

# HUX2021 / Dokumentation

HA Empirische Methoden der Sprachwissenschaft / Meinschäfer/Schirakowski SS20 FUB

St. Schwarz

2023-06-11

## Contents

<b>1</b>	<b>draft abstract</b>	<b>1</b>
<b>2</b>	<b>Zum Experiment</b>	<b>1</b>
<b>3</b>	<b>Theorierahmen</b>	<b>2</b>
<b>4</b>	<b>Dokumentation</b>	<b>2</b>
4.1	Design . . . . .	2
4.2	Zum Inhalt des Tests . . . . .	3
4.3	Auswertung . . . . .	3
4.4	Studienansicht des Experiments . . . . .	8
<b>B. REF:</b>		<b>8</b>

---

to self compile the RMarkdown source to this document in RStudio cf. Schwarz (2021)

---

## 1 draft abstract

Ich werde in dieser Arbeit den Versuch unternehmen, ein im Rahmen empirischer Studien durchgeführtes psycholinguistisches Experiment ausführlich zu dokumentieren. Die Studie wurde in einem Forschungsseminar der BUA<sup>1</sup> in Kooperation Studierender von FU, TU, HU unter der Aufsicht der Charite durchgeführt. Es handelt sich in diesem Teil um die partielle Replikation des von Paula Rubio-Fernandez (Rubio-Fernández, Cummins, and Tian 2016) unternommenen Versuchs, das Metaphernverständnis von schizophrenen Personen zu untersuchen. Dazu wurde in der Kontrollgruppe ein *self paced reading* Experiment veranstaltet und ausgewertet.

## 2 Zum Experiment

Das Experiment hat in einer Laufzeit von zwei Wochen Daten von 46 Versuchspersonen erhoben, die hauptsächlich in der Statistiksprache R ausgewertet wurden. Basis der Daten ist die Messung von Lesezeiten in einem set von 24 items<sup>2</sup>, die in vierfacher Ausprägung literale und figurative Elemente in kurzen Zusammenhängen verbanden.

---

<sup>1</sup>Berlin University Alliance

<sup>2</sup>*item* hier als ein in vier Ausprägungen entworfener Text mit durchschnittlicher Länge von 0 Zeichen.

### 3 Theorierahmen

Mit Carston (Carston 2010) sind wir davon ausgegangen, dass im Experiment zu untersuchende *single* (SM = Metapher eingebettet in literalen Kontext) und *extended metaphors* (EM = erweiterte Metapher, eingebettet in weitgehend figurativen Kontext) unterschiedlich verarbeitet werden, wobei durch die notwendige Aktivierung und Reaktivierung der literalen Bedeutung (ad hoc concept) an SM ein höherer Aufwand nötig sei, der im Gegensatz zu EM in einer längeren Lesezeit resultiert. Diese Annahme führt zu folgenden Fragestellungen, die in Anlehnung an ein Experiment (Verifikation/Replikation) von Paula Rubio-Fernandez et al. (Rubio-Fernández, Cummins, and Tian 2016) hier aufgenommen wurden:

1. unterscheiden sich die Lesezeiten kurzer Texte, die nach den Annahmen folgendem Muster single und extended metaphors enthalten oder rein literal (LC) aufgebaut sind?
2. wie unterscheidet sich eine vierte Variante, bei der die Struktur der SM invertiert wurde (ISM = literales Element eingebettet in weitgehend figurativen Kontext), hinsichtlich ihrer Verarbeitungszeit?

Die zu überprüfenden Hypothesen lauten wie folgt:

1. Lesezeit SM > Lesezeit EM
2. Lesezeit EM = Lesezeit LC
3. Lesezeit ISM = SM > EM = LC

Zur Überprüfung der Hypothesen wurde ein *self paced reading* Experiment durchgeführt, das die Lesezeit in den unterschiedlichen Kategorien erfasste.

