

03 CSI online aphasia: Spoken - Descriptives

Kirsten Stark

26 September, 2024

Load packages

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(tidyr)

rm(list = ls())
```

Load and preprocess data

```
options( "encoding" = "UTF-8" )

# input
input <- "aphasia_final.csv"

# load data
df <- read.csv(here::here("data", "transient_data_files", input),
               sep = ",", na = "")
```

Duration of the experiment

```
print("Total duration, not outlier corrected")

## [1] "Total duration, not outlier corrected"
```

```
# mean(df$timetotal, na.rm=TRUE) # 33.13 min
mean(df$time_correct)
```

```
## [1] 28.59683
```

```
sd(df$timetotal) # 21.04 min
```

```
## [1] 21.13194
```

```
range(df$timetotal) # 15 to 126 min
```

```
## [1] 12.68333 167.58333
```

```
print("Total duration, split by session")
```

```
## [1] "Total duration, split by session"
```

```
df %>% group_by(type,session) %>% summarise(mean = mean(timetotal),
                                              median=median(timetotal),
                                              sd = sd(timetotal),
                                              min = min(timetotal),
                                              max = max(timetotal))
```

```
## 'summarise()' has grouped output by 'type'. You can override using the
## '.groups' argument.
```

```
## # A tibble: 6 x 7
## # Groups:   type [2]
##   type    session mean median    sd   min   max
##   <chr>    <int> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 PWA         1  39.4  29.5 23.0  21.1 104.
## 2 PWA         2  32.0  27.0 22.8  19.4 126.
## 3 PWA         3  29.2  24.6 13.6  15.0  74.6
## 4 control     1  28.3  19.2 32.3  13.8 168.
## 5 control     2  18.2  17.5  3.62 12.7  26.1
## 6 control     3  19.2  16.4  8.40 13.8  52.0
```

Description of participants

Gender:

```
(df %>% mutate(gender_char = case_when(gender == 1 ~ "female",
                                         gender == 2 ~ "male")) %>%
  group_by(type, gender_char) %>% count(gender_char) %>%
  mutate(n=n/160/3) -> gender)
```

```
## # A tibble: 4 x 3
## # Groups:   type, gender_char [4]
##   type   gender_char     n
##   <chr>   <chr>         <dbl>
## 1 PWA     female           3
## 2 PWA     male            17
## 3 control female           3
## 4 control male            17
```

```
# 1 = female, 2 = male, 3 = diverse
```

```
print("percentage female:")
```

```
## [1] "percentage female:"
```

```
sum(df$gender == 1)/nrow(df)
```

```
## [1] 0.15
```

Age:

```
print('age:')
```

```
## [1] "age:"
```

```
df %>% group_by(type) %>%
  summarise(mean=round(mean(age),2), sd=round(sd(age),2), min=min(age), max=max(age))
```

```
## # A tibble: 2 x 5
##   type   mean    sd   min   max
##   <chr> <dbl> <dbl> <int> <int>
## 1 PWA    53.4  5.43   39    62
## 2 control 53.2  5.72   38    62
```

Handedness:

```
# 1 = left handed, 2 = right handed, 3 = ambidexter/both
(handedness <- df %>% mutate(handedness_char = case_when(handedness == 1 ~ "left-handed",
                                                         handedness == 2 ~ "right-handed",
                                                         handedness==3 ~ "ambidexter/both"))) %>%
  group_by(type) %>% count(handedness_char))
```

```
## # A tibble: 5 x 3
## # Groups:   type [2]
##   type   handedness_char     n
##   <chr>   <chr>         <int>
## 1 PWA     left-handed     1600
## 2 PWA     right-handed    8000
## 3 control ambidexter/both   960
## 4 control left-handed    1120
## 5 control right-handed   7520
```

```

print("percentage right-handed:")

## [1] "percentage right-handed:"

print('PWA')

## [1] "PWA"

sum(df$handedness[df$type=="PWA"] == 2)/nrow(df[df$type=="PWA",])

## [1] 0.8333333

print('control')

## [1] "control"

sum(df$handedness[df$type=="control"] == 2)/nrow(df[df$type=="control",])

