

Appendix D. Structure of Multiple Regression Models

The main paper reports results of multivariate analyses that inferred models for estimating six target variables: target *PMLUg* per word, produced *PMLUg* per word, *PWPg*, ratio of produced syllables to produced *PMLUg*, ratio of produced consonants to produced *PMLUg*, and ratio of correct consonants to produced *PMLUg*. The fitted models were linear regression equations with multiple explanatory factors, as well as the basis for estimates of how much of the variation in the target variable is explained by variation in the factors and significance tests for each factor effect. The explanatory factors were the child's language, the child's gender, the child's age, the number of words elicited from the child, the ratio of syllables per word in the target words elicited in the session, and the ratio of consonants to syllables in the target words elicited in the session.

Each multiple regression model was estimated using data from groups of children from six language populations and ten child sessions per language group, resulting in a linear equation derived using traditional least-squares methods. The regression equations from these analyses all shared the following structure (where \hat{Y} is the point estimate for the target variable of the model and the b_i values are the model coefficients):

$$\hat{Y} = b_0 + b_1l_D + b_2l_E + b_3l_J + b_4l_P + b_5l_S + b_6g + b_7z_a + b_8z_n + b_9z_s + b_{10}z_c$$

The language effect in the model is represented by five explanatory variables that employed effect coding:

$l_D = 1$ if child spoke Dutch, -1 if child spoke Xi'iùy, 0 otherwise

$l_E = 1$ if child spoke English, -1 if child spoke Xi'iùy, 0 otherwise

$l_J = 1$ if child spoke Japanese, -1 if child spoke Xi'iùy, 0 otherwise

$l_P = 1$ if child spoke Portuguese, -1 if child spoke Xi'iùy, 0 otherwise

$l_S = 1$ if child spoke Spanish, -1 if child spoke Xi'iùy, 0 otherwise

The gender effect is defined by a single explanatory variable that employed effect coding:

$g = 1$ if child was female, -1 if child was male

The remaining explanatory effects were transformed to language-relative z -scores as follows:

$$z_{Factor} = \frac{\text{observed value for child session} - \text{average value for all child sessions in language group}}{\text{standard deviation of values for all child sessions in language group}}$$

The resulting transformed variables are:

z_a = language-relative z -score for child age at time of session in months

z_n = language-relative z -score for number of elicited words in session

z_s = language-relative z -score for ratio of target syllables per word in session

z_c = language-relative z -score for ratio of target consonants per syllable in session