Understanding Vocabulary Growth Through An Adaptive Language Learning System

Elma Kerz[1], Andreas Burgdorf[3], Daniel Wiechmann[2],
Stefan Meeger[1], Yu Qiao[1], Christian Kohlschein[1], Tobias Meisen[3]

[1] RWTH Aachen University
[2] University of Amsterdam
[3] University of Wuppertal

8th NLP4CALL - NoDaLiDa workshop
Turku, Finland, September 30 2019

Overview

1. Background
2. Introducing the AISLE System: Design Principles and Architecture
   - Optimal Language Input: Extraction and Representation of Items
   - Optimal Presentation Intervals: Learning algorithm
   - User Interface
   - Integrated individual differences tasks and measures
3. Modeling Growth Trajectories: Using AISLE System to Study Vocabulary Growth
   - What is the best longitudinal model that describes participants' vocabulary growth?
   - What is the role of a range of IDs factors in explaining variation in participants' vocabulary acquisition?
4. Conclusions

Background

- Individuals learn language at different rate.
  - Particularly apparent in the area of vocabulary acquisition *(Hart and Ris- ley, 1995; Pellicer-Sánchez, 2018)*.
- Understanding the pace of vocabulary growth is importance for:
  - Strongly related to a variety of academic, vocational and social outcomes *(Rohde and Thompson, 2007; Dollinger et al., 2008; Verhoeven et al., 2011)*. .
  - A strong predictor of reading comprehension ability *(Muter et al., 2004; Tannenbaum et al., 2006; Verhoeven and Van Leeuwe, 2008; Verhoeven et al., 2011; Cain and Oakhill, 2011)*.

Background

Technology-enhanced language learning systems and automatic analyses of generated data...

- Link statistics and machine learning with the cognitive and language sciences *(Vahdat et al., 2016)*.
- Make it possible to perform learning behavior analytics.

The Present Study

1. Introduce the **AISLE**, an adaptive language learning system designed to ...
   - track and accelerate the development of vocabulary skills and
   - to understand the dynamics of growth of individual learning trajectories
2. We present first results of a study on vocabulary growth

Introducing the AISLE System: Design principles

- Optimal Language Input
  - Extract optimal word lists from large corpus with the optimal rank order.
  - Due to the target population, we are interested in tracking and accelerating the development of academic vocabulary (AV).
  - Language requires not only knowledge of words but also multiword sequences (MWS) *(Arnon and Chris- tiansen, 2017)*
- Optimal Repetition Intervals
  - When a word is presented to a user again? When a word is considered to be learned?