### 4 Dokumentation

Ich werde im Folgenden den Aufbau und Ablauf des Experiments erklären und im nächsten Abschnitt dessen hypothesengeleitete Auswertung erläutern.

#### 4.1 Design

Die Versuchspersonen bekamen während des Tests, der online über einen link abrufbar war, ein set aus jeweils 8 items (Kontexte) die wiederum die Varianten der Variable (Einbettung der Metapher) in Ausprägungen von

- LC (Metapher literal in literalem Kontext)
- SM (Metapher figurativ in literalem Kontext)
- EM (Metapher figurativ in figurativem Kontext)
- ISM (Metapher literal in figurativem Kontext)

enthielten sowie einen Bestand aus 16 filler items, die sich ebenfalls in den Kontextvarianten unterschieden. Die Probandinnen lasen einen Text, von dem jeweils nur eine Zeile nicht maskiert war, die jeweils folgende Zeile wurde durch das Drücken der Leertaste sichtbar. So konnte die Verweildauer auf einer sichtbaren Zeile gemessen werden. Als Grundlage für die Messung wurde ein von (Rose 2022) frei verfügbares javascript benutzt, das an unsere Bedürfnisse angepasst worden war; u.a. war ursprünglich keine dauerhafte Speicherung der Daten in einer Datei vorgesehen, diese Funktion wurde von mir mit einigem Aufwand realisiert, so dass die Lesezeiten danach in einer Tabelle verfügbar waren.

Die Einbettung des Experiments in einen auf der Plattform soscisurvey als akademische Studie angemeldeten Fragebogen erforderte ebenfalls einigen Aufwand, da die Randomisierung der Itemabfrage in anderer Weise als vom script vorgesehen geschehen musste.

Die Daten selbst enthielten Werte für:

- den Start und Endzeitpunkt des durchgeführten Tests (tnid)
- die Position der einzelnen maskierten Phrasen im Satz (targetposition) (unabhängige Variable)
- die Verweildauer auf den jeweiligen Positionen (elapsed time) (abhängige Variable)
- Angaben über die Zuordnung der Phrasen zur jeweiligen Ausprägung (Kategorie) der unabhängigen Variablen

## 4.2 Zum Inhalt des Tests

Die items, die in Anlehnung an R/F nach obigem Muster von uns entworfen wurden, entsprachen im Aufbau dem set, das R/F in ihrem Test verwendet hatte; vier ihrer im Anhang zur Verfügung gestellten items hatten wir ins deutsche übersetzt und eines davon als item in vier Varianten, die restlichen als filler übernommen. Die Auswahl der verwendeten items wurde gemeinschaftlich nach Kriterien wie Konsistenz innerhalb der items, Stil und Kohärenz bestimmt. Ebenfalls nach den constraints von R/F gerichtete Vorgaben wie Wort- und Phrasenlänge (`median characters/phrase = 44`), Itemlänge (`median characters/item = 0`) und Position des targets hatten beim Entwurf eine wichtige Rolle gespielt, waren jedoch nicht immer optimal umgesetzt worden; dazu einige Bemerkungen im Auswertungsteil. Ein item sollte eine Metapher in je figurativer und literaler Bedeutung, eingebettet in einen je figurativen oder literalen Kontext, d.h. also in insgesamt vier unterschiedlichen Konfigurationen, präsentieren. Die Metapher (unser target) sollte sich im dritten Viertel des Kontextes befinden, um davor und danach auftretende Latenzen der Lesezeit auswerten zu können. Dafür wurden im script des Tests die Positionen der einzeln zu lesenden Phrasen des items mit 0 für das target und je negativen und positiven Werten für den Abstand zum target festgelegt.<sup>3</sup>

Das set eines einzelnen durch Abruf des Fragebogens aufgerufenen Tests bestand für jede Studienteilnehmerin aus einer randomisierten Auswahl einer von vier mal acht Itemvarianten (aus der Menge von 8 items), womit jeweils zweifach Meszdaten der vier Varianten pro TN erhoben werden konnten, aber niemals ein TN ein item in mehreren Varianten las. Die 16 filler items traten innerhalb des sets aus also 24 items an zufälligen Positionen auf, wurden aber bei allen TN aus demselben pool bedient und wiesen ausgeglichene Variation hinsichtlich der figurativ-literalen Konfiguration auf.

