# signdata-combined

```r
# load libraries and data
library(tidyverse)
library(readxl)
library(stringr)

eval_senses <- read_excel("semeval2007.gold.key.xlsx")
asl_lex <- read_excel("signdata.xlsx")

glimpse(eval_senses)
```

```
Rows: 455
Columns: 1
$ WordKey <chr> "d000.s000.t000 refer%2:32:01::", "d000.s000.t001 research%1:0~
```

```r
glimpse(asl_lex)
```

```
Rows: 2,723
Columns: 6
$ EntryID                   <chr> "tree", "night", "hamburger", "nephew", "c~
$ Code                      <chr> "A_01_002", "A_01_003", "A_01_004", "A_01_~
$ LexicalClass              <chr> "Noun", "Noun", "Noun", "Noun", "Noun", "A~
$ SignBankAnnotationID      <chr> "TREE", "NIGHT", "HAMBURGER", "NEPHEW", "C~
$ SignBankEnglishTranslations <chr> "forest, nature, plant, tree, woods", "nig~
$ SignBankReferenceID       <dbl> 342, 199, 1844, 3011, 2471, 2940, 388, 131~
```

```r
#filter asl_lex nouns and verbs only
#rename columns to match
noun_verbs <- asl_lex |>
  rename(WordKey = EntryID) |>
  filter(LexicalClass == "Noun" | LexicalClass == "Verb")
```

```
#mutate eval_senses to visualize individual words
eval_senses$WordKey <-
  gsub("d.*? ", "", eval_senses$WordKey)

eval_senses$WordKey <-
  gsub("%.*", "", eval_senses$WordKey)

#enlongate the data so each english translation gets its own row
nv_long <- noun_verbs |>
  separate_rows(SignBankEnglishTranslations,
                sep = ",\\s*") |>
  mutate(SignBankEnglishTranslations = str_trim(SignBankEnglishTranslations))
```

```
#inner join eval_senses and asl_lex noun_verbs

use_data <- inner_join(nv_long, eval_senses, by = "WordKey") |>
  distinct()

glimpse(use_data)
```

```
Rows: 493
Columns: 6
$ WordKey                   <chr> "read", "read", "read", "book", "book", "b~
$ Code                      <chr> "A_01_022", "A_01_022", "A_01_022", "A_01_~
$ LexicalClass              <chr> "Verb", "Verb", "Verb", "Noun", "Noun", "N~
$ SignBankAnnotationID      <chr> "READ", "READ", "READ", "BOOK", "BOOK", "B~
$ SignBankEnglishTranslations <chr> "look through", "read", "skim", "book", "b~
$ SignBankReferenceID       <dbl> 341, 341, 341, 418, 418, 418, 418, 436, 43~
```