

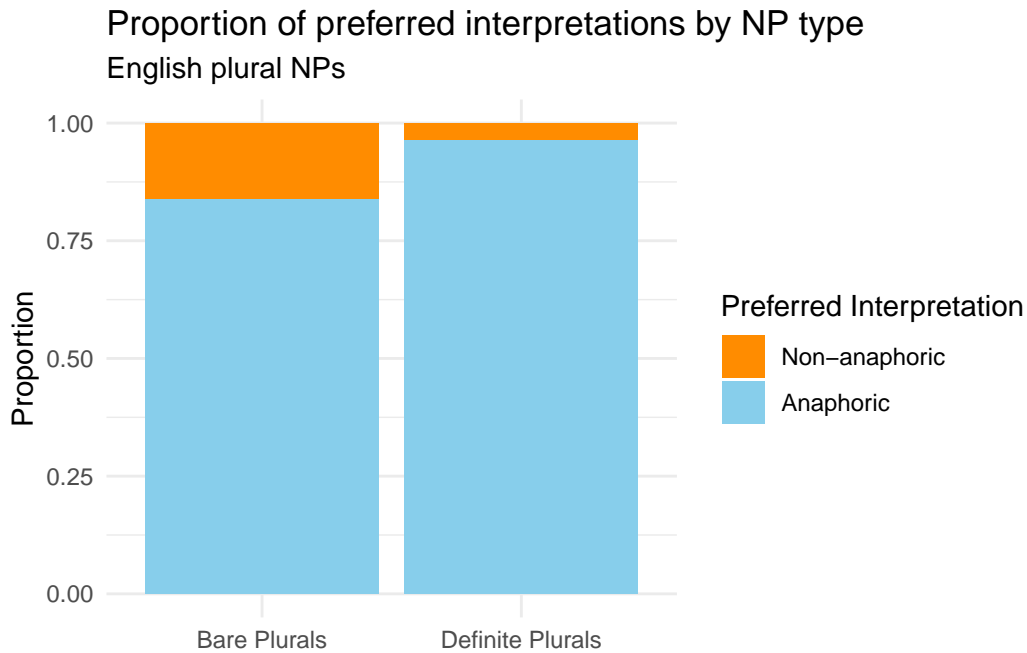
Anaphoricity Experiment Results

The results of the experiment were analyzed using mixed-effects logistic regression. First, each language was analyzed separately. The models had the following syntax: `choice_num ~ context + (1|ResponseId) + (1|item)`, which includes:

- `choice_num`: the choice of the interpretation (anaphoric vs. non-anaphoric) as a predictor,
- `context`: the type of the plural NP (bare nouns vs. definite nouns in English and bare nouns vs nouns modified with an indefinite pronoun in Russian) as a fixed effect,
- `ResponseId`: the participant identifier as a random effect
- `item`: the experimental item number as a random effect.

