Linguistic Constructs Represent the Domain Model in Intelligent Language Tutoring

Jue Hou

University of Helsinki

June 1, 2023

Outline

- Background
- Core Components
- System demo
- Data
- Modeling language mastery
  - Item Response Theory
- Challenges
- Experiments and Simulations
- Summary

Background

- helsinki.fi/revita — Online platform for language learning/tutoring *beyond the beginner level*
  - Collaboration with language teachers at several universities
  - Available for several languages
    - Finnish
    - Russian
    - Italian (β)
    - ...

Background

Main principle:

- User-selected content
  - Learner can upload arbitrary, real texts to use as learning content
- System automatically generates variety of exercises based on chosen text
  - Cloze (Fill-in-the-blank)
  - Multiple choice
  - Listening
  - ...

Background

System Structure:

- Domain model
  - Representation of real-world concepts / tasks / skills to be learned
- Student model
  - Representation of the learner's knowledge and skills
- Instruction model
  - Representation of the learning goals and the learning process

Background

Goal:

- Support personalized language learning process
- Provide feedback to learners and teachers

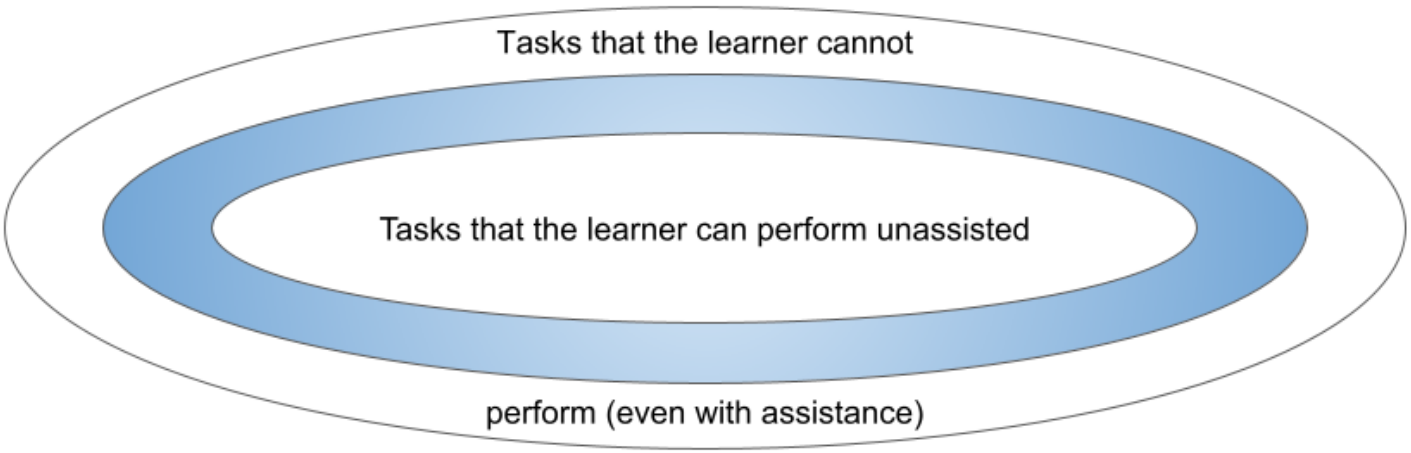


Figure: *Zone of proximal development*: blue area—tasks that the learner can perform with some assistance are those that the learner is most prepared to learn next.

Core components

Linguistic Construct as a representation of domain model

- *Constructs* are linguistic phenomena or rules, that vary in specificity
  - *Finnish verb government:* verb *tutustua* ("to become acquainted") requires its argument to be in illative case
  - *Construction Grammar:*
    - Grammatical constructs
    - Multi-word expressions (MWEs)
    - Collocations
    - Idioms
    - ...
- We engage language teachers to create constructs for their language
- Currently, Finnish and Russian have the most developed system of constructs, each with over 200 constructs.

Example of lingustic constructs

Constructs	Examples
Finnish	
(1) Necessive construction: Present passive participle, with *-ttava* ending	*Energiakriisin lähestyessä kaikki keinot on* **otettava** *käyntiin.* (With the energy crisis approaching, all means must **be taken** into action.)
(2) Transitive vs. intransitive verbs	*Voisitko* **sammuttaa** *valon?* (Could you **turn off** the light?)
(3) Verb government: translative case	*Kaupungit eivät ole muuttuneet* **energiatehokkaammiksi**. (Cities have not become **more energy efficient**.)
(4) Substitute clause: participle substitutes for "that"-relative clause	*Maija kertoi* **vanhempien asuvan** *kaupungissa.* (Maija said that **her parents live** in the city.)
Russian	
(5) Verb Conjugation Irregular	*Мы скоро* ***увидим*** *восход.* (We **will see** the sunrise soon.)
(6) Complex pronoun:	*Нам нужно* ***кое о чем*** *поговорить.* (We need to talk **about something**)
(7) Perfective vs. imperfective aspect	*Страны* ***согласовали*** *проект о будущих отношениях.* (The countries **agreed on** a draft on future relations.)
(8) Dative subject & impersonal verb	***Мне*** *необходимо поговорить с врачом.* (**I** need to talk to a doctor.)
German	
(9) Past perfect tense	*Ich* **wäre** *mit ihm* **gekommen**, *aber er wurde krank.* (I **would have come** with him, but he got sick.)
(10) Weak masculine nouns	*Ich möchte* **den Jungen** *kennenlernen.* (I want to meet **the boy**.)
(11) Prepositions governing dative case	*Wir sind* **aus dem** *Haus gelaufen.* (We ran **out of the** house.)

Core components

Construct detection

- Token level → Token features
  - HFST analyzer
  - NN-based morphological analyzer
  - ...
- Phrase/sentence level → Context features
  - Dependency parsing
  - Rule-based pattern detection

Potential exercises are based on detected constructs in text

- Selected according to learner's level and construct difficulty
- Highlighted in *Reading View*

Core components

Exercise generation: Exercise type

- Cloze → Lemma
  - Morphological analyzers
- Multiple choice → Distractor generation
  - Rules
  - Morphological generators
  - Morphological analyzers (e.g., UDAR for stress in Russian)
- Listening exercise → Context
  - Dependency parsing
  - Text-to-speech synthesis

