Grammaticality judgments in Russian Sign Language

Vadim Kimmelman

The issue

- Should we use more controlled (formal/experimental) approaches to judgment data collection, specifically in SL research?

Answers:

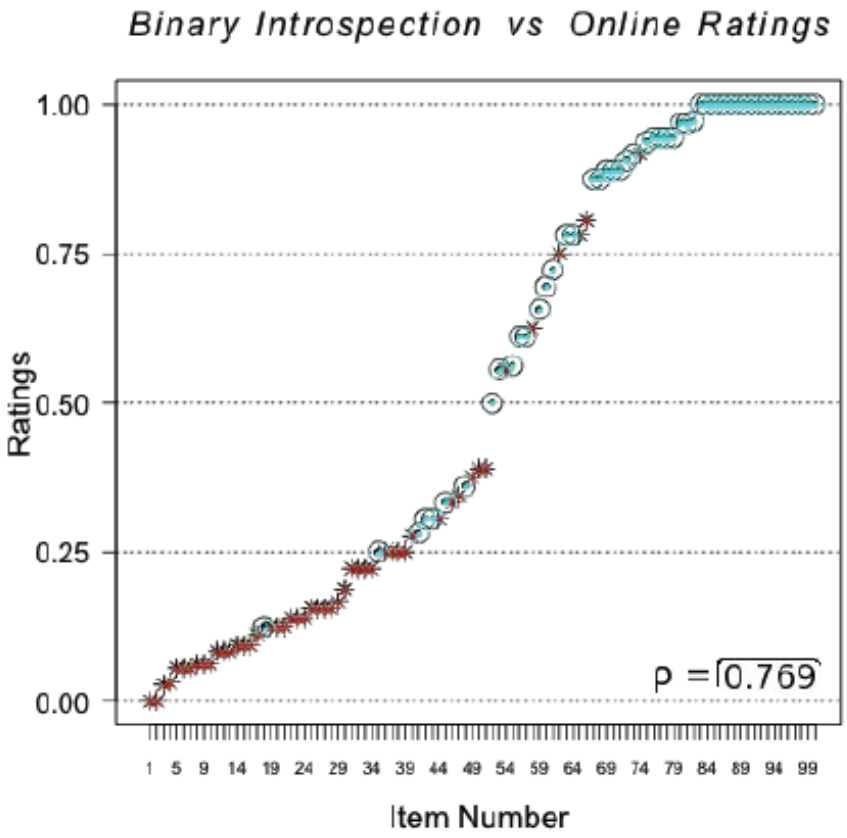
- Yes, because it is more reliable
- No, because it is equally (un)reliable and more expensive

