LEAD-ME

EYE TRACKING IN MEDIA ACCESSIBILITY RESEARCH
– METHODS, TECHNOLOGIES AND DATA ANALYSES

Using Linear Mixed Models to Analyse Subtitle Reading

Agnieszka Szarkowska
& Breno Silva
University of Warsaw

7 July 2021

Plan for today

1. Previous eye-tracking research on subtitling
2. Linear mixed-effects models (LMMs)
3. Hands-on walkthrough on how to do LMMs in SPSS based on a dataset from the SURE Project

Eye tracking research on subtitling

Scleral reflection eye-tracking technique

Muylaert, W., Nootens, J., Poesmans, D., & Pugh, A. K. (1983). Design and utilisation of subtitles on foreign language television programmes. In P. H. Nelde (Ed.), *Theorie, Methoden and Modelle der Kontaktlinguistik (pp. 201-214). Dummler.*

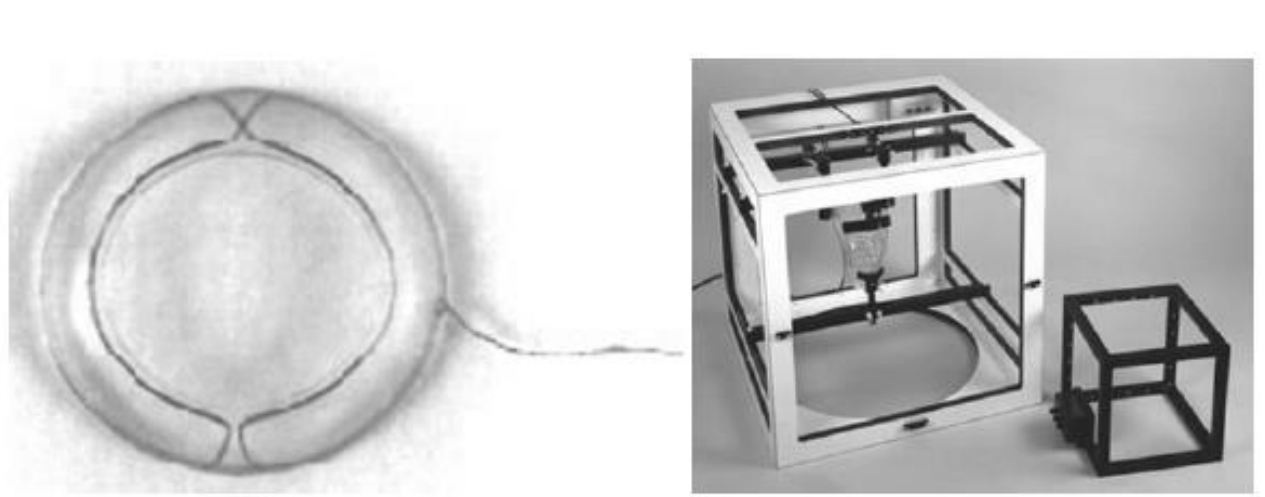


Fig. 5.2. Example of search coil embedded in contact lens and electromagnetic field frames for search coil eye movement measurement. Courtesy of Skalar Medical, Delft, The Netherlands <http://www.skalar.nl>. Reproduced with permission.

Duchowski, A. T. (2003). Eye Tracking Techniques. In *Eye Tracking Methodology: Theory and Practice (pp. 55-65). Springer London. https://doi.org/10.1007/978-1-4471-3750-4_5*

"Such data are not easily worked nor are they readily and accurately capable of relating to the objects seen so that one may say with confidence where a subject is looking."

Prof. Géry d'Ydewalle

Department of Psychology,
University of Leuven

d'Ydewalle, G. (1984). Processing TV information and eye movements research. Readings on Cognitive Ergonomics - Mind and Computers, Berlin, Heidelberg.

d'Ydewalle, G., Muylle, P., & Rensbergen, J. v. (1985). Attention shifts in partially redundant information situations. In R. Groner, G. W. McConkie, & C. Menz (Eds.), Eye Movements and Human Information Processing (pp. 375-384). Elsevier.

d'Ydewalle, G., Rensbergen, J. v., & Pollet, J. (1987). Reading a message when the same message is available auditorily in another language: the case of subtitling. In J. K. O'Regan & A. Levy-Schoen (Eds.), Eye movements: from physiology to cognition (pp. 313-321). Elsevier.

d'Ydewalle, G., Praet, C., Verfaillie, K., & Van Rensbergen, J. (1991). Watching subtitled television: automatic reading behavior. Communication Research, 18(5), 650-666.