Each exercise is associated with a construct

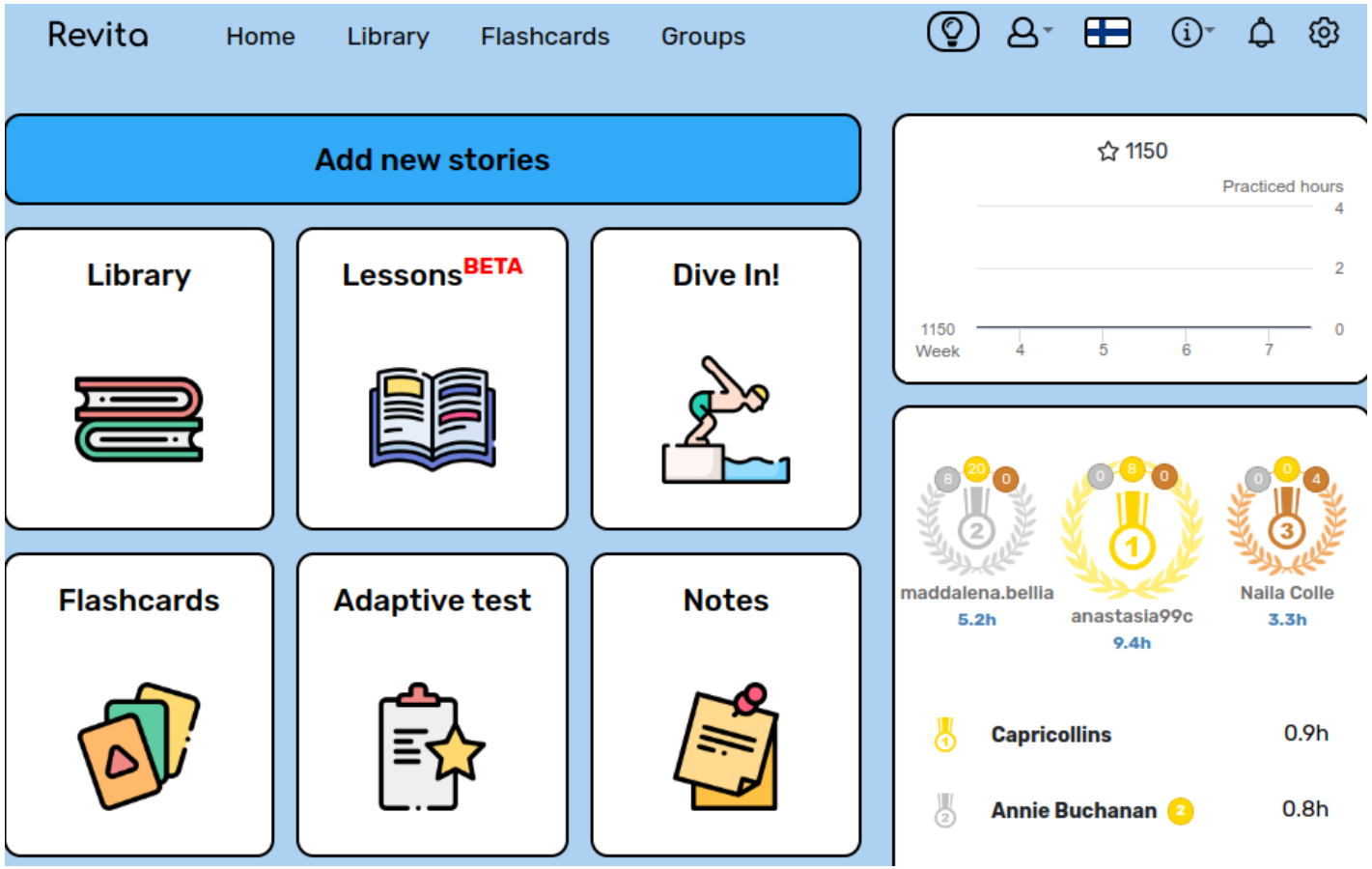
Core components

Feedback

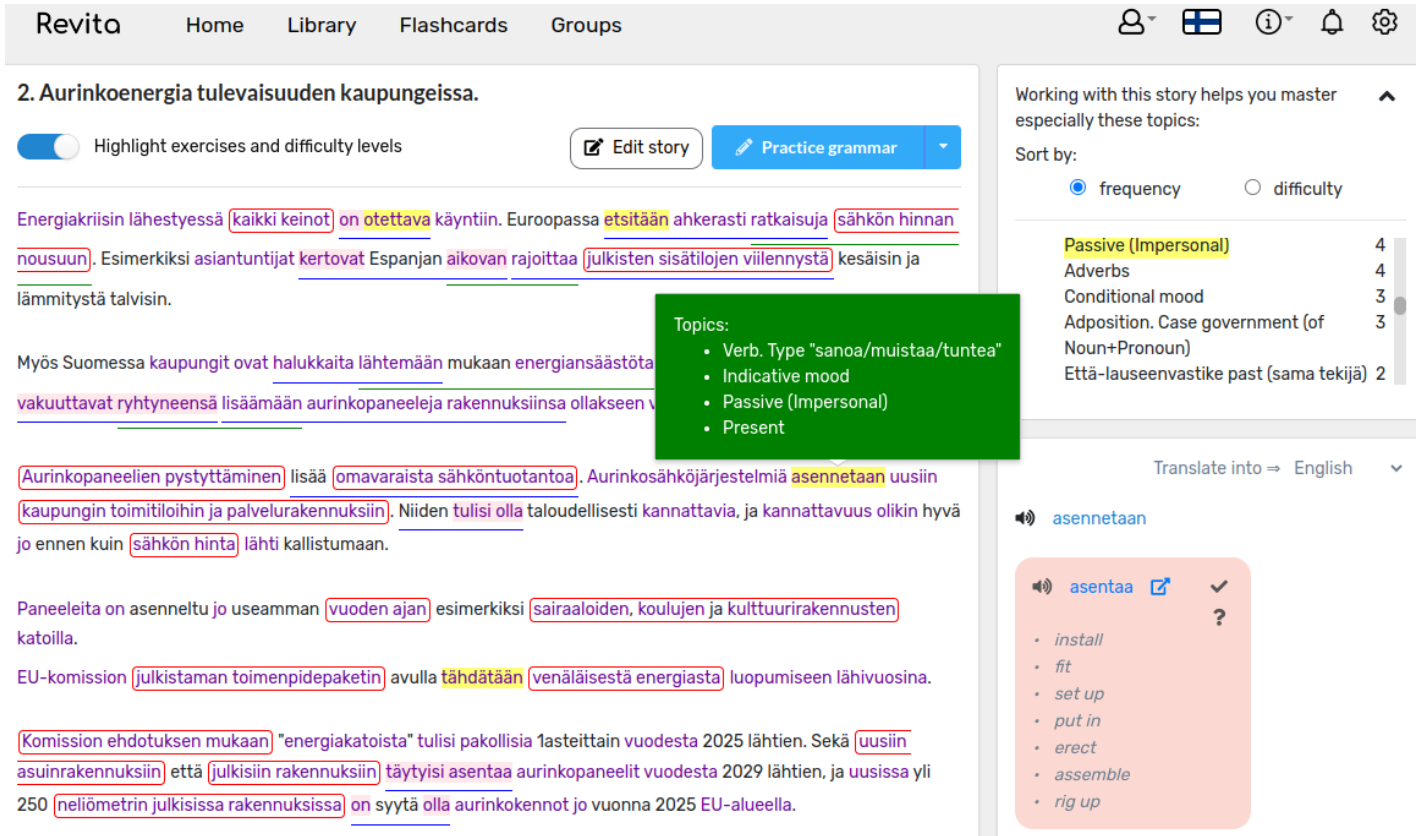
- Iterative, increasing specificity
- Based on constructs
- Based on learner's answer
  - Grammatical features
  - Potential context features
- Based on language-specific hierarchy
  - $\rightarrow$ Feature
  - $\rightarrow$ Order

