

Automatic assessment of second language proficiency

Klára Vaníčková



Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics



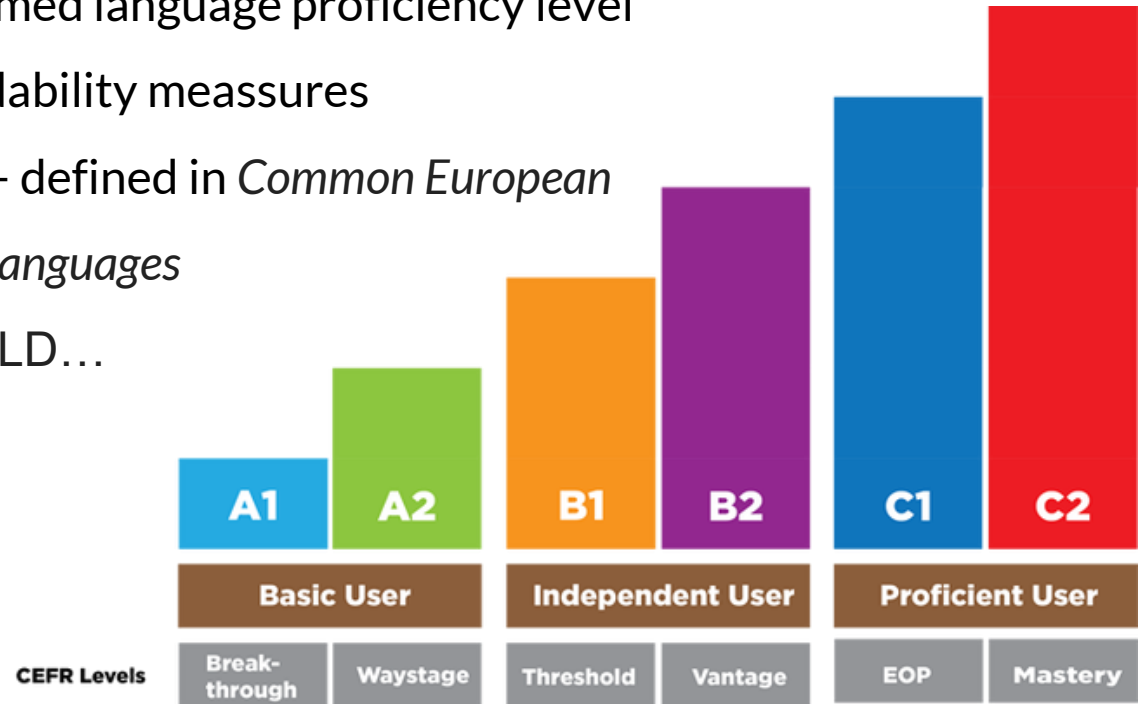
unless otherwise stated

Original plan

- Longitudinal study of students' progress in Czech as a foreign language
- Collect data every week/two weeks in preparatory courses (from 0 to B2 in 9 months)
- Complications
 - COVID-19 unclear situation about incoming students
 - Financing annotators + respondents – not found feasible system
 - failed GAUK proposal

New task

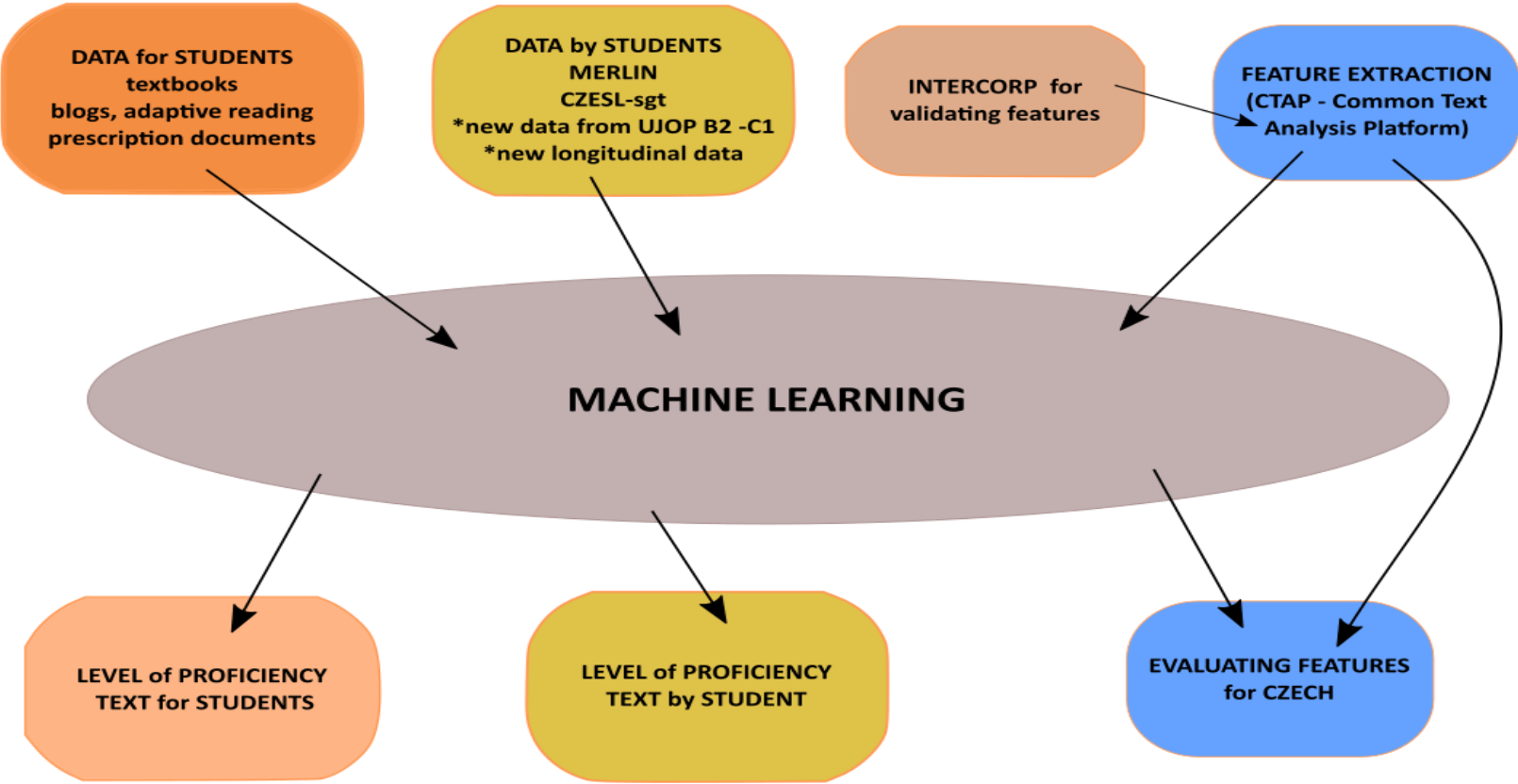
- Categorize texts **for/by** students of Czech as a foreign language according their proclaimed language proficiency level by complexity and readability measures
- Language proficiency – defined in *Common European Framework of Reference for Languages*
- Solving by TOEFL, EVALD...



