

Do languages spoken in multilingual communities converge? A case study of reflexivity marking in Mano and Kpelle

Data collection procedure and guidelines to the database

2024-01-12

Contents

1	Study design	1
1.1	Background data	1
1.2	Speaker sampling	2
1.3	Pictures	2
1.4	Coding of results	9
2	Database	11
3	Speaker sample by residence place, adoptive family, age, and gender	12
4	Modeling	15
5	Syntax by residence place	24
6	Coordinate construction	25
7	Package versions	27

1 Study design

1.1 Background data

Sociolinguistic background data was collected using a background questionnaire. The questionnaire includes questions about the father's and the mother's age; birthplace, current and former residence place; languages spoken; parents' parents' ethnic identification. The ages of all the children were also provided. In case the family has fostered children, the village of birth of the child, her parents' ethnic identification and the length of the time she has lived in the hosting family was also asked. Some background information came not from the questionnaire, but from the respondents themselves (thus, husbands tend to give an incorrect age of their spouses and children give incorrect age information about themselves; sometimes the local team member had to resort to approximation).

The following background information appears in the database. First, we enter the village of birth ("birth_place") which is coded (socio_birthplace) as being either Mano, Mano-dominant, Bilingual, Kpelle-dominant or Kpelle. No census data is available to support these classifications which are made instead according to the estimations of our local team member. Mano and Mano-dominant were regrouped during the analysis as "Mano" and Kpelle and Kpelle-dominant were regrouped as "Kpelle". Similar information about residence is put in "residence_place" and "socio_residenceplace". For the sake of anonymity, published materials include only the codes, but not actual villages. The residence history of some adults is complex. For example, for ZOW_10_M_45_1978_1 we listed all the places where he spent extended periods of time in his childhood and youth. These villages, all listed in "birth_place", include two Kpelle villages where his parents are from, a bilingual village where his maternal grandmother is from and a bilingual town where he studied. As an average, we put "Kpelle-dominant"

into “socio_birth_place”. We also include the place of current residence (“residence_place”). The place of elicitation is put in “elicitation_place”.

“Gender” is straightforward. “Age” is also straightforward (but often approximative). Because we collect data by households, “household” includes information about that, it is the code of the household’s male head.

The individuals are coded the following way: the code of the village, followed by the code of the male head of the household, followed by gender, age, date of birth and a numerical index, for example: BAN_19_M_32_1990_5.

Next, we include information about biological mother and father, the language and ethnic group they are identified with. We put “Mano” in “socio_family_biological” if both biological parents are identified as Mano, “Kpelle” if they are identified as Kpelle, and “Bilingual” if one of them is identified as Mano and another, as Kpelle. We also include similar information about adoptive family in “socio_family_adoptive” (we copy the information about biological family if the person does not have a history of fostering). The children’s family environment is the one of their current family of residence (adoptive or biological) and for adults, their family where they grew up. Because “family” reports information about a family where a person grew (or is growing) up, children and adults within the same household may not have the same value for family. For adult participants, we also put information about their spouse, Kpelle or Mano (or other), in “socio_marriage”. Finally, column “socio_varia” is reserved for relevant information which cannot be coded in other categories, especially information about fostered children, such as the relationship of the fostered child to the male head of the household. If an adult has lived only a short period of time in an adoptive family, we put this information in “socio_varia”, rather than in “socio_family_adoptive”.

1.2 Speaker sampling

Overall we analyzed 112 speakers (see the details in the following table). This number does not include one individual who was excluded from the analysis because they had not completed the entire questionnaire.

	Bilingual family	Kpelle family	Mano family	other family
Bilingual residence	8	13	13	0
Kpelle residence	20	5	14	1
Mano residence	0	0	38	0





1.3 Pictures

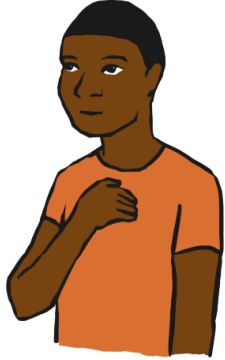

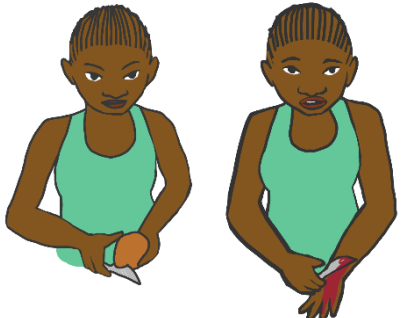
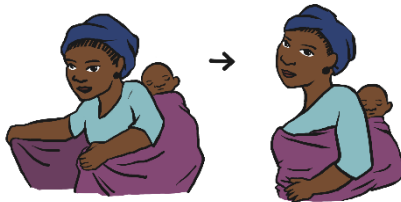
The pictures included in the questionnaire can be described the following way:



- Target images

Table 2. Target images used in the main study

Picture ID	Picture and tentative picture description	Expected position of the target form	Coding
COORD	[Researcher’s name] and her/his mother	Coordinate construction, possessive	COORD Poss

Picture ID	Picture and tentative picture description	Expected position of the target form	Coding
laver se	A woman is washing herself 	Direct object	DO
brosser dents se	A woman is brushing her teeth 	Possessor of the direct object	DO Poss
mettre couteau se	A woman is putting a knife next to herself 	Postpositional phrase	PP
mettre bol se	A woman is putting a bowl on her head 	possessor in postpositional phrase	PP




Picture ID	Picture and tentative picture description	Expected position of the target form	Coding
toucher se	A man is touching himself	Postpositional phrase	PP
			
regarder se	A woman is seeing herself in the mirror	Direct object	DO
			
blesses se	A woman cut herself	Direct object, direct object's possessor	DO, DO Poss
			
mettre enfant se	A woman is attaching a child to her back	Postpositional phrase	PP
			



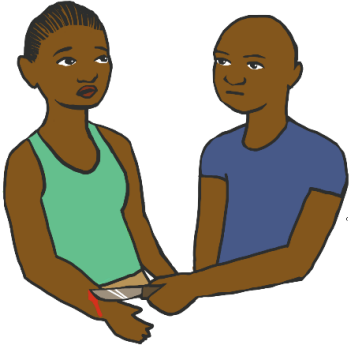
Picture ID	Picture and tentative picture description	Expected position of the target form	Coding
montrer maitresse	A hairdresser is showing a woman to herself in the mirror 	Postpositional phrase, coreference to the Direct object	PP to DO
montrer peintre	A painter is showing a woman's portrait to her 	Postpositional phrase, coreference to the Direct object's possessor	PP to DO Poss

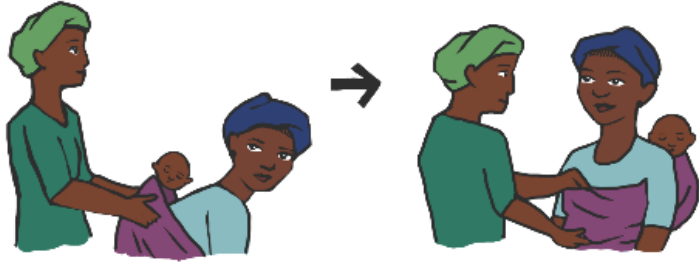

· Filler images

Table 3. Filler images used in the main study

Picture ID	Image	Picture description
laver se pl		Children are washing

Picture ID	Image	Picture description
brosser dents se pl	 An illustration of three children of African descent brushing their teeth. On the left, a girl in a red shirt. In the middle, a boy in a blue shirt. On the right, a girl in a purple shirt. They are all holding toothbrushes and looking towards the camera.	Children are brushing their teeth
laver	 An illustration of a woman in a white top and pink skirt washing a child's hair. The child is standing with their back to the viewer. The woman is using a green bucket of water. There is a pink bar of soap on the floor next to the bucket.	A woman is washing a child
brosser dents	 An illustration of a woman in a green shirt brushing a young boy's teeth. The boy is wearing a blue shirt and is smiling. The woman is holding a toothbrush to his teeth.	A woman is brushing a child's teeth

Picture ID	Image	Picture description
toucher	 An illustration of a woman with dark skin and braided hair, wearing a blue tank top, touching the face of a man with dark skin and short hair, wearing an orange t-shirt.	A woman is touching a man
regarder	 An illustration showing a woman with dark skin and red hair, wearing a purple shirt, looking into a circular mirror. To her right is another woman with dark skin and braided hair, wearing a blue shirt, shown in profile.	A woman is seeing another woman in the mirror
blesser	 An illustration of a woman with dark skin and braided hair, wearing a green tank top, and a man with dark skin and a shaved head, wearing a blue t-shirt, cutting her hair with scissors.	A man cutting a woman

Picture ID	Image	Picture description
mettre enfant		A woman is attaching a child to another woman's back
mettre couteau		A woman is putting a knife next to another woman
mettre bol		A woman is putting a bowl on a child's head

The first picture given was the one of the researcher with her/his mother. The order of the rest of the pictures was fixed following the order given by a randomizer <https://www.random.org/lists/>. The order was changed where additional fillers were added (see below).

1.4 Coding of results

The experiment was conducted in Mano by our Mano-speaking local team member. Each target image was expected to prompt a description where a pronoun, typically reflexive or basic, refers to a NP within the same clause. Such configuration we qualify as a coreferential configuration and put “SELF” in the column “orientation”. The target marker can be a simplex reflexive pronoun (*ē*), complex reflexive marker with an intensifier (*ē diè*), basic pronoun (*à*) or basic pronoun with an intensifier (*à diè*); in rare cases, the reflexive marker with a possessive intensifier was used (*ē zì*), a 3PL basic pronoun (*ō*) or a zero (zero), these labels are put in the “target marker” column. The column “analysis” classifies the target markers into types. Whenever the reflexive pronoun is used, we use a shortcut label “refl”, whenever the basic pronoun is used we use the label “basic”, for 3PL basic pronoun we use “3pl” and for zero we use “zero”.

