

# German translation of the Artificial-Social-Agent questionnaire instrument for evaluating human-agent interaction

Transformation from raw data to the input files - first translation round, second half

Boleslav Khodakov and Emma Bokel, based on Fengxiang Li

23 May, 2023

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Data file Final_ASA_German_Round_1_Second_Half_anonym.sav</b>	<b>2</b>
<b>3</b>	<b>Transformation results as input data file for further analysis</b>	<b>6</b>
3.1	File transformed_data_round_1_half_2.sav . . . . .	6

## 1 Introduction

In this document, we transform a raw data file into an input data file. Because of privacy reasons we did not release for public access the raw data file.

There are two documents to be transformed. They are divided like the questionnaire - split in two. Two groups answered either questionnaire. In the second group, which this document is about, human-ASA interaction evaluation data of the last 46 items were collected from 30 bilingual participants with German as their primary language, and English as their fluent language. Bilingual participants rated human-ASA interaction on 46 English items and corresponding German translations. In additions, they answered 14 attention control questions.

We transformed the raw data file into an input data file for further analysis, which we describe in ‘Questionnaire translation and validation paper analysis’.

We use the following packages:

```
library(haven)      # Use read_sav function
library(dplyr)      # Use select function
library(knitr)       # Get markdown file
library(tinytex)     # Use TeX environment
```

```
library(rtticles) # Use CTeX documents template
library(pander) # For pandering tables
panderOptions("table.alignment.default", "left")
```

## 2 Data file Final\_ASA\_German\_Round\_1\_Second\_Half\_anonym.sav

We read the raw data file consisting of the last 46 English ASA item scores and corresponding German translation scores. The raw scores are from a 7-point scale, ranging from -3 (disagree), 0 (neither agree nor disagree) to 3 (agree). Before we could do any analysis, we first had to “clean up” the score of items by removing any characters other than numbers (except the minus sign for negative numbers).

```
# Read the file
d_ASA_1 <- read_sav("Final_ASA_German_Round_1_Second_Half_anonym.sav")

# Only take into account people who recommend using their data for scientific purposes
d_ASA_2 <- d_ASA_1[d_ASA_1$Use_Data == 'Yes',]

# Cleanup labels (delete the actual words, only numbers remain)
clean_labels <- function(data){
  data[data=="Neither agree nor disagree\n\n0"] <- "0"
  data[data=="Disagree\n\n-3"] <- "-3"
  data[data=="Agree\n\n3"] <- "3"
  data[data=="Weder noch\n\n0"] <- "0"
  data[data=="Stimme nicht zu\n\n-3"] <- "-3"
  data[data=="Stimme zu\n\n3"] <- "3"
  data[data==" \n\n1"] <- "1"
  data[data==" \n\n2"] <- "2"
  data[data==" \n\n-1"] <- "-1"
  data[data==" \n\n-2"] <- "-2"
  return(data)
}

# Clean up labels
d_ASA_1 <- clean_labels(d_ASA_1)
d_ASA_2 <- clean_labels(d_ASA_2)

# Make columns of question items into numeric types (previously was string)
d_ASA_1 <- d_ASA_1 %>% mutate(across("Q_ITEMS_372":"Q_ITEMS_487", as.numeric))
d_ASA_2 <- d_ASA_2 %>% mutate(across("Q_ITEMS_372":"Q_ITEMS_487", as.numeric))
```