## [1] 0.7833333

```

Mother tongue (experiment was restricted to native German speakers): This seems to have worked

```

table(df$language) # 1 = yes (mother tongue is German), 2 = no

##
##      1
## 19200

```

Attention checks

1) Item vs. non-item

```

## Item vs. non-item
# CH01_01 (Taube), CH01_02 (Apfel), CH02_01 (Luftballon) and CH02_02 (Biene) are items and 2 should be .
# CH01_03 (Radio), CH01_04 (Sparschwein), CH02_03 (Laptop) and CH02_04 (Wattestäbchen) are non-items an
## Did participants cheat
# CH03 = 1 - yes, I worked through it till the end,
# CH03 = 2 - no, I stopped or cheated midway
# CH03 = -9 - no answer

attcheck <- data.frame(subject = unique(df$subject))

df <- df %>% mutate(itemvsnonitem1 =
  case_when(CH01_01==2 & CH01_02==2 & CH01_03==1 & CH01_04==1 ~2, # all correct
            CH01_01==2 | CH01_02==2 ~1, # one correct
            CH01_01!=2 & CH01_02!=2 ~0)) %>% # none correct
  dplyr::mutate(itemvsnonitem2 =
    case_when(CH02_01==2 & CH02_02==2 & CH02_03==1 & CH02_04==1 ~2,
              CH02_01==2 | CH02_02==2 ~1,
              CH02_01!=2 & CH02_02!=2 ~0))
df %>% group_by(type, session) %>% count(itemvsnonitem1) %>% mutate(n=n/160)

```

```
## # A tibble: 12 x 4
## # Groups:   type, session [6]
##   type      session itemvsnonitem1      n
##   <chr>      <int>      <dbl> <dbl>
## 1 PWA          1          0      1
## 2 PWA          1          1      6
## 3 PWA          1          2     13
## 4 PWA          2          1      1
## 5 PWA          2          2     19
## 6 PWA          3          1      2
## 7 PWA          3          2     18
## 8 control      1          1      4
## 9 control      1          2     16
## 10 control     2          2     20
## 11 control     3          1      1
## 12 control     3          2     19
```

```
df %>% group_by(type, session) %>% count(itemvsnonitem2) %>% mutate(n=n/160)
```

```
## # A tibble: 13 x 4
## # Groups:   type, session [6]
##   type      session itemvsnonitem2      n
##   <chr>      <int>      <dbl> <dbl>
## 1 PWA          1          1      2
## 2 PWA          1          2     18
## 3 PWA          2          0      1
## 4 PWA          2          1      1
## 5 PWA          2          2     18
## 6 PWA          3          0      1
## 7 PWA          3          1      2
## 8 PWA          3          2     17
## 9 control      1          1      1
## 10 control     1          2     19
## 11 control     2          1      1
## 12 control     2          2     19
## 13 control     3          2     20
```

```
# table(df$itemvsnonitem1)/160
# table(df$itemvsnonitem2)/160
```

All had at least one item selected correctly in the attention test

2) Cheating

```
df <- df %>% mutate(CH03 = case_when(CH03 == 1 ~
  " Ja, ich habe alles bis zum Ende bearbeitet.",
  CH03 == 2 ~
  "Nein, ich habe zwischendurch aufgehört oder geschummelt."))
df %>% group_by(type, session) %>% count(CH03) %>% mutate(n=n/160)
```

```
## # A tibble: 9 x 4
## # Groups:   type, session [6]
##   type      session CH03
```

n

```
##   <chr>      <int> <chr>                                     <dbl>
## 1 PWA        1 " Ja, ich habe alles bis zum Ende bearbeitet."      19
## 2 PWA        1 "Nein, ich habe zwischendurch aufgehört oder geschumme~  1
## 3 PWA        2 " Ja, ich habe alles bis zum Ende bearbeitet."      20
## 4 PWA        3 " Ja, ich habe alles bis zum Ende bearbeitet."      19
## 5 PWA        3 "Nein, ich habe zwischendurch aufgehört oder geschumme~  1
## 6 control    1 " Ja, ich habe alles bis zum Ende bearbeitet."      20
## 7 control    2 " Ja, ich habe alles bis zum Ende bearbeitet."      20
## 8 control    3 " Ja, ich habe alles bis zum Ende bearbeitet."      19
## 9 control    3 "Nein, ich habe zwischendurch aufgehört oder geschumme~  1
```

