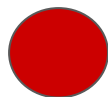


Workshop on Profiling L2 vocabulary and grammar

April 20-21, 2023. University of Gothenburg, Sweden

Organizers



- Elena Volodina, University of Gothenburg, Gothenburg, Sweden
- Therese Lindström Tiedemann, University of Helsinki, Helsinki, Finland
- David Alfter, Gothenburg Research Infrastructure in Digital Humanities (GRIDH), University of Gothenburg, Gothenburg, Sweden



Elena Volodina

University of Gothenburg, Sweden
NLP, L2 Swedish



David Alfter

University of Gothenburg, Sweden
ICALL, lexical complexity

**Therese Lindström
Tiedemann**

University of Helsinki, Finland
Scandinavian linguistics & general
Linguistics

welcome

shutterstock.com · 1721907820

- ≈120 registrations, incl. organizers and invited speakers
- ≈ 25 onsite
- Interest to the topic!

First things first

Hierarchy of needs



- Maslow's hierarchy of needs

First things first

Hierarchy of needs



- Eduroam / wifi
- Sockets, and charging
- Coffee breaks, lunch, dinner
- Bathrooms
- Name tags / stickers / signed “icons” on Zoom
- Workshop program

Day 1, 20 April, 2023	
Venue: University of Gothenburg, Humanisten campus, J330	
Lexical profiling - approaches, resources, applications	
9:00-9:10	Welcome and information [Elena Volodina]
9:10-9:55	Invited talk [Chair: Elena Volodina] <i>Jelena Kallas, Institute of the Estonian language (Estonia)</i> Aligning learners' dictionaries with the CEFR: the case of Estonian Vocabulary and Grammar Profiles
9:55-10:00	Short break
	Session 1. Chair: Elena Volodina
10:00-10:20	<i>Nina Hicks, University of Fribourg (Switzerland) (online)</i> Lexical features in adolescents' writing: Insights from the trilingual parallel corpus SWIKO
10:20-10:40	Coffee break
	Session 2. Chair: Julia Prentice
10:40-11:00	<i>Kris Heylen, Dutch Language Institute (The Netherlands)</i> <i>Ilan Kernerman, Lexicala by K dictionaries (Israel)</i> <i>Carole Tiberius, Dutch Language Institute (The Netherlands)</i> Linking CEFR-based learner profiles to lexicographic data
11:00-11:20	<i>Bernardo Stearns, University of Galway (Ireland)</i> Using Learner language models for lexical profile
11:20-11:40	<i>Mojca Stritar Kučuk, University of Ljubljana (Slovenia)</i> A cross-section of linguistic competence of South Slavic university students learning Slovene as L2
11:40-11:45	Short break
11:45-12:30	Organizer presentation and demo [Chair: Jelena Kallas] <i>David Alfter, University of Gothenburg (Sweden)</i> Swedish Lexical profile
12:30-14:00	Lunch University canteen

Jelena Kallas, Eesti Keele Instituut, Estonia

Title: *Aligning learners' dictionaries with the CEFR: Estonian Vocabulary and Grammar Profiles*

Abstract: In the talk, we will introduce [the Estonian Vocabulary Profile](#) and [the Estonian Grammar Profile](#), which are designed to support the CEFR illustrative descriptors scales of linguistic competence with language-specific descriptions. We will focus on the methodology and corpora that were used for the development, trial and validation of [the Estonian Grammar Profile](#). Currently, the profile provides descriptions of grammar competence on the morphology, derivation, phrase and sentence levels, from the pre-A1 level up to the B2 level for young learners, and from A1 to C1 for adult learners. All descriptions are equipped with example sentences compiled either by experts or taken from the coursebook and learner's corpora.

In addition, we will address the issues related to our attempt to combine this resource with the Estonian learners' dictionary [Sõnaveeb for Learners](#). The dictionary is compiled in the Dictionary Writing System [Ekilex](#), whose long-term goal is to have a single data source that provides consistent and comprehensive information about Estonian, including CEFR labels. We will report on the work in progress from the point of view of data modelling. Given a construction-based and usage-based understanding of L2 acquisition, we assume that linguistic knowledge at a particular proficiency level is not best described as a set of words and a set of grammatical structures, as is the current practice, but rather as a set of combinations of particular word meanings and forms with particular schematic constructions. This means that the lexicographic resource must include descriptions of grammatical constructions, and that the language proficiency level should be attributed not to lemmas and constructions, but to particular word meanings in particular forms and in particular constructions.