Optimal Language Input: NLP Pipeline

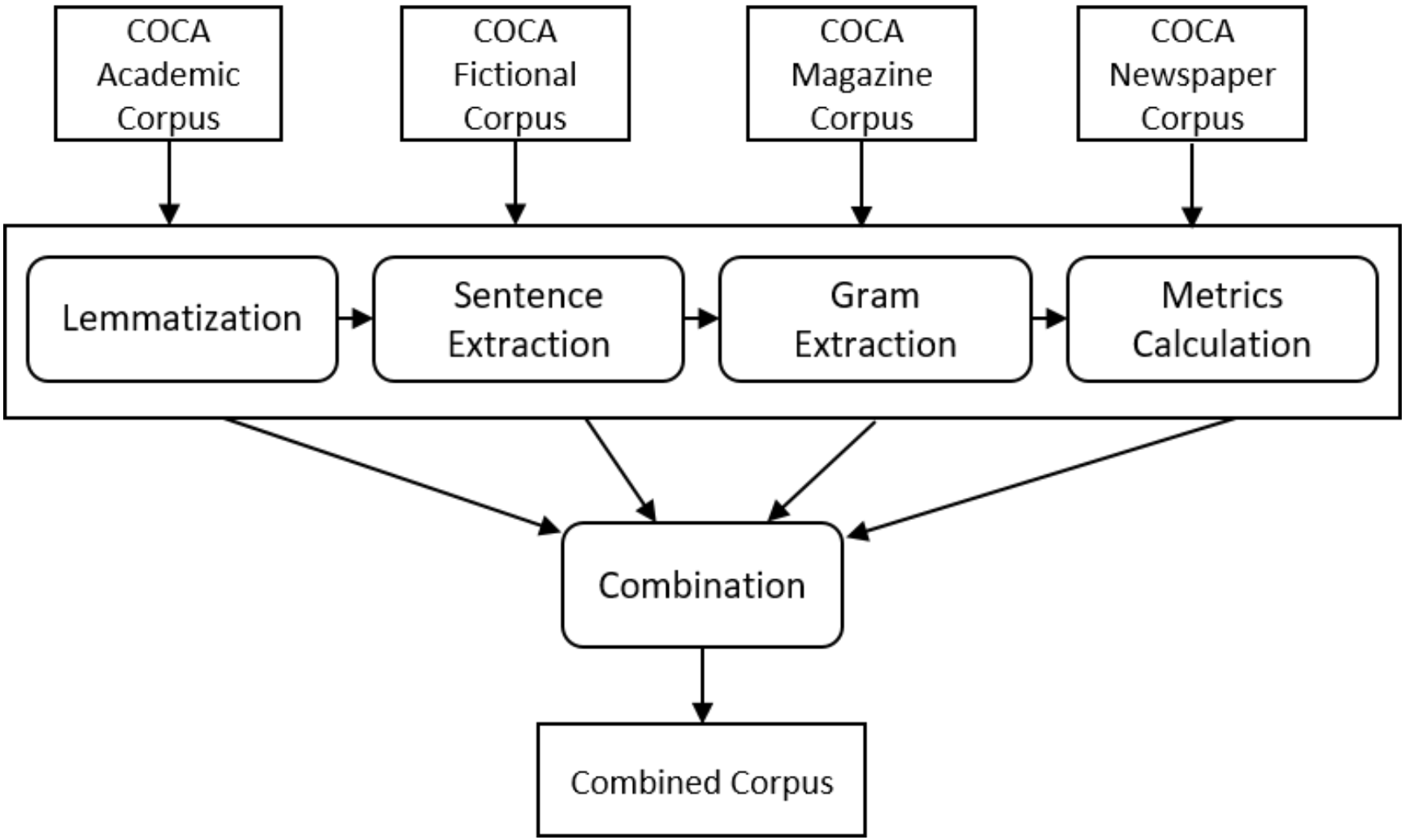


Figure: NLP pipeline for extracting statistically relevant vocabulary items

Optimal Language Input: Metrics

$$f_{gen}(i_n) := \frac{\sum_{k\in\{a,b,c\}} \#i_n(T_k)}{\sum_{k\in\{a,b,c\}} |N_n(T_k)|} \quad (1)$$

$$d_{gen}(i_n) := \frac{\sum_{k\in\{a,b,c\}} \sigma i_n(T_k)}{\sum_{k\in\{a,b,c\}} |N_n(T_k)|} \quad (2)$$