Motivation

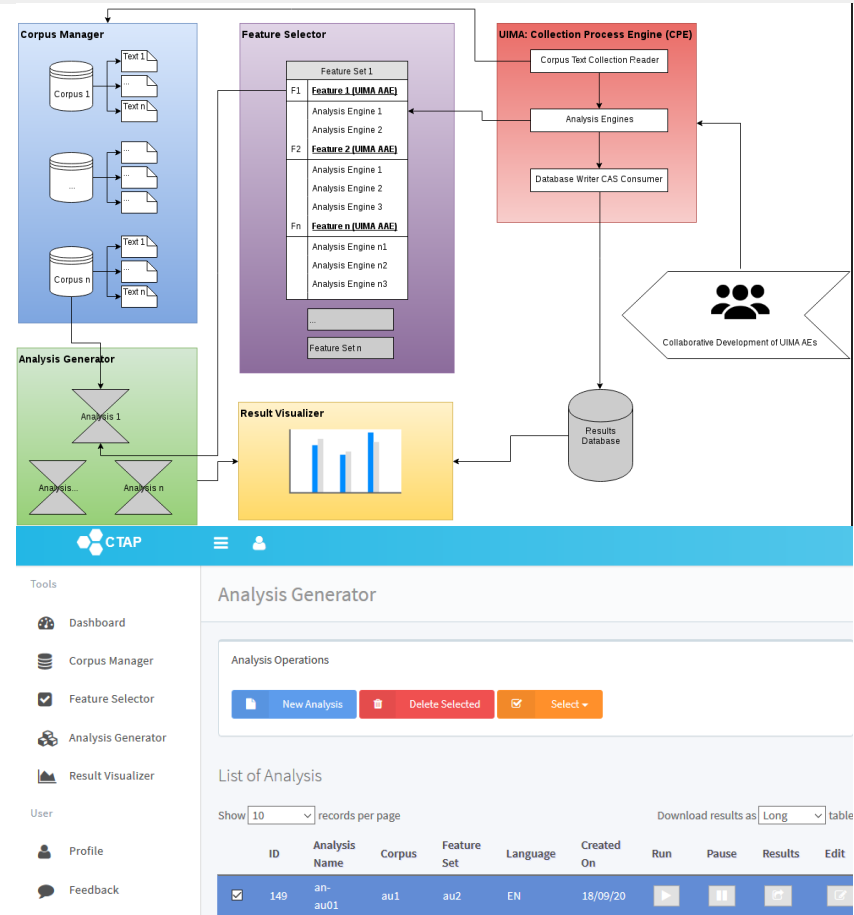
- Level assesment tool for purpose of exams
 - Certified exams
 - Exam for permanent residency
 - Exam for citizenship
- Self-assesment tool
- Scientific pedagogical tool (selection of appropriate reading texts)

Realization



Common Text Analysis Platform (CTAP)

- Chen, X., Meurers, D. (2016). CTAP: A Web-Based Tool Supporting Automatic Complexity Analysis. In *Proceedings of The Workshop on Computational Linguistics for Linguistic Complexity*. Osaka, Japan
- Collected around 600 complexity features: EN, DE, IT, FR, NL, PT, **CS**
- plain text -> NLP "annotators" -> rule-based feature "extractors" -> analysis -> visualization
- infrastructure ready, just plug in CS NLP , write "annotators", re-use (adjust?) extractors available, add own extractors (?)
- User-friendly web GUI



Adaptation of international complexity formulas to Czech

$$RE = 206.835 - (1.015 \times ASL) - (84.6 \times ASW)$$

RE = Readability Ease

ASL = Average Sentence Length (i.e., the number of words divided by the number of sentences)

ASW = **Average number of syllables per word** (i.e., the number of syllables divided by the number of words)

90-100 : Very Easy

80-89 : Easy

70-79 : Fairly Easy

60-69 : Standard

50-59 : Fairly Difficult

30-49 : Difficult

0-29 : **Very Confusing**

text	Flesch RE
jsem-blizky-cloveku-ve-vykonu-trestu	25
jsem-v-exekuci	6
prava-a-povinnosti-dluznika/	32
/web/msp/podani-stiznosti-k-eslp	1
ministerstvo-spravedlnosti	-29

very simple texts for laymen!

rather clear and concise press release

Thank you!