Bio: Jelena Kallas is a Senior Computational Lexicographer – Project Manager at Eesti Keele Instituut ([Institute of the Estonian Language](#), Tallinn, Estonia). She has been involved in various lexicographic projects, including monolingual and bilingual dictionaries, and SLA projects. She is leading the **Estonian L2 profile project called "Teacher Toolkit"**. She is a holder of national team grant "Expanding the scope of a multi-purpose lexicographic resource to grammar and L2 competence (2023-2027)" where she and the other project members will work on the development of a theoretical and methodological framework for the description of grammatical constructions and L2 linguistic competence in an lexicographic resource, relying on a usage-based and construction-based approach to linguistic theory, language acquisition and lexicography.

Grammatical profiling - approaches, resources, applications

14:00-14:45	Invited talk [Chair: Therese Lindström Tiedemann] <i>Geraldine Mark, Cardiff University (Wales)</i> Building on insights from the English Grammar Profile: From <i>really good</i> to <i>painfully obvious</i>
14:45-14:50	Short break
	Session 3. Chair: Aleksandrs Berdicevskis
14:50-15:10	<i>Annekatriin Kaivapalu, University of Helsinki (Finland)</i> Profiling learner Finnish and Estonian: interaction of frequency and accuracy as an indicator of language skills
15:10-15:30	<i>David Alfter, University of Gothenburg (Sweden)</i> French Verb profile
15:30-16:00	Coffee break
	Session 4. Chair: Aleksandrs Berdicevskis
16:00-16:20	<i>Nicolas Ballier, Université Paris Cité (France)</i> Grammatical profiling with UD annotation (WIP)
16:20-16:40	<i>Ekaterina Vlasova, University of Helsinki (Finland)</i> Prepositions in L2 Russian
16:40-16:45	Short break
16:45-17:30	Organizer presentation and demo [Chair: Geraldine Mark] <i>Therese Lindström Tiedemann, University of Helsinki (Finland)</i> Swedish Grammatical Profile
17:30-17:40	Rounding off Instructions how to get to dinner place [Elena Volodina]
18:00 -...	Dinner at <i>Berzelius Bar & Matsal</i> . Address: Södra Vägen 20, 412 54 Göteborg

Geraldine Mark, Cardiff University, Wales

Title: Building on insights from the English Grammar Profile: From *really good* to *painfully obvious*

Abstract: The English Grammar Profile (EGP) Project was a four-year quasi-longitudinal study investigating learner grammar from the Cambridge Learner Corpus (CLC). The main output of the research is the EGP, a free educational online database, which provides a profile of over 1,200 corpus-based grammar competency statements about learner grammar use across the six CEFR levels. In the first part of this talk I'll describe the methodology that we developed to build the EGP, discuss the key insights from the study and show how the investigation has enhanced our understanding of the developmental nature of grammar acquisition and use. I'll then look at further ways to explore the data taking a usage-based (UB) approach. UB studies have shown that language users are sensitive to the statistics of repeated patterns in language and that we figure out 'structural regularities' in language as we subconsciously tune into mappings of form and meaning (Ellis et al. 2016). Using this large scale proficiency-levelled data I will look at how we can use corpus tools to investigate if and how structural regularities develop in L2 English and how this might offer further insight into learner language development.

Ellis, N. C., Römer, U., & O'Donnell, M. B. (2016). Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of construction grammar. Oxford: Wiley.

Bio: Geraldine Mark is an applied corpus linguist with experience in research, teaching and learning, publishing and materials design. Her principal interests are in corpus linguistics and its diverse applications, particularly in relation to language development and usage in L1 and L2, data-driven learning, and multi-modal interaction. She is a Visiting Lecturer at the University of Malta, and advises on the FoRCE project, building and analysing a corpus of Maltese English. She is a Research Associate on a multi-modal project, IVO (www.ivohub.com), funded by the UKRI Arts and Humanities Research Council and the Irish Research Council, examining online workplace multi-modal interaction. She is co-author of English Grammar Today (2011, Cambridge University Press, with Ronald Carter, Michael McCarthy and Anne O'Keeffe) and co-principal researcher (with Anne O'Keeffe) of the English Grammar Profile, an online resource profiling L2 grammar development.