In column “synt_position” we code the syntactic position where the target marker occurred. The coding principles are the following: whenever the target marker occurs in the direct object or possessor within direct object position, it is coded respectively DO and DO Poss. The following examples illustrate these two positions:

(1a) *lē ē jèē gè-pèlè gààzù yí*
 3SG.EXI 3SG.REFL face see-INF mirror in
 ‘She is looking at her face in the mirror.’
 Target marker: *ē*; syntactic position: DO Poss. [BAN_27_F_50_1972_1, regarder_se]

(1b) *lē ē diè jèē gè-pèlè gààzù yí*
 3SG.EXI 3SG.REFL INT face see-INF mirror in
 ‘She is looking at her face in the mirror.’
 Target marker: *ē diè*; syntactic position: [DO BAN_15_F_7_2015_1, regarder_se]

Coding in postpositional phrases is not straightforward because there is no clear-cut distinction between postpositions and locative nouns with arguments, in which case the target marker occurs not in the argument of postposition position, but rather possessor of the locative noun position. On the surface and in syntactic behavior, however, at least as far as this questionnaire data is concerned, there is no difference between the two interpretations. We qualified as postpositions any elements which take the target markers directly as their arguments, including “true” postpositions, such as *sónó* ‘next to’ or *lèē* ‘to’ and spatial and bodypart nouns in the locative function, such as *dīā* ‘back’ or *ɲwíí* ‘top’. We qualified as constructions with a postposition and its possessor any sequences with two elements, such as *zò là* ‘heart on’. The following examples illustrate these two position, PP and PP Poss:

(2a) *mí tóō bēē āā pōō yà ē ɲwíí*
 person:H DEM.PROX.VIS also 3SG.PRF thing:PL sit 3SG.REFL head
 ‘This person has put things on her head.’
 Target marker: *ē*, syntactic position: [PP GOD_22_M_4_2018_4, mettre bol_se]

(2b) *āā pōō yà ē diè ɲwíí*
 3SG.PRF thing:PL sit 3SG.REFL INT head
 ‘She has put things on her head.’
 Target marker: *ē diè*, syntactic position: [PP GOD_31_M_64_1958_1, mettre bol_se]

(2c) *āā pōō yà à ɲwíí*
 3SG.PRF thing:PL sit 3SG head
 ‘She has put things on her head.’
 Target marker: *à*, syntactic position: [PP BAN_13_F_49_1973_1, mettre bol_se]

(3a) *āā ē kò kpó à zò mò*
 3SG.PRF 3SG.REFL hand put 3SG heart on
 ‘He put his hand on his heart.’
 Target marker: *à*, syntactic position: [PP Poss BAN_13_F_49_1973_1, toucher_se]

(3b) *āā ē kò kpó ē zò là*
 3SG.PRF 3SG.REFL hand put 3SG.REFL heart on
 ‘He put his hand on his heart.’
 Target marker: *ē*, syntactic position: [PP Poss BAN_13_M_8_2014_1, toucher_se]

Finally, we coded separately all instances where the antecedent is found outside the subject position, such as in the DO, DO Poss or Subj Poss positions. In all such cases, the target is in the PP or PP Poss position and the position of the antecedent is specified. For example, for the example (4), where the target marker is found in the PP and the antecedent is in the DO position, coding is “PP to DO”:

- (4) *ō wáà à diè zḡḡ à lèē gààzù yí*
 3PL.EXI 3PL.JNT 3SG INT show:JNT 3SG to mirror in
 ‘They are showing her to herself.’
 Target marker: à, syntactic position: [PP to DO BAN_6_F_30_1992_1 montrer_peintre]

The construction used and the position of the target marker may be different from the one intended by the picture. For example, the picture *montrer_maitresse* is designed to elicit a PP to DO. A typical response however is “She is looking at herself in the mirror” and is coded as DO. The following description is completely off target, but it still contains a coreferential configuration (DO Poss) and is coded as such.

- (5) *lē ē lū gē-pèlè gààzù yí*
 3SG.EXI 3SG.REFL elder.sibling see-INF mirror in
 ‘She is seeing her elder sister in the mirror.’ [BAN_9_M_6_2016_3]

If a participant’s description of the picture did not contain a coreferential configuration, she was prompted to use a desired construction and sometimes an additional explanation of the picture was provided.

- (6a) *pḡḡ lāā kèlè āā yà?*
 thing.PL 3SG.EXI>3SG hand 3SG.PRF>3SG put
 Interviewer: ‘She has things, she put them...?’

- (6b) *ē ḡwíí*
 3SG.POSS head
 Interviewee: ‘On her head.’ [BAN_19_M_32_1990_5, mettre_bol_se]

However, the expectations are not always met, and the participant may choose a construction with no coreference relation. In this case, if a pronoun is used without antecedent in the same clause, as in ex. 7, we put “OTHER” in the column “orientation”. If there is no pronoun at all but a lexical expression, the construction is coded as “other construction”. Both cases are not included in the counts.

- (7) *lē gbāā yàá à ḡwíí*
 3SG.EXI now put_on.GER.WITH 3SG head
 ‘It is now put on her head’ [GBA_14_M_43_1980_2, mettre_bol_se]

If the same response contains more than one instance of the coreferential configuration, or if a participant gives more than one non-identical response, we list all of them and code separately. There are two types of filler items. The first two items, *laver_se_pl* and *brosser_dents_se_pl*, represent situations with three participants, not one (a group of children washing and a group of children brushing their teeth), so a plural pronoun is expected. The latter eight items represent situations similar to some of those the target items, but with a modification: it is an other-oriented action, rather than self-oriented. For example, picture *laver* represents a woman washing a child, while the picture *laver_se* represents a woman washing herself. Responses to filler items are not included in the counts even if they contain target items.

The experimental procedure was not set in place once and for all, and there was a period of trial and error.

- Picture with id *montrer_peintre* was added only in April 2022, so 11 out of 35 respondents from the bilingual village were not presented with the picture.
- Filler items with other-oriented actions were also added later in the process. As a result, only one family out of nine received a full toolkit. However, the choice between a reduced or full toolkit is unlikely to have influenced the choice of the reflexive vs basic pronoun to describe self-oriented actions. Indeed, we have tested two families from the Kpelle villages with two years apart, first, as part of the trial, with a reduced toolkit which did not include images depicting other-oriented actions, and then a full toolkit. They had approximately the same rate of reflexive markers (some exactly the same, some slightly lower, some slightly higher). Therefore, the usage of a full vs reduced toolkit does not influence the results.

Some issues during the experimental procedure must be flagged. Sometimes the target form was provided before the respondent had time to give his response. It could be provided by the interviewee in the prompting process, although he generally avoided it, or, more often, by another person present at the experimental session. Although the interviewer tried his best to make sure there was only the respondent present, sometimes it was not possible.

(8a) *léé wāā lē né yà-pèlè ē dīā āà yà?*
 woman:H DEM 3SG.EXI child sit-INF 3SG.REFL back 3SG.PRF>3SG sit

(Interviewer:) ‘The woman is putting the child on her back, she is putting it?’

(8b) *ē dīā*
 3SG.REFL back

(Interviewee:) ‘On her back.’ [GOD_22_M_58_1964_3, mettre_enfant_se]

We have documented all such cases of priming but do not signal them in the published database. The effects of priming have not been studied.

In certain cases the respondents did not understand the picture the intended way. For example, the picture `mettre_bol_se` could be understood as involving another person who put the bowl on the woman’s head:

(9) *āā pōō yà ē lūú ŋwí*
 3SG.JNT thing.PL put_on 3SG.REFL daughter head
 ‘She put the objects on her daughter’s head.’ [GOD_31_F_6_2016_1, mettre_bol_se]

We count these items in the database because they still illustrate the use of reflexive construction, even if the image was read in a different way.

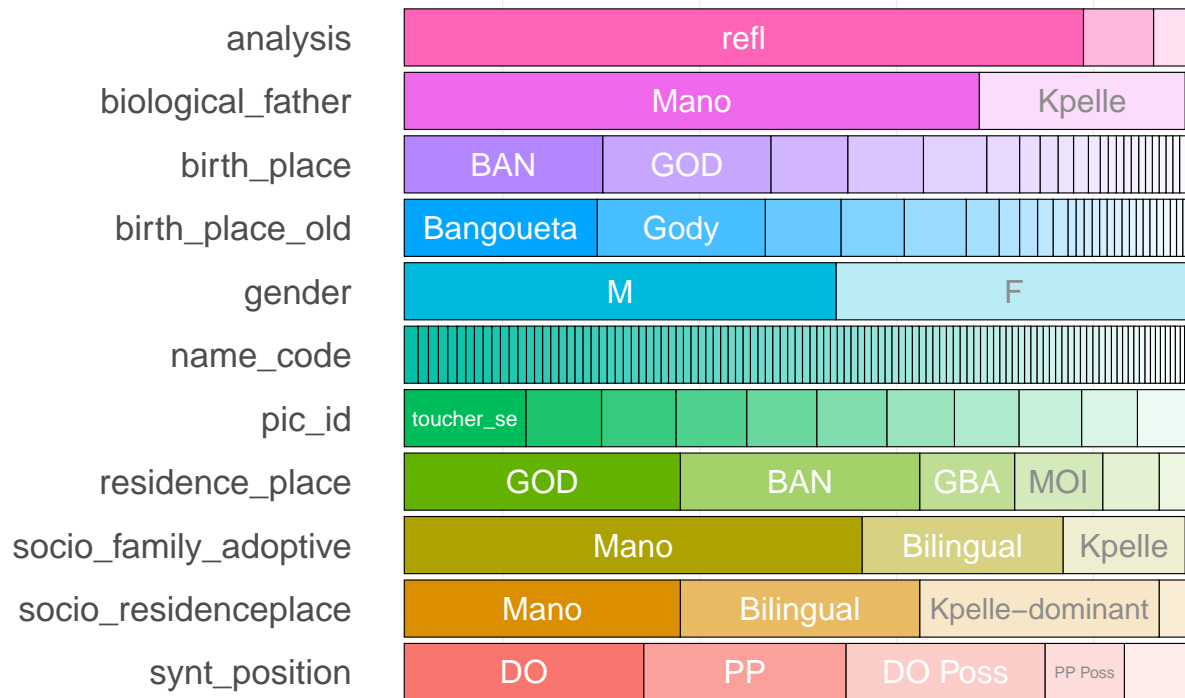
2 Database

The database (1270 observations) contains the following variables:

- `name_code` — code for the informant;
- `birth_place` — code for the birthplace;
- `residence_place` — code for the residence place;
- `socio_residenceplace` — annotations of the residence place, possible values are Mano, Mano-dominant, Bilingual, Kpelle-dominant or Kpelle;
- `socio_family_adoptive` — Mano, Kpelle or Maninka;
- `biological_father` — self identification of the father Mano, Kpelle or Maninka;
- `age`;
- `household` — code for the household;
- `analysis` — see the Coding results section;
- `synt_position` — syntactic position, see the Coding results section;
- `pic_id` — identifier for stimuli pictures.