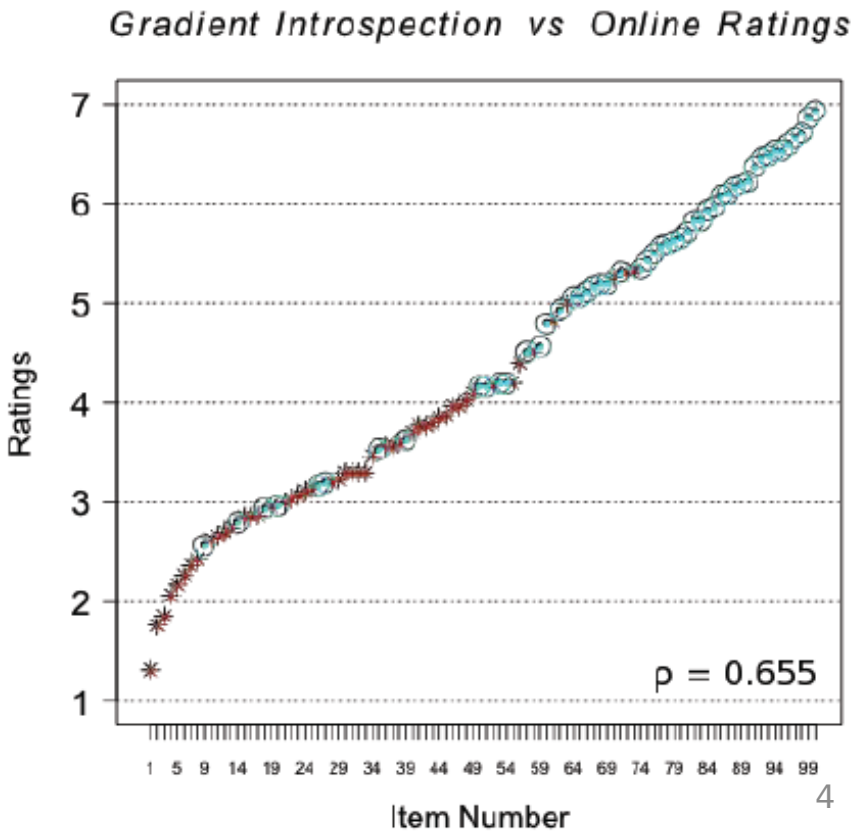
Motivation 1: the debate

- **Schütze** 1996/2016, Featherston 2007 i.a.:
  - Theoretical and empirical arguments in favor of a more controlled approach to grammaticality/acceptability judgements
- Sprouse, **Schütze** & Almeida 2013:
  - Tested 296 data points from LI 2001-2010
  - 936 participants in three formal tasks
  - 95% (±5%) convergence rates between formal and informal methods

Motivation 1: the debate

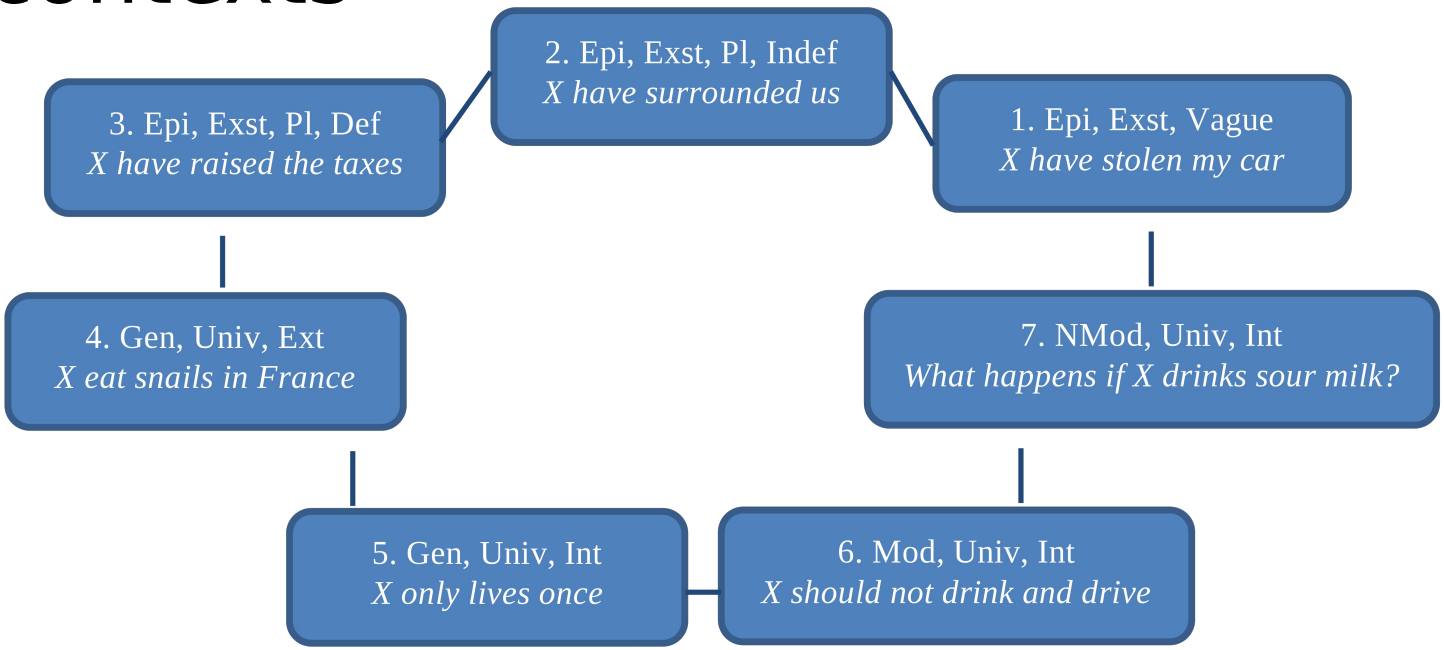
- Häussler, Juzek 2016; Häussler, Juzek, Wasow 2016:





Motivation 2: my data

- Topic: impersonal reference in RSL
- Gast & Van der Auwera 2013: contexts



Motivation 2: my data

- Data collection:
  - 4 native signers
  - Discussion of examples in contexts, with glosses as suggestions for lexical content

Context: My friend travels a lot. He tells me about crazy habits in different countries. For example, he said:

FRANCE EAT SNAILS

  - Contexts presented in Russian (2 signers) or RSL

Motivation 2: my data

sentence	S1	S2	S3	S4
Someone is ringing the bell.	someone *prodrop	someone prodrop	*someone prodrop	someone ?prodrop
Someone stole my bike.	someone prodrop	someone prodrop	*someone prodrop	?someone prodrop
In France, **they** eat snails.	prodrop *they	prodrop they	prodrop *they	prodrop *they
If **you** drink, you cannot drive	prodrop *you	prodrop you	prodrop you	prodrop *you

Additional information:

- Most patterns also found in corpus data, but not enough data for a clear picture

Experimental setup

- Three conditions:
  - Pro-drop vs. someone in existential impersonals
  - Pro-drop vs. they in universal impersonals
  - Pro-drop vs. you in conditionals
- Token set (same context):
  1. RING (pro-drop)
  2. SOMEONE RING (someone)

Experimental setup

- General guidelines (Gries 2013):
  - Stimuli in a concrete token set differ only in the relevant variable
  - More than one token set per condition
  - Every subject sees only one level of the concrete token set
  - Every subject sees every level combination, more than once and equally frequently
  - Every item presented to multiple subjects (equally many subjects)
  - Distractors, ideally two/three times more than test items
  - The order is pseudorandomised
- More than the guidelines in Featherston (2007)

Experimental setup

- Stimuli in a concrete token set differ only in the relevant variable check (although not perfectly identical productions)
- More than one token set per condition check, 4 (one exception)
- Every subject sees only one level of the concrete token set check
- Every subject sees every level combination, more than once and equally frequently check, twice (this is too little for individual analysis)
- Every item presented to multiple subjects (equally many subjects) check
- Distractors, ideally two/three times more than test items check, 24 fillers per 4 items, other RQ investigated in the same test
- The order is pseudorandomised check, 8 different pseudorandomized orders

Experimental setup

- Additional considerations:
  - Instruction and examples completely in RSL (video)
  - The context followed by a black screen followed by the test item
  - Scale: 5-point colored smiley scale

“In this study I will be describing different situations in sign language. You will be evaluating whether it was correct or not. I will be telling a short story, then there will be a dark screen, and then one more sentence connected to the same story. You look at this second sentence (not the first part) and evaluate whether it is correct or not. Below there are five buttons. The green one means everything is correct, the signs are correct. The next one is maybe you can say that. The third one means it is difficult to say, maybe correct maybe not, do not know. The fourth one means, no, it is bad. The last one, the red one, means no, terrible, impossible. If you do not understand a video, you can replay it. There are 36 stories in total.”