## 4.3 Auswertung

O.a. Daten wurden in R (RStudio and PBC 2022) 1. deskriptiv und 2. mittels des Paketes lme4 (Bates et al. 2015) zur Erhebung kovariater Abhängigkeiten (linear mixed model) analog der Vorgaben (Rubio-Fernández, Cummins, and Tian 2016) ausgewertet. Das script dazu kann hier, sowie eine visualisierte, vom user nach eigenen Parametern anzupassende deskriptive Auswertung hier nachvollzogen werden. Für eine ausschnittshaften Visualisierung cf. (Figure 1, 2, 3).

### 4.3.1 a.

Zur Auswertung wurden die erhobenen durch ein php-script im flachen comma-separated Format gespeicherten Lesezeiten in R importiert. Um zu gewährleisten, dass Abweichungen hinsichtlich der Zeichenanzahl der zeitgemessenen Phrasen keine unerwünschten Effekte in der Berechnung der Latenzen zeitigen, wurde bei jeder weiteren Berechnung dafür ein von der Zeichenanzahl/Phrase anhängiger Koeffizient einbezogen, der in der ersten, deskriptiven Auswertung einfach aus der Zeichenanzahl, bei der zweiten kovariaten Auswertung mittels lme4 bestimmt (Fine et al. 2013) und als Residual (hier: korrigierte Lesezeit) einbezogen wurde. R/F hatten ihre Daten mit dem R-package lme4 unter multivariaten Gesichtspunkten analysiert; so konnte auch ich (in dieser Phase ging die Experiment-Arbeitsgruppe getrennte Wege) nach Feststellung des Effekts der Phrasenlänge und Berücksichtigung desselben als korrigierte Lesezeit (hier: die abhängige Variable in der lmer Formel) einen signifikanten Effekt der Itemkategorie auf die Lesezeiten von SM und ISM (im Dataset “MM”) feststellen. Bei der Multivarianzanalyse wurden hier ein *main effect* der Kategorie (LC, SM, EM,

<sup>3</sup>da diese Positionen erst nach: Start des Tests konsequent im script festgelegt wurden, musste bei der Auswertung einiges manuell nachgearbeitet werden, um die Daten konsistent zu halten.

ISM = group) unter *random effects* von TN (participant) und Kategorie auf TN (group by participant) vorausgesetzt.

#### 4.3.2 b.

Deskriptiv konnten am target keine signifikanten Unterschiede in der durchschnittlichen Lesezeit an den verschieden eingebetteten Metaphern festgestellt werden. Da mir noch wesentliche Grundlagen zur Interpretation der Ergebnisse der Multivarianzanalyse fehlen und es sich hier nur um die kurze Dokumentation des Experiments im Seminarkontext handeln soll, die endlich abgeschlossen werden soll, werde ich darauf verzichten, meine Ergebnisse der multivariaten Berechnungen zur Grundlage einer dadurch ansatzweisen Bestätigung der 2. und 3. Hypothesen ( $ISM > LC = EM$ ) zu stilisieren. Sie seien hier jedoch kurz angeführt - die Hierarchie der Lesezeiten wäre dann (model estimates) wie folgt zu lesen:  $ISM > EM = LC > SM$ , in Kontrast zur Hypothese  $ISM = SM > EM = LC$  (cf. lmer summary):

```
## rtc ~ XvsGr + (1 | item) + (1 | participant) + (1 + XvsGr | participant)
## <environment: 0x7fb38d690290>

## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: lmeforms$sold[[1]]
## Data: dset
##
## REML criterion at convergence: 13279.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.6774 -0.6315 -0.1277  0.4835  3.8334
##
## Random effects:
## Groups      Name      Variance Std.Dev. Corr
## participant (Intercept) 102555   320.2
##              XvsGr2EM   338946   582.2   -0.41
##              XvsGr3LC    90446   300.7   -0.11  0.81
##              XvsGr4MM   19995   141.4    0.61  0.41  0.54
## participant.1 (Intercept) 432559   657.7
## item          (Intercept)  8760    93.6
## Residual              550737   742.1
## Number of obs: 819, groups: participant, 39; item, 8
##
## Fixed effects:
##              Estimate Std. Error    df t value Pr(>|t|)
## (Intercept)  -317.38     131.85   38.96  -2.407   0.0209 *
## XvsGr3LC      147.50      90.30   28.00   1.633   0.1136
## XvsGr2EM      151.68     120.56   34.53   1.258   0.2168
## XvsGr4MM      202.74      75.51   35.93   2.685   0.0109 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr) XvG3LC XvG2EM
## XvsGr3LC -0.287
## XvsGr2EM -0.242  0.572
## XvsGr4MM -0.188  0.378  0.480
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00234563 (tol = 0.002, component 1)
```

