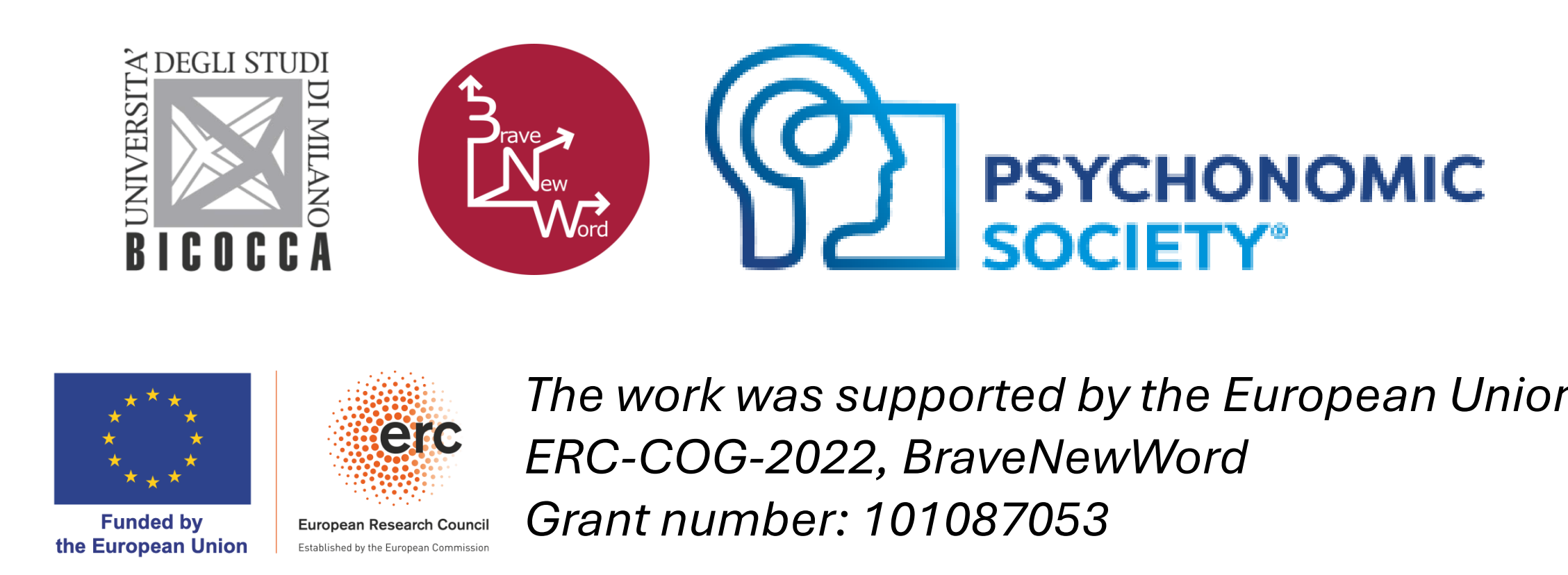


# Individual differences in language experience and their effects on the semantic processing of nonwords

Fabio Marson<sup>1\*</sup>, Rolando Bonandrini<sup>1</sup>, Simona Amenta<sup>1</sup>, Marco Marelli<sup>1</sup>

<sup>1</sup>Psychology Department, University of Milano-Bicocca, Milan, Italy

\*✉: fabio.marson@unimib.it



## Theoretical Background

Nonwords are strings of letters that are not attested in any vocabulary (e.g., "mohalk"). These kind of stimuli were considered devoid of meaning for a long time, thus used as control "meaningless" strings in psycholinguistic studies. Recent studies (Hendrix and Sun, 2021; Gatti et al., 2023) showed that **nonwords carry semantic information** that can affect reaction times in behavioral tasks like semantic priming or lexical decision tasks. According to usage-based theories of language experience allows generalization of rules, thus, greater **exposition to language and larger vocabulary size** lead to increased **sensitivity to semantic information** (Kidd et al., 2018).