De Bruycker, W., & d'Ydewalle, G. (2003). Reading native and foreign language television subtitles in children and adults. In J. Hyönä, R. Radach, & H. Deubel (Eds.), The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research (pp. 671-684). North-Holland. https://doi.org/10.1016/B978-044451020-4/50036-0

d'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. European Psychologist, 12(3), 196-205. https://doi.org/10.1027/1016-9040.12.3.196

D'Ydewalle's ET research on subtitling

- Subtitle reading
  - Automatic reading behaviour
  - Visual attention distribution between subtitle and image
- Subtitle reading depends on:
  - Subtitle characteristics (speed, language, number of lines)
  - Viewer characteristics (age, familiarity with subtitles, language proficiency)

Equipment

PDP11 computer

TV sets in the 1980s

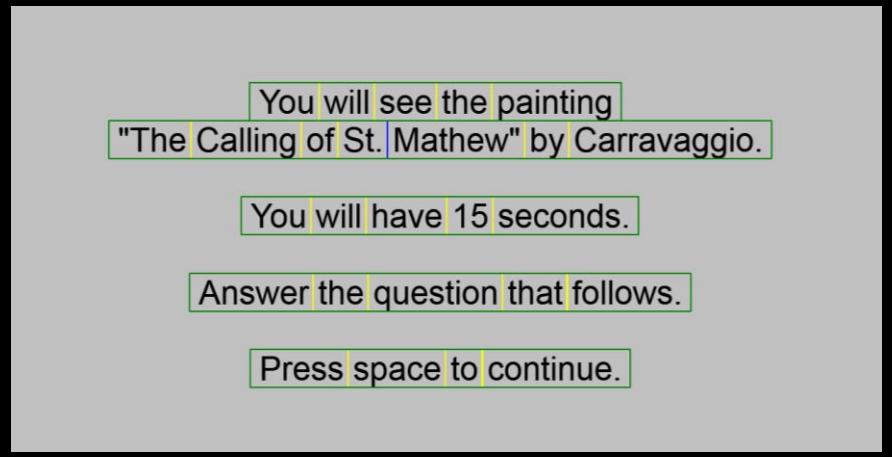
http://zgastelewizor.blogspot.com/2015/08/koledzy-rubina-z-poki.html

The equipment (DEBIC 80), based on the pupil-center corneal-reflection method, provided a registration sampling rate of 50 Hertz (a sampling every 20 milliseconds). A head-tracking mirror system gave the subject considerable freedom to seat in a relaxed way in front of a regular TV screen (approximately 180 cm distance).

SMI
SensoMotoric Instruments

Precision in studies on subtitling

- AOIs on subtitles
- AOIs on words
- Threshold line: subtitle vs. image
  (Perego et al. 2010, Bisson et al. 2012)
- Automatic AOIs on text:



Manual AOI on entire subtitle

Manual AOIs on words

EyeLink and Python-based scripts

Automatic AOIs for:

- Each subtitle
- Each word in the subtitle

Courtesy of Jan-Louis Kruger

Original eye tracking experiment

SMI RED
250 mobile

Original experimental design

2 x 3 mixed factorial design

- Within-subject variable: subtitling speed
  - 12 cps (slow)
  - 20 cps (fast)
- Between-subject variable: mother tongue
  - Polish
  - Spanish
  - English
- Mixed ANOVA analyses

Experimental design for today's LMM analysis

- Only Polish participants
- Repeated measures design
- Independent variable: subtitling speed
  - 12 cps (slow)
  - 20 cps (fast)
- Dependent variable
  - Proportional reading time (PRT)

Proportional reading time (PRT)

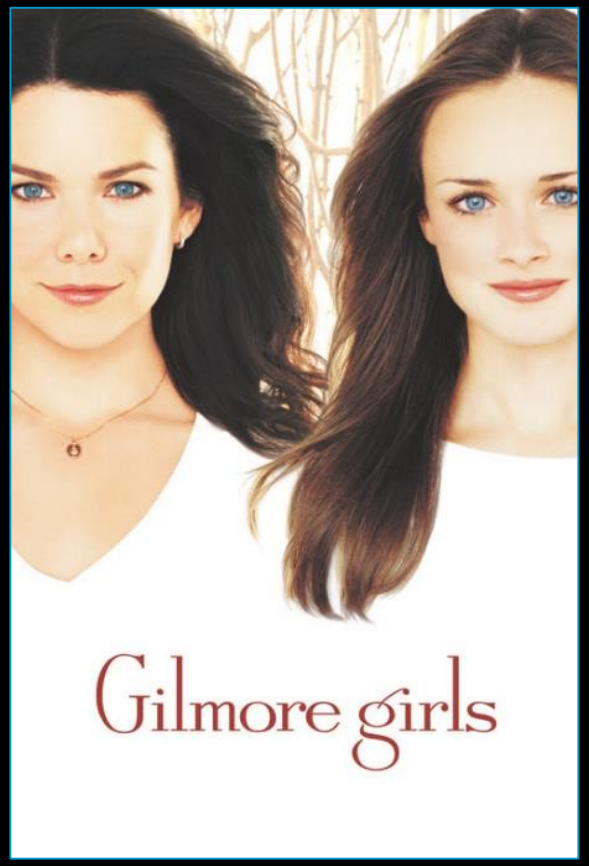
- The percentage of dwell time a participant spent in the AOI as a function of subtitle display time