In the first translation round, the labels on Qualtrics were arbitrary. To make things easier for the reader, question items are renamed to their respective names from the ASA questionnaire. The prefix ‘Q\_E\_’ means that the question item is English, the prefix ‘Q\_DE\_’ means that the item is German. Suffixes with numbers (e.g. ‘\_2’) for English items indicate a duplicate. Suffixes with numbers (e.g. ‘\_2’) for German items indicate an alternative translation. This allows us to have a 1:1 match - each German item matches one English item. Attention checks are also renamed to have a unified ‘AC’ in their name.

```
rename_attention <- function(data){
  # Rename German Attention Checks
  colnames(data)[colnames(data) == "Q_ITEMS_300"] = "AttentionCheck_DE_1"
  colnames(data)[colnames(data) == "Q_ITEMS_301"] = "AttentionCheck_DE_2"
  colnames(data)[colnames(data) == "Q_ITEMS_302"] = "AttentionCheck_DE_3"
  colnames(data)[colnames(data) == "Q_ITEMS_303"] = "AttentionCheck_DE_4"
  colnames(data)[colnames(data) == "Q_ITEMS_304"] = "AttentionCheck_DE_5"
  colnames(data)[colnames(data) == "Q_ITEMS_305"] = "AttentionCheck_DE_6"
  colnames(data)[colnames(data) == "Q_ITEMS_306"] = "AttentionCheck_DE_7"
  # Rename English Attention Checks
  colnames(data)[colnames(data) == "Q_ITEMS_292"] = "AttentionCheck_E_1"
  colnames(data)[colnames(data) == "Q_ITEMS_293"] = "AttentionCheck_E_2"
  colnames(data)[colnames(data) == "Q_ITEMS_294"] = "AttentionCheck_E_3"
  colnames(data)[colnames(data) == "Q_ITEMS_295"] = "AttentionCheck_E_4"
  colnames(data)[colnames(data) == "Q_ITEMS_296"] = "AttentionCheck_E_5"
  colnames(data)[colnames(data) == "Q_ITEMS_297"] = "AttentionCheck_E_6"
  colnames(data)[colnames(data) == "Q_ITEMS_298"] = "AttentionCheck_E_7"
  return(data)
}
```

```
# Rename English items to constructs names
rename_english <- function(data){
  colnames(data)[colnames(data) == "Q_ITEMS_372"] = "Q_E_UE1"
  colnames(data)[colnames(data) == "Q_ITEMS_374"] = "Q_E_UE2"
  colnames(data)[colnames(data) == "Q_ITEMS_376"] = "Q_E_UE3"
  colnames(data)[colnames(data) == "Q_ITEMS_379"] = "Q_E_UT1"
  colnames(data)[colnames(data) == "Q_ITEMS_381"] = "Q_E_UT2"
  colnames(data)[colnames(data) == "Q_ITEMS_382"] = "Q_E_UT3"
  colnames(data)[colnames(data) == "Q_ITEMS_384"] = "Q_E_UAL1"
  colnames(data)[colnames(data) == "Q_ITEMS_385"] = "Q_E_UAL2"