We analyzed publicly available data about nonwords collected in the **English Crowdsourcing Project** (Mandera et al., 2020) to test whether:

- 1) individual differences in experience with language affect the sensitivity to semantic information carried by nonwords
- 2) Sensitivity to semantic information for words could predict the sensitivity to semantic information carried by nonwords

## Materials and methods

**English Crowdsourcing Project:** Megastudy about known vocabulary in a given language (e.g., American English)

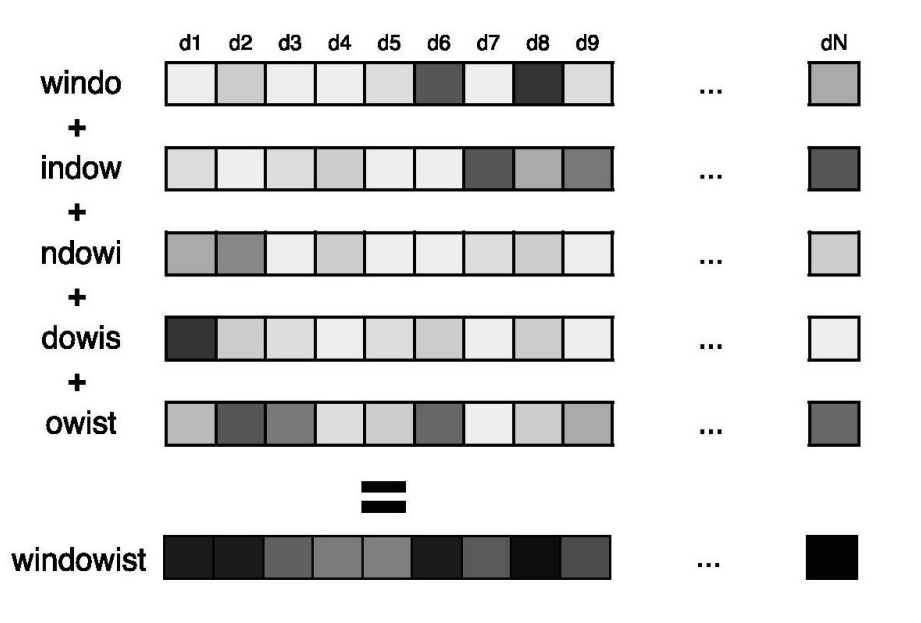
Task: non-speeded vocabulary test



Stimuli:  
➢ Words (70% of trials)  
➢ Nonwords (30% of trials)  
303k nonwords generated with Wuggy [e.g., "duth", "memigabize", "culebrent", "tidgings", "fapar", "monarbiance", "povemino"]

Slower RTs for NWs rejection: **interference** of semantic information = greater **word-like activation**

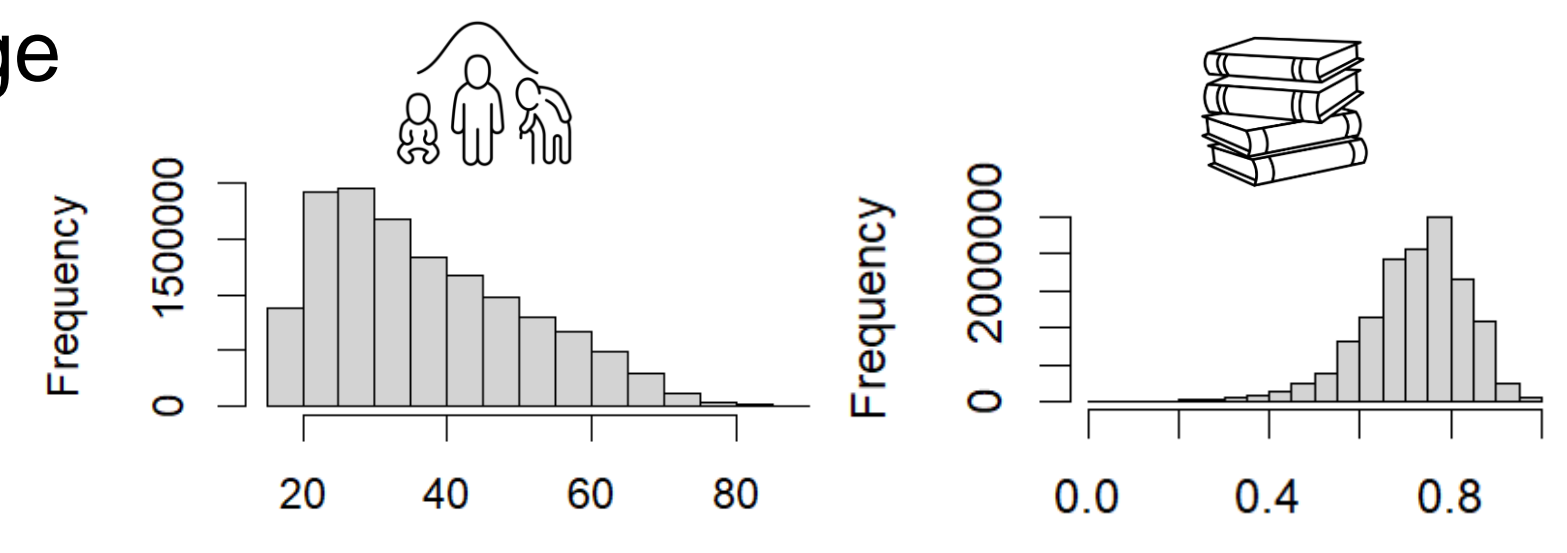
**FastText**, a word embedding technique that allows the generation of semantic vectors for infrequent or unattested string of letters, was used to generate semantic vectors for nonwords.