Figure 1. Distribution of categorical variables

```
library(inspectdf)
database |>
  select(-age) |>
  inspect_cat() |>
  show_plot() +
  labs(title = "", subtitle = "")
```



On the Figure 1 above we can see the distribution of categorical variables in our dataset.

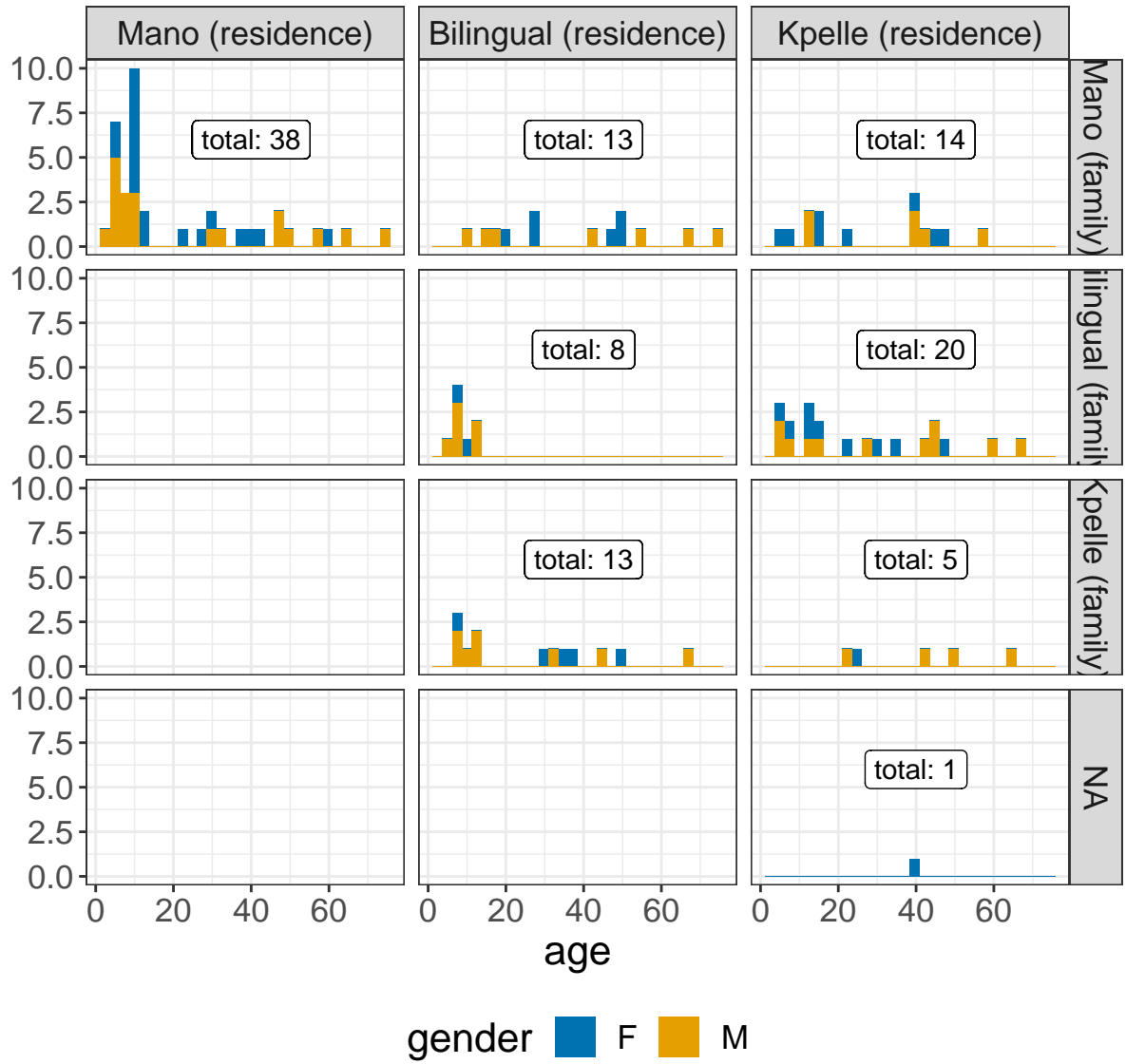
3 Speaker sample by residence place, adoptive family, age, and gender

Here we can see the code and the result of visualization of distribution of people by residence place, adoptive family, age and gender:

Figure 2. Speaker sample by residence place, adoptive family, age, and gender

```
for_visualization |>
  mutate(socio_family_adoptive = str_c(socio_family_adoptive, " (family)"),
         socio_residenceplace = str_c(socio_residenceplace, " (residence)"),
         socio_residenceplace = factor(socio_residenceplace, levels = c("Mano (residence)",
                               "Bilingual (residence)",
                               "Kpelle (residence)")),
         socio_family_adoptive = factor(socio_family_adoptive, levels = c("Mano (family)",
                               "Bilingual (family)",
                               "Kpelle (family)"))) |>

  add_count(socio_family_adoptive, socio_residenceplace) |>
  ggplot(aes(age)) +
  geom_histogram(aes(fill = gender)) +
  geom_label(x = 40, y = 6, aes(label = str_c("total: ", n))) +
  facet_grid(socio_family_adoptive ~ socio_residenceplace) +
  labs(y = "", color = "adoptive family") +
  theme(legend.position = "bottom") +
  scale_fill_manual(values = c("#0072B2", "#E69F00"))
```



```
ggsave("age_by_res_place_and_adoptive_fam.png", width = 9, height = 9, bg = "white")
```

Here is the same data in the table format:

Table 4. Speaker sample by residence place, adoptive family, age, and gender

```
for_visualization |>
  count(biological_father, socio_family_adoptive, socio_residenceplace, gender) |>
  pivot_wider(names_from = gender, values_from = n, values_fill = 0) |>
  rename(`biological father` = biological_father,
         `adoptive family` = socio_family_adoptive,
         `residence place` = socio_residenceplace) |>
  knitr::kable()
```

biological father	adoptive family	residence place	F	M
Kpelle	Mano	Kpelle	1	2
Kpelle	Bilingual	Bilingual	1	2

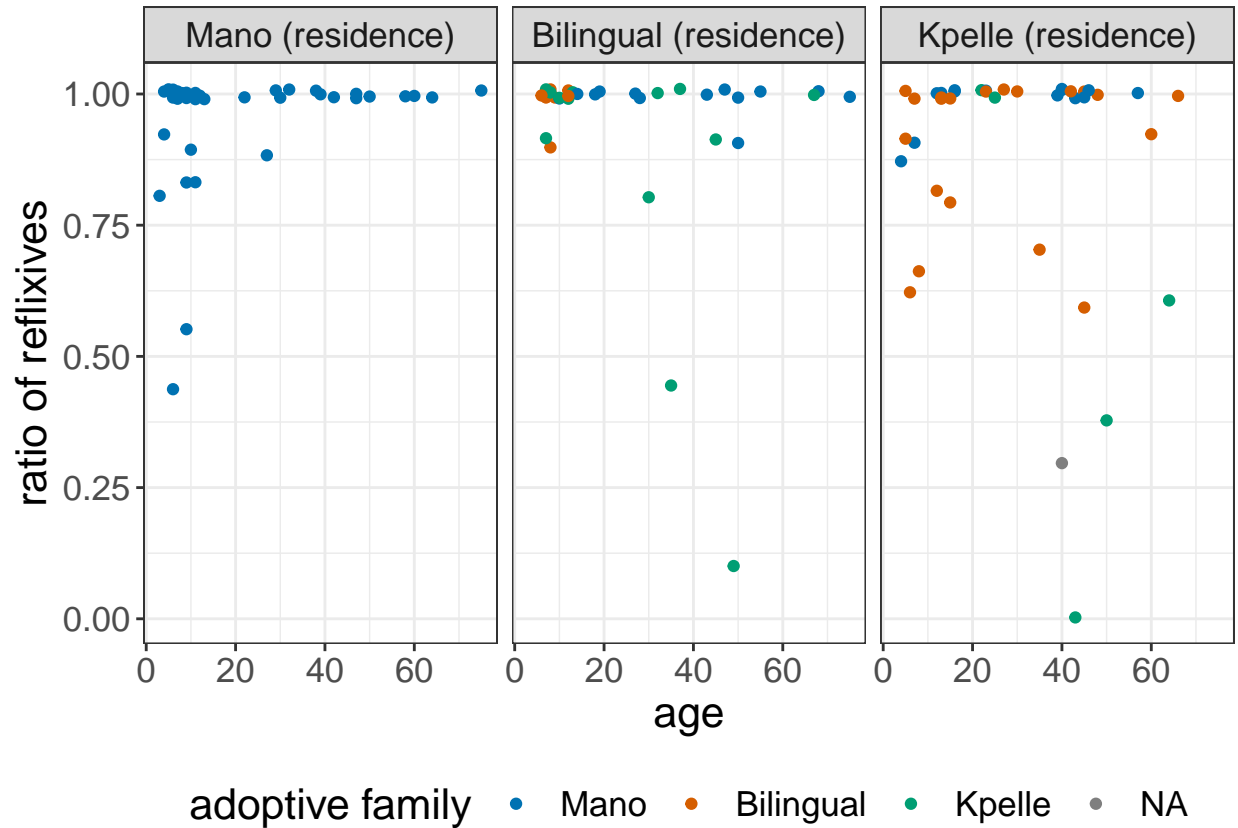
biological father	adoptive family	residence place	F	M
Kpelle	Bilingual	Kpelle	2	6
Kpelle	Kpelle	Bilingual	5	7
Kpelle	Kpelle	Kpelle	0	3
Kpelle	NA	Kpelle	1	0
Maninka	Bilingual	Bilingual	0	1
Mano	Mano	Mano	18	20
Mano	Mano	Bilingual	6	7
Mano	Mano	Kpelle	7	4
Mano	Bilingual	Bilingual	1	3
Mano	Bilingual	Kpelle	7	5
Mano	Kpelle	Bilingual	0	1
Mano	Kpelle	Kpelle	1	1