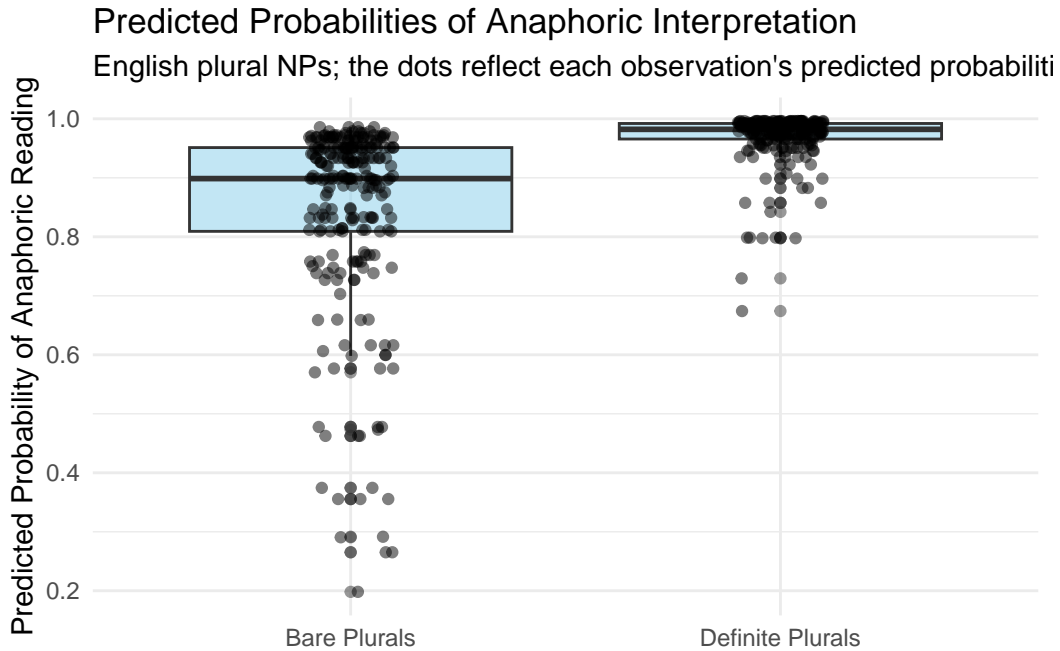
English

English bare nouns were interpreted anaphorically in 83.75% of trials, 95% CI[78.46, 88.18], and non-anaphorically in 16.25% of trials, 95% CI[11.86, 21.53]. English definite nouns were interpreted anaphorically in 96.25% of trials, 95% CI[93, 98.27], and non-anaphorically in 3.75% of trials, 95% CI[1.73, 6.99]. The proportions of chosen interpretations are illustrated by the plot below.



A mixed-effects logistic regression showed that the choice of anaphoric vs. non-anaphoric interpretation depends on whether the plural is definite or not. In other words, participants are significantly more likely to interpret definite plurals anaphorically compared to bare plurals, given the context is the same ($\beta_{def} = 2.02$, $SE = 0.62$, $p < 0.001$). The intercept of the model was estimated as 2.32, $SE = 0.46$, and its significance level is $p < 0.001$. The β coefficient reflects a significant positive relationship between the log odds of choosing the anaphoric interpretation for a bare plural and the log odds of choosing anaphoric interpretation for a definite plural in English. The odds of choosing the anaphoric interpretation were 7.5 times higher in English definite plurals than in English bare plurals. Such effect size can be considered relatively large.

The predicted probabilities of the anaphoric interpretation are illustrated in the plot below.

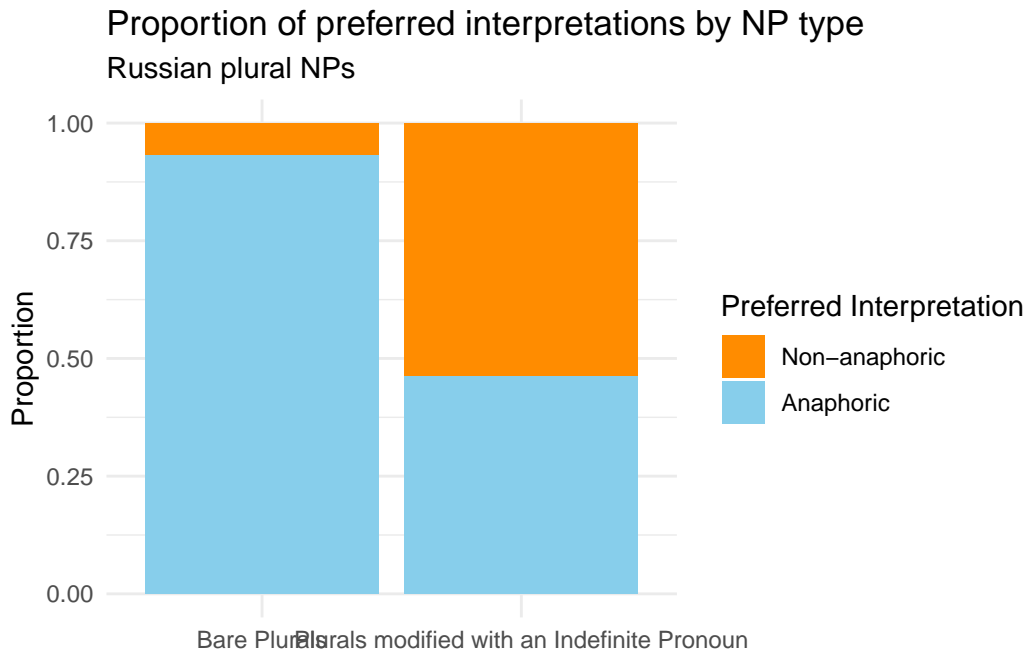


14.65% of the variance of the choice between anaphoric non-anaphoric is explained by the plural type (bare vs. definite). The model also included random intercepts for both participant (`ResponseId`) and experimental item (`item`) in order to account for variability due to these factors. The variance attributed to `ResponseId` was 1.51 (47.3% of total variance in the outcome), which flags that the variability between participants is substantial. The raw variance explained by `item` was 1.19 (35.7% of total variance), indicating substantial variability between different experimental items as well. No clear patterns were observed in the distribution of random effects.

In total, fixed and random effects explained 52.72% of the variance in the outcome, which indicates that the model is a reasonably good fit for a behavioral experiment. There is no overdispersion, as well as no clear pattern in the distribution of the model's residuals, which indicates no heteroscedasticity or auto-correlation. While the residuals were not normally distributed, the large sample size (480 observations) accounts for non-normality.