  Subtitle duration = 4 seconds (4000 ms)
  participant spent 2 seconds (2000 ms)
  PRT = 2000/4000 ms = 50%
  = while the subtitle was displayed for 4 seconds,
  the participant was looking at that subtitle for 50% of the time

- Compare the proportion of time spent in the subtitles between 12 and 20 cps
- If PRT approaches 100% → participants spent most of their time reading subtitles and do not have time to look at onscreen action

Experimental videos

GF



GG

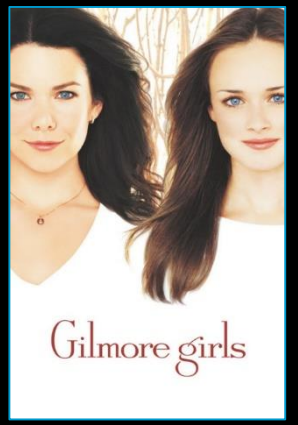
Comparability of the clips

Grace and Frankie (GF)	Gilmore Girls (GG)
4 mins 22 sec	4 min 41 sec
▪ Past-paced dialogue-heavy ▪ 2-4 people talking ▪ Similar readability scores — 1-2 on 7-point scale in Jasnopis — 8-9 points in FOG index	

COUNTERBALANCING

PARTICIPANT	Grace and Frankie	Gilmore Girls
P01	12 cps	20 cps
P02	20 cps	12 cps

GF



GG

		Speed		
Clip	Number of lines	12 cps	20 cps	Total
Grace and Frankie	1 line	254	225	479
	2 lines	297	241	538
Gilmore Girls	1 line	202	135	337
	2 lines	443	599	1042
Total	1 line	456	360	816
	2 lines	740	840	1580
	Total	1196	1200	2396

Subtitle data selection

- Tracking ratio: min. 80%
- Only subtitles which were looked at
- Fixation duration: between 80 ms and 500 ms

Traditional statistical analyses used in previous eye-tracking studies on subtitling

- T-test & Mann-Whitney U test
  - Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010, 2010/08/31). The cognitive effectiveness of subtitle processing. Media Psychology, 13(3), 243-272.
  - Kruger, J.-L., Hefer, E., & Matthew, G. (2013). *Measuring the impact of subtitles on cognitive load: eye tracking and dynamic audiovisual texts* Proceedings of the 2013 Conference on Eye Tracking South Africa, Cape Town, South Africa.
- ANOVA
  - Bisson, M.-J., Van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. Applied Psycholinguistics, 35(02), 399-418.
  - Szarkowska, A., & Gerber-Morón, O. (2018). Viewers can keep up with fast subtitles: Evidence from eye movements. PLoS ONE, 13(6).

Data in most previous studies

- Averaging data from all subtitles into one data point per participant
  - Data from different words
  - Data from different subtitles
- Wide data format

Participant	PRT_12cps	PRT_20cps
P01pl	23,91	35,11
P03pl	35,49	37,13
P04pl	35,74	13,43
P05pl	35,81	45,95
P06pl	51,46	42,45
P08pl	31,87	28,01
P09pl	18,93	36,71
P10pl	45,11	51,87
P11pl	27,67	39,11
P12pl	42,53	44,68
P14pl	41,61	46,11
P15pl	38,32	60,35
P16pl	36,37	30,89
P17pl	38,42	56,53
P18pl	41,45	48,85
P19pl	51,55	68,27
P20pl	28,95	27,55
P21pl	25,50	51,49

Problems with these analyses

- Lack of granularity
- Disregarding differences between particular subtitles and words
- Insufficient accounting for individual differences between participants and subtitles

Long data format for linear mixed models (LMMs) analyses

Participant	Speed	Actual_speed	Clip	Subtitle	Lines	PRT
P01pl	12 cps	11,33	Grace and Frankie	Sol próbuje teraz powiedzieć,	1 line	6,09
P01pl	20 cps	20,19	Gilmore Girls	znajdującego się w sumiew ...	2 lines	8,46
P01pl	20 cps	20,77	Gilmore Girls	- Co my tu właściwie robimy...	2 lines	9,30
P01pl	12 cps	12,80	Grace and Frankie	Nie. To nie ma sensu.	1 line	9,51
P01pl	12 cps	12,90	Grace and Frankie	Cześć, kochanie.	1 line	12,56
P01pl	12 cps	11,80	Grace and Frankie	Rozdział się skończy.Przej...	2 lines	13,37
P01pl	12 cps	11,95	Grace and Frankie	Cóż jest aż tak ważnego,że...	2 lines	14,07
P01pl	12 cps	12,50	Grace and Frankie	Jestem zakochany z Solu.S...	2 lines	14,33
P01pl	12 cps	13,64	Grace and Frankie	- Ja powiem.- Nie.	2 lines	14,85
P01pl	12 cps	12,16	Grace and Frankie	Wy jesteście wspólnikami,n...	2 lines	15,22
P01pl	12 cps	12,22	Grace and Frankie	Zaczynam coś odczuwać.	1 line	16,45
P01pl	20 cps	19,53	Gilmore Girls	Czyżby to było za często?	1 line	16,81
P01pl	20 cps	20,00	Gilmore Girls	- Przygotuj się.- Na co?	2 lines	16,99
P01pl	12 cps	12,93	Grace and Frankie	A także zmiany.	1 line	17,24
P01pl	20 cps	20,00	Gilmore Girls	- A dla kogo?- Dla mnie.	2 lines	17,34
P01pl	12 cps	12,16	Grace and Frankie	Gdy zaczyna się nowy rozd...	1 line	17,98
P01pl	12 cps	12,86	Grace and Frankie	- Proszę.- Robert.	2 lines	18,28
P01pl	12 cps	12,30	Grace and Frankie	- Chcemy się pobrać.- Pobr...	2 lines	18,85
P01pl	12 cps	12,38	Grace and Frankie	- Chłopaków jeszcze nie ma...	2 lines	18,89
P01pl	20 cps	19,77	Gilmore Girls	- Marnuję swoje życie.- Niep...	2 lines	20,00
P01pl	20 cps	19,23	Gilmore Girls	Ale ty to co innego.	1 line	20,38