$$f_{ac}(i_n) := \frac{\#i_n(T_{(d)})}{|N_n(T_{(d)})|} \quad (3)$$

$$d_{ac}(i_n) := \frac{\sigma i_n(T_{(d)})}{|N_n(T_{(d)})|} \quad (4)$$

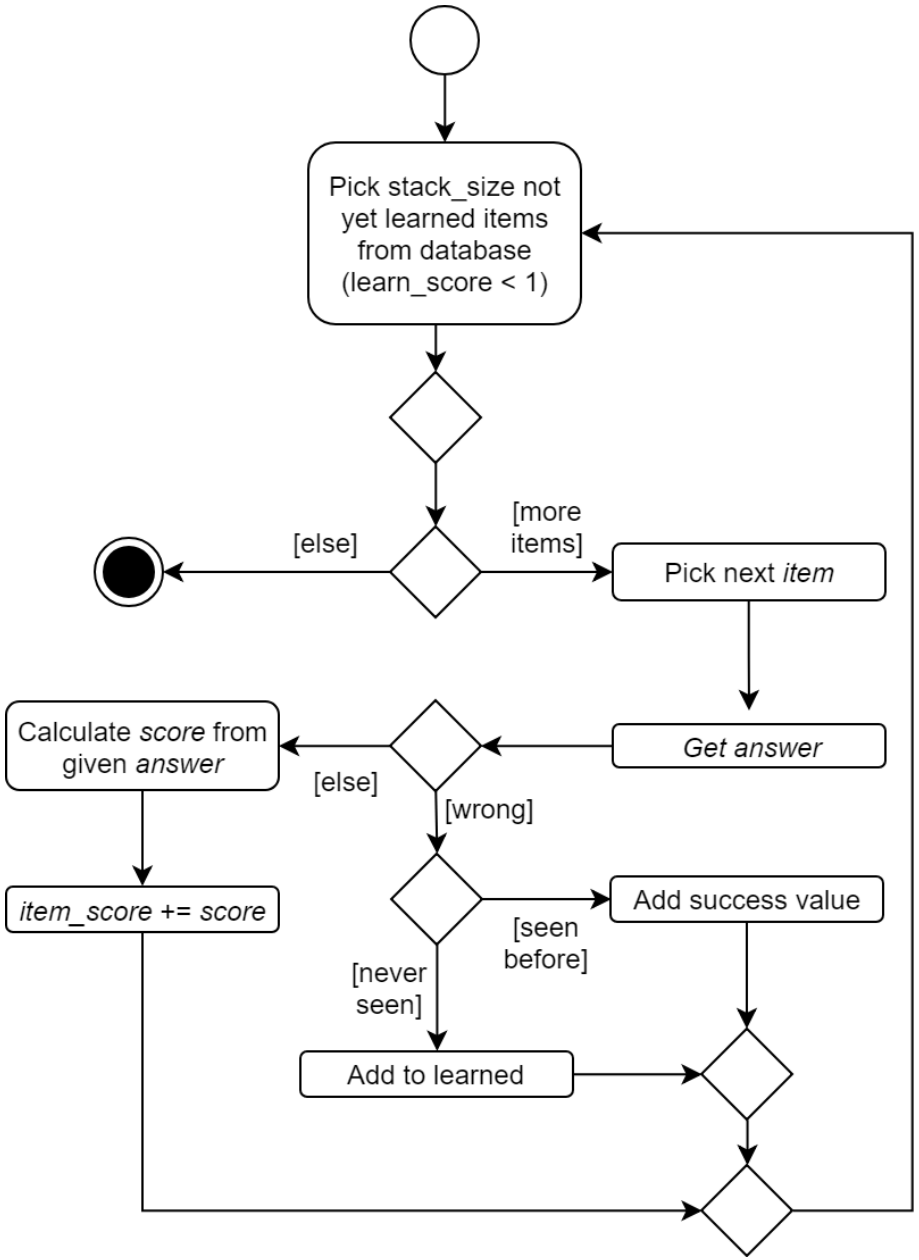
$$\frac{f_{ac}(i_n)}{f_{gen}(i_n)} > k_f \quad (5)$$