```
table(df$CH03)/160
```

```
##
##           Ja, ich habe alles bis zum Ende bearbeitet.
##                                     117
## Nein, ich habe zwischendurch aufgehört oder geschummelt.
##                                     3
```

Comments

Comments don't indicate any problems that should lead to participant exclusion:

```
table(df$comments)/160
```

```
## numeric(0)
```

Arrays

```
table(df$array)/160
```

```
##
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
##  5  4  3  3  4  4  5  5  4  3  4  4  5  3  4  3  7  5  5  3  3  4  3  3  5  4
## 27 28 29 30
##  4  3  3  5
```

```
x <- df %>% group_by(subject, session) %>% count(array) %>% mutate(n=n/160)
array_rep = 0
for(i in 2:nrow(x)){
  if(x$subject[i-1] == x$subject[i] & x$array[i-1] == x$array[i]){
    array_rep = array_rep+1
  }
}
print(paste0(array_rep, " time an array was repeated within the same participant"))
```

```
## [1] "1 time an array was repeated within the same participant"
```

```
df %>% group_by(type, session) %>% count(array) %>% mutate(n=n/160) %>% arrange(array)
```

```
## # A tibble: 94 x 4
## # Groups:   type, session [6]
##   type      session array      n
##   <chr>      <int> <int> <dbl>
## 1 PWA          1      1      3
## 2 PWA          3      1      1
## 3 control      2      1      1
## 4 PWA          2      2      1
## 5 PWA          3      2      2
## 6 control      2      2      1
## 7 control      1      3      1
## 8 control      2      3      1
## 9 control      3      3      1
## 10 PWA          1      4      1
## # i 84 more rows
```

```
df %>% count(array) %>% mutate(n=n/160)
```

```
##   array n
## 1     1 5
## 2     2 4
## 3     3 3
## 4     4 3
## 5     5 4
## 6     6 4
## 7     7 5
## 8     8 5
## 9     9 4
## 10    10 3
## 11    11 4
## 12    12 4
## 13    13 5
## 14    14 3
## 15    15 4
## 16    16 3
## 17    17 7
## 18    18 5
## 19    19 5
## 20    20 3
## 21    21 3
## 22    22 4
## 23    23 3
## 24    24 3
## 25    25 5
## 26    26 4
## 27    27 4
## 28    28 3
## 29    29 3
## 30    30 5
```

Get an overview of the computer setups

Information was read-out automatically from SoSci Survey SD22_PRV (provider): 1 = Android 2 = Apple 3 = BlackBerry 4 = PlayBook 5 = Kindle 6 = Microsoft -1 = Sonstige oder unbekannt

SD22_OS (OS): 10 = Windows NT 11 = Windows Vista 12 = Windows 7 13 = Windows 8 15 = Windows 10 20 = Linux 30 = MacOS X 40 = iOS 50 = Android -1 = Sonstiges -2 = unbekannt

SD22_BID (Browser-Identification (HTTP_USER_AGENT))

SD22_BNM (Browser) 1 = Android WebView 2 = Chrome 3 = Chromium 4 = Coast 5 = Firefox 6 = IE 7 = Opera 8 = Safari 9 = Edge 10 = Safari WebView 11 = IE (mobile) 12 = Edge (mobile) -1 = Sonstiges -2 = Default Browser -3 = unbekannt

SD22_BVS (Browser-version)

SD22_FmF (Format) 1 = Computer 2 = Fernsehgerät 3 = Tablet 4 = Mobilgerät 5 = Smartphone -2 = unbekannt

SD22_ScW (screen width [Pixel]) SD22_ScH (screen [Pixel])

SD22_QnW (questionnaire width [Pixel])