LMMs allow for subtitle-based analyses

- Number of words per subtitle
- Number of characters per subtitle
- Duration of subtitle
- Number of lines (1 or 2)

LMMs allow for word-based analyses

- Word frequency
- Word length
- Word type (lexical vs. grammatical)
- Concreteness
- Cognateness
- Part of speech

Eye-tracking studies with LMMs

- Orrego-Carmona, D. (2015). *The reception of (non)professional subtitling.* Universitat Rovira i Virgili. PhD Thesis. Tarragona.
- Ragni, Valentina (2016) *More than Meets the Eye: A Reception Study on the Effects of Translation on Noticing and Memorisation of L2 Reverse Subtitles.* PhD thesis. University of Leeds.
- Liao, S., Yu, L., Reichle, E. D., & Kruger, J.-L. (2020). Using Eye Movements to Study the Reading of Subtitles in Video. *Scientific Studies of Reading*, 1-19. https://doi.org/10.1080/10888438.2020.1823986
- Lång, J., Vrzakova, H., & Mehtätalo, L. (2021). Modelling Gaze Behaviour in Subtitle Processing: The Effect of Structural and Lexical Properties. *Journal of Audiovisual Translation, 4*(1), 71-95. https://doi.org/10.47476/jat.v4i1.2021
- Szarkowska, A., Silva, B., & D. Orrego-Carmona (submitted) Effects of subtitle speed on viewers' gaze: re-analysing proportional reading time with linear mixed-effects models.

Linear mixed models (LMMs): Fixed effects AND random effects

Fixed effects

- Variables you would add in a regression or ANOVA
- Usually the variables of primary interest that would be used again if the experiment was replicated. Examples:
  - Factor (2 levels): Speed (e.g., Subtitles at 12 cps or 20 cps)
  - Factor (3 levels): Language of audio (e.g., Polish subtitles with no audio, English audio, Polish audio)
  - Covariates (subtitle level): number of words/characters per subtitle, duration of subtitle
  - Covariates (word level): part of speech, frequency, length.