$$\frac{d_{ac}(i_n)}{d_{gen}(i_n)} > k_d \quad (6)$$

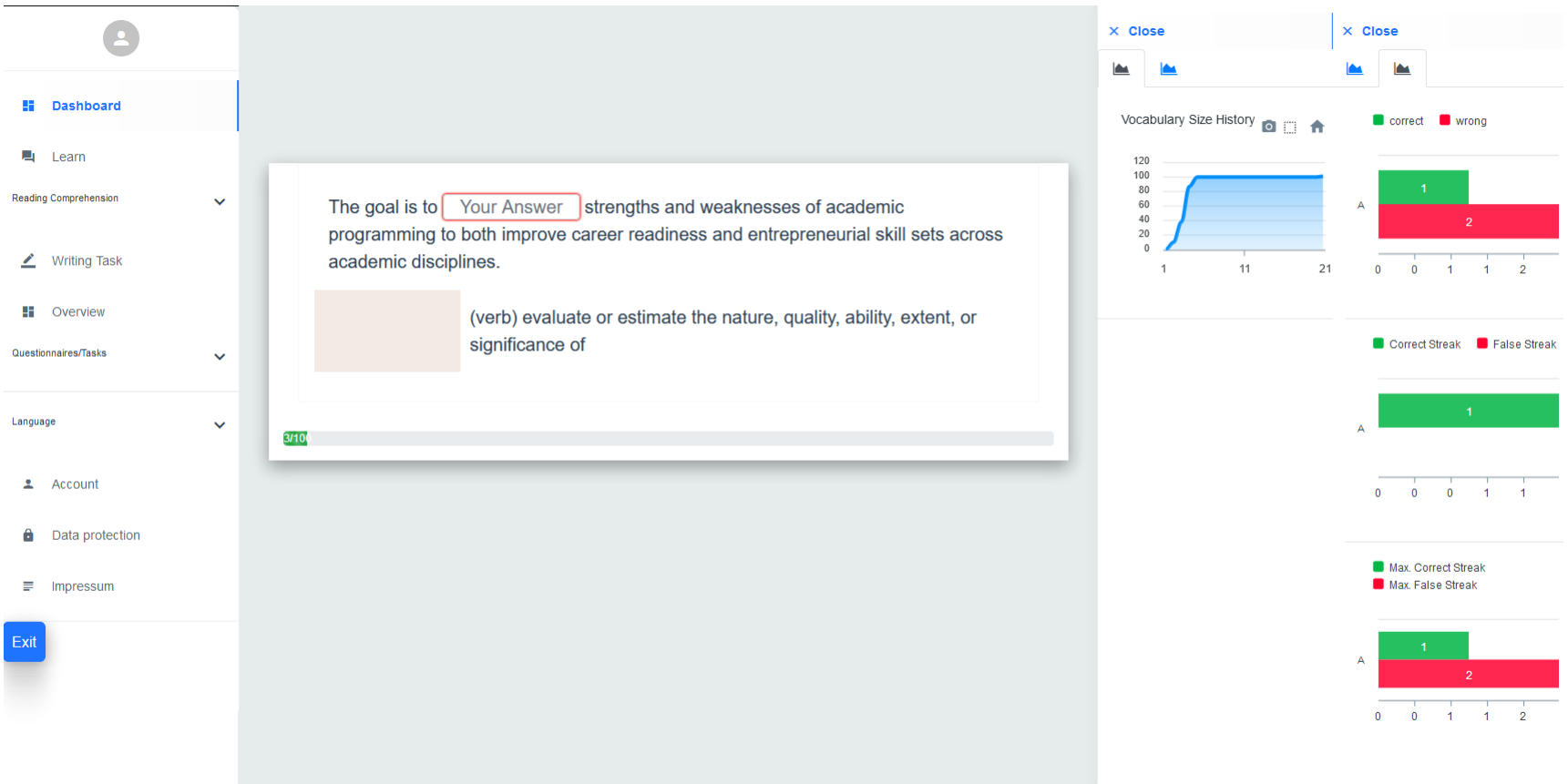
$$rank = \begin{cases} 0 & \text{if } d_{ac} < MIN_D \\ 0 & \text{if } \frac{d_{ac}}{d_{gen}} < MIN_R \\ \frac{f_{ac}}{f_{gen}} \cdot \frac{d_{ac}}{d_{general}} & \text{else} \end{cases} \quad (7)$$

Figure: Metrics used in the identification of statistically relevant vocabulary items