### 4.3.3 c.

Grob liesze sich durch die allein deskriptive Auswertung die Aussage machen, dasz sich bei target = -1, 0, 1, Vernachlässigung ungültiger Fälle = 1, outliers discarded = 1 & korrigierter RT (selection = -1, 0, 1, 1, 1, 0) folgende Hierarchie der Lesezeiten ergibt:

LC > ISM > SM > EM. Dabei gilt es zu beachten, dasz der hier grösste Unterschied der zwischen EM und LC mit 416ms zu messen ist, übersetzt unterhalb einer halben Sekunde liegt. (Zum Vergleich: ein ununterbrochenes Weiterklicken der lesbaren Zeilen mittels der space-Taste zeigt ca. 200ms Latenz.)

Siehe dazu den Ausschnitt **TI + rtc = length corrected reading time**

	RT	group	LZ	dif_SM	dif_EM	dif_LC	dif_ISM
6	TI + rtc	EM	1927	-10	NA	-416	-389
5	TI + rtc	SM	1937	NA	10	-406	-379
8	TI + rtc	ISM	2316	379	389	-27	NA
7	TI + rtc	LC	2343	406	416	NA	27

aus der folgenden Tabelle:<sup>4</sup>

RT	group	LZ	dif_SM	dif_EM	dif_LC	dif_ISM
TimeInterval	SM	2256	NA	117	-242	-190
TimeInterval	EM	2139	-117	NA	-359	-307
TimeInterval	LC	2499	242	359	NA	53
TimeInterval	ISM	2446	190	307	-53	NA
TI + rtc	SM	1937	NA	10	-406	-379
TI + rtc	EM	1927	-10	NA	-416	-389
TI + rtc	LC	2343	406	416	NA	27
TI + rtc	ISM	2316	379	389	-27	NA
TI char	SM	2449	NA	4	-2	-24
TI char	EM	2445	-4	NA	-6	-27
TI char	LC	2451	2	6	NA	-22
TI char	ISM	2473	24	27	22	NA

Bei isoliertem target (target = 0, 0, 0) ergibt sich folgende Hierarchie der Lesezeiten:

EM < SM < LC < ISM mit dem grössten Unterschied zwischen EM und ISM mit 400ms.

	RT	group	LZ	dif_SM	dif_EM	dif_LC	dif_ISM
6	TI + rtc	EM	1388	-157	NA	-364	-400
5	TI + rtc	SM	1545	NA	157	-207	-243
7	TI + rtc	LC	1753	207	364	NA	-36
8	TI + rtc	ISM	1788	243	400	36	NA

### 4.3.4 d. LEGENDE

- **LZ:** berechnete Lesezeit (response time)
- **TimeInterval:** uncorrected response time (TI)
- **TI + RTC:** TimeInterval + lmer residuals of TI dependent on phrase length
- **TI char:** TimeInterval corrected by mean phrase length (TI/characters\*mean(characters))

<sup>4</sup>Für weitere Ansichten und Studiensample besuchen Sie bitte o.a. interaktive Visualisierungen.