Here we can see the code and the result of visualization of distribution of ratio reflexive forms across different residence places and adoptive families (since dots overlay, we jittered dots along y axis):

Figure 3. Ratio of reflexive pronouns according to residence place and adoptive family

```
for_visualization |>
  mutate(socio_residenceplace = str_c(socio_residenceplace, " (residence)"),
         socio_residenceplace = factor(socio_residenceplace, levels = c("Mano (residence)",
                                "Bilingual (residence)",
                                "Kpelle (residence)"))) |>

  ggplot(aes(age, ratio, color = socio_family_adoptive))+
  geom_jitter(width = 0, height = 0.01)+
  facet_wrap(~socio_residenceplace)+
  labs(y = "ratio of reflexives", color = "adoptive family")+
  theme(legend.position = "bottom")+
  scale_color_manual(values = c("#0072B2", "#D55E00", "#009E73"))
```



```
ggsave("raw_ratio_by_res_place_and_adoptive_fam.png", width = 9, height = 9, bg = "white")
```

4 Modeling

In order to model the probability of reflexive forms, we employ Bayesian logistic mixed effects model as implemented in the R (R Core Team, 2023) package `brms` (Bürkner, 2017) using default priors. Input data were 1157 answers collected from 112 speakers from 6 residences that were interviewed during the 2021-2022 period. First, filtered only subject-oriented reflexive configurations, where the target marker corefers with the subject. The target marker appears in Direct object (DO), Postpositional phrase (PP), Possessor of Direct object (DO Poss), Possessor within Postpositional phrase (PP Poss) positions as illustrated by examples 3a, 4a, 4b. These configurations account for the majority of the data and we call them main syntactic positions. Another position was possessor within a coordinate phrase (ex. 5, 6) and NP within a PP whose antecedent is the DO, the subject's or the object's possessor (ex. 7 illustrating the subject's possessor). We account for the coordinate construction and for other cases of non-subject orientation separately. We ran the model in four chains, with 2500 warm-ups and 5000 iterations.

Here is our model (bayesian beta mixed effect model):

```
library(brms)
database >|>
  filter(analysis %in% c("basic", "refl"),
         synt_position != "COORD Poss") >|>
  mutate(socio_residenceplace = str_remove(socio_residenceplace, "-dominant"),
         socio_residenceplace = factor(socio_residenceplace, levels = c("Mano", "Bilingual", "Kpelle")),
         socio_family_adoptive = factor(socio_family_adoptive, levels = c("Mano", "Bilingual", "Kpelle"))) ->
  for_regression
```

```
for_regeression |>
  brm(analysis ~ socio_residenceplace+socio_family_adoptive + (1|residence_place) + (1|name_code),
      family=bernoulli, iter = 5000, backend = "cmdstanr", cores = 16, data = _) ->
  logistic_model
```

```
# Running MCMC with 4 chains, at most 16 in parallel...
```

```
#
# Chain 1 Iteration: 1 / 5000 [ 0%] (Warmup)
# Chain 2 Iteration: 1 / 5000 [ 0%] (Warmup)
# Chain 3 Iteration: 1 / 5000 [ 0%] (Warmup)
# Chain 4 Iteration: 1 / 5000 [ 0%] (Warmup)
# Chain 1 Iteration: 100 / 5000 [ 2%] (Warmup)
# Chain 2 Iteration: 100 / 5000 [ 2%] (Warmup)
# Chain 3 Iteration: 100 / 5000 [ 2%] (Warmup)
# Chain 4 Iteration: 100 / 5000 [ 2%] (Warmup)
# Chain 1 Iteration: 200 / 5000 [ 4%] (Warmup)
# Chain 2 Iteration: 200 / 5000 [ 4%] (Warmup)
# Chain 3 Iteration: 200 / 5000 [ 4%] (Warmup)
# Chain 1 Iteration: 300 / 5000 [ 6%] (Warmup)
# Chain 4 Iteration: 200 / 5000 [ 4%] (Warmup)
# Chain 2 Iteration: 300 / 5000 [ 6%] (Warmup)
# Chain 3 Iteration: 300 / 5000 [ 6%] (Warmup)
# Chain 1 Iteration: 400 / 5000 [ 8%] (Warmup)
# Chain 2 Iteration: 400 / 5000 [ 8%] (Warmup)
# Chain 3 Iteration: 400 / 5000 [ 8%] (Warmup)
# Chain 4 Iteration: 300 / 5000 [ 6%] (Warmup)
# Chain 1 Iteration: 500 / 5000 [ 10%] (Warmup)
# Chain 2 Iteration: 500 / 5000 [ 10%] (Warmup)
# Chain 3 Iteration: 500 / 5000 [ 10%] (Warmup)
# Chain 4 Iteration: 400 / 5000 [ 8%] (Warmup)
# Chain 1 Iteration: 600 / 5000 [ 12%] (Warmup)
# Chain 2 Iteration: 600 / 5000 [ 12%] (Warmup)
# Chain 3 Iteration: 600 / 5000 [ 12%] (Warmup)
# Chain 4 Iteration: 500 / 5000 [ 10%] (Warmup)
# Chain 2 Iteration: 700 / 5000 [ 14%] (Warmup)
# Chain 1 Iteration: 700 / 5000 [ 14%] (Warmup)
# Chain 3 Iteration: 700 / 5000 [ 14%] (Warmup)
# Chain 4 Iteration: 600 / 5000 [ 12%] (Warmup)
# Chain 2 Iteration: 800 / 5000 [ 16%] (Warmup)
# Chain 3 Iteration: 800 / 5000 [ 16%] (Warmup)
# Chain 1 Iteration: 800 / 5000 [ 16%] (Warmup)
# Chain 4 Iteration: 700 / 5000 [ 14%] (Warmup)
# Chain 2 Iteration: 900 / 5000 [ 18%] (Warmup)
# Chain 3 Iteration: 900 / 5000 [ 18%] (Warmup)
# Chain 4 Iteration: 800 / 5000 [ 16%] (Warmup)
# Chain 2 Iteration: 1000 / 5000 [ 20%] (Warmup)
# Chain 1 Iteration: 900 / 5000 [ 18%] (Warmup)
# Chain 3 Iteration: 1000 / 5000 [ 20%] (Warmup)
# Chain 4 Iteration: 900 / 5000 [ 18%] (Warmup)
# Chain 2 Iteration: 1100 / 5000 [ 22%] (Warmup)
# Chain 1 Iteration: 1000 / 5000 [ 20%] (Warmup)
# Chain 3 Iteration: 1100 / 5000 [ 22%] (Warmup)
# Chain 2 Iteration: 1200 / 5000 [ 24%] (Warmup)
# Chain 4 Iteration: 1000 / 5000 [ 20%] (Warmup)
```



```

# Chain 1 Iteration: 1100 / 5000 [ 22%] (Warmup)
# Chain 3 Iteration: 1200 / 5000 [ 24%] (Warmup)
# Chain 2 Iteration: 1300 / 5000 [ 26%] (Warmup)
# Chain 4 Iteration: 1100 / 5000 [ 22%] (Warmup)
# Chain 1 Iteration: 1200 / 5000 [ 24%] (Warmup)
# Chain 2 Iteration: 1400 / 5000 [ 28%] (Warmup)
# Chain 3 Iteration: 1300 / 5000 [ 26%] (Warmup)
# Chain 4 Iteration: 1200 / 5000 [ 24%] (Warmup)
# Chain 1 Iteration: 1300 / 5000 [ 26%] (Warmup)
# Chain 3 Iteration: 1400 / 5000 [ 28%] (Warmup)
# Chain 2 Iteration: 1500 / 5000 [ 30%] (Warmup)
# Chain 4 Iteration: 1300 / 5000 [ 26%] (Warmup)
# Chain 1 Iteration: 1400 / 5000 [ 28%] (Warmup)
# Chain 3 Iteration: 1500 / 5000 [ 30%] (Warmup)
# Chain 2 Iteration: 1600 / 5000 [ 32%] (Warmup)
# Chain 1 Iteration: 1500 / 5000 [ 30%] (Warmup)
# Chain 3 Iteration: 1600 / 5000 [ 32%] (Warmup)
# Chain 2 Iteration: 1700 / 5000 [ 34%] (Warmup)
# Chain 4 Iteration: 1400 / 5000 [ 28%] (Warmup)
# Chain 1 Iteration: 1600 / 5000 [ 32%] (Warmup)
# Chain 3 Iteration: 1700 / 5000 [ 34%] (Warmup)
# Chain 2 Iteration: 1800 / 5000 [ 36%] (Warmup)
# Chain 4 Iteration: 1500 / 5000 [ 30%] (Warmup)
# Chain 1 Iteration: 1700 / 5000 [ 34%] (Warmup)
# Chain 3 Iteration: 1800 / 5000 [ 36%] (Warmup)
# Chain 2 Iteration: 1900 / 5000 [ 38%] (Warmup)
# Chain 4 Iteration: 1600 / 5000 [ 32%] (Warmup)
# Chain 1 Iteration: 1800 / 5000 [ 36%] (Warmup)
# Chain 3 Iteration: 1900 / 5000 [ 38%] (Warmup)
# Chain 2 Iteration: 2000 / 5000 [ 40%] (Warmup)
# Chain 4 Iteration: 1700 / 5000 [ 34%] (Warmup)
# Chain 1 Iteration: 1900 / 5000 [ 38%] (Warmup)
# Chain 3 Iteration: 2000 / 5000 [ 40%] (Warmup)
# Chain 2 Iteration: 2100 / 5000 [ 42%] (Warmup)
# Chain 1 Iteration: 2000 / 5000 [ 40%] (Warmup)
# Chain 4 Iteration: 1800 / 5000 [ 36%] (Warmup)
# Chain 3 Iteration: 2100 / 5000 [ 42%] (Warmup)
# Chain 2 Iteration: 2200 / 5000 [ 44%] (Warmup)
# Chain 3 Iteration: 2200 / 5000 [ 44%] (Warmup)
# Chain 4 Iteration: 1900 / 5000 [ 38%] (Warmup)
# Chain 1 Iteration: 2100 / 5000 [ 42%] (Warmup)
# Chain 2 Iteration: 2300 / 5000 [ 46%] (Warmup)
# Chain 3 Iteration: 2300 / 5000 [ 46%] (Warmup)
# Chain 1 Iteration: 2200 / 5000 [ 44%] (Warmup)
# Chain 2 Iteration: 2400 / 5000 [ 48%] (Warmup)
# Chain 4 Iteration: 2000 / 5000 [ 40%] (Warmup)
# Chain 1 Iteration: 2300 / 5000 [ 46%] (Warmup)
# Chain 3 Iteration: 2400 / 5000 [ 48%] (Warmup)
# Chain 4 Iteration: 2100 / 5000 [ 42%] (Warmup)
# Chain 2 Iteration: 2500 / 5000 [ 50%] (Warmup)
# Chain 2 Iteration: 2501 / 5000 [ 50%] (Sampling)
# Chain 1 Iteration: 2400 / 5000 [ 48%] (Warmup)
# Chain 3 Iteration: 2500 / 5000 [ 50%] (Warmup)
# Chain 3 Iteration: 2501 / 5000 [ 50%] (Sampling)