Experimental setup

Experimental setup

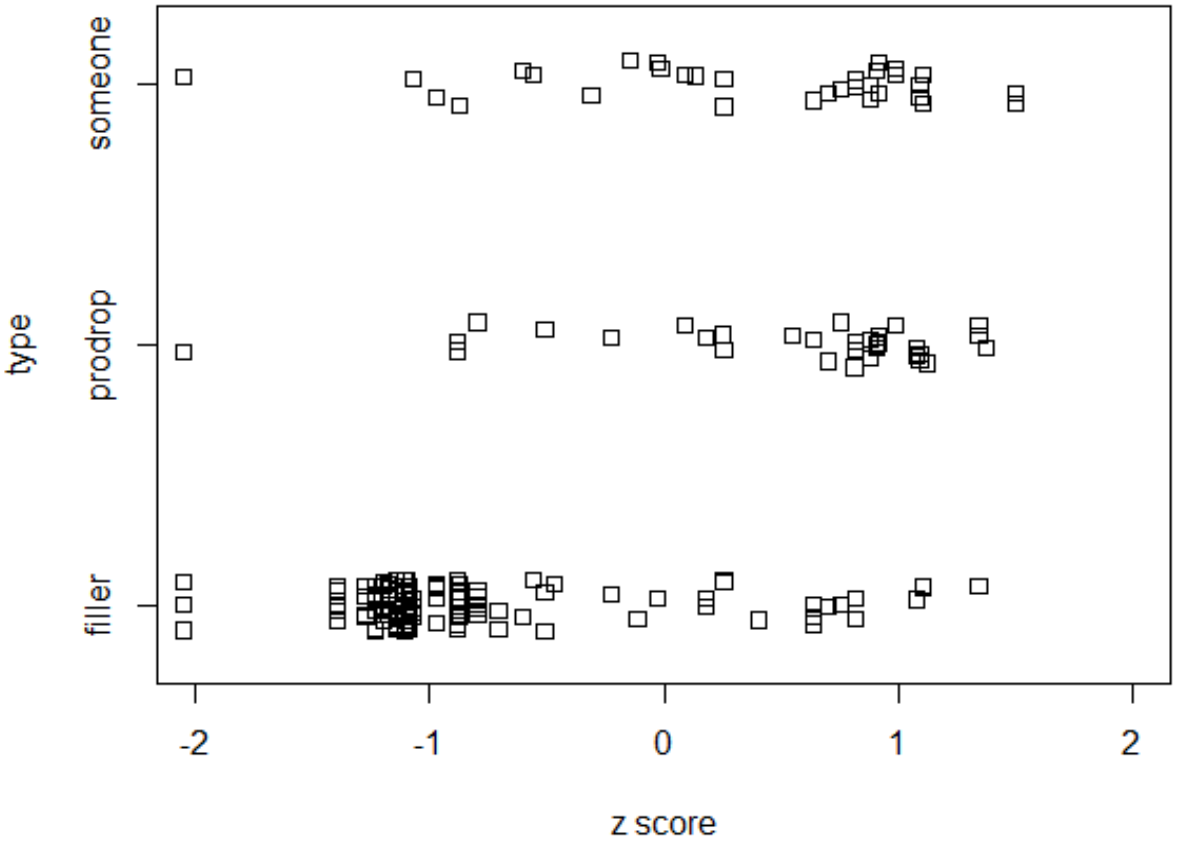
- 17 signers, one excluded
  - 4 male, 12 female
  - Mean age: 40
  - 11/16 native, but no differences in judgments
- Power analysis:
  - Schütze & Sprouse (2014): for the Likert scale task 35 judgments per condition are necessary to achieve 80% power
  - In my case for each (correct) condition: 32 judgments

Analysis

- Raw scores (1 to 5) transferred to z-scores per participant to eliminate potential scale bias
- Checked participants for abnormal behavior based on clearly grammatical and ungrammatical fillers
  - All participants give a median z-score of 1 to grammatical items and a median z-score of -1 to ungrammatical items
- Statistics: a linear regression model to compare conditions to fillers, and the U-test to compare conditions to each other