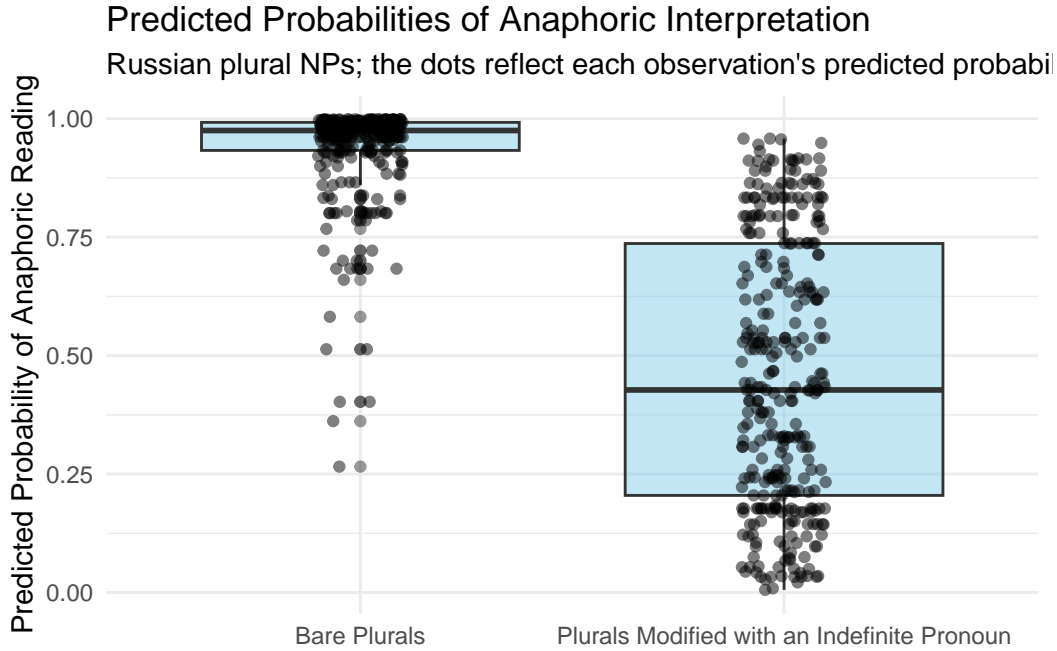
Russian

The Russian version of the experiment combined bare plurals and plurals modified by an indefinite pronoun *kakie-to*. Russian bare plurals were interpreted anaphotically in 93.13% of trials, 95% CI[89.78, 95.64], and non-anaphorically in 6.88% of trials, 95% CI[4.36, 10.22]. In contrast, nouns modified with *kakie-to* were interpreted anaphorically in only 46.25% of trials, 95% CI[40.69, 51.88], and non-anaphorically in 53.75% of trials, 95% CI[48.12, 59.31]. The proportions are illustrated by the plot below.



A mixed-effects logistic regression (a generalized linear mixed model with a binomial link function) showed that these proportions are significantly different. Participants were more likely to interpret a bare noun anaphorically compared to a noun modified with *kakie-to* ($\beta_{kakie-to} = -4.19$, $SE = 0.55$, $p < 0.001$). This indicates a significant negative relationship between the log odds of a participant choosing anaphoric interpretation for a bare noun and the log odds of a participant choosing the anaphoric interpretation for a noun modified with an indefinite pronoun. The intercept of the model was estimated as 3.95, $SE = 0.52$, and its significance level is $p < 0.001$. The odds for a plural modified by *kakie-to* to be interpreted anaphorically decrease by 98.5% compared to a bare plural, which indicates a very strong effect size.

The predicted probabilities of obtaining an anaphoric interpretation are presented in the plot below.



38.69% of the total variance in the interpretation (anaphoric vs. non-anaphoric) is explained by the noun type (bare vs. modified with an indefinite pronoun). The model also included random intercepts for both participant (**ResponseId**) and experimental item (**item**) in order to account for variability due to these factors. The variance attributed to **ResponseId** was $\sigma^2 = 2.68$ (38.5% of total variance), indicating substantial variability between participants. The variance attributed to **item** was $\sigma^2 = 0.99$ (14.2% of total variance), which means moderate variability between items. There are no clear patterns in the distribution of both random effects.

In total, fixed and random effects explain 71.03% of total variance, which indicates that the applied model is a good fit. The residuals of the model are not heteroscedastic or auto-correlated, and the large sample size (640 observations) allows us to relax the normality assumption for residuals.

The comparison of the two languages

In order to analyze the difference between different plural types across the two languages, Pearson's Chi-squared test was run. The proportions of the two interpretations were found to be significantly different for English bare nouns vs. Russian bare nouns, $\chi(1, 560) = 11.5$, $p = < 0.001$, with a small effect size ($V = 0.14$). The proportions of the two interpretations for English definite plurals and Russian bare plurals were not found to significantly differ, $\chi(1, 560) = 2$, $p = 0.16$.