Each hint is associated with a construct as well

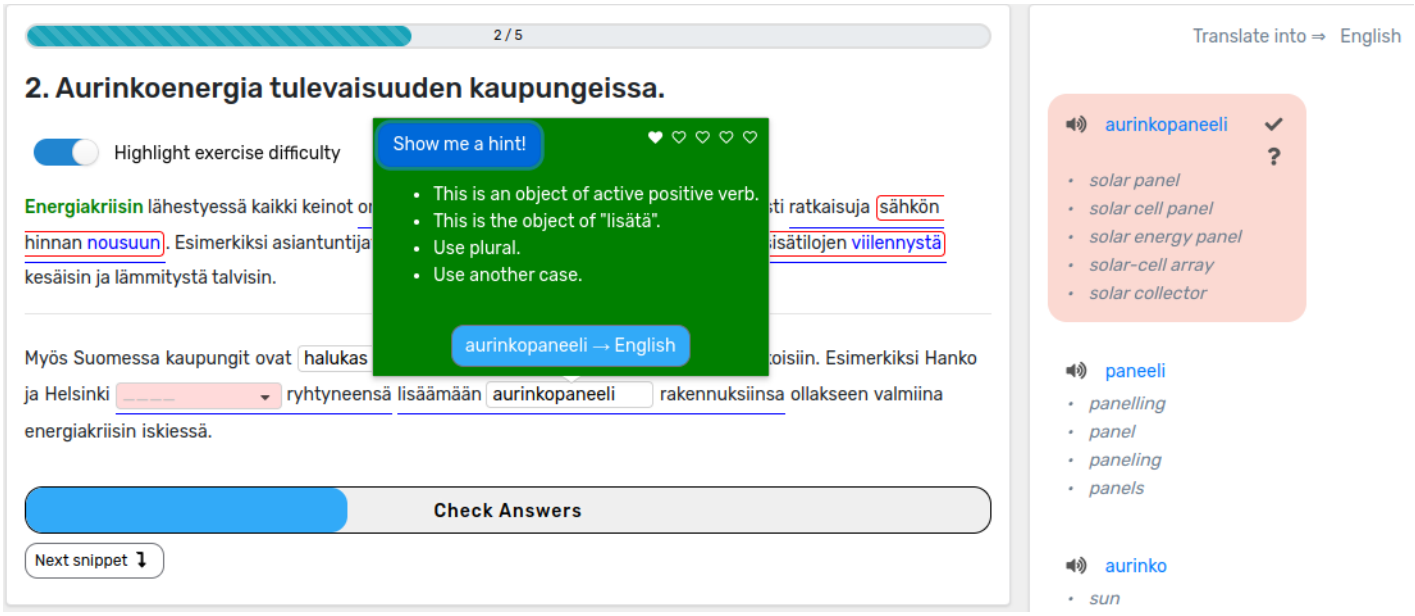
System demo


Revita
Home
Library
Flashcards
Groups
Add new stories
Library
Lessons BETA
Dive In!
Flashcards
Adaptive test
Notes
1150
Practiced hours
4
2
0
1150
Week
4
5
6
7
maddalena.bellia
5.2h
anastasia99c
9.4h
Naila Colle
3.3h
Capricollins
0.9h
Annie Buchanan
0.8h