```

```

# Chain 4 Iteration: 2200 / 5000 [ 44%] (Warmup)
# Chain 2 Iteration: 2600 / 5000 [ 52%] (Sampling)
# Chain 1 Iteration: 2500 / 5000 [ 50%] (Warmup)
# Chain 1 Iteration: 2501 / 5000 [ 50%] (Sampling)
# Chain 4 Iteration: 2300 / 5000 [ 46%] (Warmup)
# Chain 3 Iteration: 2600 / 5000 [ 52%] (Sampling)
# Chain 2 Iteration: 2700 / 5000 [ 54%] (Sampling)
# Chain 4 Iteration: 2400 / 5000 [ 48%] (Warmup)
# Chain 1 Iteration: 2600 / 5000 [ 52%] (Sampling)
# Chain 3 Iteration: 2700 / 5000 [ 54%] (Sampling)
# Chain 2 Iteration: 2800 / 5000 [ 56%] (Sampling)
# Chain 1 Iteration: 2700 / 5000 [ 54%] (Sampling)
# Chain 2 Iteration: 2900 / 5000 [ 58%] (Sampling)
# Chain 3 Iteration: 2800 / 5000 [ 56%] (Sampling)
# Chain 4 Iteration: 2500 / 5000 [ 50%] (Warmup)
# Chain 4 Iteration: 2501 / 5000 [ 50%] (Sampling)
# Chain 1 Iteration: 2800 / 5000 [ 56%] (Sampling)
# Chain 4 Iteration: 2600 / 5000 [ 52%] (Sampling)
# Chain 2 Iteration: 3000 / 5000 [ 60%] (Sampling)
# Chain 3 Iteration: 2900 / 5000 [ 58%] (Sampling)
# Chain 1 Iteration: 2900 / 5000 [ 58%] (Sampling)
# Chain 4 Iteration: 2700 / 5000 [ 54%] (Sampling)
# Chain 2 Iteration: 3100 / 5000 [ 62%] (Sampling)
# Chain 3 Iteration: 3000 / 5000 [ 60%] (Sampling)
# Chain 1 Iteration: 3000 / 5000 [ 60%] (Sampling)
# Chain 4 Iteration: 2800 / 5000 [ 56%] (Sampling)
# Chain 2 Iteration: 3200 / 5000 [ 64%] (Sampling)
# Chain 3 Iteration: 3100 / 5000 [ 62%] (Sampling)
# Chain 1 Iteration: 3100 / 5000 [ 62%] (Sampling)
# Chain 2 Iteration: 3300 / 5000 [ 66%] (Sampling)
# Chain 3 Iteration: 3200 / 5000 [ 64%] (Sampling)
# Chain 4 Iteration: 2900 / 5000 [ 58%] (Sampling)
# Chain 1 Iteration: 3200 / 5000 [ 64%] (Sampling)
# Chain 2 Iteration: 3400 / 5000 [ 68%] (Sampling)
# Chain 4 Iteration: 3000 / 5000 [ 60%] (Sampling)
# Chain 3 Iteration: 3300 / 5000 [ 66%] (Sampling)
# Chain 1 Iteration: 3300 / 5000 [ 66%] (Sampling)
# Chain 2 Iteration: 3500 / 5000 [ 70%] (Sampling)
# Chain 4 Iteration: 3100 / 5000 [ 62%] (Sampling)
# Chain 3 Iteration: 3400 / 5000 [ 68%] (Sampling)
# Chain 1 Iteration: 3400 / 5000 [ 68%] (Sampling)
# Chain 2 Iteration: 3600 / 5000 [ 72%] (Sampling)
# Chain 3 Iteration: 3500 / 5000 [ 70%] (Sampling)
# Chain 4 Iteration: 3200 / 5000 [ 64%] (Sampling)
# Chain 1 Iteration: 3500 / 5000 [ 70%] (Sampling)
# Chain 2 Iteration: 3700 / 5000 [ 74%] (Sampling)
# Chain 3 Iteration: 3600 / 5000 [ 72%] (Sampling)
# Chain 4 Iteration: 3300 / 5000 [ 66%] (Sampling)
# Chain 1 Iteration: 3600 / 5000 [ 72%] (Sampling)
# Chain 2 Iteration: 3800 / 5000 [ 76%] (Sampling)
# Chain 3 Iteration: 3700 / 5000 [ 74%] (Sampling)
# Chain 4 Iteration: 3400 / 5000 [ 68%] (Sampling)
# Chain 1 Iteration: 3700 / 5000 [ 74%] (Sampling)
# Chain 2 Iteration: 3900 / 5000 [ 78%] (Sampling)

```

```

# Chain 3 Iteration: 3800 / 5000 [ 76%] (Sampling)
# Chain 4 Iteration: 3500 / 5000 [ 70%] (Sampling)
# Chain 1 Iteration: 3800 / 5000 [ 76%] (Sampling)
# Chain 2 Iteration: 4000 / 5000 [ 80%] (Sampling)
# Chain 3 Iteration: 3900 / 5000 [ 78%] (Sampling)
# Chain 4 Iteration: 3600 / 5000 [ 72%] (Sampling)
# Chain 1 Iteration: 3900 / 5000 [ 78%] (Sampling)
# Chain 2 Iteration: 4100 / 5000 [ 82%] (Sampling)
# Chain 3 Iteration: 4000 / 5000 [ 80%] (Sampling)
# Chain 4 Iteration: 3700 / 5000 [ 74%] (Sampling)
# Chain 1 Iteration: 4000 / 5000 [ 80%] (Sampling)
# Chain 3 Iteration: 4100 / 5000 [ 82%] (Sampling)
# Chain 2 Iteration: 4200 / 5000 [ 84%] (Sampling)
# Chain 4 Iteration: 3800 / 5000 [ 76%] (Sampling)
# Chain 1 Iteration: 4100 / 5000 [ 82%] (Sampling)
# Chain 3 Iteration: 4200 / 5000 [ 84%] (Sampling)
# Chain 2 Iteration: 4300 / 5000 [ 86%] (Sampling)
# Chain 4 Iteration: 3900 / 5000 [ 78%] (Sampling)
# Chain 1 Iteration: 4200 / 5000 [ 84%] (Sampling)
# Chain 3 Iteration: 4300 / 5000 [ 86%] (Sampling)
# Chain 2 Iteration: 4400 / 5000 [ 88%] (Sampling)
# Chain 4 Iteration: 4000 / 5000 [ 80%] (Sampling)
# Chain 1 Iteration: 4300 / 5000 [ 86%] (Sampling)
# Chain 2 Iteration: 4500 / 5000 [ 90%] (Sampling)
# Chain 3 Iteration: 4400 / 5000 [ 88%] (Sampling)
# Chain 4 Iteration: 4100 / 5000 [ 82%] (Sampling)
# Chain 1 Iteration: 4400 / 5000 [ 88%] (Sampling)
# Chain 2 Iteration: 4600 / 5000 [ 92%] (Sampling)
# Chain 3 Iteration: 4500 / 5000 [ 90%] (Sampling)
# Chain 4 Iteration: 4200 / 5000 [ 84%] (Sampling)
# Chain 1 Iteration: 4500 / 5000 [ 90%] (Sampling)
# Chain 2 Iteration: 4700 / 5000 [ 94%] (Sampling)
# Chain 3 Iteration: 4600 / 5000 [ 92%] (Sampling)
# Chain 4 Iteration: 4300 / 5000 [ 86%] (Sampling)
# Chain 1 Iteration: 4600 / 5000 [ 92%] (Sampling)
# Chain 2 Iteration: 4800 / 5000 [ 96%] (Sampling)
# Chain 3 Iteration: 4700 / 5000 [ 94%] (Sampling)
# Chain 4 Iteration: 4400 / 5000 [ 88%] (Sampling)
# Chain 1 Iteration: 4700 / 5000 [ 94%] (Sampling)
# Chain 2 Iteration: 4900 / 5000 [ 98%] (Sampling)
# Chain 3 Iteration: 4800 / 5000 [ 96%] (Sampling)
# Chain 4 Iteration: 4500 / 5000 [ 90%] (Sampling)
# Chain 1 Iteration: 4800 / 5000 [ 96%] (Sampling)
# Chain 2 Iteration: 5000 / 5000 [100%] (Sampling)
# Chain 3 Iteration: 4900 / 5000 [ 98%] (Sampling)
# Chain 2 finished in 25.1 seconds.
# Chain 4 Iteration: 4600 / 5000 [ 92%] (Sampling)
# Chain 1 Iteration: 4900 / 5000 [ 98%] (Sampling)
# Chain 3 Iteration: 5000 / 5000 [100%] (Sampling)
# Chain 3 finished in 25.7 seconds.
# Chain 4 Iteration: 4700 / 5000 [ 94%] (Sampling)
# Chain 1 Iteration: 5000 / 5000 [100%] (Sampling)
# Chain 1 finished in 26.3 seconds.
# Chain 4 Iteration: 4800 / 5000 [ 96%] (Sampling)