  colnames(data)[colnames(data) == "Q_ITEMS_386"] = "Q_E_UAL3"
  colnames(data)[colnames(data) == "Q_ITEMS_387"] = "Q_E_UAL4"
  colnames(data)[colnames(data) == "Q_ITEMS_390"] = "Q_E_UAL5"
```

```

colnames(data)[colnames(data) == "Q_ITEMS_391"] = "Q_E_UAL6"
colnames(data)[colnames(data) == "Q_ITEMS_393"] = "Q_E_AA1"
colnames(data)[colnames(data) == "Q_ITEMS_394"] = "Q_E_AA2"
colnames(data)[colnames(data) == "Q_ITEMS_395"] = "Q_E_AA3"
colnames(data)[colnames(data) == "Q_ITEMS_398"] = "Q_E_R_AC1"
colnames(data)[colnames(data) == "Q_ITEMS_400"] = "Q_E_R_AC2"
colnames(data)[colnames(data) == "Q_ITEMS_401"] = "Q_E_R_AC3"
colnames(data)[colnames(data) == "Q_ITEMS_402"] = "Q_E_R_AC4"
colnames(data)[colnames(data) == "Q_ITEMS_403"] = "Q_E_AI1"
colnames(data)[colnames(data) == "Q_ITEMS_404"] = "Q_E_AI2"
colnames(data)[colnames(data) == "Q_ITEMS_405"] = "Q_E_R_AI3"
colnames(data)[colnames(data) == "Q_ITEMS_407"] = "Q_E_AI4"
colnames(data)[colnames(data) == "Q_ITEMS_408"] = "Q_E_AT1"
colnames(data)[colnames(data) == "Q_ITEMS_409"] = "Q_E_AT2"
colnames(data)[colnames(data) == "Q_ITEMS_412"] = "Q_E_R_AT3"
colnames(data)[colnames(data) == "Q_ITEMS_414"] = "Q_E_SP1"
colnames(data)[colnames(data) == "Q_ITEMS_415"] = "Q_E_SP2"
colnames(data)[colnames(data) == "Q_ITEMS_417"] = "Q_E_SP3"
colnames(data)[colnames(data) == "Q_ITEMS_418"] = "Q_E_IIS1"
colnames(data)[colnames(data) == "Q_ITEMS_419"] = "Q_E_IIS2"
colnames(data)[colnames(data) == "Q_ITEMS_421"] = "Q_E_IIS3"
colnames(data)[colnames(data) == "Q_ITEMS_422"] = "Q_E_IIS4"
colnames(data)[colnames(data) == "Q_ITEMS_423"] = "Q_E_AEI1"
colnames(data)[colnames(data) == "Q_ITEMS_424"] = "Q_E_AEI2"
colnames(data)[colnames(data) == "Q_ITEMS_425"] = "Q_E_R_AEI3"
colnames(data)[colnames(data) == "Q_ITEMS_426"] = "Q_E_AEI4"
colnames(data)[colnames(data) == "Q_ITEMS_427"] = "Q_E_R_AEI5"
colnames(data)[colnames(data) == "Q_ITEMS_428"] = "Q_E_UEP1"
colnames(data)[colnames(data) == "Q_ITEMS_429"] = "Q_E_UEP2"
colnames(data)[colnames(data) == "Q_ITEMS_430"] = "Q_E_UEP3"
colnames(data)[colnames(data) == "Q_ITEMS_432"] = "Q_E_UEP4"
colnames(data)[colnames(data) == "Q_ITEMS_English_489"] = "Q_E_UAI1"
colnames(data)[colnames(data) == "Q_ITEMS_English_491"] = "Q_E_UAI2"
colnames(data)[colnames(data) == "Q_ITEMS_English_492"] = "Q_E_UAI3"
colnames(data)[colnames(data) == "Q_ITEMS_English_493"] = "Q_E_UAI4"
return(data)
}