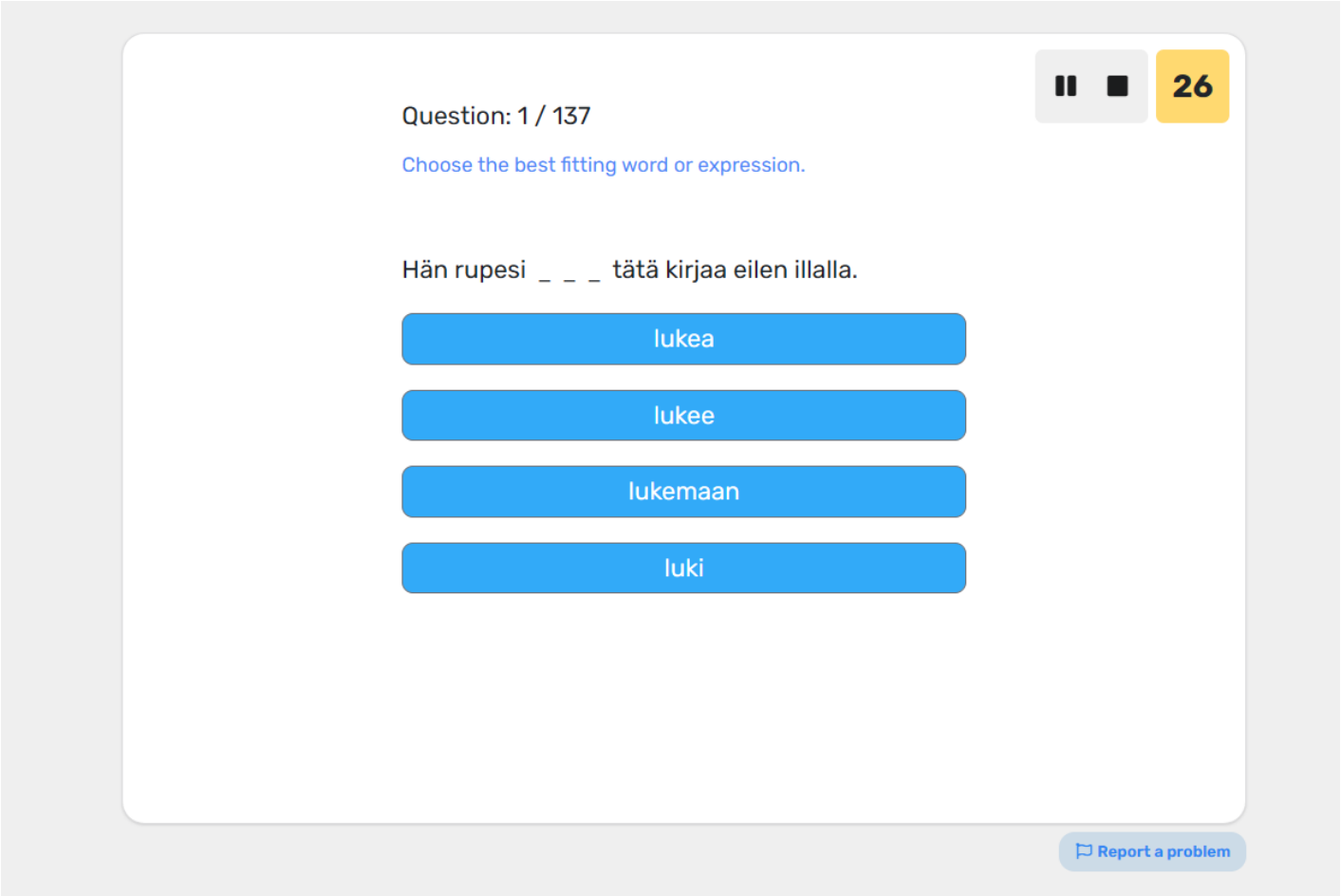
System demo


Revita
Home
Library
Flashcards
Groups
2. Aurinkoenergia tulevaisuuden kaupungeissa.
Highlight exercises and difficulty levels
Edit story
Practice grammar
Energiakriisin lähestyessä kaikki keinot on otettava käyntiin. Euroopassa etsitään ahkerasti ratkaisuja sähkön hinnan nousuun. Esimerkiksi asiantuntijat kertovat Espanjan aikovan rajoittaa julkisten sisätilojen viilennystä kesäisin ja lämmitystä talvisin.
Myös Suomessa kaupungit ovat halukkaita lähtemään mukaan energiansäästöta
vakuuttavat ryhtyneensä lisäämään aurinkopaneeleja rakennuksiinsa ollakseen
Aurinkopaneelien pystyttäminen lisää omavaraista sähköntuotantoa. Aurinkosähköjärjestelmiä asennetaan uusiin kaupungin toimitiloihin ja palvelurakennuksiin. Niiden tulisi olla taloudellisesti kannattavia, ja kannattavuus olikin hyvä jo ennen kuin sähkön hinta lähti kallistumaan.
Paneeleita on asenneltu jo useamman vuoden ajan esimerkiksi sairaaloiden, koulujen ja kulttuurirakennusten katoilla.
EU-komission julkistaman toimenpidepaketin avulla tähdätään venäläisestä energiasta luopumiseen lähivuosina.
Komission ehdotuksen mukaan "energiakatoista" tulisi pakollisia 1asteittain vuodesta 2025 lähtien. Sekä uusiin asuinrakennuksiin että julkisiin rakennuksiin täytyisi asentaa aurinkopaneelit vuodesta 2029 lähtien, ja uusissa yli 250 neliömetrin julkisissa rakennuksissa on syytä olla aurinkokennot jo vuonna 2025 EU-alueella.
Topics:
Verb. Type "sanoa/muistaa/tuntea"
Indicative mood
Passive (Impersonal)
Present
Working with this story helps you master especially these topics:
Sort by:
frequency
difficulty
Passive (Impersonal) 4
Adverbs 4
Conditional mood 3
Adposition. Case government (of Noun+Pronoun) 3
Että-lauseenvastike past (sama tekijä) 2
Translate into ⇒ English
asennetaan
asentaa
install
fit
set up
put in
erect
assemble
rig up

System demo


2 / 5
2. Aurinkoenergia tulevaisuuden kaupungeissa.
Highlight exercise difficulty
Show me a hint!
This is an object of active positive verb.
This is the object of "lisätä".
Use plural.
Use another case.
aurinkopaneeli → English
Energiakriisin lähestyessä kaikki keinot o
ti ratkaisuja sähkön
hinnan nousuun. Esimerkiksi asiantuntija
isätilojen viilennystä
kesäisin ja lämmitystä talvisin.
Myös Suomessa kaupungit ovat halukas
oisiin. Esimerkiksi Hanko
ja Helsinki ryhtyneensä lisäämään aurinkopaneeli rakennuksiinsa ollakseen valmiina
energiakriisin iskiessä.
Check Answers
Next snippet
Translate into ⇒ English
aurinkopaneeli
solar panel
solar cell panel
solar energy panel
solar-cell array
solar collector
paneeli
panelling
panel
paneling
panels
aurinko
sun

System demo


26
Question: 1 / 137
Choose the best fitting word or expression.
Hän rupesi _ _ _ tätä kirjaa eilen illalla.
lukea
lukee
lukemaan
luki
Report a problem

Data

Russian:

- Exercise
  - Generated from any arbirary text
    - No explicit item bank
  - Selected according to learner's level
  - 214K exercise responses from 1.5K learners
    - Including information about hints
    - Involve multiple constructs
- Test
  - Dichotoumous (correct/incorrect) multiple choice questions
  - Exhausitve assessment follows a fixed template (300 questions)
  - 750K test responses from 1.8K learners
  - Manual difficulty labeled by teachers

Modeling language mastery

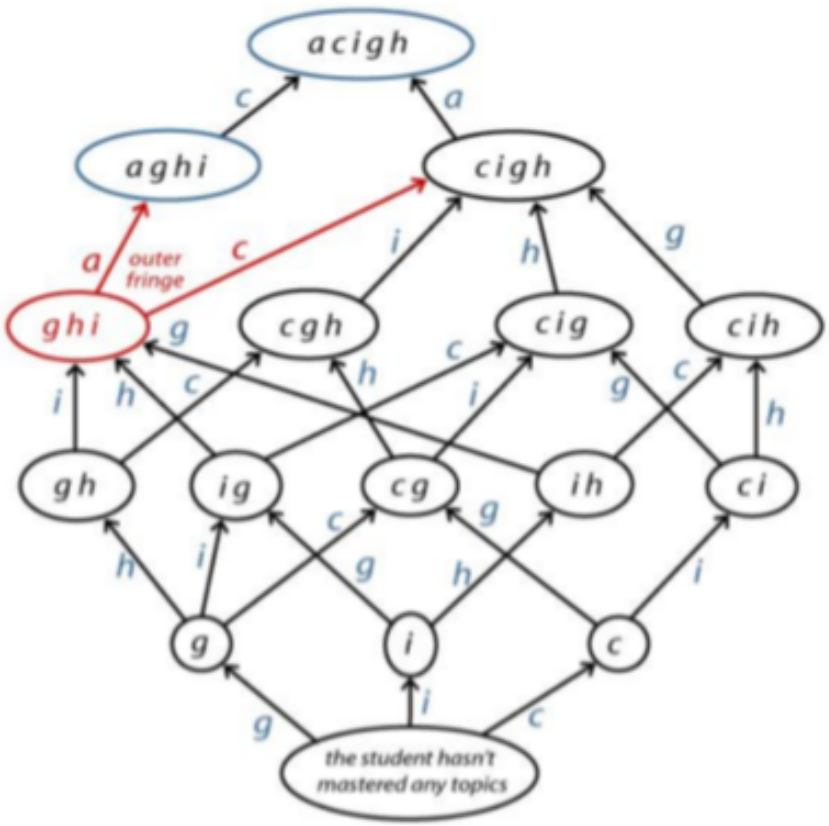
- Techniques to model user's proficiency:
  - Bayesian Knowledge Tracing
    - model learner's mastery in a Hidden Markov Model as latent variables

