We carried out a set of model comparisons to determine which n-grams should be included in our subsequent logistic regression models for this project. This involved selecting predictors at each n-gram level separately, starting at the unigram level before moving to the bigram level, followed by the trigram level.

1) We checked for multicollinearity at each n-gram level using variance inflation factor analyses; no item reached a score of 2.
2) Starting at the unigram level, we used a leave-one-out procedure to see which predictors explained variance over and above that explained by any other variable. The full/baseline model at this level included random effects of the first 5 unigrams (by child) as well as fixed effects for all 5 unigrams. This was then compared to five subsequent models, each leaving out the fixed effect term for a different unigram (random effects by child were included for every unigram in each model). Removal of only the first two unigrams harmed model fit to a significant extent, according to log-likelihood tests. Thus, these two unigrams were held over for the next level of model comparisons.
3) The same procedure described in step #2 was then carried out for the first four bigrams, but with random (by child) and fixed effects for the first two unigrams included in each model. Removal of only the first two bigrams harmed model fit to a statistically significant extent, according to the log-likelihood tests. Thus, the first two unigrams and first two bigrams were held over for the final set of model comparisons.
4) The same procedure (with random and fixed effects for the first two unigrams and first two bigrams) was carried out for the first three trigrams. Removal of the first two trigrams harmed model fit to a significant extent. Thus, the final set of predictors included the first two n-grams at each level.