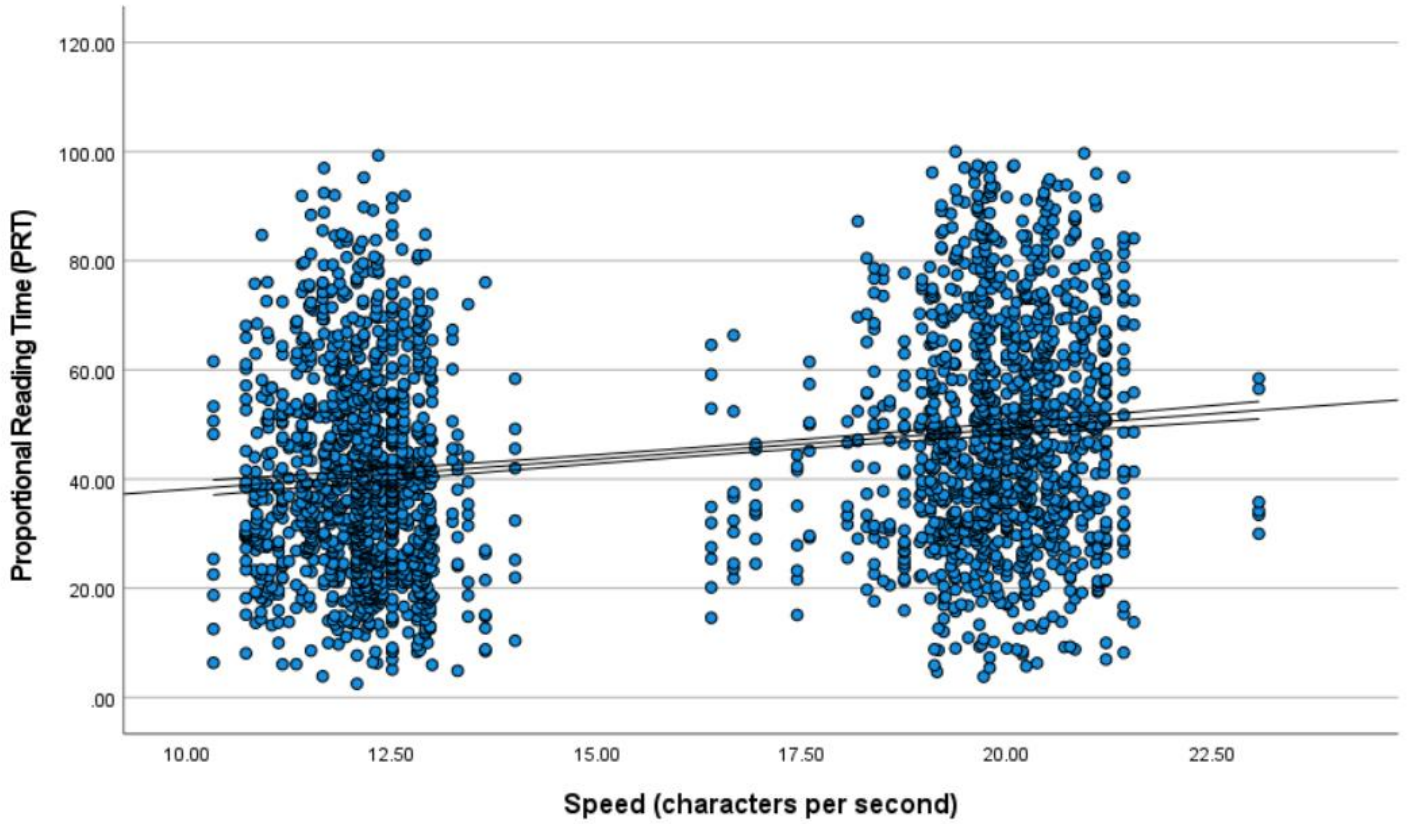
Random effects (the advantage of LMMs)

- Contextual or grouping factors (e.g., subtitles, participants, location of the experiment, country of provenance, clips)
- Helps explain more variance (error) in the data that fixed effects cannot explain
  - Some sources of variance are unknown
  - Some we can't/didn't measure
  - Some were measured, but do not explain 100%

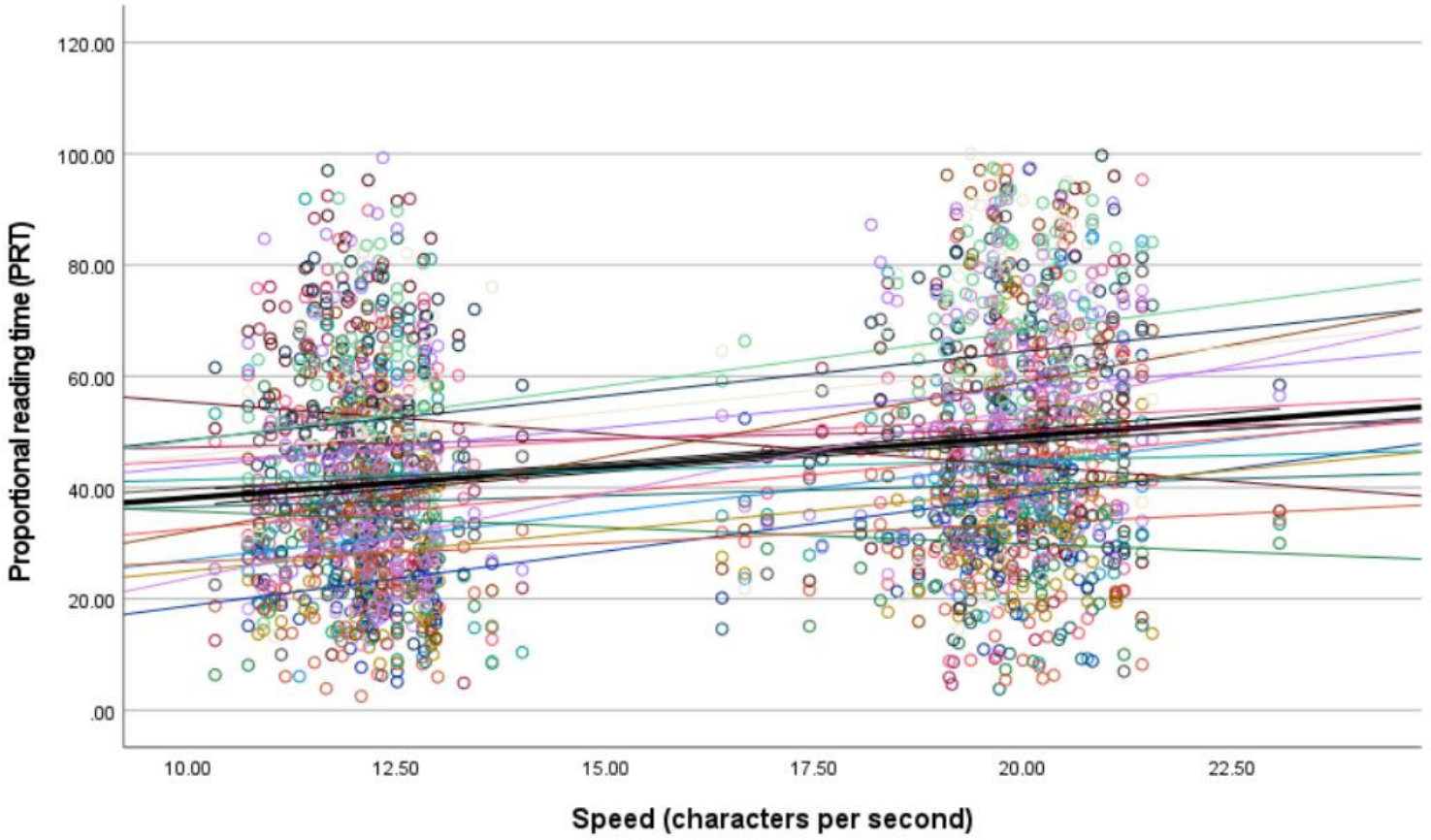
Random effects (the advantage of LMMs)