Modeling language mastery

- Techniques to model user's proficiency:
  - Bayesian Knowledge Tracing
    - model learner's mastery in a Hidden Markov Model as latent variables
  - Knowledge Space Theory
    - combine concepts / skills into knowledge states
    - build a graph (knowledge space) of states to represent learning path

Modeling language mastery

- Techniques to model user's proficiency:
  - Bayesian Knowledge Tracing
    - model learner's mastery in a Hidden Markov Model as latent variables
  - Knowledge Space Theory
    - combine concepts / skills into knowledge states
    - build a graph (knowledge space) of states to represent learning path



Modeling language mastery

- Techniques to model user's proficiency:
  - Bayesian Knowledge Tracing
    - model learner's mastery in a Hidden Markov Model as latent variables
  - Knowledge Space Theory
    - combine concepts / skills into knowledge states
    - build a graph (knowledge space) of states to represent learning path
  - Dynamic Key-value Memory Network
    - model learner's mastery with single-head attention

Modeling language mastery

- Techniques to model user's proficiency:
  - Bayesian Knowledge Tracing
    - model learner's mastery in a Hidden Markov Model as latent variables
  - Knowledge Space Theory
    - combine concepts / skills into knowledge states
    - build a graph (knowledge space) of states to represent learning path
  - Dynamic Key-value Memory Network
    - model learner's mastery with single-head attention
  - Item Response Theory
  - ...

Item Response Theory

- Item Response Theory (IRT) — psychometric theory that models the relationship the latent trait and observed performance
- IRT is applied in many settings including stress testing, psychological and medical testing, etc.
  - Anxiety
  - Neurosis
  - Personality
  - Language proficiency

Item Response Theory

3PL: "Three-parameter logistic model"
Probability that student s with current ability estimate θ_s will give a correct answer to Q_i— Question item i.
The probability function is expressed as:

$$P(\theta_s, Q_i) = c_i + (1 - c_i) \cdot \frac{1}{1 + exp(-a_i(\theta_s - b_i))} \tag{1}$$

where the parameters—the properties of Q_i—are:

- a_i: discrimination factor,
- b_i: estimate of difficulty,
- c_i: probability that a random guess is correct.

Item Response Theory

Item information: measures the amount of information a question Q_i yields, based on the learner's current ability estimate θ_s

$$I(\theta_s, Q_i) = a_i^2 \frac{1 - P(\theta_s, Q_i)}{P(\theta_s, Q_i)} \left[\frac{P(\theta_s, Q_i) - c_i}{1 - c_i} \right]^2 \tag{2}$$

Information function: used during the adaptive test to select the most informative item, for given value of ability θ_s.

Challenges: Test

- Test question as an item
  - Easy to assign credit
- Long and exhausting process
  - 300 questions overall
  - 15 sec for each question
  - Stressful for students

Challenges: Test

- Test question as an item
  - Easy to assign credit
- Long and exhausting process
  - 300 questions overall
  - 15 sec for each question
  - Stressful for students

Research questions:

- **RQ1**: Does imperfect learner data still provide robust assessment of learner ability?
- **RQ2**: How do estimates of ability from a model trained on learner data compare with estimates of ability based on question difficulty assessed *manually* by teachers?

Challenges: Exercises

- Compared with test items:
  - *Item* not as clearly defined as test questions
  - Not clear judgement on answers when assigning credit and penalty

Challenges: Exercises

- Compared with test items:
  - *Item* not as clearly defined as test questions
  - Not clear judgement on answers when assigning credit and penalty
- Construct as an item
  - Map exercise to constructs
  - Detect learner's error as constructs
  - 1-N mapping
- Rely on NLP components
  - Dependency parser
  - Morphological analyzer
  - Rule-based pattern matching
  - ...

Challenges: Exercises

- Compared with test items:
  - *Item* not as clearly defined as test questions
  - Not clear judgement on answers when assigning credit and penalty
- Construct as an item
  - Map exercise to constructs
  - Detect learner's error as constructs
  - 1-N mapping
- Rely on NLP components
  - Dependency parser
  - Morphological analyzer
  - Rule-based pattern matching
  - ...

Research questions:

- **RQ3**: Can we reliably model learner ability based on the learner responses to exercises — without testing?

Experiments and Simulations: RQ1

- **RQ1**: Does imperfect learner data still provide robust assessment of learner ability?
- **RQ2**: How do estimates of ability from a model trained on learner data compare with estimates of ability based on question difficulty assessed *manually* by teachers?
- **RQ3**: Can we reliably model learner ability based on the learner responses to exercises?

