurrence of embedders in their narrations. Gender by itself is not the direct cause of this linguistic practice. However, the observer paradox must not be excluded.

Given that the first author is a male, and therefore socially empowered in terms of the Shawi gender stratification, it could be the case that women feel less comfortable when narrating a story in front of a man. This could also be the case even when our Shawi assistants carry out the semi-structured interview. Studies carried out by women will confirm or reject our first observations.

Briefly, given our results, we argue that bilingualism (Amazonian Spanish-Shawi) and education, indirectly restricted by complex gender differences in the communities, play a significant role in the establishment of linguistic preferences in narration. In addition, we argue that the use of embedding reflects the impact of the Spanish-speaking mestizo society from the nineteenth century until today in the Shawi area, reshaping not only the demographics of the Shawi communities, but also their linguistic practices.

FINAL REMARKS AND FUTURE AVENUES OF RESEARCH

This is the first time a systematic quantitative study on syntactic variation involving a number of social parameters has been carried out in Peruvian northwestern Amazonia. We have demonstrated that the frequent use of embedding in Shawi is or was socially motivated. In particular, this study allowed us to understand the impact that the influx of western mestizo society has had since the nineteenth century in a constantly reshaping culture such as the Shawi.

(i) We found that education—the access to both schooling and academic Spanish—is a significant variable beyond the main effect of utterance length. Embedding strategies in Shawi narrations tend to be more frequent the higher the education level of the consultant (in line with Dąbrowska 1997). This is indirectly related to gender, since the ones who typically have access to education are men. We interpreted this finding as a result of a linguistic practice common to the educated Shawi, resorting to Spanish-associated ways of narrating to accomplish a particular communicative goal while indexing educational status (Matras 2009, 2011).

(ii) We also found that the frequency of use of embedders in narration in Cahuapanas was significantly different from other regions. In our study, people from Cahuapanas used embedders more often than people from Balsapuerto and Sillay. We hypothesised that this constitutes a *sociolinguistic indicator* or first order index, reflecting the bilingual practices of Cahuapanas that go all the way back to the first migrations of people from the highlands of Moyobamba to Cahuapanas territory from the late nineteenth century until the mid-twentieth century. From a contact linguistics perspective, this would not have entailed any structural borrowing, that is, the Shawi system would have been preserved. However, following Aikhenvald (2006:22) contact with Spanish would have enhanced the frequency and productivity of some of these embedders, as embedding itself is a shared feature of both Spanish and Shawi.

However, several questions remained open:

1. Is the pervasiveness of embedders found in our corpus directly related to Shawi-Spanish bilingualism, or is it a product of contact with Cahuapanas speakers, them being the high-prestige Shawi speakers?
2. What is the frequency of use of embedders in the Amazonian Spanish spoken by the Shawi in all regions?
3. What is the frequency of use of embedders in the Quechua still spoken in some islets of the Shawi speaking area by elders? Could missionary Quechua have also played a role in the incursion of embedding in narrations given its frequent use in the Shawi zone as the language in which Catholic missionaries preached?

Regarding question 1, contact between Shawi varieties is a topic that needs to be studied from scratch in the future. The current dynamics of socialisation in the universities and institutes of the cities of Yurimaguas and San Lorenzo would need to be explored in more detail, especially in the context of the programmes for teaching Shawi in intercultural bilingual education schools, where speakers from different regions of the Shawi speaking area get to know each other and interact. In addition, careful ethnographic research that studies more textual styles than the one and only presented in this study will help answer these questions in the future.

On a general level, this study contributes to the discussion of the role of social motivations in the pervasiveness of embedding in language. Although we do not claim that embedding in Shawi is a product of westernisation, Spanish-Shawi bilingual education has certainly been a contributing factor.

From a methodological perspective, this study has demonstrated the value of Android or IOS applications such as the MPI Field Kit, which clearly boosts data collection in ways that were not conceivable until recently in Amazonian linguistics. Today, it is possible to collect a systematic corpus with large populations in a relatively short period of time. We hope that this first study inspires more fieldwork-oriented linguists to conduct similar experiments. Collection of large corpora enables statistical testing of many of our observations in the field and, therefore, is an invaluable method for the analysis of linguistic phenomena beyond the lens of traditional elicitation.

APPENDIX: LIST OF ABBREVIATIONS

3	third person
A	subject of transitive
ADDIT	additive
ALIEN	alienable
ALL	allative case
ANDAT	andative
AUG	augmented

BEN	benefactive case
COM	commitative
COMP	comparative
COP	copula
DIM	diminutive
ERG	ergative case
EXCL	exclusive
MIN	minimal
N.FUT	non-future
NMLZ	nominaliser
O	object of transitive
OFF	'offspring' marker
POSS	possessive
PROG	progressive
PURP	purposive
S	subject of intransitive
SEQ	sequential
SIM	simultaneous
SOC.CAUS	sociative causative
VM	valency modifier

NOTES

[1]According to the Oxford English Dictionary, a mestizo, is '[a] person of mixed European (esp. Spanish or Portuguese) and non-European parentage; spec. (originally) a man with a Spanish father and an American Indian mother; (later) a person of mixed American Spanish and American Indian descent' (see *OED Online*, http://www.oed.com/view/Entry/117138). In this study, we rather refer to a mestizo as a member of the dominant Peruvian society.