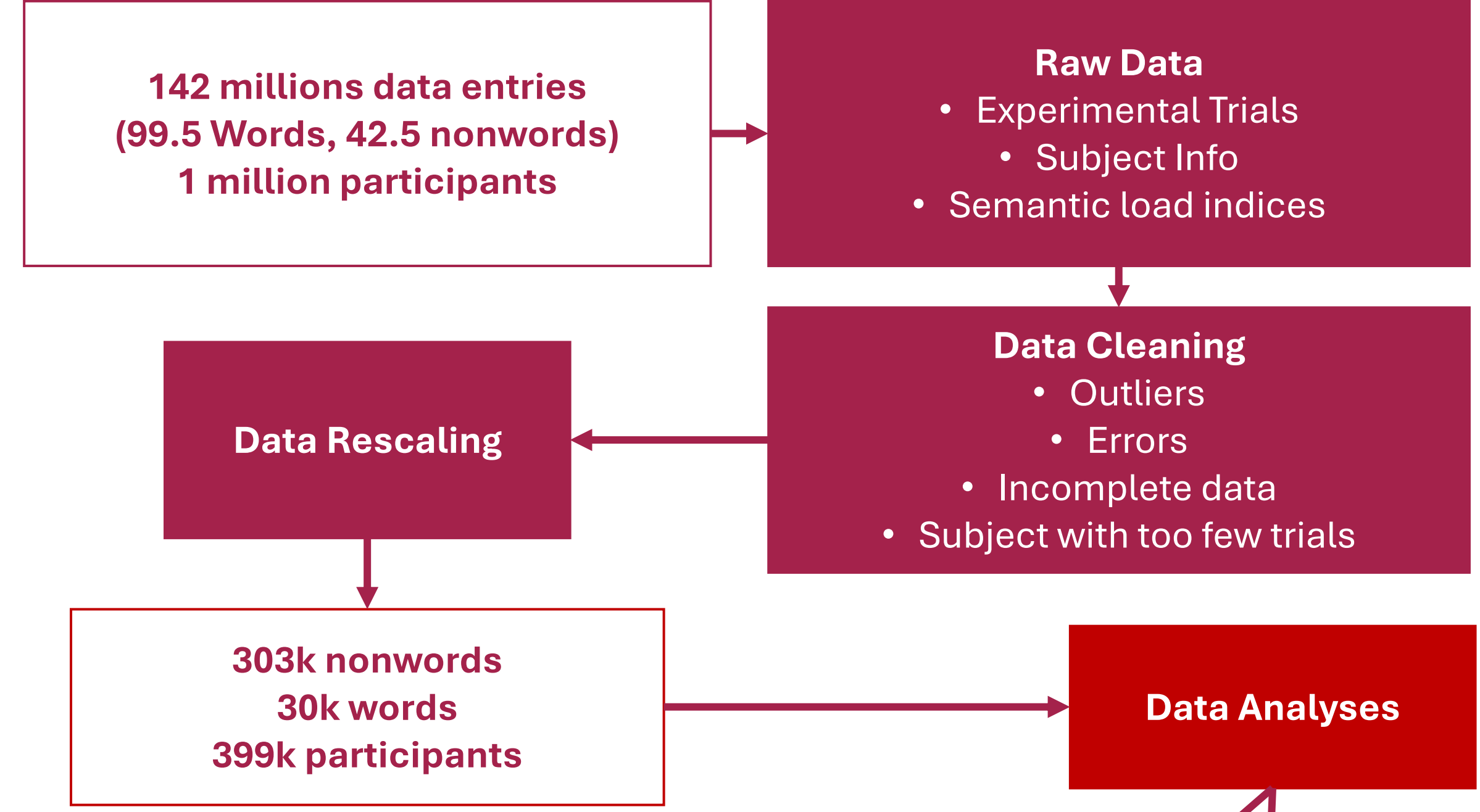
Two semantic load indices were computed for nonwords:  
○ **OSC** (Orthography-to-Semantic Consistency)  
○ **SND** (Semantic Neighborhood Density)