Experiments and Simulations: RQ1

Simulation process: *Adaptive test*

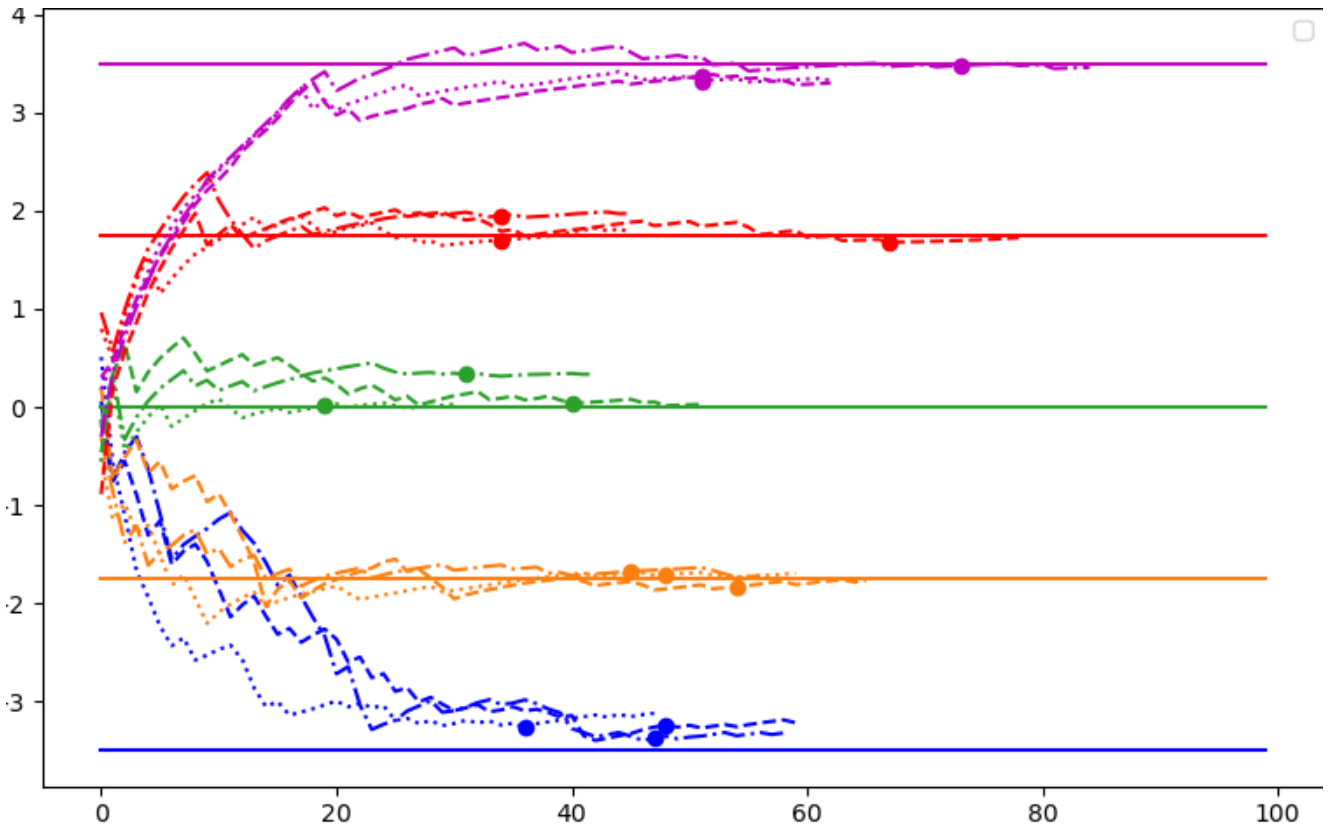
1. Initialize ability θ_0 randomly
2. Pick the most informative question from item bank
   - $i = \arg\max_i I(\theta_n, Q_i)$
3. User answers selected question $\rightarrow$ re-estimate θ_{n+1}
4. Repeat from step 2 and 3 until θ_n converges

Experiments and Simulations: RQ1

- Trained with 750K test responses
- Simulate with "artificial" user
  - 5 different ability levels
  - 3 simulations each

Experiments and Simulations: RQ1

- Trained with 750K test responses
- Simulate with "artificial" user
  - 5 different ability levels
  - 3 simulations each

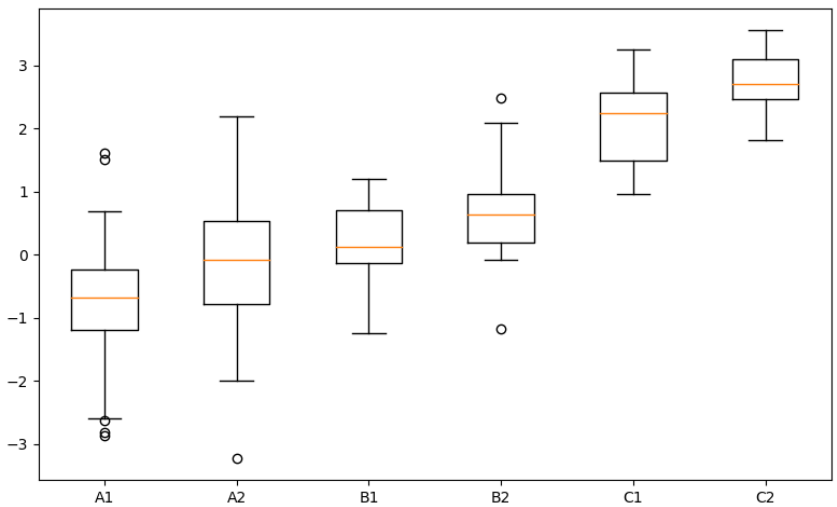


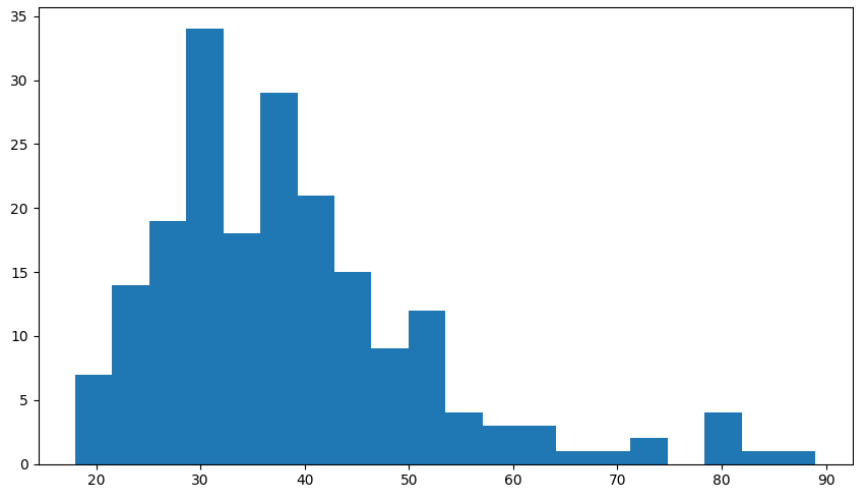
Experiments and Simulations: RQ1

- Simulate with real data
  - $\sim$ 200 students with grades assigned by teachers
  - pick question from previous test session instead of entire item bank

Experiments and Simulations: RQ1

- Simulate with real data
  - $\sim$ 200 students with grades assigned by teachers
  - pick question from previous test session instead of entire item bank





Experiments and Simulations: RQ1

Does imperfect learner data still provide robust assessment of learner ability?

- Imperfect exhaustive test process
- Feasible for IRT
  - Correlates well with manually assigned grade
- More efficient than exhaustive testing
  - Vast majority of tests converge in 60 questions or less.
  - Default exhaustive test length is 300 questions.