```

```
# Chain 4 Iteration: 4900 / 5000 [ 98%] (Sampling)
# Chain 4 Iteration: 5000 / 5000 [100%] (Sampling)
# Chain 4 finished in 27.3 seconds.
#
# All 4 chains finished successfully.
# Mean chain execution time: 26.1 seconds.
# Total execution time: 27.6 seconds.
```

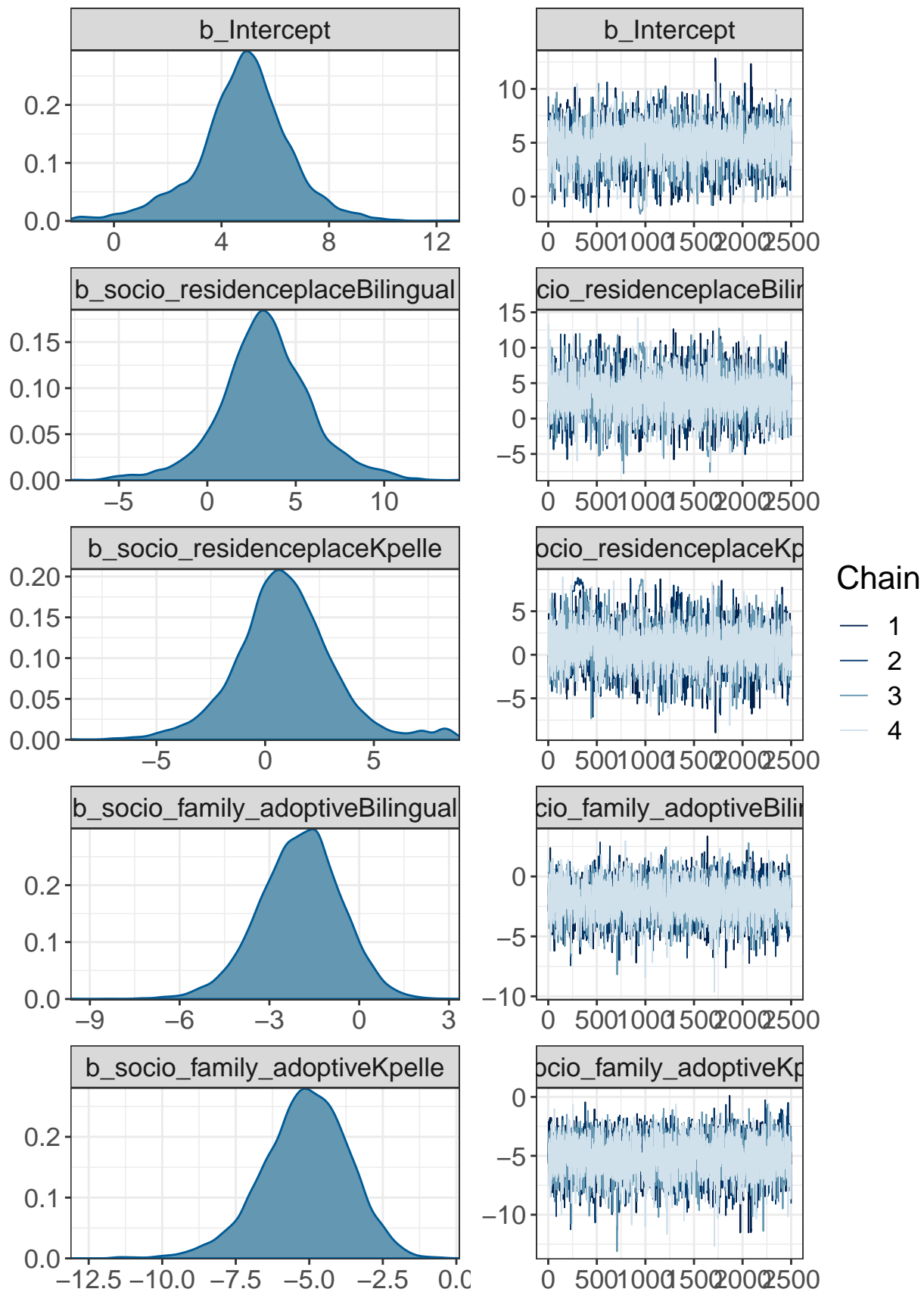
Inspect the model and traceplots:

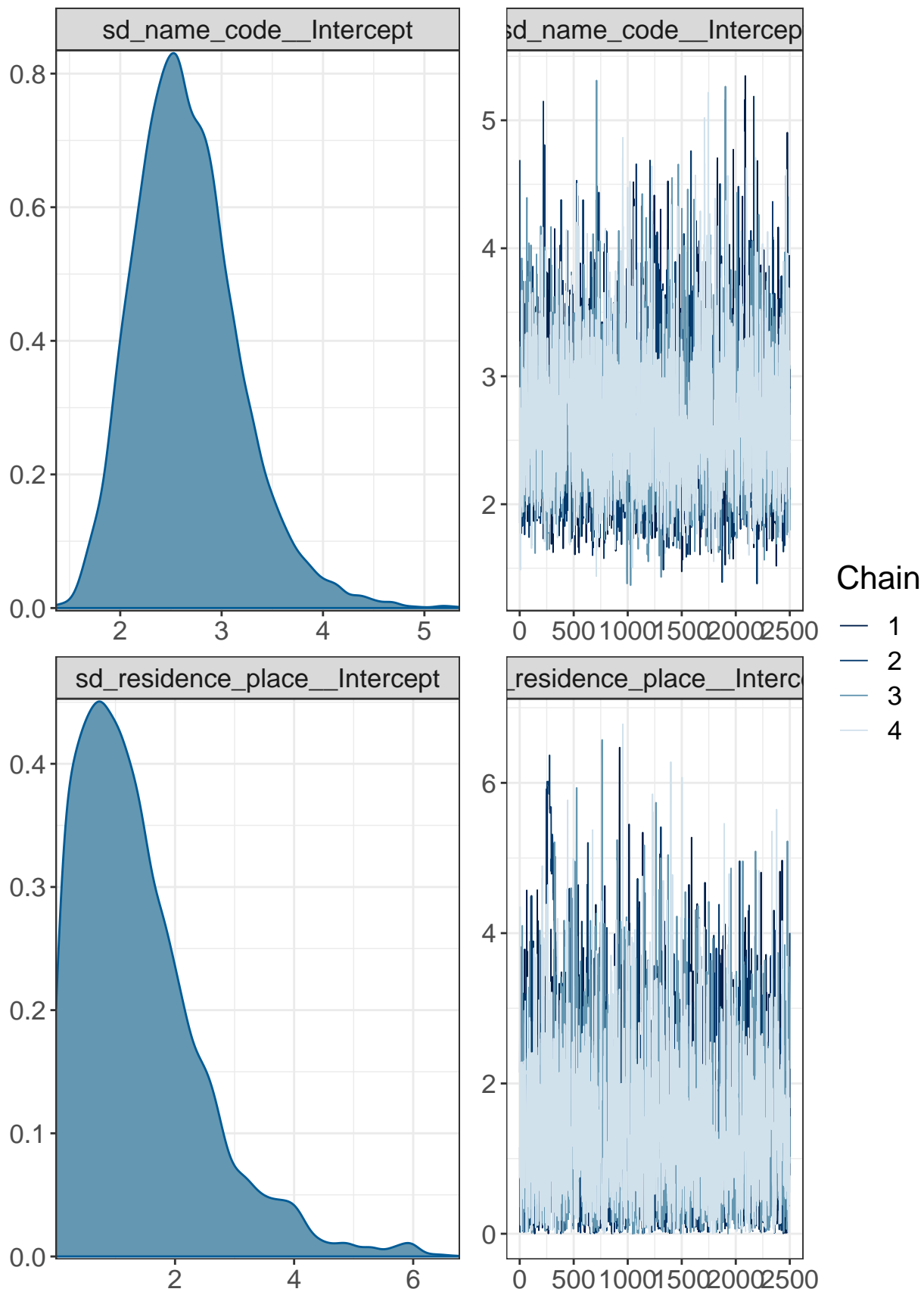
Figure 4. Model estimates and chains

```
logistic_model

# Family: bernoulli
# Links: mu = logit
# Formula: analysis ~ socio_residenceplace + socio_family_adoptive + (1 | residence_place) + (1 | name_code)
# Data: for_regeression (Number of observations: 1147)
# Draws: 4 chains, each with iter = 5000; warmup = 2500; thin = 1;
#       total post-warmup draws = 10000
#
# Group-Level Effects:
# ~name_code (Number of levels: 111)
#       Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
# sd(Intercept)    2.67    0.51    1.82    3.85 1.00    3088    4829
#
# ~residence_place (Number of levels: 6)
#       Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
# sd(Intercept)    1.39    1.07    0.06    4.05 1.01     824     429
#
# Population-Level Effects:
#
#       Estimate Est.Error 1-95% CI u-95% CI Rhat
# Intercept           4.81    1.73    0.87    8.14 1.00
# socio_residenceplaceBilingual    3.30    2.63   -2.21    8.88 1.00
# socio_residenceplaceKpelle       0.92    2.31   -3.69    6.53 1.00
# socio_family_adoptiveBilingual   -1.99    1.38   -4.82    0.58 1.00
# socio_family_adoptiveKpelle     -5.13    1.49   -8.33   -2.44 1.00
#
#       Bulk_ESS Tail_ESS
# Intercept           1648    807
# socio_residenceplaceBilingual    1930    1607
# socio_residenceplaceKpelle       826    323
# socio_family_adoptiveBilingual    3639    5034
# socio_family_adoptiveKpelle      3427    4875
#
# Draws were sampled using sample(hmc). For each parameter, Bulk_ESS
# and Tail_ESS are effective sample size measures, and Rhat is the potential
# scale reduction factor on split chains (at convergence, Rhat = 1).
```

```
plot(logistic_model)
```