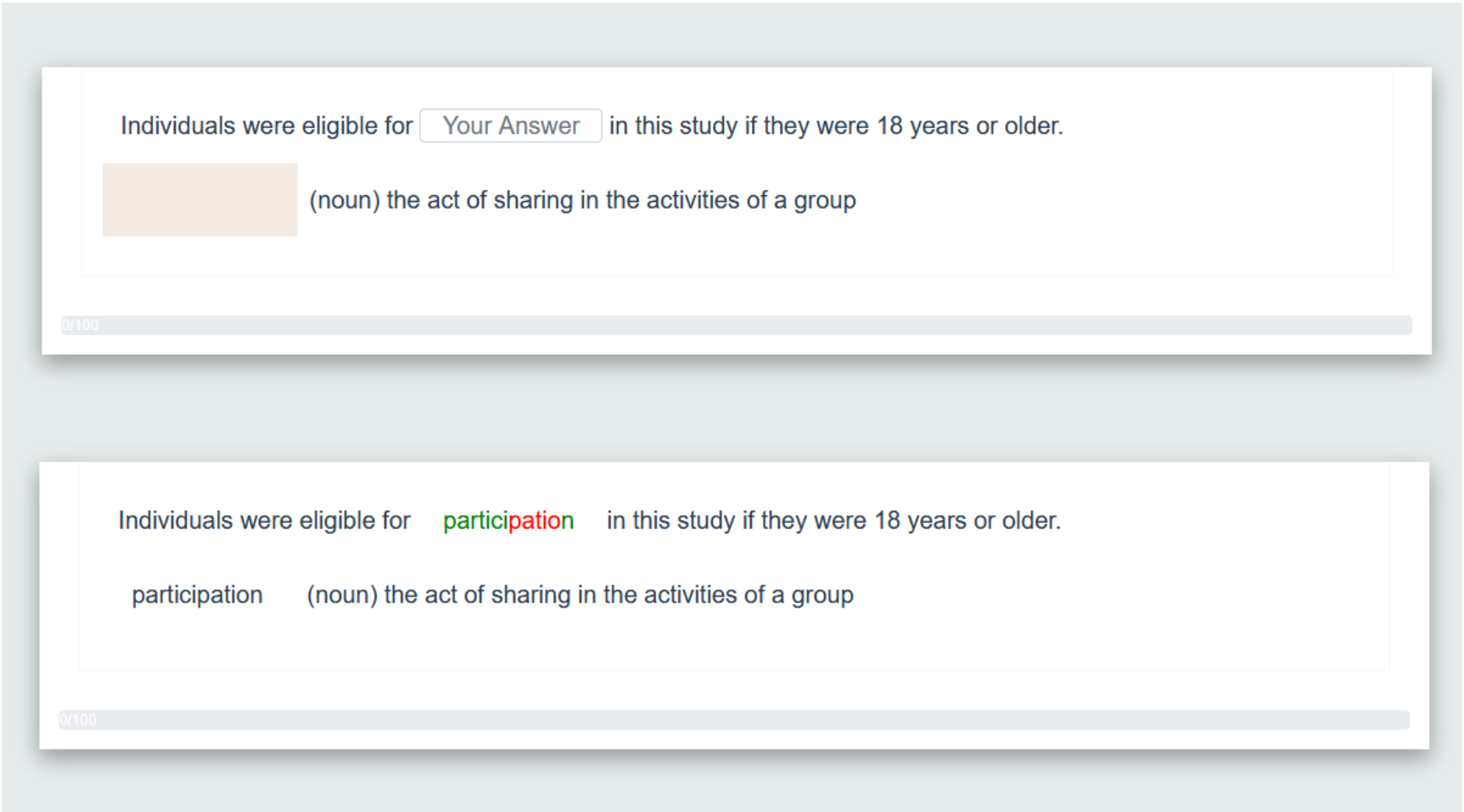
Optimal Presentation Intervals: Learning algorithm