Individual differences related to language experience of interest were:

- Age
- Vocabulary Size



## Data processing and analysis



**Hypothesis 1: individual differences and semantic load metrics affect Nonwords RTs**

Linear mixed effect model:  
 $\log(\text{RTs}) \sim \text{length} + 3\text{grams frequency} + 5\text{grams frequency} + \text{Coltheart N} + \text{OLD20} + \text{age} + \text{vocabulary size} + \text{OSC} + \text{SND} + \text{age} \times \text{OSC} + \text{vocabulary size} \times \text{OSC} + \text{age} \times \text{SND} + \text{vocabulary size} \times \text{SND} + (1|\text{subject})$

**Hypothesis 2: sensitivity to semantics for Words processing affect RTs during Nonwords processing**

**Step1:** A model was fit for each subject to find sensitivity to semantics during Words processing:  
 $\log(\text{RTs}) \sim \text{length} + \text{frequency} + \text{Coltheart N} + \text{OLD20} + \text{OSC} + \text{SND}$

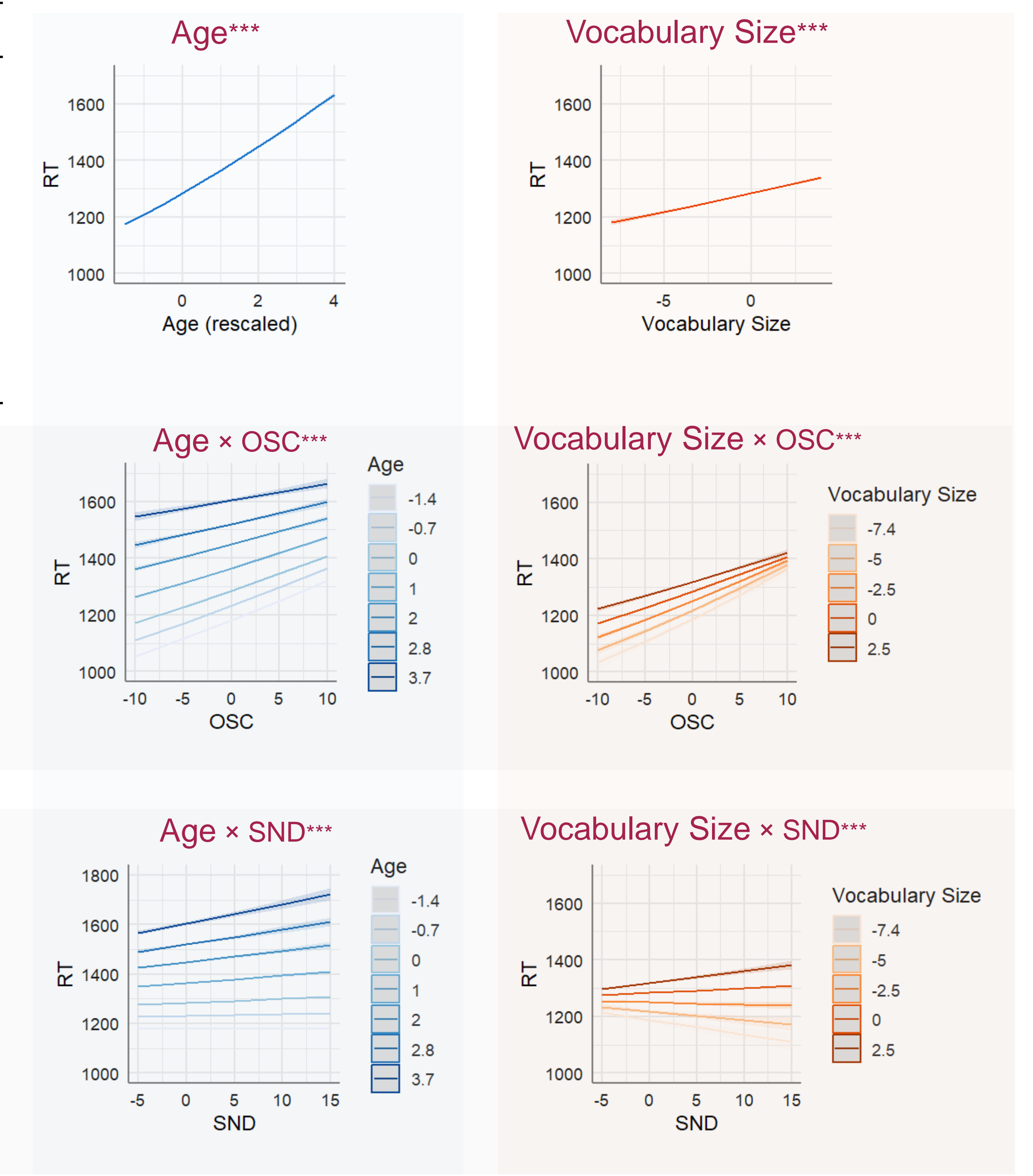
Coefficients for **OSC** and **SND** were extracted as indices of sensitivity to semantics of Words

**Step2:** Fit a new model on Nonwords RTs with words-based semantic coefficients in interaction with semantic load indices:  
 $\log(\text{RTs}) \sim \text{length} + 3\text{grams frequency} + 5\text{grams frequency} + \text{Coltheart N} + \text{OLD20} + \text{OSC}(\text{nonword}) + \text{SND}(\text{nonword}) + \text{OSC}(\text{word}) + \text{SND}(\text{word}) + \text{OSC}(\text{nonword}) \times \text{OSC}(\text{word}) + \text{OSC}(\text{nonword}) \times \text{SND}(\text{word}) + \text{SND}(\text{nonword}) \times \text{SND}(\text{word}) + \text{SND}(\text{nonword}) \times \text{OSC}(\text{word}) + (1|\text{subject})$