[2]Previous studies within usage-based linguistics had already made a similar claim to the study by Everett (2005). See Dąbrowska (1997) for an experimental approach to perception of complex constructions from the generative literature by groups of speakers from diverse educational backgrounds.

[3]Q.v. Levinson (2013), for a detailed discussion on the preponderance of different levels of recursion in discourse. Apparently, recursion is far more prominent in interaction than in syntax alone.

[4]The so-called *reducciones* were a type of settlement established by the Jesuits in the region in the seventeenth century. These included people from different ethnic and linguistic background and were envisaged to improve the Christian Catholic missionary enterprise.

[5]Barbasco is the common name given to *Paullinia pinnata*, a plant used for fishing due to its toxic components that narcotise fish. The cow tree is famous for its milky liquid that is used in pottery, the manufacturing of canoes, and plastics.

[6]Translation from the Spanish original by the first author: 'Las oleadas migratorias provenientes de la selva alta (moyobambinos) van a dar origen progresivamente a una sociedad "blanco-mestiza" como una esfera separada y dominante, superpuesta a una esfera "indígena", replegada en sí misma y

conformada por los antiguos pueblos-reducciones de Mainas. Esta separación étnica—que persiste hasta la actualidad—se produce a pesar e independientemente del tránsito de personas entre una y otra, fundamentalmente a través del mestizaje biológico. Así se explica la presencia de apellidos de origen mestizo en las comunidades chayahuita, manteniéndose sin embargo la identidad nativa'. (Fuentes 1988:31)

[7]As opposed to Balsapuerto men, who are all bilingual and travel frequently to the trade markets in the city of Yurimaguas.

[8]The situation was such that Yris Barraza de García, during her fieldwork stays in Santa María de Cahuapanas in the late nineties, thought that Shawi would eventually vanish from these communities (Barraza de García p.c.). This never happened. Shawi-Spanish bilingualism in Cahuapanas seems to be the status quo.

[9]See that nominalisations 'serve as complements of abstract matrix clause verbs' (Gildea 2008:57), hence embedding.

[10]This is an example worth looking at to see whether we are facing parataxis or hypotaxis. If we place (6) under a higher operator of negation, the sentence would be *Yu amitemu'tukawatun wa'washa ku pa'-sarinwe* 'After placing the child on his head, the deer did not go'.

[11]The data collection would not have been as fruitful as it was without the help of our Shawi assistants Elio Yumi Pizango (Balsapuerto), Víctor Pizango Púa (Balsapuerto), Moisés Pinedo Escobedo (Cahuapanas), Abimael Huiñapi Tangoa (Pueblo Chayahuita), and Segundo Pinedo Escobedo (Cahuapanas).

[12]This operationalisation was shown to be the best (as compared to the number of words or log-transformed values of either the number of morphemes or words) in terms of linearity of distribution as well as in terms of explanational power in the mixed models analyses.

[13]Because of the large number of zero counts in our dependent variable, we also ran a zero-inflated generalised linear mixed model with the same variables. The results of the zero-inflated model were very similar to the results of the Poisson model, and given that zero-inflated models are more difficult to interpret, we present here only the results of the Poisson model.

[14]Based on the AIC-values, log-likelihood, and deviance.

[15]Some villages, like Barranquita, seem to have lost their vernacular, and have become Spanish monolingual villages. This seems to be a trend the closer one gets to the city of San Lorenzo del Datem del Marañón.

[16]A pejorative way of referring to the inhabitants of Balsapuerto.

[17]Something to be noted is that Shawi speakers are not calquing any morphosyntactic structure from Spanish. All embedders are originally Shawi. As such, there is no slight structural borrowing (q.v. Thomason & Kaufman 1988:78), whereby we would assume that the embedded constructions have entered into the repertoire of Shawi speakers through borrowing. The frequency of their use as native Shawi elements is what is in play in this study.

[18]Here we are describing the situation most common to Shawi communities along the Sillay, Paranapura, and Cachiyacu. In Cahuapanas, bilingualism is far more pervasive.

REFERENCES

Aikhenvald, Alexandra Y. (2006). Grammars in contact: A cross-linguistics perspective. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Grammars in contact: A cross-linguistics typology*, 1–66. Oxford: Oxford University Press.

Barraza de García, Yris (2005). *El sistema verbal en la lengua Shawi*. Recife: Universidade Federal de Pernambuco PhD thesis.

Bates, Douglas; Martin Mächler; Ben Bolker; & Steve Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). Online: https://doi.org/10.18637/jss.v067.i01.

Dąbrowska, Ewa (1997). The LAD goes to school: A cautionary tale for nativists. *Linguistics* 35:735–66.