#### 4.3.5 p. visuals

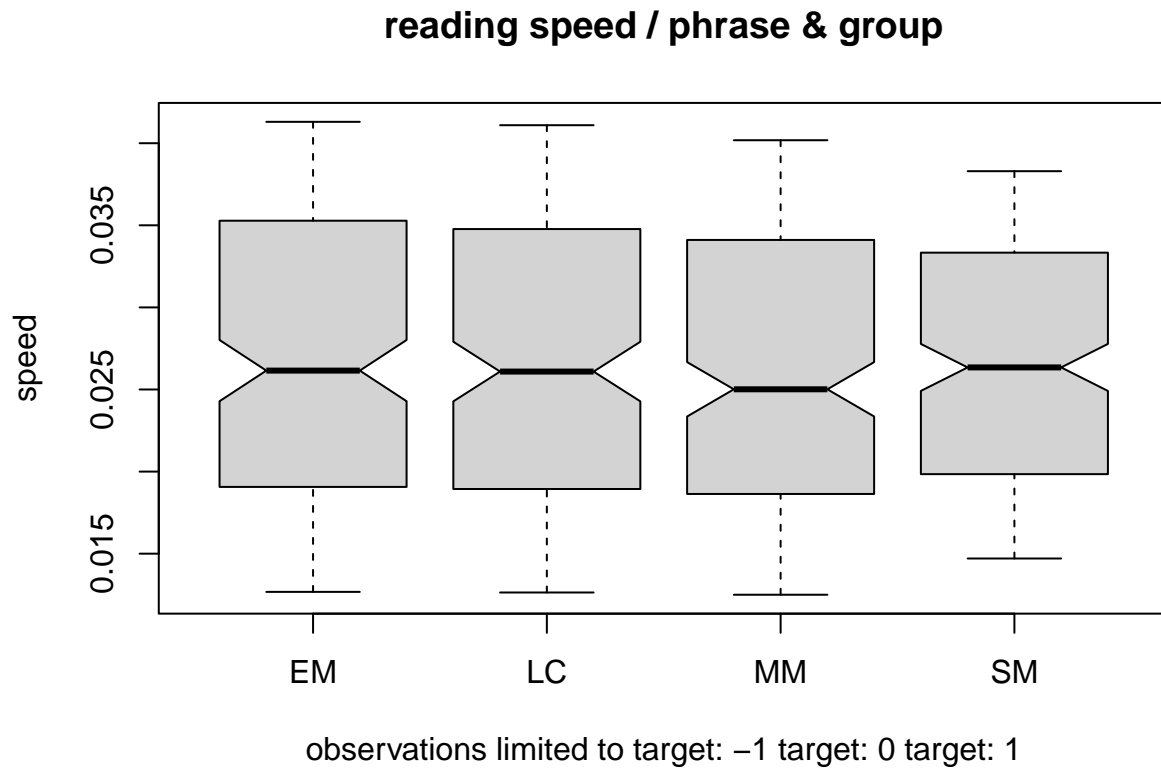


Figure 1: evaluate target, target-preceding, target-following phrase, not length corrected