```

*# Rename German items to constructs names*

```
rename_german <- function(data){  
  colnames(data)[colnames(data) == "Q_ITEMS_German_489"] = "Q_DE_UE1"  
  colnames(data)[colnames(data) == "Q_ITEMS_440"] = "Q_DE_UE2"  
  colnames(data)[colnames(data) == "Q_ITEMS_441"] = "Q_DE_UE2_2"  
  colnames(data)[colnames(data) == "Q_ITEMS_442"] = "Q_DE_UE3"  
  colnames(data)[colnames(data) == "Q_ITEMS_443"] = "Q_DE_UT1"  
  colnames(data)[colnames(data) == "Q_ITEMS_444"] = "Q_DE_UT2"  
  colnames(data)[colnames(data) == "Q_ITEMS_445"] = "Q_DE_UT2_2"  
  colnames(data)[colnames(data) == "Q_ITEMS_488"] = "Q_DE_UT2_3"  
  colnames(data)[colnames(data) == "Q_ITEMS_446"] = "Q_DE_UT3"  
  colnames(data)[colnames(data) == "Q_ITEMS_447"] = "Q_DE_UAL1"  
  colnames(data)[colnames(data) == "Q_ITEMS_448"] = "Q_DE_UAL1_2"  
  colnames(data)[colnames(data) == "Q_ITEMS_449"] = "Q_DE_UAL2"  
  colnames(data)[colnames(data) == "Q_ITEMS_450"] = "Q_DE_UAL3"  
  colnames(data)[colnames(data) == "Q_ITEMS_451"] = "Q_DE_UAL4"  
  colnames(data)[colnames(data) == "Q_ITEMS_452"] = "Q_DE_UAL5"  
  colnames(data)[colnames(data) == "Q_ITEMS_453"] = "Q_DE_UAL6"  
  colnames(data)[colnames(data) == "Q_ITEMS_454"] = "Q_DE_AA1"  
  colnames(data)[colnames(data) == "Q_ITEMS_455"] = "Q_DE_AA2"  
  colnames(data)[colnames(data) == "Q_ITEMS_456"] = "Q_DE_AA3"  
  colnames(data)[colnames(data) == "Q_ITEMS_457"] = "Q_DE_R_AC1"  
  colnames(data)[colnames(data) == "Q_ITEMS_458"] = "Q_DE_R_AC2"  
  colnames(data)[colnames(data) == "Q_ITEMS_459"] = "Q_DE_R_AC3"  
  colnames(data)[colnames(data) == "Q_ITEMS_460"] = "Q_DE_R_AC4"  
  colnames(data)[colnames(data) == "Q_ITEMS_461"] = "Q_DE_AI1"  
  colnames(data)[colnames(data) == "Q_ITEMS_462"] = "Q_DE_AI2"  
  colnames(data)[colnames(data) == "Q_ITEMS_463"] = "Q_DE_R_AI3"  
  colnames(data)[colnames(data) == "Q_ITEMS_464"] = "Q_DE_AI4"  
  colnames(data)[colnames(data) == "Q_ITEMS_465"] = "Q_DE_AT1"  
  colnames(data)[colnames(data) == "Q_ITEMS_466"] = "Q_DE_AT2"  
  colnames(data)[colnames(data) == "Q_ITEMS_467"] = "Q_DE_R_AT3"  
  colnames(data)[colnames(data) == "Q_ITEMS_468"] = "Q_DE_SP1"  
  colnames(data)[colnames(data) == "Q_ITEMS_469"] = "Q_DE_SP2"  
  colnames(data)[colnames(data) == "Q_ITEMS_470"] = "Q_DE_SP3"  
  colnames(data)[colnames(data) == "Q_ITEMS_471"] = "Q_DE_IIS1"  
  colnames(data)[colnames(data) == "Q_ITEMS_472"] = "Q_DE_IIS2"  
  colnames(data)[colnames(data) == "Q_ITEMS_473"] = "Q_DE_IIS3"  
  colnames(data)[colnames(data) == "Q_ITEMS_474"] = "Q_DE_IIS4"
```

```

colnames(data)[colnames(data) == "Q_ITEMS_475"] = "Q_DE_AEI1"
colnames(data)[colnames(data) == "Q_ITEMS_476"] = "Q_DE_AEI2"
colnames(data)[colnames(data) == "Q_ITEMS_477"] = "Q_DE_R_AEI3"
colnames(data)[colnames(data) == "Q_ITEMS_478"] = "Q_DE_AEI4"
colnames(data)[colnames(data) == "Q_ITEMS_479"] = "Q_DE_R_AEI5"
colnames(data)[colnames(data) == "Q_ITEMS_480"] = "Q_DE_UEP1"
colnames(data)[colnames(data) == "Q_ITEMS_481"] = "Q_DE_UEP2"
colnames(data)[colnames(data) == "Q_ITEMS_482"] = "Q_DE_UEP3"
colnames(data)[colnames(data) == "Q_ITEMS_483"] = "Q_DE_UEP4"
colnames(data)[colnames(data) == "Q_ITEMS_484"] = "Q_DE_UAI1"
colnames(data)[colnames(data) == "Q_ITEMS_485"] = "Q_DE_UAI2"
colnames(data)[colnames(data) == "Q_ITEMS_486"] = "Q_DE_UAI3"
colnames(data)[colnames(data) == "Q_ITEMS_487"] = "Q_DE_UAI4"
return(data)
}

```

```

#Rename columns for "all" and "recommended" dataframes
d_ASA_1 <- rename_attention(d_ASA_1)
d_ASA_1 <- rename_english(d_ASA_1)
d_ASA_1 <- rename_german(d_ASA_1)
d_ASA_2 <- rename_attention(d_ASA_2)
d_ASA_2 <- rename_english(d_ASA_2)
d_ASA_2 <- rename_german(d_ASA_2)

```

### 3 Transformation results as input data file for further analysis

#### 3.1 File transformed\_data\_round\_1\_half\_2.sav

We removed scores of attention control questions and other irrelevant data ('Finished' and 'rand\_id'), retaining ratings of English items and corresponding German translations. In case some of participants didn't complete the questionnaire survey, the resulting null values were omitted by omitting data rows of these participants all-together. The code excluding people failing attention checks (from the Chinese translation codebase) is still in place. However, for the current study we already excluded participants failing the checks on Qualtrics. **Please note: Since these participants are not paid for their submission, we cannot use or publish their data in any way.**

```

# Select only question items from dataframe (incl. attention checks)
# all participants
dd1 <- data.frame(select(d_ASA_1, AttentionCheck_E_1:Q_DE_UAI4))
# "recommended" participants

