**Research highlights**

* We examined CHILDES, the major source of data on children’s linguistic environment for potential sampling bias in terms of SES, urbanization, family structure, and language.
* We found that the 48 countries represented in CHILDES overrepresented rich, educated, urbanized countries with small nuclear families.
* Within these countries, corpora overrepresented rich, educated, urban, nuclear, and non-bilingual families.
* Interpretation of studies using CHILDES should acknowledge these biases.