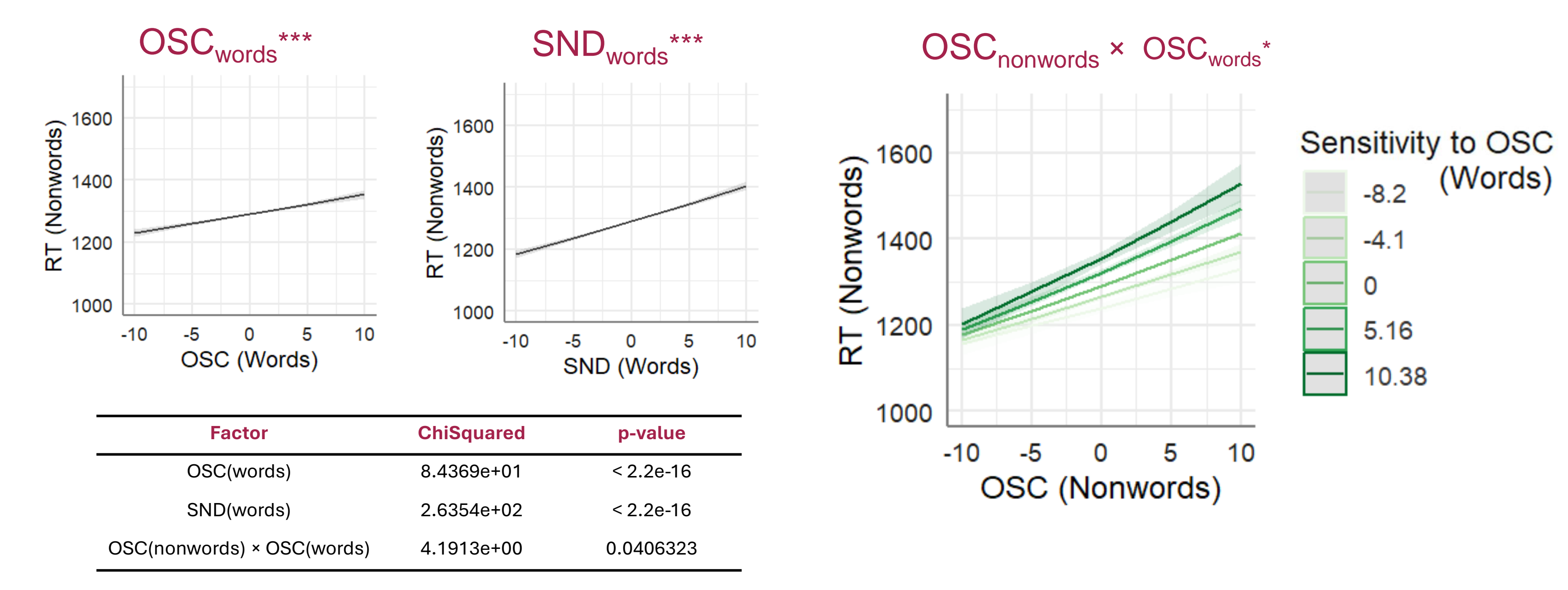
## Results

Hypothesis 1: individual differences affect sensitivity to nonwords' semantic

Factor	ChiSquared	p-value
Age	1.4425e+04	< 2.2e-16
Vocabulary Size	4.4767e+02	< 2.2e-16
OSC	5.1872e+03	< 2.2e-16
SND	8.7741e+01	< 2.2e-16
Age × OSC	1.4278e+02	< 2.2e-16
Voc Size × OSC	2.5674e+01	4.041e-07
Age × SND	5.6155e+01	6.698e-14
Voc Size × SND	3.5762e+01	2.229e-09



Hypothesis 2: sensitivity to words' semantic affects sensitivity to nonwords' semantic



## Discussion

Our results showed that behavioral responses to nonwords in a non-speeded vocabulary task were affected by individual differences in age and vocabulary size as well as by nonwords' semantic contents. Their interactions showed that the interference effect of **OSC decreased** with age and vocabulary size; while the interference effect of **SND increased** with age and vocabulary size. These results support the idea that language is a flexible and dynamic process affected by experiential factors (see Pexman & Yap, 2018). Finally, the sensitivity to **words** semantic content was observed to affect the RTs for nonwords processing. Interestingly, the sensitivity to OSC for words was observed to interact with OSC for nonwords, suggesting that the mechanism underlying processing of OSC from both words and nonwords could be a semantic extraction process generalized beyond stimuli type.

## Take home message

Our experience with linguistic materials affect how we process semantic information carried by nonwords. Generally, the more experience we have with language, the more we are sensitive to semantic information. The mechanisms underlying semantic processing could be generalized across known and unknown strings of letters.

Hendrix, P., & Sun, C. C. (2021). Journal of Experimental Psychology: Learning, Memory, and Cognition, 47(1), 157. Gatti, D., Marelli, M., & Rinaldi, L. (2023). Journal of Experimental Psychology: General, 152(3), 851. Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Trends in cognitive sciences, 22(2), 154-169. Mandera, P., Keuleers, E., & Brysbaert, M. (2020). Behavior Research Methods, 52, 741-760. Pexman, P. M., & Yap, M. J. (2018). Journal of Experimental Psychology: Learning, Memory, and Cognition, 44(7), 1091.