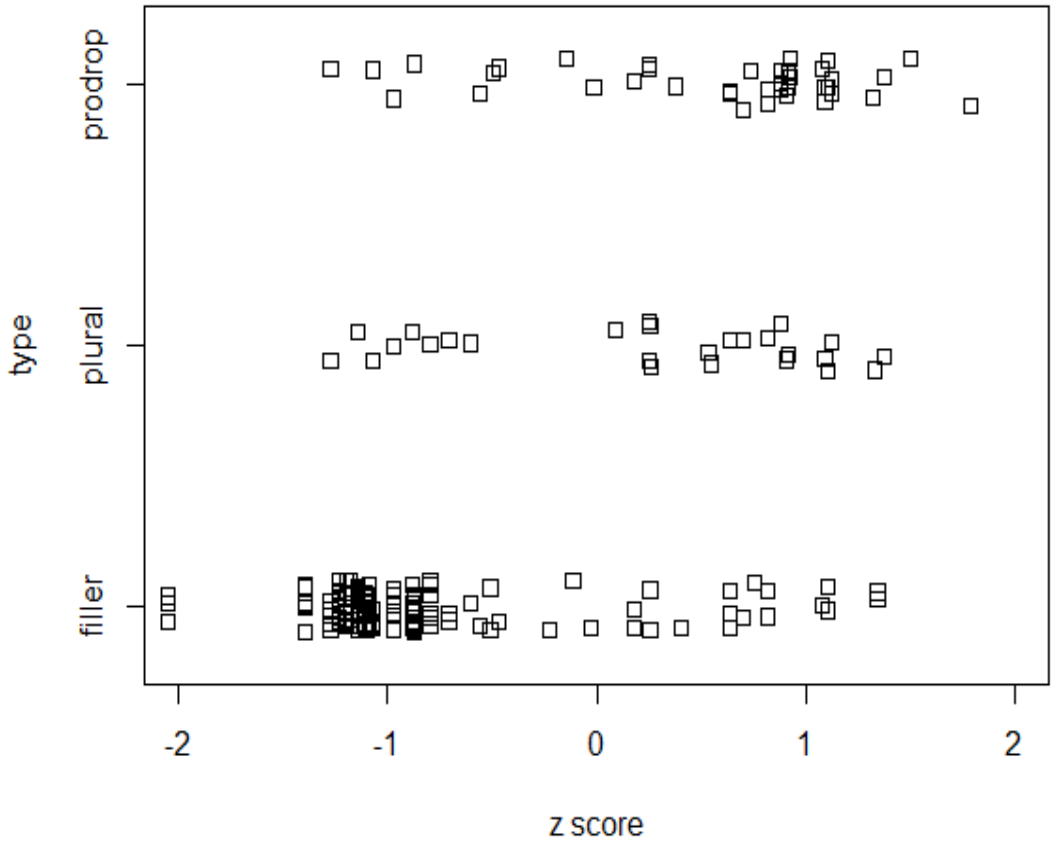
Results

- Existential contexts (someone rings the bell)