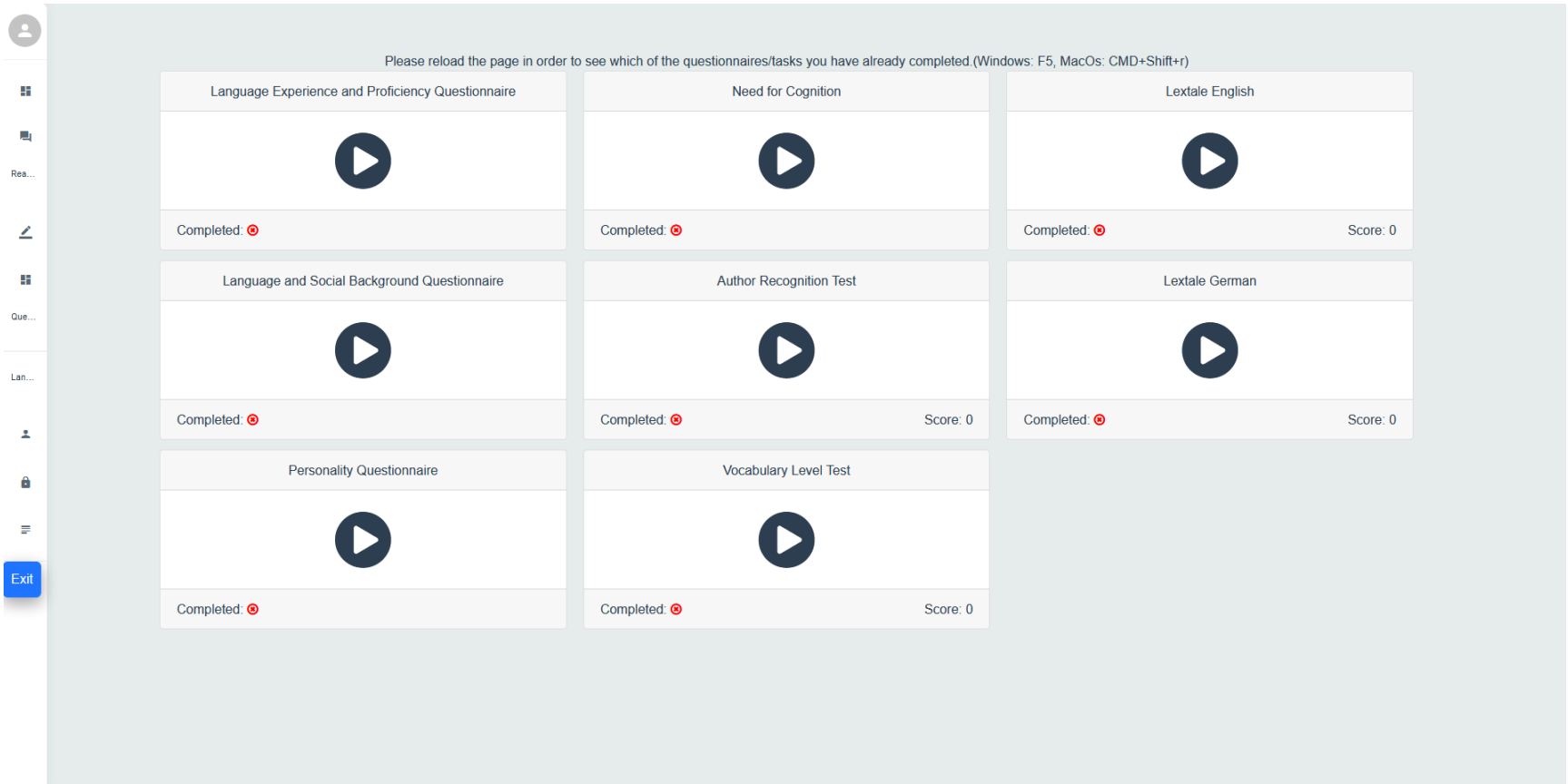
User Interface



User Interface



Integrated individual differences (IDs) tasks and measures



Modeling Growth Trajectories: Using AISLE System to Study Vocabulary Growth

Two research questions:

1. What is the best longitudinal model that describes participants' vocabulary growth and how much variation is there in growth rates?
2. What is the role of a range of IDs factors in explaining variation in participants' vocabulary acquisition?

Participants

- The data come from forty-six second language (L2) learners of English (25 female and 21 male, M= 22.98 years, SD = 3.32).
- All participants were university students from the RWTH Aachen University studying towards a BA or MA degree

RQ1: What is the best longitudinal model that describes participants' vocabulary growth

- Focus on the individual words (1-grams).
- Vocabulary growth within a four-hour engagement with the AISLE system
- Growth curve analysis.
  - Fitted linear, quadratic, and cubic growth models to the data
  - Use orthogonal polynomials of 'number of interactions' as our 'Time' variables
  - Models are compared using Akaike's Information Criterion

RQ1 - Results

- The best fitting model: cubic growth model

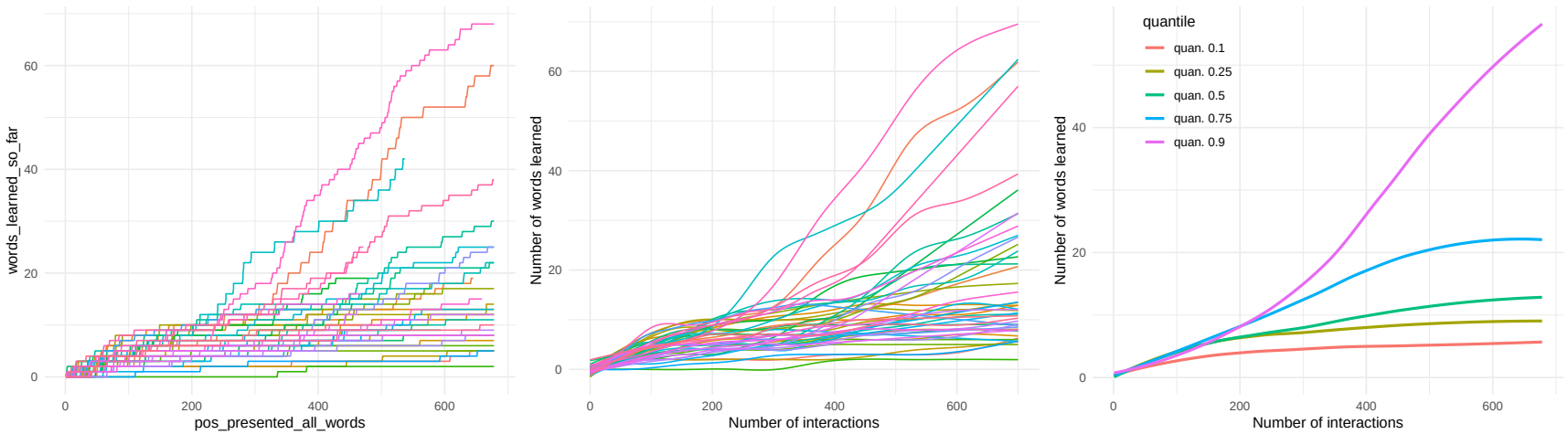


Figure: Plots of empirical growth trajectories (left), predicted growth trajectories from cubic model (center), and average predicted vocabulary growth at the 10th, 25th, 50th, 75th, and 90th percentiles

