CONFIDENTIAL - FOR PEER-REVIEW ONLY
Morphological convergence in Hungarian OTKA (#87324)

Created: 02/07/2022 10:25 AM (PT)

This is an anonymized copy (without author names) of the pre-registration. It was created by the author(s) to use during peer-review.
A non-anonymized version (containing author names) should be made available by the authors when the work it supports is made public.

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
Do people converge to patterns of conversation partners in a word matching game using variable Hungarian morphology?
Is this driven by the type of morphology? The specific lexical items they are exposed to? Time passed after exposure (short or long-term effect)?

3) Describe the key dependent variable(s) specifying how they will be measured.
Responses in a word matching game hosted online. Responses are always binary: variant A or variant B.

4) How many and which conditions will participants be assigned to?
Across-participant conditions:
- TYPE of morphological variation in MAIN task. (i) 1sg.indef verb variation, (ii) variation in verb stem epenthesis, (iii) variation in front-harmony in noun suffixation.
- RATE of use in MAIN task: the co-player uses a large amount of variant A or a large amount of variant B.
- Lexical TYPICALITY in MAIN task: the co-player uses variant A with prompts that are more likely to have variant A or the co-player uses variant B with prompts that are more likey to have variant A.
Within-participant conditions:
- Individual prompts in MAIN task are pre-rated by separate participants in a BASELINE task to ascertain where they are in the variation spectrum.
- Word-matching TEST with co-player followed by POST-TEST with no co-player in MAIN task.
- After MAIN task selected participants are re-tested in later FOLLOW-UPs.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
Binary responses in post-test analysed. Bayesian generalised linear regression with weakly informative priors. Formula: Response is variant A ~ co-player RATE + co-player TYPICALITY + word prompt BASELINE rate of variant A + is this the POST-TEST or a follow-up + (1 + word BASELINE rate of variant A + is this the POST-TEST or a follow-up | participant) + (1 + co-player RATE + co-player TYPICALITY | word)

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
Prompt words are assessed in a BASELINE rating task. Participants with 100% use of variant A or B in TEST or POST-TEST in main task are removed from analysis. Participants outside median +- 3 mean absolute deviation of condition-specific log odds of use of variant A / B also removed.
Trials with a response time outside outside median +- 3 mean absolute deviation of condition-specific response type distribution removed.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
50 prompts in TEST and 50 prompts in POST TEST of main task. additional batches of 50 in follow-ups. 25 participants per across-participant condition in MAIN TASK after outliers removed, based on power analysis. Means 300 in total. FOLLOW-UPS likely to fall apart.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
Data and code archived here: https://github.com/petyaracz/Racz2024
Github repo will contain data and point to gitlab repos containing psychopy code running the experiment on gitlab. This means reviewer can cross-reference timestamps on pushes in repo with timestamps of data collection to make sure there is no post-hoc tinkering besides whatever we admit to in paper(s).