Day 2, 21 April, 2023

Venue: University of Gothenburg, Humanisten campus, J330

Word families, Morphological families and Morphological profiling - approaches, resources, applications

9:00-9:45	Invited talk [Chair: Therese Lindström Tiedemann] <i>Gabriele Pallotti, University of Modena and Reggio Emilia (Italy)</i> Profiling complexity: methodological issues and applications to L2 morphology
9:45-9:50	Short break
/td>	Session 5. Chair: Kris Heylen
9:50-10:10	<i>Maria Belén Díez-Bedmar, University of Jaén (Spain)</i> The FineDesc learner corpus: Making the CEFR/CV more user-friendly: fine-tuning descriptors with Learner Corpus Research results
10:10-10:30	<i>Francesca La Russa, Sapienza Università di Roma (Italy)</i> <i>Maria Roccaforte, Sapienza Università di Roma (Italy)</i> Using a learner corpus to design a phraseological syllabus of Italian collocations
10:30-10:50	Coffee break
/td>	Session 6. Chair: Kris Heylen
10:50-11:10	<i>Christina Lindqvist, University of Gothenburg (Sweden)</i> <i>Mårten Ramnäs, University of Gothenburg (Sweden)</i> A Digital Dictionary of Romance Word Families
11:10-11:30	<i>Isidora Glišić, University of Iceland (Iceland)</i> From corpus to profiles: Icelandic L2 corpus
11:30-11:40	Short break
11:40-12:25	Organizer presentation and demo [Chair: Christina Lindqvist] <i>Elena Volodina, University of Gothenburg (Sweden)</i> <i>Morphological profile</i>
12:25-13:40	Lunch University canteen

Gabriele Pallotti, University of Modena and Reggio Emilia, Italy

Title: *Profiling complexity: methodological issues and applications to L2 morphology*

Abstract: In this talk I will first discuss how linguistic complexity should be theoretically defined and practically operationalized, in a wider context of interlanguage analysis and linguistic profiling. In particular, I will argue that it needs to be kept apart from other constructs such as processing difficulty or developmental timing. Then, I will present an approach to empirically measuring morphological complexity, its conceptual and methodological challenges and how they were addressed in the development of an online morphological complexity analyzer.

Bio: Gabriele Pallotti is a professor of Language teaching methodology at the University of Modena and Reggio Emilia. His research focusses on interlanguage development, linguistic complexity, morphology, L2 interaction, methodology and epistemology in applied linguistics. He coordinates the project Observing interlanguage and is the associate editor of the Eurosla Studies Series (Language Science Press). He has led several national and international projects on language learning and teaching, funded by the National Ministry of Education and the European Union.

Linguistic complexity – tying it all together

13:40-14:25	Invited talk [Chair: Gabriele Pallotti] <i>Aleksandrs Berdicevskis, University of Gothenburg (Sweden)</i> We need to know more about relative complexity and learnability
14:25-14:45	Coffee break
14:45-15:30	Discussion session (University canteen for onsite and Zoom rooms for online participants) Discussion leaders and instructions: coming soon
15:30-15:45	Move back to the main room and Zoom
15:45-16:00	Sum up from small groups [Chair: Therese Lindström Tiedemann]
16:00-16:30	Rounding off: prospects, next steps [Elena Volodina]

[Aleksandrs Berdicevskis](#), University of Gothenburg, Sweden

Title: *We need to know more about relative complexity and learnability*

Abstract: For the last two decades, language typology and related fields have witnessed a hot debate about language complexity. Several influential theories have emerged that claim that languages are not equally complex, and that the distribution of complexity depends on social factors, such as number of speakers, degree of language contact and number of non-native speakers. I will briefly review some recent evidence in favour and against those theories and argue that whether the theories are correct or not, they make interesting non-trivial hypotheses about mechanisms of language learning and language change. I will then make my main point, which is that these hypotheses cannot be properly addressed without a deep understanding of second language acquisition, most crucially, the concepts of relative complexity ("what is difficult for whom") and learnability. The talk will mostly focus on morphological complexity.