Dradi, María Pía (1987). *La mujer chayahuita: ¿un destino de marginación?; análisis de la condición femenina en una sociedad indígena de la Amazonía*. Lima: Instituto Nacional de Planificación.

Eckert, Penelope, & William Labov (2017). Phonetics, phonology and social meaning. *Journal of Socio-linguistics* 21(4):467–96. Online: https://doi.org/10.1111/josl.12244.

ELAN (2018). ELAN (version 5.2) [computer software]. Max Planck Institute for Psycholinguistics. Online: https://tla.mpi.nl/tools/tla-tools/elan/.

Everett, Daniel L. (2005). Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology* 46(4):621–46. Online: https://doi.org/10.1086/431525.

Fuentes, Aldo (1988). *Porque las piedras no mueren*. Lima: Centro Amazónico de Antropología y Aplicación Práctica.

Gildea, Spike (2008). Explaining similarities between main clauses and nominalized phrases. *Amerindia* 32:57–75.

Hart, Helen (1988). *Diccionario Chayahuita-Castellano Castellano Chayahuita: Canponanquë nisha nisha nonacaso'*. Lima: SIL International.

Karlsson, Fred (2010). Syntactic recursion and iteration. In van der Hulst, 43–68. Online: https://doi.org/10.1515/9783110219258.43.

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

Laury, Ritva, & Tsuyoshi Ono (2010). Recursion in conversation: What speakers of Finnish and Japanese know how to do. In van der Hulst, 69–92. Online: https://doi.org/10.1515/9783110219258.69.

Levinson, Stephen C. (2013). Recursion in pragmatics. *Language* 89(1):149–62. Online: https://doi.org/10.1353/lan.2013.0005.

Matras, Yaron (2009). *Language contact*. Cambridge: Cambridge University Press.

——— (2011). Explaining convergence and the formation of linguistic areas. In Osamu Hieda, Christa König, & Hiroshi Nakagawa (eds.), *Geographical typology and linguistic areas: With special reference to Africa*, 143–60. Amsterdam: John Benjamins.

Mayer, Mercer (1969). *Frog, where are you?* New York: Dial Books for Young Readers.

Mithun, Marianne (2010). The fluidity of recursion and its implications. In van der Hulst, 17–42. Online: https://doi.org/10.1515/9783110219258.17.

Ochoa-Gilonne, Nancy (2007). *Entre plusieurs mondes: Les Chayahuita de l'Amazonie*. Paris: EHESS PhD thesis.

O'Connor, Loretta, & Pieter Muysken (eds.) (2014). *The native languages of South America: Origins, development, typology*. Cambridge: Cambridge University Press.

R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Online: https://www.R-project.org/.

Rojas-Berscia, Luis Miguel (2019a). Nominalization in Shawi (Chayahuita). In Roberto Zariquiey, David W. Fleck, & Masayoshi Shibatani (eds.), *Nominalization in the languages of the Americas*, 491–514. Amsterdam: John Benjamins.

Rojas-Berscia, Luis Miguel (2019b). *From Kawapanan to Shawi: Topics in language variation and change*. Nijmegen: Radboud Universiteit Nijmegen PhD dissertation.

Rojas-Berscia, Luis Miguel, & Corentin Bourdeau (2018). 'Optional' or syntactic ergativity in Shawi: Distribution and possible origins. *Linguistic Discovery* 15(1):50–65.

Sakel, Jeanette, & Eugenie Stapert (2010). Pirahã – in need of recursive syntax? In van der Hulst, 1–16. Online: https://doi.org/10.1515/9783110219258.1.

Seuren, Pieter A. M. (2021). Embedding. In Luis Miguel Rojas-Berscia & Pieter A. M. Seuren (eds.), *Case studies in semantic syntax*. Nijmegen: Max Planck Institute for Psycholinguistics, ms.

Silverstein, Michael (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication* 23(3–4):193–229. Online: https://doi.org/10.1016/S0271-5309(03)00013-2.

Thomason, Sarah Grey, & Terrence Kaufman (1988). *Language contact, creolization and genetic linguistics*. Amsterdam: John Benjamins.

Trudgill, Peter (2011). *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.

Valenzuela, Pilar M. (2015). ¿Qué tan 'amazónicas' son las lenguas kawapana? Contacto con las lenguas centro-andinas y elementos para un área lingüística intermedia. *Lexis* 39(1):5–56.

van der Hulst, Harry (ed.) (2010). *Recursion and human language*. New York: De Gruyter Mouton.

van Gijn, Rik, & Pieter Muysken (2020). Highland-lowland language relations. In Adrian J. Pearce, David G. Beresford-Jones, & Paul Heggarty (eds.), *Rethinking the Andes-Amazonia divide: A cross-disciplinary exploration*, 178–210. London: UCL Press.

Withers, P. (2015). *Fieldkit* (version 1.0) [English; Android]. Max Planck Institute for Psycholinguistics.

Address for correspondence:
Luis Miguel Rojas-Berscia
Gordon Greenwood Building (32)
The University of Queensland
St Lucia QLD 4072, Australia
lmrojasb@pucp.edu.pe