# Guidelines for selecting words

## What we need

* ~400 words (out of ~3500)
* suitable for children between 2 and 7 years of age
* equally spread out in terms of age of acquisition
* some very easy words (spoken by most 2-year-olds) and some very difficult ones (spoken by very few 7-year-olds)
* roughly 50% nouns, 25% adjectives, 25% verbs

## Data files

* there are 10 data files, each with 344 words
* in each file, select roughly 17 nouns, 9 verbs and 9 adjectives
* to select, put an x into the column select

## Columns in the data files

* german : word in German
* select : type x to select the word
* aoa\_german\_comb : age of acquisition rating for german words
* word\_freq : how frequent is the word in the language (normalized lemma frequency per million)
* aoa\_mor0 : age at which children could name pictures according to Morrison et al. (1997)
* clt : is the word included in the CLT?
* english : British English translation

## Things to consider

* age of acquisition ratings are not correct in terms of absolute age. However, they tend to be correct in terms of relative difficulty. That is, a word that has a rated age of acquisition of 2.5 years is probably not learned at 2.5 years of age, but it’s likely to be learned earlier than a word that has a rating of 3.5 years
* the age range in age of acquisition ratings is wider than the target age range. Consider all words in the lists, not just those that fall between 2 and 7.

## Select words based on the following principles

* select words that are indicative of language abilities more broadly
* select words that are *different* from one another. That is words that are **not** learned in the same context; for example, if you select mouse, you should avoid cat, or if you select good avoid bad.
* select words equally spread out across the range in each file. That is, do **not** select words only from the top or the bottom of each file.
* all words that are part of the CLT (already marked in the select column)
* avoid words that are likely to be used by only a subgroup of children (e.g. words that are associated with a certain hobby)
* avoid words that have changed their meaning recently