- Including random effects means the following:
  - More error explained = more precision = more reliable results
  - You're able to check if the fixed effects (e.g., different Speed) are significant OVER AND ABOVE differences between, e.g., participants and subtitles.
  - "*If a fixed effect [e.g., Speed] is significant in such a model, this means it is significant after the variance associated with subject and items [subtitles] is simultaneously controlled for*". (Jaeger, 2008, p. 444).

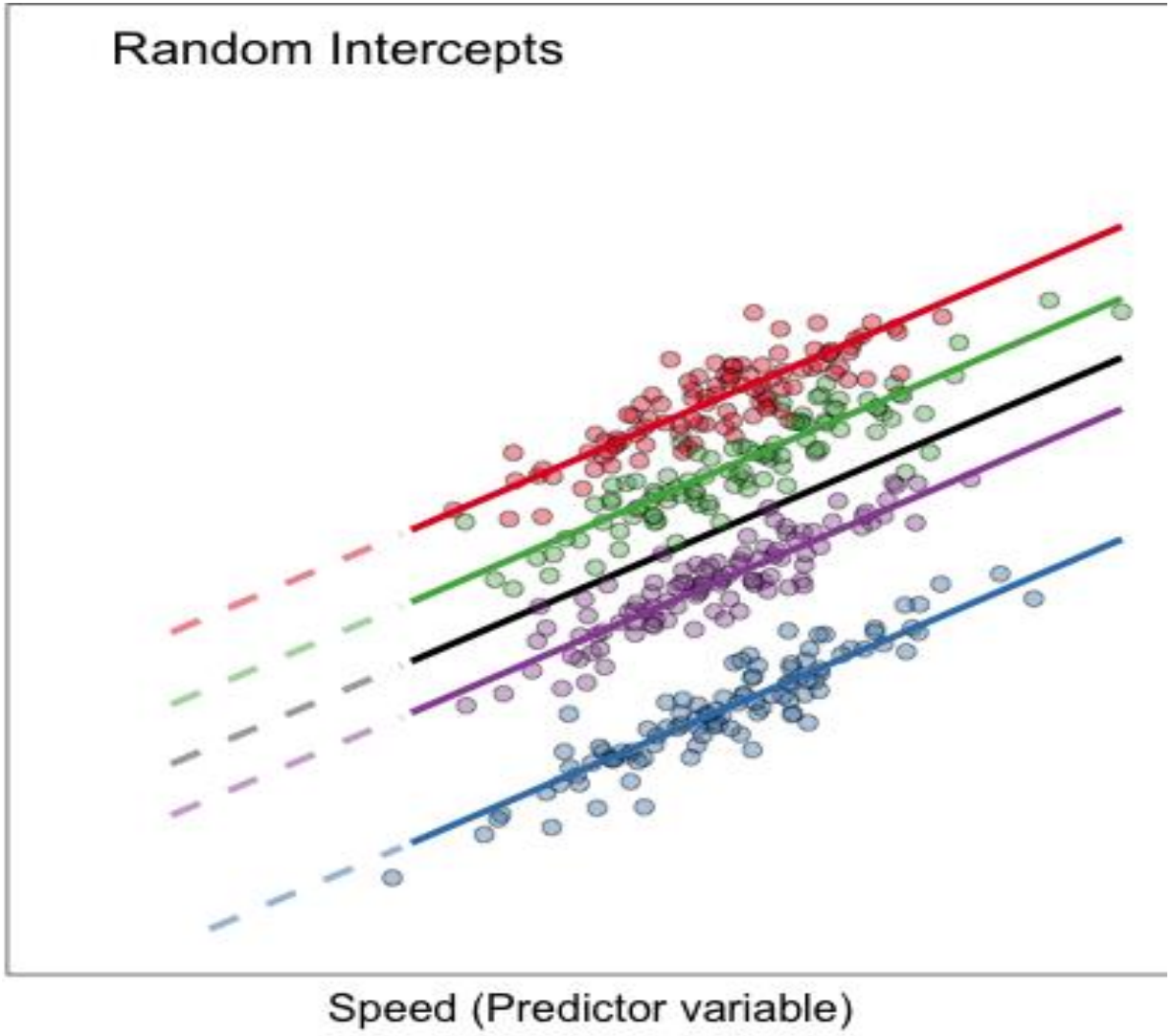
A linear regression …