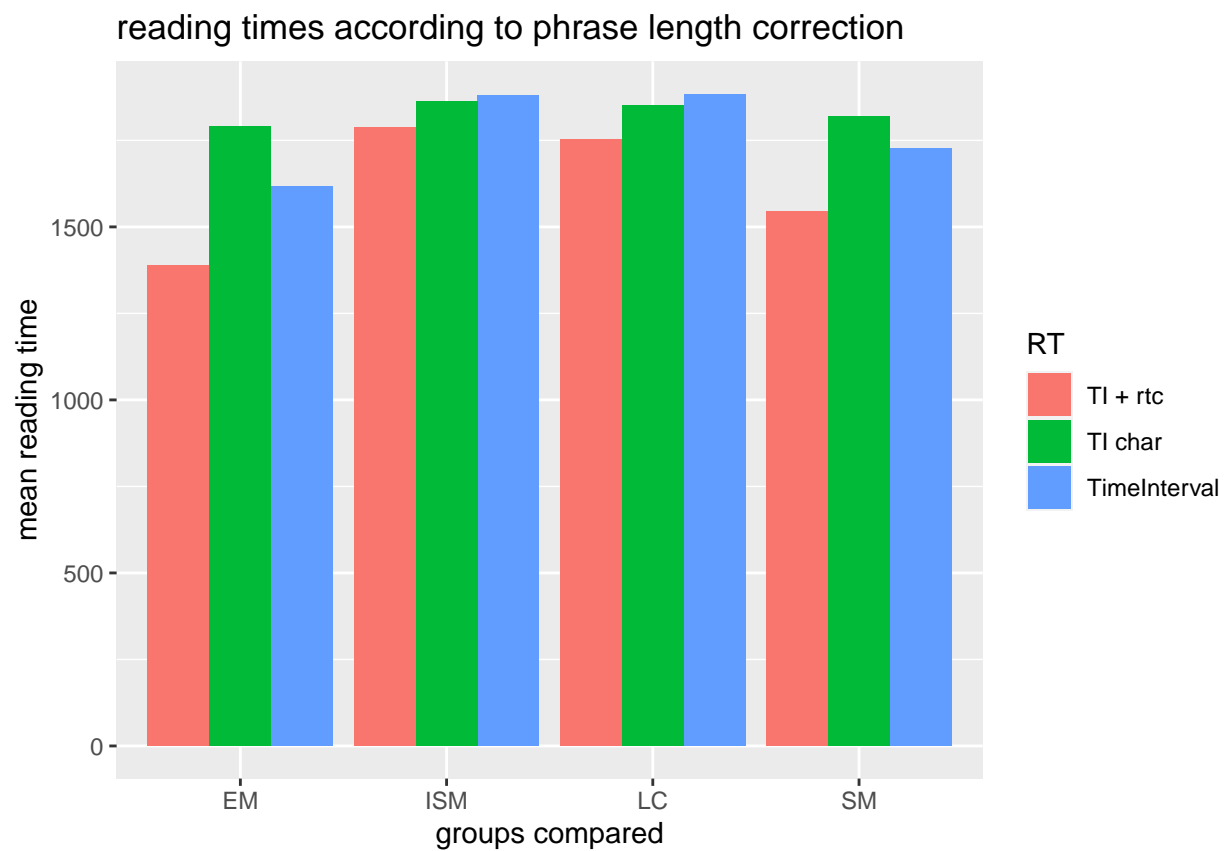


Figure 2: reading times (uncorrected = timeinterval) and with 2 different approaches of processing the phrase length