Experiments and Simulations: RQ2

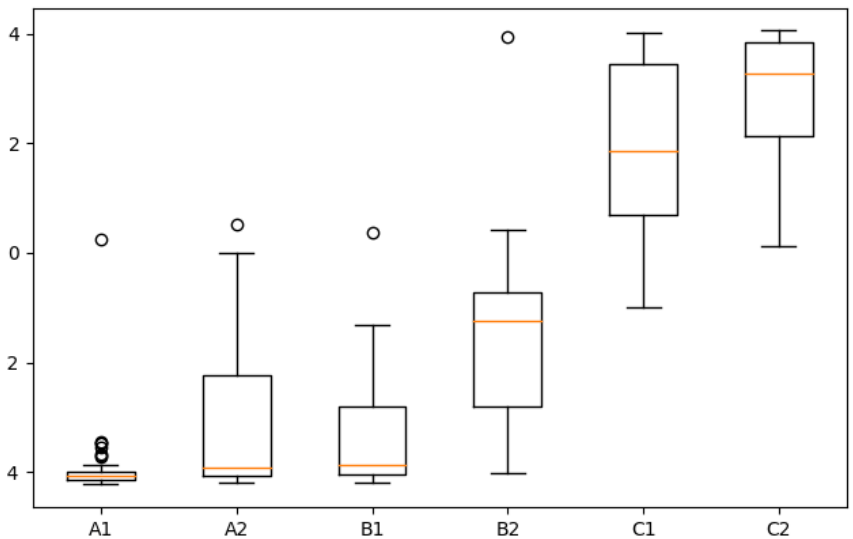
- **RQ1**: Does imperfect learner data still provide robust assessment of learner ability?
- **RQ2**: How do estimates of ability from a model trained on learner data compare with estimates of ability based on question difficulty assessed *manually* by teachers?
- **RQ3**: Can we reliably model learner ability based on the learner responses to exercises?

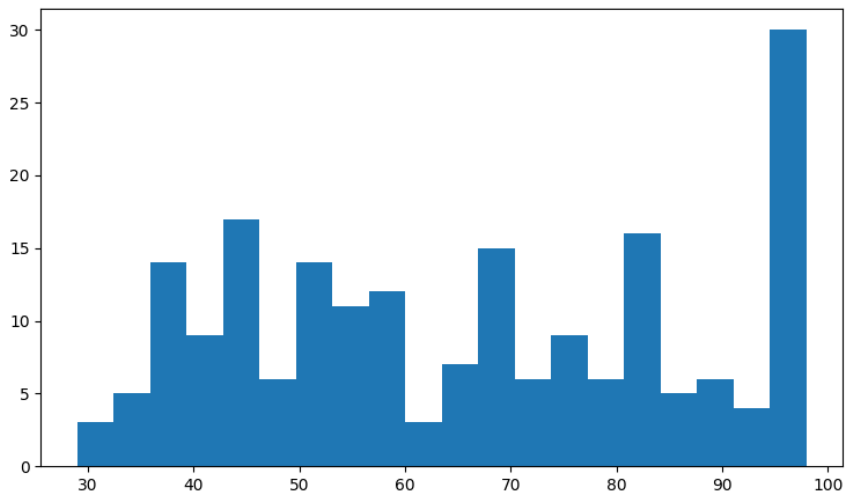
Experiments and Simulations: RQ2

- Item difficulty — manually set by teachers
- Similar simulation as RQ1
- Cutoff: 100 questions

Experiments and Simulations: RQ2

- Item difficulty — manually set by teachers
- Similar simulation as RQ1
- Cutoff: 100 questions





Experiments and Simulations: RQ2

- Item difficulty — manually set by teachers
- "Full benefit of the doubt" $\rightarrow$ estimate with full test session

Experiments and Simulations: RQ2

- Item difficulty — manually set by teachers
- "Full benefit of the doubt" $\rightarrow$ estimate with full test session

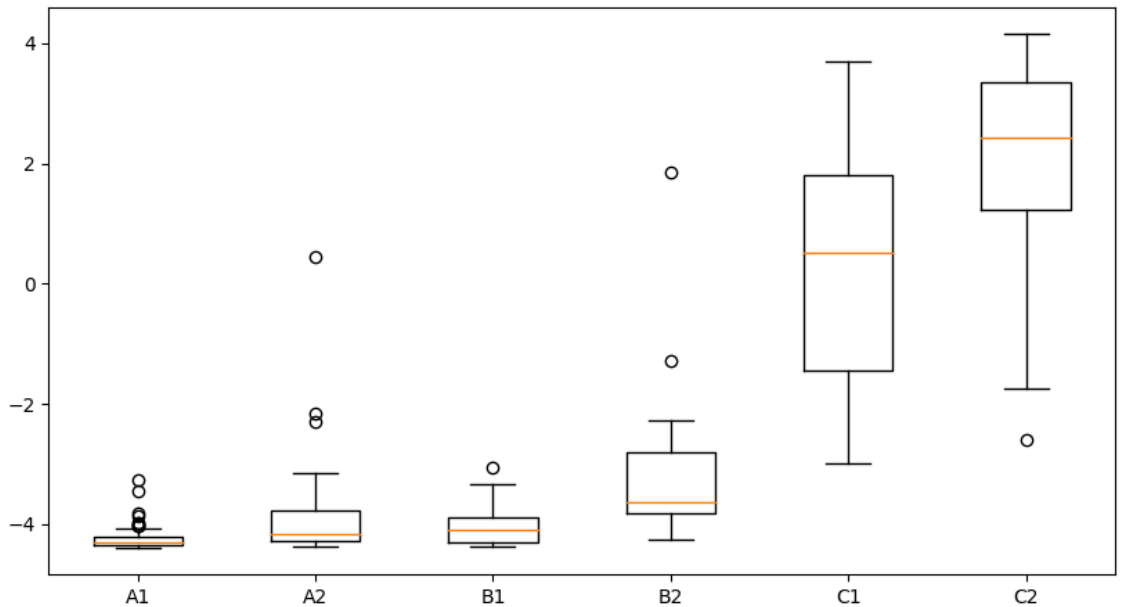


Figure: X-axis—the 6 CEFR levels; Y-axis—ability estimate

Experiments and Simulations: RQ2

How do estimates of ability from a model trained on learner data compare with estimates of ability based on question difficulty assessed *manually* by teachers?