The variation between participants

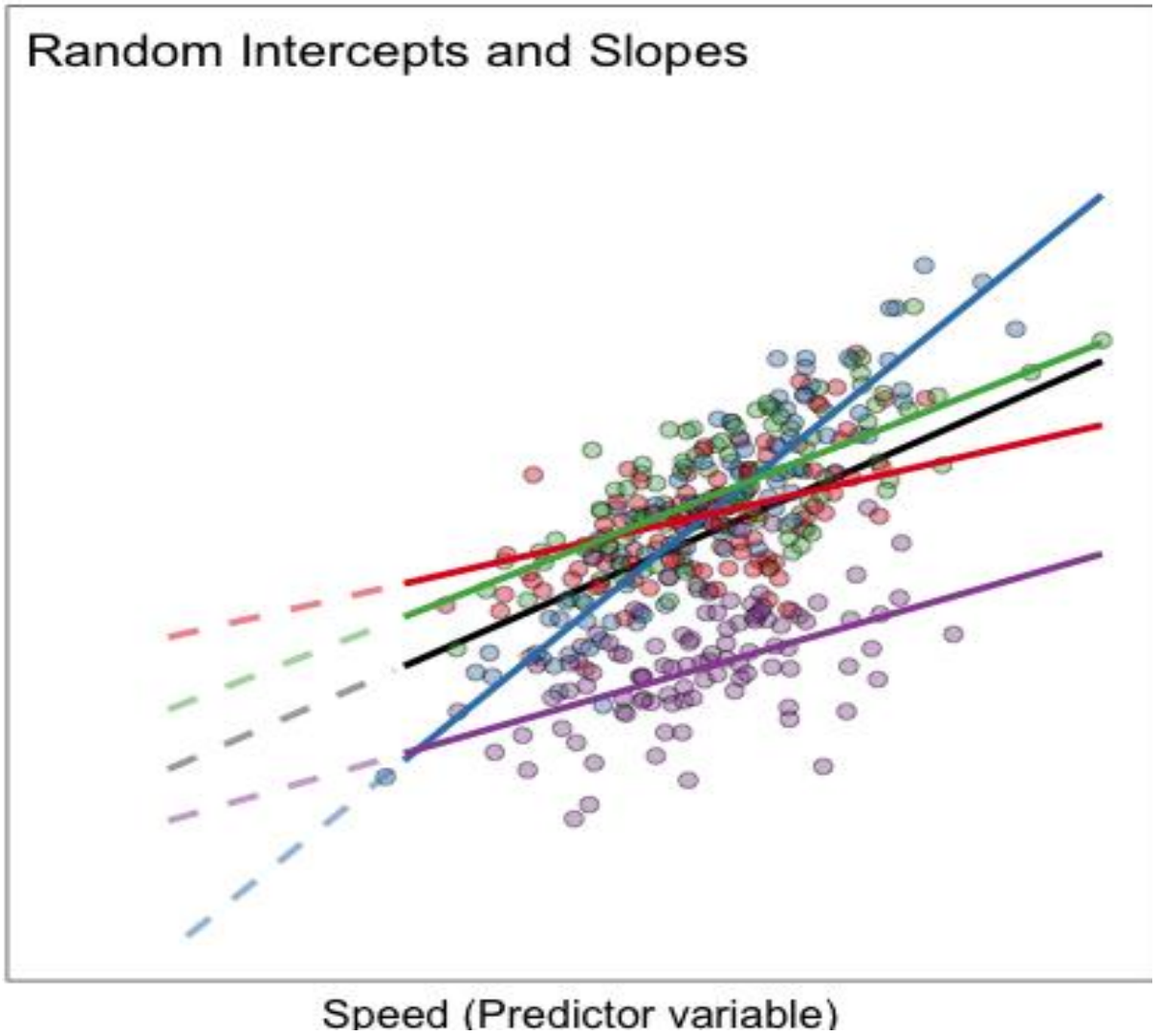


Random effects (**intercepts**)



Participants means allowed to vary, but model assumes all participants have a common slope between speeds (here, same increase rate)

Random effects (intercepts and slopes)



Participants means allowed to vary differently across different speeds (different slopes)

What random effects can we have?

- Participants intercept
- Speed | Participants slope
- Clip | Participants slope
- Lines | Participants slope

- What about Clips?

