```
┌─────────────────────────────────────────────┐
│     Collect daylong recordings in the        │
│          United States and Bolivia           │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│   Chop each recording into 30 second clips   │
│    (15832 total clips across 10 children)    │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│     Run vocal activity detector over clips    │
│     (Usoltsev, 2015) and report activity     │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│    Manually identify portions of recording    │
│        where child is sleeping               │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│   0% vocal activity and sleeping clips not    │
│           drawn for annotation                │
└─────────────────────────────────────────────┘
          │                        │
          ▼                        ▼
┌──────────────────────┐  ┌──────────────────────┐
│  Clips containing vocal│  │  Every other clip    │
│  activity randomly     │  │  with vocal activity │
│  selected, with        │  │  annotated for       │
│  replacement, and      │  │  speaker,            │
│  annotated for speaker,│  │  addressee, and      │
│  addressee, and language│ │  language            │
└──────────────────────┘  └──────────────────────┘
          │
          ▼
┌──────────────────────┐
│  Annotate 1) minimum of│
│  n=80 clips and 2)     │
│  until ratio between   │
│  categories asymptotes │
│  (976 total clips      │
│  annotated across 10   │
│  children)             │
└──────────────────────┘
```