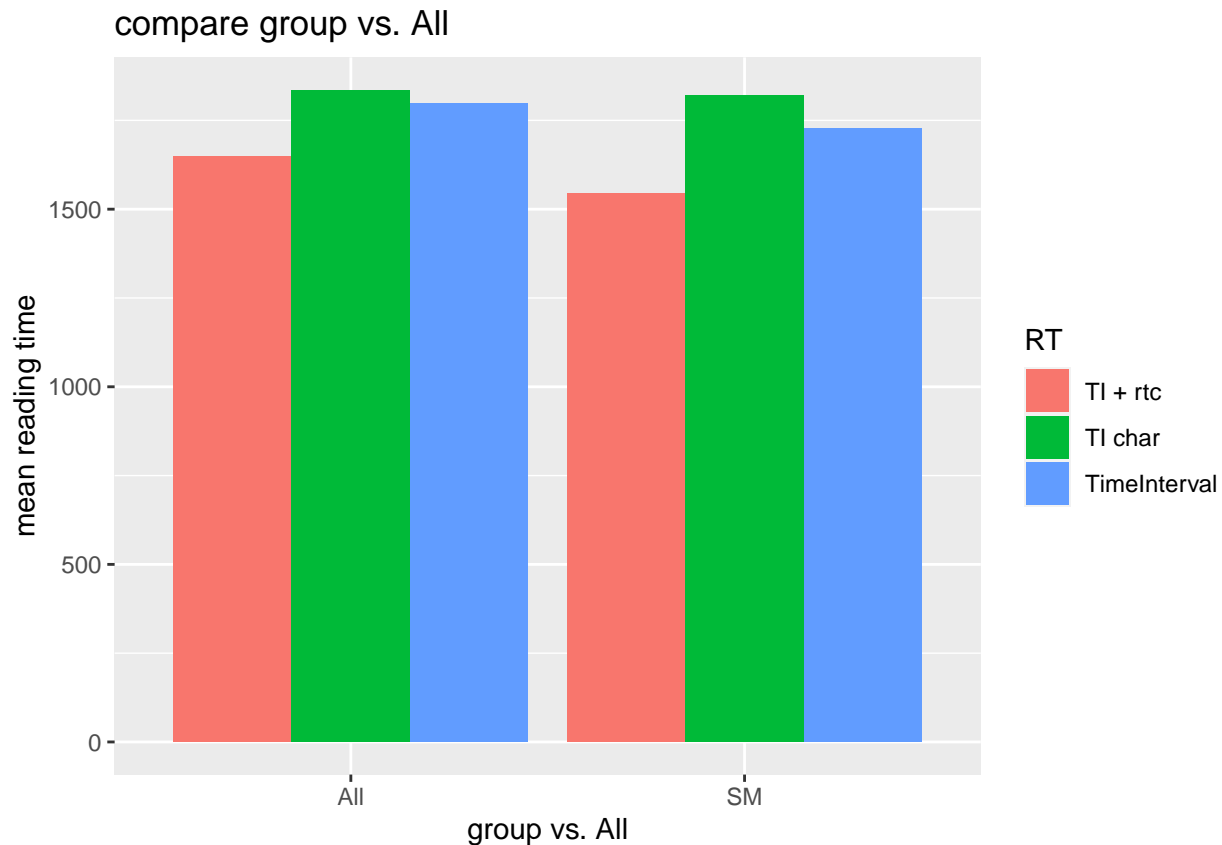


Figure 3: comparing SM vs. other

#### 4.4 Studienansicht des Experiments

cf. (Figure 4)

wenn Sie sich in der .pdf-Ausgabe der Arbeit befinden, können Sie hier die Ansicht aufrufen.

sonst hier ein Auszug aus der verwendeten Version, der im .pdf nicht erscheint.

## B. REF:

- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using Lme4.” *Journal of Statistical Software* 67 (1): 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Carston, Robyn. 2010. “XIII-Metaphor: Ad Hoc Concepts, Literal Meaning and Mental Images.” *Proceedings of the Aristotelian Society* 110 (3pt3): 295–321. <https://doi.org/10.1111/j.1467-9264.2010.00288.x>.
- Fine, Alex B., T. Florian Jaeger, Thomas A. Farmer, and Ting Qian. 2013. “Rapid Expectation Adaptation During Syntactic Comprehension.” *PLOS ONE* 8 (10): e77661. <https://doi.org/10.1371/journal.pone.0077661>.
- Rose, Ralph. 2022. “JESPR.” <https://github.com/fildpauz/jespr>.
- RStudio, and PBC. 2022. “RStudio IDE.” <https://www.rstudio.com/products/rstudio/download/>.
- Rubio-Fernández, Paula, Chris Cummins, and Ye Tian. 2016. “Are Single and Extended Metaphors Processed Differently? A Test of Two Relevance-Theoretic Accounts.” *Journal of Pragmatics* 94 (March): 15–28. <https://doi.org/10.1016/j.pragma.2016.01.005>.
- Schwarz, St. 2021. “RMarkdown Source to This Document.” [https://github.com/esteschwarz/DH\\_essais/blob/main/sections/hux2021/hux\\_HA/hux\\_ha\\_main.Rmd](https://github.com/esteschwarz/DH_essais/blob/main/sections/hux2021/hux_HA/hux_ha_main.Rmd).