- Far worse than applying item difficulty learned by the model
- Item parameters learned from data are more accurate than the question levels labeled by experts in language teaching

Experiments and Simulations: RQ3

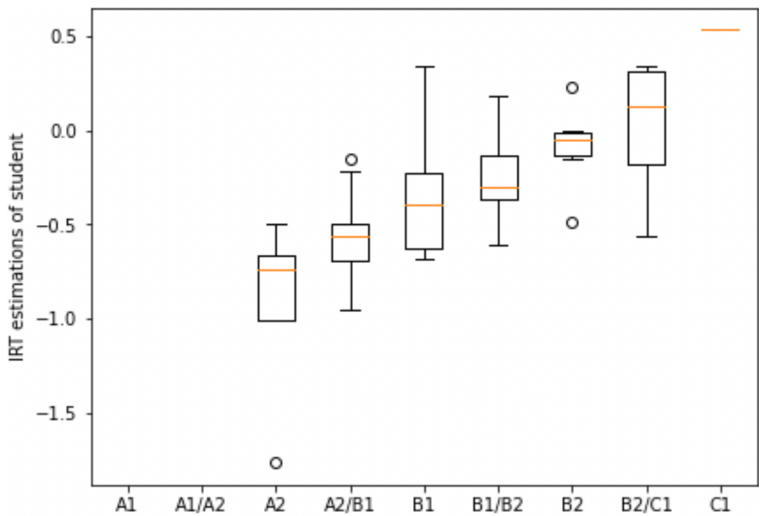
- **RQ1**: Does imperfect learner data still provide robust assessment of learner ability?
- **RQ2**: How do estimates of ability from a model trained on learner data compare with estimates of ability based on question difficulty assessed *manually* by teachers?
- **RQ3**: Can we reliably model learner ability based on the learner responses to *exercises*?

Experiments and Simulations: RQ3

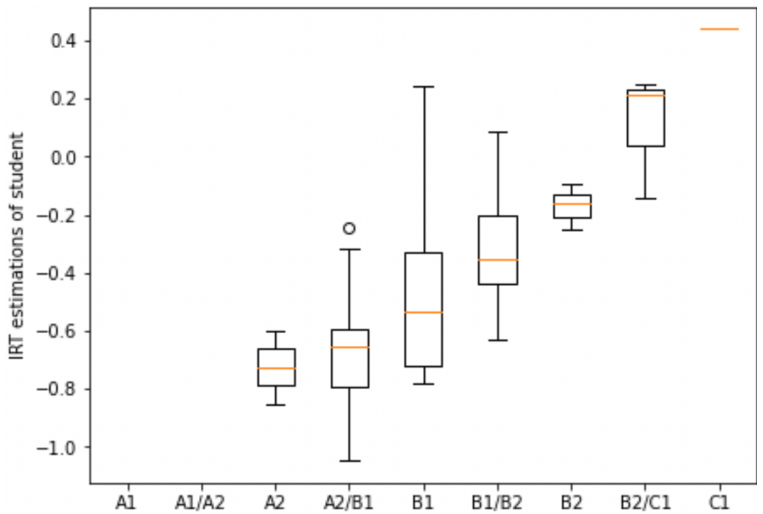
- Train model with data from reliable students
  - Data from students who have done over min_{exer} exercises

Experiments and Simulations: RQ3

- Train model with data from reliable students
  - Data from students who have done over min_{exer} exercises



(a) $min_{exer} = 50$, $min_{constr} = 4$, $\rho = .663$



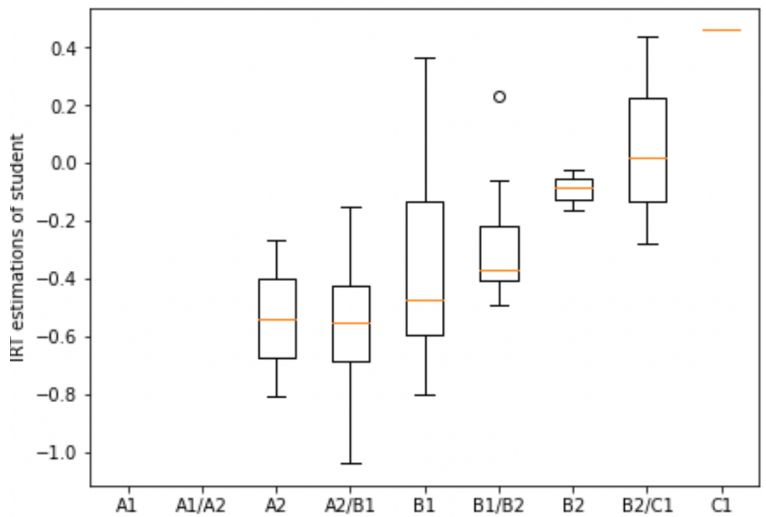
(b) $min_{exer} = 100$, $min_{constr} = 4$, $\rho = .724$

Experiments and Simulations: RQ3

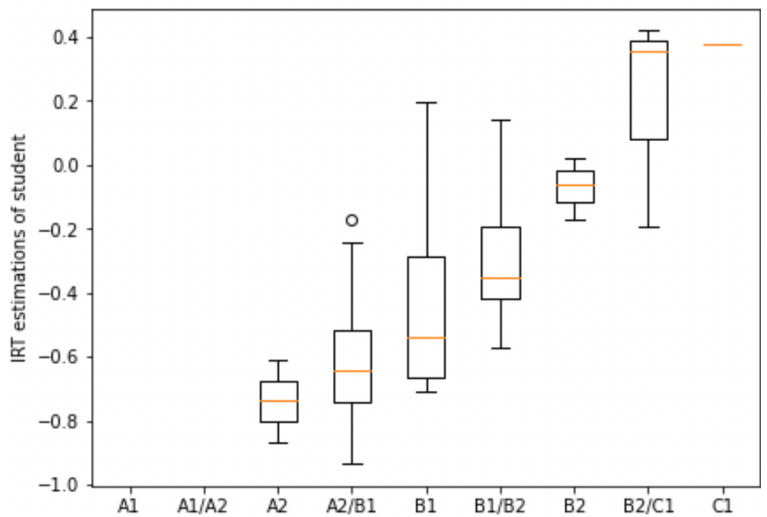
- Evaluate with reliable constructs
  - Estimate with constructs that have over min_{constr} responses

Experiments and Simulations: RQ3

- Evaluate with reliable constructs
  - Estimate with constructs that have over min_{constr} responses



(a) $min_{exer} = 100$, $min_{constr} = 1$, $\rho = .608$



(b) $min_{exer} = 100$, $min_{constr} = 7$, $\rho = .75$

Experiments and Simulations: RQ3

Can we reliably model learner ability based on the learner responses to exercises?

- No explicit item bank as we have for test
- As good as scores from adaptive test
- Higher min_{exer} and min_{constr} $\rightarrow$ better model

Summary

- Present language learning platform Revita
- Linguistic constructs as Revita's domain model
- Illustrate the use of IRT to model learner mastery
  - Imperfect learner data from tests is usable to build a reliable adaptive model
  - Item parameters learned from data are more accurate than manual item difficulty
  - Model learner ability from responses to exercises
    - No explicit item

Thank you!

revita.cs.helsinki.fi