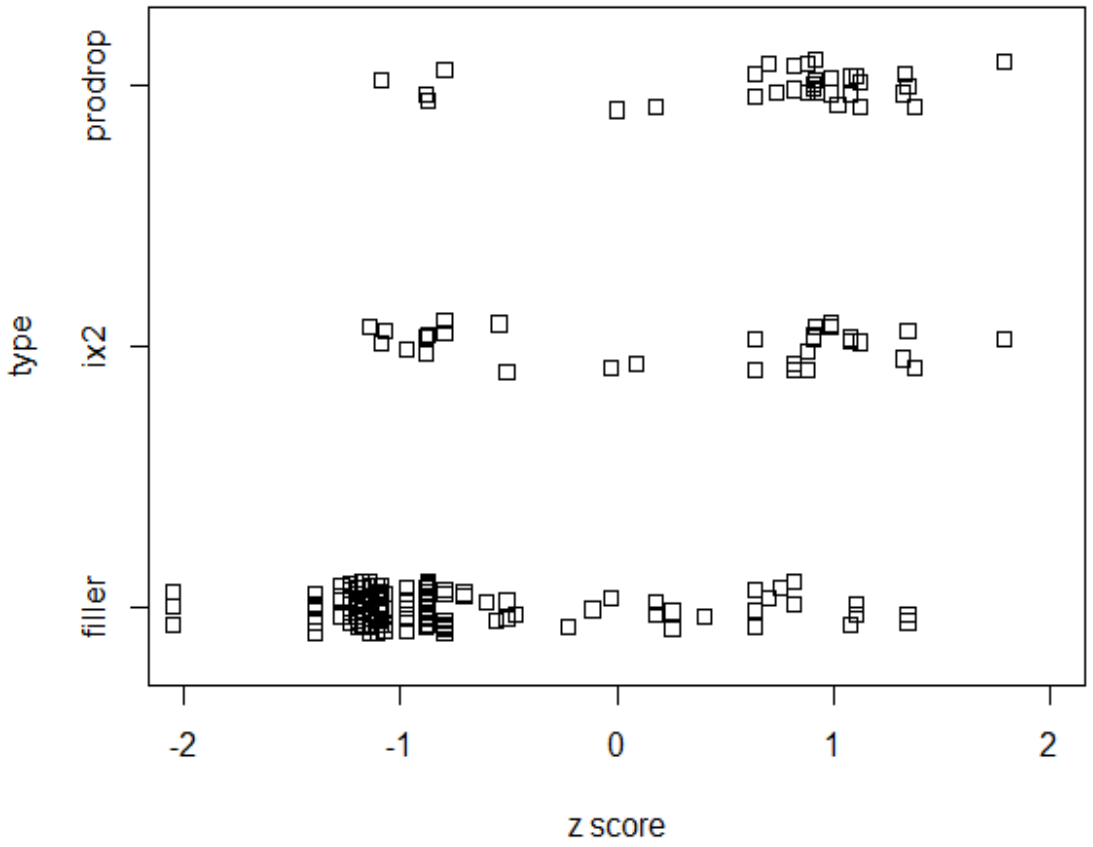
Results

- Universal contexts (they eat snails in France)



Results

- Conditionals (if you drink, you cannot drive)



Discussion

- For existential contexts, pro-drop and *someone* are equally grammatical
- For universal contexts, both pro-drop and *they* are grammatical, but for *they* there is more variation
- For condition contexts, both pro-drop and *you* are grammatical, but for *you* there is more variation

Discussion

- Possible explanation for *they:*
  - Some signers use the plural pronoun as a universal quantifier (all)
  - Those who do, accept it in universal impersonals
- Possible explanation for *you:*
  - This construction is borrowed from Russian
  - Some signers evaluate it as a foreign element and others do not

General discussion

- The overall results are very similar to the conclusions from the informal task
  - Grammaticality of *you* contrary to introspection
  - No new insights into the reasons for variation
  - Reliability (statistical testing) as an advantage?
- What do grammaticality judgments asses?
  - Performance factors
  - Ease of corrections
  - For RSL:
    - Iconicity violations
    - Interaction with Signed/Spoken Russian?

Questions? Comments?
Useful references?

References

Featherston, Sam. 2007. Data in generative grammar: The stick and the carrot. *Theoretical Linguistics* 33(3).

Gast, Volker & Johan Van Der Auwera. 2013. Towards a distributional typology of human impersonal pronouns, based on data from European languages. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries*. Berlin: Mouton de Gruyter.

Gries, Stefan Thomas. 2013. *Statistics for linguistics with R: a practical introduction*. 2nd revised edition. Berlin: De Gruyter Mouton.

Juzek, Tom S. & Jana Häussler. 2016. To be Grammatical or not to be Grammatical – Is that the Question? Evidence for gradience. Tübingen.

Juzek, Tom S., Jana Häussler & Tom Wasow. 2016. To be Grammatical or not to be Grammatical – Is that the Question? Evidence for gradience. Washington DC.

Schütze, Carson T. & Jon Sprouse. 2014. Judgment Data. In Devyani Sharma & Rob Podesva (eds.), *Resarch Methods in Linguistics.*, 27–50. Cambridge: Cambridge University Press.

Schütze, Carson. 2016. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology.* Language Science Press.

Sprouse, Jon, Carson T. Schütze & Diogo Almeida. 2013. A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001–2010. *Lingua* 134. 219–248.