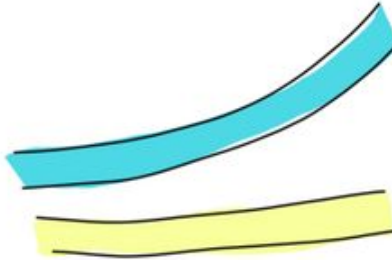
Bio: Aleksandrs Berdicevskis is a researcher in computational linguistics at Språkbanken Text, University of Gothenburg. His research focuses on explanatory approaches to language change and language typology, with a particular attention to language complexity. He currently leads the project [Cassandra: Explaining and predicting short-term language change in Modern Swedish](#).

Why do we bother to organize this event?

1. to establish some common practices, among others, how to **use L2 profiles practical** and how to **apply them to different research questions**
2. to showcase **scenarios** for using them in research & teaching and outline **benefits** of their use
3. to exchange **experiences and lessons learnt** between projects that have created such resources, reflecting on the ways to improve user-friendliness and identifying missing functionalities
4. to inspire other languages to work on similar resources stimulating the creation of a new family of resources. Projects of that kind can hardly be "re-run" to fix errors that are discovered post-factum, and therefore planning such projects "right from the start" is a prerequisite. With the experiences from the L2 profile projects for Swedish, Estonian and English, we intend to offer our expertise (à la "**advisory role**") and help researchers in other countries set up their projects.

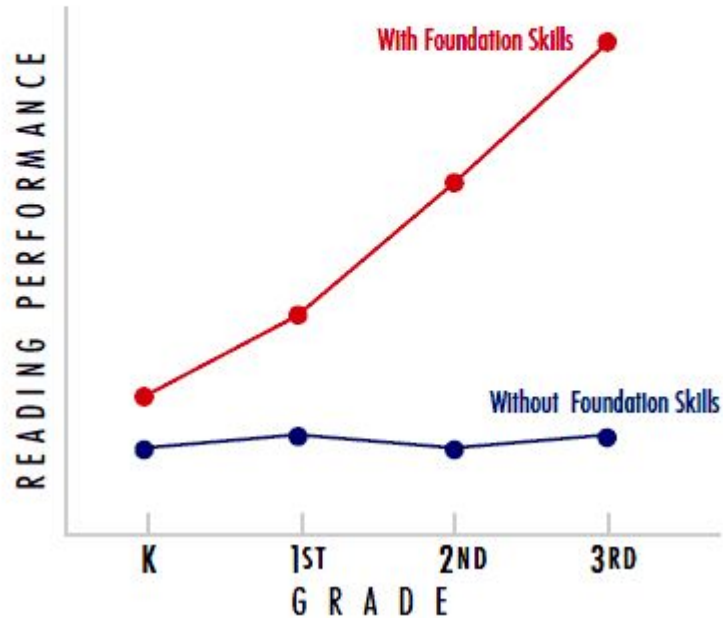
MATTHEW EFFECT

THE OLD ADAGE "FOR TO HIM WHO HAS,
MILL MORE BE GIVEN." THOSE WITH
EXISTING STATUS, PRIVILEGE, WEALTH, ETC.
STAND TO BENEFIT EVEN MORE FROM IT.
IN OTHER WORDS, "THE RICH GET RICHER
AND THE POOR GET POORER."



Matthew Effect in Reading

The Rich Get Richer and the Poor Get Poorer!



True of many disciplines:

- Education (e.g. reading)
- Wealth
- Project funding
- English vs other languages
- ...



Should We Ban English NLP for a Year?

Anders Søgaard

Abstract

Around two thirds of NLP research at top venues is devoted exclusively to developing technology for speakers of English, most speech data comes from young urban speakers, and most texts used to train language models come from male writers. These biases feed into consumer technologies to widen existing inequality gaps, not only within, but also across, societies. Many have argued that it is almost impossible to mitigate inequality amplification. I argue that, on the contrary, it is quite simple to do so, and that counter-measures would have little-to-no negative impact, except for, perhaps, in the very short term.

Over to the first invited talk

Jelena Kallas, Institute of the Estonian Language, Tallinn

Aligning learners' dictionaries with the CEFR:

Estonian Vocabulary and Grammar Profile



...over to the results of the discussion, day 2

Ways forward

- See results of the brainstorming on a [jamboard](#), page 2
- Publication (special issue on “Best practice guide” / “How-to construct new profiles”)?
- Recurrent event (e.g. every 2nd year at LCR conference) ?
- Add our resources/projects to [CoE RLDS](#) ?
- ...