```
#colnames(df)

df <- df %>% mutate(SD22_OS=as.character(as.numeric(SD22_OS))) %>%
  mutate(os_system = case_when(
    SD22_OS==10 ~ "Windows NT",
    SD22_OS==11 ~ "Windows Vista",
    SD22_OS==12 ~ "Windows 7",
    SD22_OS==13 ~ "Windows 8",
    SD22_OS==15 ~ "Windows 10",
    SD22_OS==20 ~ "Linux",
    SD22_OS==30 ~ "MacOS X",
    SD22_OS==40 ~ "iOS",
    SD22_OS==50 ~ "Android",
    SD22_OS==-1 ~ "other",
    SD22_OS==-2 ~ "unknown")) %>%
  mutate(os_system_raw=case_when(
    SD22_OS==10 ~ "Windows",
    SD22_OS==11 ~ "Windows",
    SD22_OS==12 ~ "Windows",
    SD22_OS==13 ~ "Windows",
    SD22_OS==15 ~ "Windows",
    SD22_OS==20 ~ "Linux",
    SD22_OS==30 ~ "Apple",
    SD22_OS==40 ~ "Apple",
    SD22_OS==50 ~ "Android",
    SD22_OS==-1 ~ "other",
    SD22_OS==-2 ~ "unknown"))
df %>% group_by(os_system) %>% count() %>% mutate(n=n/40/3)
```

```
## # A tibble: 5 x 2
## # Groups:   os_system [5]
##   os_system      n
##   <chr>         <dbl>
## 1 MacOS X       12
```



```
## 2 Windows 10      32
## 3 Windows 7       4
## 4 Windows 8       4
## 5 unknown      108
```

```
df %>% group_by(os_system_raw) %>% count() %>% mutate(n=n/40/3)
```

```
## # A tibble: 3 x 2
## # Groups:   os_system_raw [3]
##   os_system_raw      n
##   <chr>          <dbl>
## 1 Apple           12
## 2 Windows          40
## 3 unknown         108
```

```
# df <- df %>% mutate(SD22_PRV=as.character(as.numeric(SD22_PRV))) %>%
#   mutate(provider = case_when(
#     SD22_PRV==1 ~ "Android",
#     SD22_PRV==2 ~ "Apple",
#     SD22_PRV==3 ~ "BlackBerry",
#     SD22_PRV==4 ~ "PlayBook",
#     SD22_PRV==5 ~ "Kindle",
#     SD22_PRV==6 ~ "Microsoft",
#     SD22_PRV==7 ~ "other or unknown"))
# df %>% group_by(provider) %>% count() %>% mutate(n=n/40/3)
```

```
#df %>% group_by(SD22_BID) %>% count() %>% mutate(n=n/40/3)
```

```
df <- df %>% mutate(SD22_BNM=as.character(as.numeric(SD22_BNM))) %>%
  mutate(browser = case_when(
    SD22_BNM==1 ~ "Android WebView",
    SD22_BNM==2 ~ "Chrome",
    SD22_BNM==3 ~ "Chromium",
    SD22_BNM==4 ~ "Coast",
    SD22_BNM==5 ~ "Firefox",
    SD22_BNM==6 ~ "IE",
    SD22_BNM==7 ~ "Opera",
    SD22_BNM==8 ~ "Safari",
    SD22_BNM==9 ~ "Edge",
    SD22_BNM==10 ~ "Safari WebView",
    SD22_BNM==11 ~ "IE (mobile)",
    SD22_BNM==12 ~ "Edge (mobile)",
    SD22_BNM==13 ~ "other",
    SD22_BNM==14 ~ "default browser",
    SD22_BNM==15 ~ "unknown"))
```

```
#df %>% group_by(SD22_BVS) %>% count() %>% mutate(n=n/40/3) # browser version
```

```
df <- df %>% mutate(SD22_FmF=as.character(as.numeric(SD22_FmF))) %>%
  mutate(system_format = case_when(
    SD22_FmF==1 ~ "Computer",
    SD22_FmF==2 ~ "Television",
```

```

SD22_FmF==3 ~ "Tablet",
SD22_FmF==4 ~ "Mobile device",
SD22_FmF==5 ~ "Smartphone",
SD22_BNM==2 ~ "unknown"))
df %>% group_by(system_format) %>% count() %>% mutate(n=n/40/3)

```

```

## # A tibble: 2 x 2
## # Groups:   system_format [2]
##   system_format      n
##   <chr>          <dbl>
## 1 Computer      153.
## 2 Tablet         6.67

```