- Subtitles intercept
- ~~Speed | subtitles slope~~?
- ~~Clip | subtitles slope?~~
- ~~Lines | subtitles slope?~~

- You need at least 5-6 levels (Clips)! We only have 2

Useful reading (introductory)

- Field, A. (2017). *Discovering statistics using IBM SPSS statistics* (5th Ed.). London: Sage Publications.
  - Chapter 21 – Multilevel Linear Models. Well explained and easier to understand than most manuals. I recommend reading Chapter 9, The Linear Model, before dealing with Chapter 21. This is because a good understanding of regression models is recommended in order to deal with mixed models.
- Carson, R. J., & Beeson, C. M. L. (2013). Crossing language barriers: Using crossed random effects modelling in psycholinguistics research. *Tutorials in Quantitative Methods for Psychology, 9*(1), 25-41.
  - Also introductory paper. Shows the step-by-step process of building a model (from the "Empty model" to the final model) in SPSS. It also teaches how to interpret the results and to calculate effect sizes.
- Cunnings, I., & Finlayson, I. (2015). Mixed effects modelling and longitudinal data analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 159-181). New York: Routledge.
  - Chapter in edited book. Very clearly explained; it also shows the step-by-step process of building a model. However, the model is built in Rstudio, not SPSS. Still, the R syntax is simple and clearly broken down and explained.
- Garson, G. D. (2020). *Multilevel modelling: Applications in Stata, IBM SPSS, SAS, R, and HLM.* London: Sage.
  - To my mind, the language starts becoming slightly more complicated to understand, but still "easy". Each chapter builds a different model, starting with very simple models. Each model is built in all software covered in the book (see title) and the results are compared.

Useful reading (introductory)

- Winter, B. (2020). *Statistics for Linguists: An introduction using R*. New York: Routledge.
  - Again, not with SPSS, but with Rstudio. Still, Chapters 14, 15 and 16 cover mixed models in a relatively simple way.
- Singmann, H., & Kellen, D. (2019). An Introduction to Mixed Models for Experimental Psychology. In D. H. Spieler & E. Schumacher (Eds.), *New Methods in Cognitive Psychology* (pp. 4–31). Psychology Press.
  - A bit more complicated, but understandable if you've read something else before.
- Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language, 112*, 1-22.
  - Same as above. Although it's considered introductory, it's more complicated to follow (more technical language). Please read other papers/chapters before coming here. However, it makes very useful recommendation based on a large review of the literature.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.
  - Baayen is a respected name in the field. Again, probably not as introductory as the title suggests, but simple enough and every useful. Chapters 6 and 7 deal with regression modelling and mixed models, respectively.

Useful reading (not introductory)

- Heck, R. H., Thomas, S. L., & Tabata, L. N. (2014). *Multilevel and longitudinal modelling with IBM SPSS (2nd ed)*. New York: Routledge.
  - The language is a bit more technical, and the book is very detailed. Very useful if you want to explore models, the model building process, and model interpretation in highly useful detail.
- Heck, R. H., Thomas, S. L., & Tabata, L. N. (2012). *Multilevel modelling of categorical outcomes using IBM SPSS*. New York: Routledge.
  - Similar to the previous book but deals solely with dependent variables that are categorical (Binomial, multinomial, Poisson, Gaussian etc). Very useful.
- Scott, M. A., Simonoff, J. S., & Marx., B. D. (Eds.) (2013). *The SAGE handbook of multilevel modelling*. Los Angeles: Sage.
  - A mix of introductory and more detailed edited chapters. Covers, for example, data collection, cleaning and analysis; longitudinal and generalized mixed models; smoothing, robust methods etc. The book consists of 33 chapters.
- Barr, D. J., Levy, R., Scheepers, C. and Tily, H. J. (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.
  - Seminal paper by leading researchers. Advocates the use of the maximal model (keep all fixed and random effects you hypothesized to affect the results, irrespective of model fit)

Useful reading (not introductory)

- Bates, D., Kliegl, R., Vasishth, S., Baayen, R. H. (2018*). Parsimonious mixed models*. Retrieved from https://arxiv.org/pdf/1506.04967.pdf [Accessed26 November 2020].
  - A reply of sorts to Barr et al (2013). Leading researchers (Bates is one of the creator of the lmer function for mixed models in Rstudio). Essentially, it explains why the maximal model should be reduced in favour of parsimony and gives you step-by-step instructions to do so. A version from 2015 is also available.
- Matuschek, H., Kliegl., R., Vasishth, S., Baayen, H., Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305-315.
  - Another paper advocating for parsimonious models via simulations. Maximal models (Barr et al. 2013) lead to a loss of power without any apparent benefit.
- Brysbaert, M. & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. Journal of Cognition, 1(1): 9, pp. 1–20, DOI: https://doi.org/10.5334/joc.10.
  - As the title suggests, it's a guide on power analysis for mixed models. It recommends a minimum of 40 participants and 40 items per participant (e.g., subtitles) for a mixed model to achieve sufficient power.

Now, to SPSS …