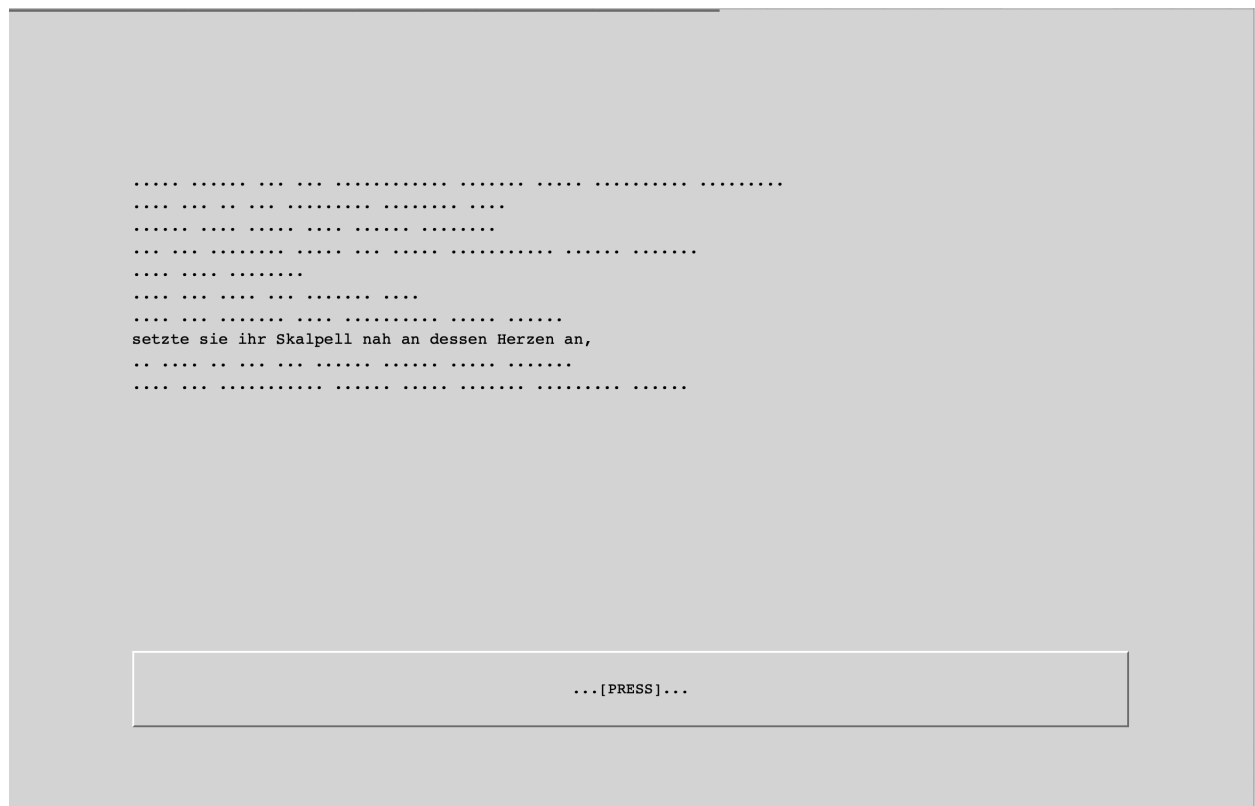


Figure 4: Figure 4: Experiment screenshot