```

df <- df %>% mutate(screen_width=as.character(as.numeric(SD22_ScW)))
df %>% group_by(screen_width) %>% count() %>% mutate(n=n/40/3)

```

```

## # A tibble: 10 x 2
## # Groups:   screen_width [10]
##   screen_width      n
##   <chr>          <dbl>
## 1 1024           4
## 2 1280          12
## 3 1366         58.7
## 4 1368           4
## 5 1440           8
## 6 1536          16
## 7 1600          12
## 8 1768           4
## 9 1920         37.3
## 10 2560           4

```

```

df <- df %>% mutate(screen_height=as.character(as.numeric(SD22_ScH)))
df %>% group_by(screen_height) %>% count() %>% mutate(n=n/40/3)

```

```

## # A tibble: 12 x 2
## # Groups:   screen_height [12]
##   screen_height      n
##   <chr>          <dbl>
## 1 1000           4
## 2 1080         21.3
## 3 1200          16
## 4 1440           4
## 5 720           5.33
## 6 768          62.7
## 7 800           6.67
## 8 864           16
## 9 900           12
## 10 912           4
## 11 960           4
## 12 992           4

```

```
df <- df %>% mutate(questionnaire_width=as.character(as.numeric(SD22_QnW)))
df %>% group_by(questionnaire_width) %>% count() %>% mutate(n=n/40/3)
```

```
## # A tibble: 2 x 2
## # Groups:   questionnaire_width [2]
##   questionnaire_width      n
##   <chr>                <dbl>
## 1 799                    4
## 2 800                   156
```

```
(df %>% group_by(subject, session) %>% select(SD22_BID, os_system_raw,
                                              os_system, browser,
                                              system_format,
                                              screen_width, screen_height,
                                              questionnaire_width) %>%
  unique()) -> setup_overview)
```

```
## Adding missing grouping variables: 'subject', 'session'
```

```
## # A tibble: 120 x 10
## # Groups:   subject, session [120]
##   subject session SD22_BID      os_system_raw os_system browser system_format
##   <int>   <int> <chr>          <chr>          <chr>   <chr>   <chr>
## 1     101     1 Mozilla/5.0 (W~ unknown      unknown default~ Computer
## 2     101     2 Mozilla/5.0 (W~ unknown      unknown default~ Computer
## 3     101     3 Mozilla/5.0 (W~ unknown      unknown default~ Computer
## 4     102     1 Mozilla/5.0 (W~ unknown      unknown default~ Computer
## 5     102     2 Mozilla/5.0 (W~ unknown      unknown default~ Computer
## 6     102     3 Mozilla/5.0 (W~ unknown      unknown default~ Computer
## 7     103     1 Mozilla/5.0 (W~ unknown      unknown default~ Computer
## 8     103     2 Mozilla/5.0 (W~ unknown      unknown default~ Computer
## 9     103     3 Mozilla/5.0 (W~ unknown      unknown default~ Computer
## 10    104     1 Mozilla/5.0 (M~ Apple      MacOS X   Firefox Computer
## # i 110 more rows
## # i 3 more variables: screen_width <chr>, screen_height <chr>,
## #   questionnaire_width <chr>
```

```
## save this table and anonymize by hand
write.csv(setup_overview,here::here("data", "transient_data_files", "setup_overview_all.csv"),
  row.names = FALSE)
```

Load manually cleaned overview and create table

```
setups <- read.csv2(here::here("data", "transient_data_files",
                              "setup_overview_cleaned.csv"))
setups %>% group_by(OS) %>% count()
```

```
## # A tibble: 6 x 2
## # Groups:   OS [6]
##   OS                n
##   <chr>          <int>
```

```
## 1 MacOS X 4
## 2 Samsung Galaxy Tab A 9.7 1
## 3 Windows 10 30
## 4 Windows 10 (Android on day2 and day8) 1
## 5 Windows 7 1
## 6 Windows 8 3
```

```
setups %>%
  mutate(group=case_when(subject<200 ~ "PWA", subject >200 ~"Control")) %>%
  group_by(group,browser) %>% count()
```