RQ2: What is the role of a range of IDs factors in explaining variation in participants' vocabulary acquisition

- Total of 17 individual differences (IDs) measures
  - 4 experience-based measures
  - 5 personality indicators
  - 8 cognitive measures
- All IDs variables were dichotomized based on median splits (high vs. low)
- The best fitting (minimal adequate) model was identified using a forward model selection procedure based on likelihood ratio tests.

RQ2 - Results

- 9 out of the 17 IDs variables were significant predictors of growth trajectories, NFC was the strongest
- The best fitting (minimal adequate) model contained the participants scores on the Need for Cognition (NFC) scale as well as scores on the Openness to Experience personality trait.

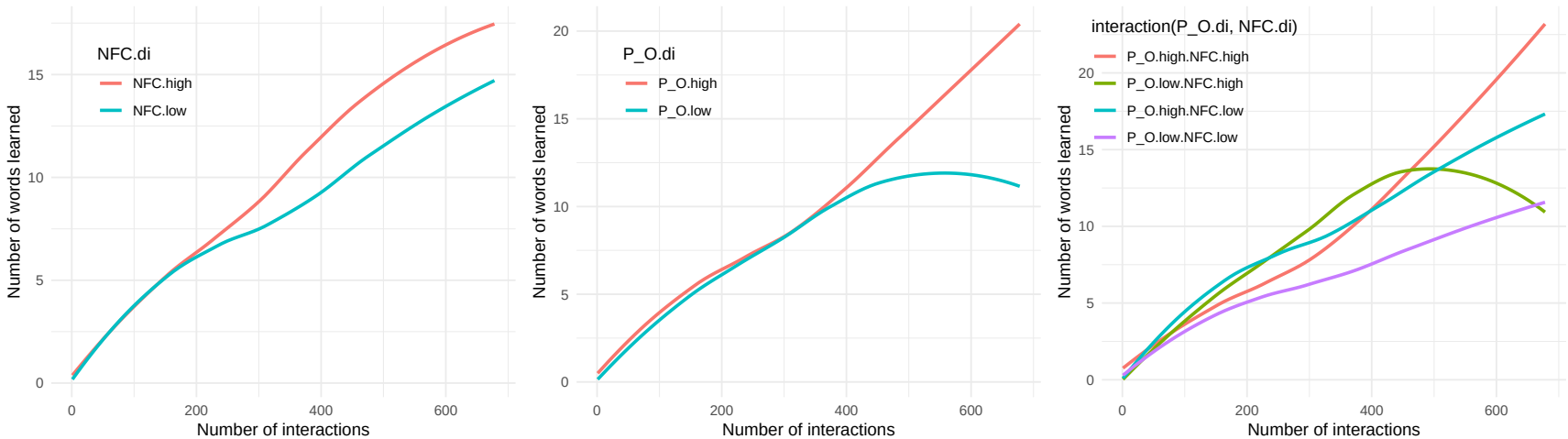


Figure: Predicted growth trajectories for participants with higher or lower NFC scores (left) and higher and lower scores on the openness to experience personality dimension (center). The plot on the right displays the results of the final model containing the effects of both NFC and Openness.

Summary

- Understanding the acquisition of language is important.
- The emergence of technology-enhanced language learning systems and automatic analyses of educational data obtained by such systems provide us with the opportunities.
- **In this presentation, we did 2 things:**
  - 1. Introduced the AISLE system
  - 2. Report in results of our study, answered 2 RQs
- **Quick results:**
  - There is substantial variation in VG.
  - This variation is syst. linked to experience-related, personality-based and cognitive IDs factors
- **Overall:** The data obtained from an adaptive language learning system such as AISLE have the potential to transform our current understanding of vocabulary growth and to provide a new window into the mechanisms and principles underlying language development in general.