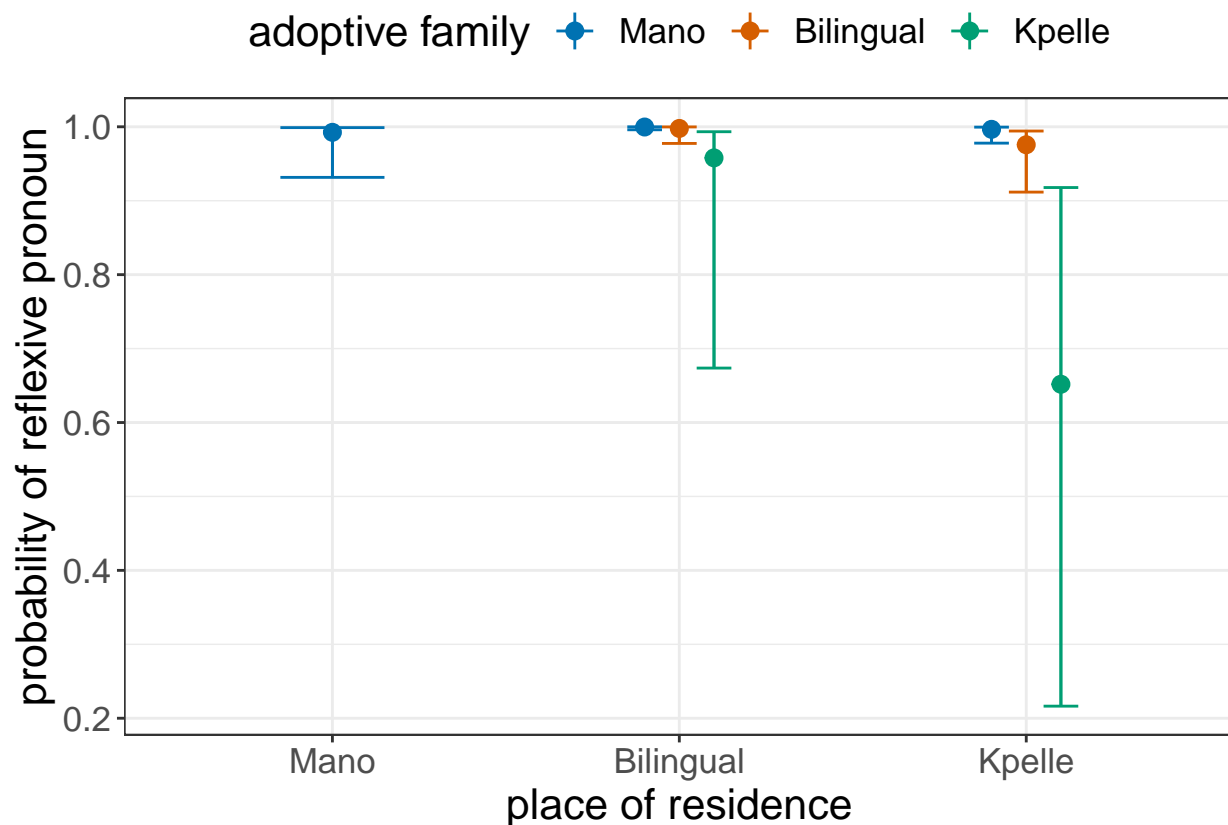




Values for categorical variables are the sum of Intercept values and variable values. However for numeric variable it is the pure variable value that counts, and as we see its mode is located around zero. We created the model's effect-plot that shows model predictions for variables "place of residence" and "adoptive family" (Figure 2 in the main article) with 80% credible interval:

Figure 5. Effect plots of the model (Figure 2 in the main article)

```
conditional_effects(logistic_model,
  prob = 0.8,
  effects = c("socio_residenceplace:socio_family_adoptive"))$socio_residenceplace |>
  filter(!(socio_residenceplace == "Mano" & socio_family_adoptive == "Bilingual"),
    !(socio_residenceplace == "Mano" & socio_family_adoptive == "Kpelle")) |>
  mutate(socio_residenceplace = factor(socio_residenceplace, levels = c("Mano", "Bilingual", "Kpelle")),
    socio_family_adoptive = factor(socio_family_adoptive, levels = c("Mano", "Bilingual", "Kpelle"))) |>
  ggplot(aes(socio_residenceplace, estimate__, ymin = lower__, ymax = upper__, color = socio_family_adoptive))+
  geom_errorbar(width = 0.3, position = position_dodge())+
  geom_pointrange(position = position_dodge(width = 0.3))+
  labs(x = "place of residence", y = "probability of reflexive pronoun", color = "adoptive family")+
  theme(legend.position = "top")+
  scale_color_manual(values = c("#0072B2", "#D55E00", "#009E73"))
```



```
ggsave("prob_by_res_place_and_adoptive_fam.png", width = 9, height = 7, bg = "white")
```

From this plot we can see the similarity of values in individuals residing in Mano or bilingual settlements, regardless of family, and those who grew up in Mano-speaking families, regardless of settlement. On the other hand, we can see that values are lower in individuals who were born or adopted into Kpelle and bilingual families in Kpelle settlements.

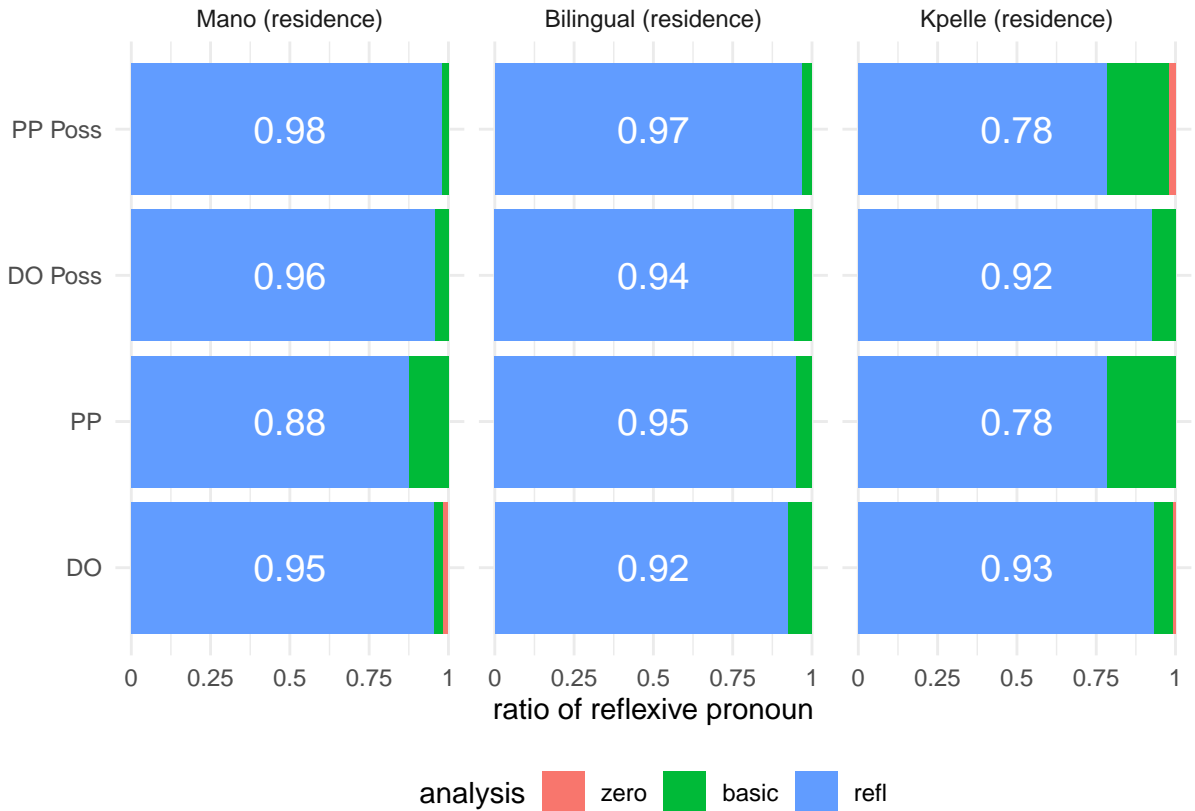
5 Syntax by residence place

Figure 6. Ratio of reflexive pronoun by place of residence and syntactic position (Figure 3 in the main article)

```
database |>
  filter(synt_position %in% c("DO", "DO Poss", "PP", "PP Poss")) |>
  mutate(socio_residenceplace = case_when(socio_residenceplace == "Kpelle-dominant" ~ "Kpelle",
                                          socio_residenceplace == "Mano-dominant" ~ "Mano",
                                          TRUE ~ socio_residenceplace),
         socio_residenceplace = str_c(socio_residenceplace, " (residence)"),
         socio_residenceplace = factor(socio_residenceplace, levels = c("Mano (residence)",
                               "Bilingual (residence)",
                               "Kpelle (residence)"))) |>

  count(synt_position, analysis, socio_residenceplace) |>
  group_by(synt_position, socio_residenceplace) |>
  mutate(ratio = n/sum(n),
         analysis = factor(analysis, levels = c("zero", "basic", "refl")),
         synt_position = factor(synt_position, levels = c("DO", "PP", "DO Poss", "PP Poss")),
         lable = ifelse(analysis == "refl", as.character(round(ratio, 2)), NA)) |>

  rowwise() |>
  select(-n) |>
  ggplot(aes(ratio, synt_position, fill = analysis, label = lable))+
  geom_col()+
  geom_text(x = 0.5, color = "white", size = 5)+
  facet_wrap(~socio_residenceplace)+
  labs(y = "", x = "ratio of reflexive pronoun")+
  theme_minimal()+
  theme(legend.position = "bottom")+
  scale_x_continuous(labels = c(0, 0.25, 0.50, 0.75, 1))
```

```
ggsave("overall_syntax_by_residenceplace.png", width = 9, height = 7, bg = "white")
```