```
## # A tibble: 9 x 3
## # Groups:   group, browser [9]
##   group browser      n
##   <chr>   <chr>   <int>
## 1 Control Chrome     5
## 2 Control Edge      3
## 3 Control Firefox    9
## 4 Control Opera      2
## 5 Control SamsungBrowser 1
## 6 PWA      Chrome     2
## 7 PWA      Chrome (Edge on day8) 1
## 8 PWA      Edge     13
## 9 PWA      Firefox    4
```

```
setups %>% group_by(system_format) %>% count()
```

```
## # A tibble: 3 x 2
## # Groups:   system_format [3]
##   system_format      n
##   <chr>           <int>
## 1 Computer        38
## 2 Computer (Tablet on day2 and day8) 1
## 3 Tablet          1
```

```
setups %>% group_by(test_location) %>% count()
```

```
## # A tibble: 2 x 2
## # Groups:   test_location [2]
##   test_location      n
##   <chr>           <int>
## 1 clinic         12
## 2 home           28
```

```
setups %>% group_by(screen_width) %>% count()
```

```
## # A tibble: 11 x 2
## # Groups:   screen_width [11]
##   screen_width      n
##   <chr>           <int>
## 1 1024            1
```

```
## 2 1280 2
## 3 1366 14
## 4 1366 (1280 on day 8) 1
## 5 1368 1
## 6 1440 2
## 7 1536 4
## 8 1600 3
## 9 1768 1
## 10 1920 10
## 11 2560 1
```

```
setups %>%
  mutate(screen_width=case_when(screen_width=="1366 (1280 on day 8)" ~ 1366,
                                TRUE ~ as.numeric(as.character(screen_width)))) %>%
  summarise(min=min(screen_width), max=max(screen_width),
            mean=mean(screen_width), sd=sd(screen_width))
```

```
## Warning: There was 1 warning in 'mutate()'.
## i In argument: 'screen_width = case_when(...)'.
## Caused by warning:
## ! NAs introduced by coercion
```

```
##      min max    mean      sd
## 1 1024 2560 1569.85 295.9897
```

```
setups %>% group_by(screen_height) %>% count()
```

```
## # A tibble: 13 x 2
## # Groups:   screen_height [13]
##   screen_height      n
##   <chr>          <int>
## 1 1000            1
## 2 1080            6
## 3 1200            4
## 4 1440            1
## 5 720            1
## 6 768           15
## 7 768 (720 on day8) 1
## 8 800            1
## 9 864            4
## 10 900            3
## 11 912            1
## 12 960            1
## 13 992            1
```

```
setups %>%
  mutate(screen_height=case_when(screen_height=="768 (720 on day8)" ~ 768,
                                TRUE ~ as.numeric(as.character(screen_height)))) %>%
  summarise(min=min(screen_height), max=max(screen_height),
            mean=mean(screen_height), sd=sd(screen_height))
```

```
## Warning: There was 1 warning in 'mutate()'.
## i In argument: 'screen_height = case_when(...)'.
## Caused by warning:
## ! NAs introduced by coercion
```

```
## i In argument: 'screen_height = case_when(...)'
## Caused by warning:
## ! NAs introduced by coercion
```

```
##   min  max  mean      sd
## 1  720 1440 913.7 174.4639
```

```
setups %>% group_by(questionnaire_width) %>% count()
```

```
## # A tibble: 2 x 2
## # Groups:   questionnaire_width [2]
##   questionnaire_width    n
##           <int> <int>
## 1             799     1
## 2             800    39
```

Fully anonymize data and reduce data frame

```
# df_a <- df %>% dplyr::select(!"gender" & !starts_with("language") &
#                               !starts_with("handedness") &
#                               !starts_with("fingers") &
#                               !starts_with("KB") &
#                               !starts_with("CHO") &
#                               !"comments" & !"type" &
#                               #!"gender_char" &
#                               !"itemsnonitem1" &
#                               !"itemsnonitem2" &
#                               !starts_with("MC") &
#                               !starts_with("SD") &
#                               !c("os_system", "os_system_raw",
#                                   "browser", "screen_width",
#                                   "system_format",
#                                   "questionnaire_width",
#                                   "screen_height", "OR02_01", "array") &
#                               ! "age"
#                               # !"name" & !"time_wo_outlier"
#                               )
```

```
#write.csv(df_a, here::here("data", "data_long_anonymous.csv"))
```