```

```

dd2 <- data.frame(select(d_ASA_2, AttentionCheck_E_1:Q_DE_UAI4))

# Filter out participant entries based on failed attention checks
attention_check_filtering <- function(data){
  # Select attention check questions
  # Note: for German translation rounds,
  # no single row should be removed based
  # on failed attention checks. This is already done in Qualtrics.
  # Note: This code chunk remains for 1) the final evaluation 2) legacy compatibility
  i <- grep("AttentionCheck",colnames(data))

  # Select desired answers for attention check questions
  Atten <- c(-3,3,3,-3,0,3,-3,0,-3,3,2,3,-3,1)

  # The following code and comments still remain from the Chinese translation code
  # with minor adjustments:

  x <- NULL # Row number of participant who failed the attention check
  for (j in (1:nrow(data))){
    # Find participants who failed attention check in 'dd1'
    count <- 0
    # The number of incorrectly answered attention control questions of each participant
    for (k in 1:14){

      if (as.numeric(data[[i[k]]][j])!=Atten[k]) # Check whether each participant's

        # attention control question answers are consistent with the correct answers
        count <- count+1

    }

    if (count>2)
      # Row number of the participant who failed more than two
      # attention control questions were added to 'x'
      x <- append(x,j)
      # Participants who failed more than two attention control questions
  }
  m <- length(x) # The number of participants who failed attention check
  if (m!=0)

```

```

data <- data[-x,] # Participants who failed attention check were excluded

# Print whether any participants failed attention checks
cat("Amount of participants who failed attention checks: ", m,
    "\nAmount of participants left after filtering based on attention checks: ", nrow(data), "\n")
return(data)
}

```

```

#Perform filtering on "all" and "recommended" dataframes

```

```
dd1 <- attention_check_filtering(dd1)
```

```
## Amount of participants who failed attention checks: 0
```

```
## Amount of participants left after filtering based on attention checks: 30
```

```
dd2 <- attention_check_filtering(dd2)
```

```
## Amount of participants who failed attention checks: 0
```

```
## Amount of participants left after filtering based on attention checks: 28
```

All participants' evaluation data were included as none of the participants failed any of the attention control questions (logically). For the formative translations, we did not yet reverse scores of questionnaire items and German translations with R indication. This can be omitted during the translation step however. The code is still there for the summative evaluation, where it will be used. Thus, we obtained the output data `transformed_data_round_1_half_2.sav` for further evaluation.

```

# Select scores of 46 English items and corresponding German translations

```

```

# all participants

```

```
d1 <- as.data.frame(select(dd1, Q_E_UE1:Q_E_UAI4, Q_DE_UE1:Q_DE_UAI4))
```

```

# "recommended" participants

```

```
d2 <- as.data.frame(select(dd2, Q_E_UE1:Q_E_UAI4, Q_DE_UE1:Q_DE_UAI4))
```

```

# For English items which have multiple German translations,

```

```

# duplicate the English items x times (x = amount of extra German translations)

```

```
duplicate <- function(data){
```

```
  data <- data %>%
```

```
    mutate(Q_E_UE2_2 = Q_E_UE2, .after = "Q_E_UE2")
```

```
  data <- data %>%
```

```
    mutate(Q_E_UT2_2 = Q_E_UT2, .after = "Q_E_UT2")
```

```
  data <- data %>%
```

```
    mutate(Q_E_UT2_3 = Q_E_UT2_2, .after = "Q_E_UT2_2")
}
```



```

data <- data %>%
  mutate(Q_E_UAL1_2 = Q_E_UAL1, .after = "Q_E_UAL1")
return(data)
}

# Perform duplication on both dataframes
d1 <- duplicate(d1)
d2 <- duplicate(d2)

# ASIMO is hardcoded to 0
d1$AgentID <- 0 # Add a column 'AgentID' to facilitate analysis for comparison
d2$AgentID <- 0 # Add a column 'AgentID' to facilitate analysis for comparison

```