6 Coordinate construction

Figure 7. Ratio of pronouns in the coordinate by place residence (Figure 4 in the main article)

```
database %>%
  filter(synt_position %in% c("COORD Poss")) %>%
  mutate(socio_residenceplace = case_when(socio_residenceplace == "Kpelle-dominant" ~ "Kpelle",
    socio_residenceplace == "Mano-dominant" ~ "Mano",
    TRUE ~ socio_residenceplace),
    socio_residenceplace = str_c(socio_residenceplace, " (residence)"),
    socio_residenceplace = factor(socio_residenceplace, levels = c("Mano (residence)",
      "Bilingual (residence)",
      "Kpelle (residence)"))) %>%

  count(synt_position, analysis, socio_residenceplace) %>%
  group_by(synt_position, socio_residenceplace) %>%
  mutate(ratio = n/sum(n),
    analysis = factor(analysis, levels = rev(c("refl", "basic", "3pl", "zero"))),
    synt_position = factor(synt_position, levels = c("DO", "PP", "DO Poss", "PP Poss")),
    label = str_c(as.character(round(ratio, 2)), " (", n, ")"),
    shade = case_when(analysis == "refl" ~ 1,
      analysis == "basic" ~ 1,
      analysis == "zero" ~ 0.95,
      analysis == "3pl" ~ 0.95)) %>%

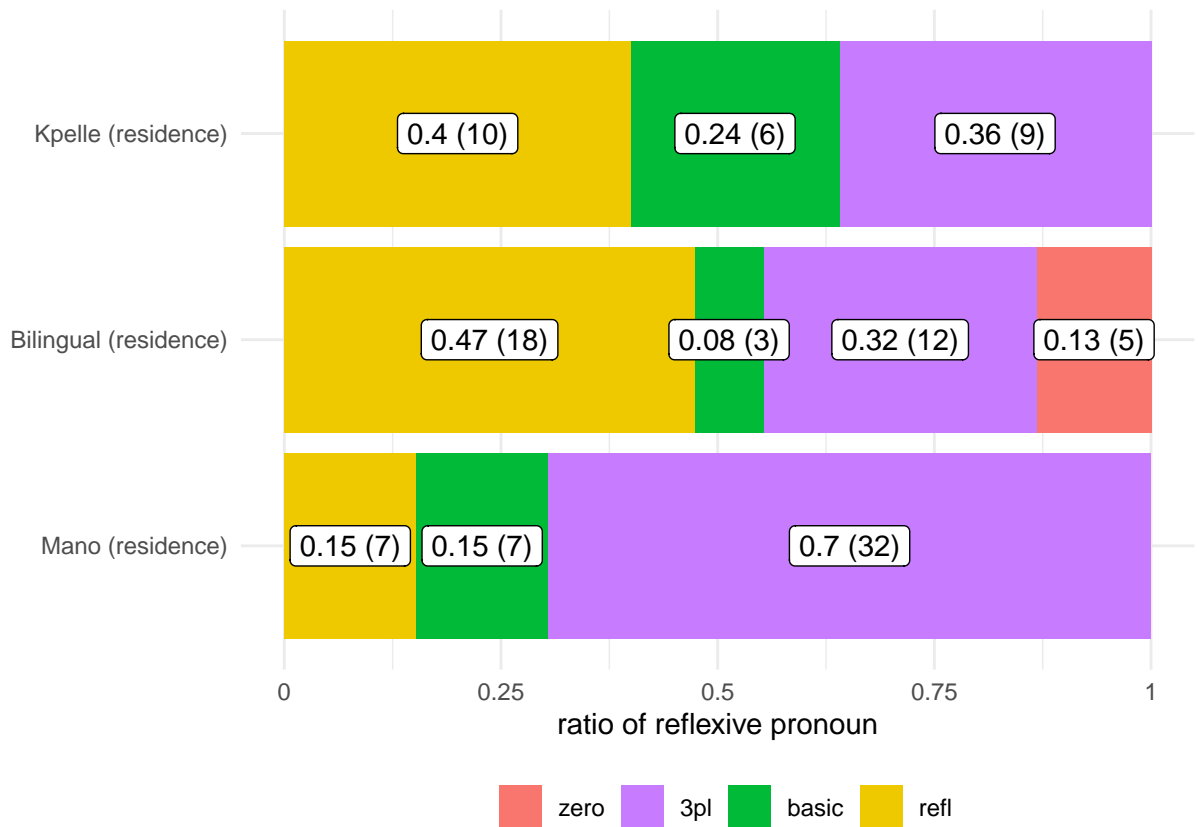
  rowwise() %>%
```

```

select(-n) ->
for_plot

for_plot %>%
  ggplot(aes(ratio, socio_residenceplace, group = analysis, label = label))+
  geom_col(aes(fill = analysis))+
  geom_label(position = position_fill(vjust = 0.5))+
  labs(y = "", x = "ratio of reflexive pronoun", fill = "")+
  scale_x_continuous(labels = c(0, 0.25, 0.50, 0.75, 1))+
  theme_minimal()+
  theme(legend.position = "bottom")+
  scale_fill_manual(values = c("#F8766D", "#C77CFF", "#00BA38", "#EEC900"))

```

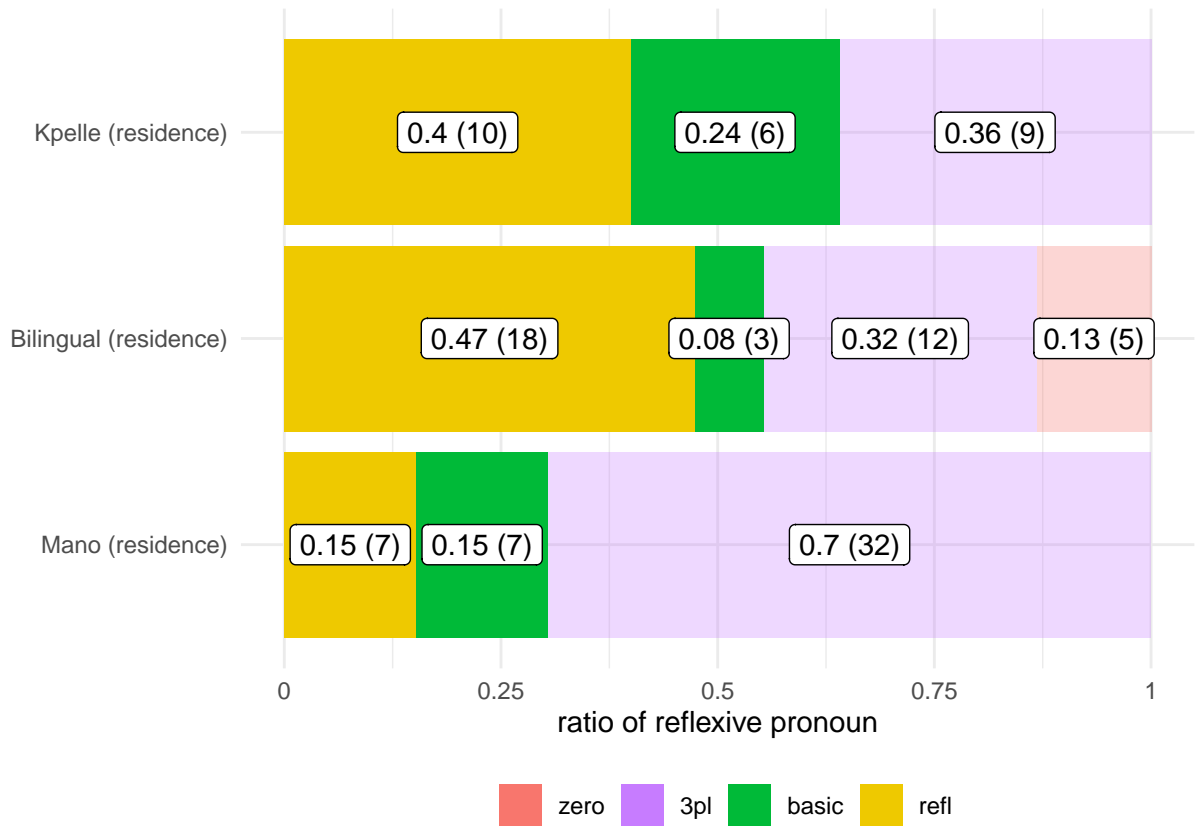


```

ggsave("coordinate_construction.png", width = 9, height = 7, bg = "white")

for_plot %>%
  ggplot(aes(ratio, socio_residenceplace, group = analysis, label = label))+
  geom_col(alpha = 0.3, aes(fill = analysis))+
  geom_col(aes(fill = analysis), data = for_plot %>% filter(analysis %in% c("refl", "basic")))+
  geom_label(position = position_fill(vjust = 0.5))+
  labs(y = "", x = "ratio of reflexive pronoun", fill = "")+
  scale_x_continuous(labels = c(0, 0.25, 0.50, 0.75, 1))+
  theme_minimal()+
  theme(legend.position = "bottom")+
  scale_fill_manual(values = c("#F8766D", "#C77CFF", "#00BA38", "#EEC900"))

```



```
ggsave("coordinate_construction_transparant.png", width = 9, height = 7, bg = "white")
```

7 Package versions

We performed our statistical analysis and visualisation with the following package versions:

package	version	citation
brms	2.20.4	Bürkner (2017)
inspectdf	0.0.12	Rushworth (2022)
knitr	1.42	Xie (2023)
rmarkdown	2.25	Xie, Dervieux, Riederer (2020)
tidyverse	2.0.0	Wickham (2019)
R	4.3.2	R Core Team (2023)