Exploring Finno-Ugric linguistics through solving IT problems

Jezikovne tehnologije in digitalna humanistika

20 September 2018, Ljubljana, Slovenia

Tobias Weber & Jeremy Bradley

Ludwig Maximilian University of Munich & University of Vienna

Weber.Tobias@campus.lmu.de & jeremy.moss.Bradley@univie.ac.at

URALIC LANGUAGES

F Finno-Ugric

FO Baltic-Finnic
- FO1 Finnish
- FO2 Karelian
- FO3 Veps
- FO4 Ingrian
- FO5 Estonian
- FO6 Votic
- FO7 Livonian

FS Sami languages
- FS1 Western Sami
- FS2 Central Sami
- FS3 Eastern Sami

FU Ugric
- FU1 Hungarian
- FU2 Mansi
- FU3 Khanty

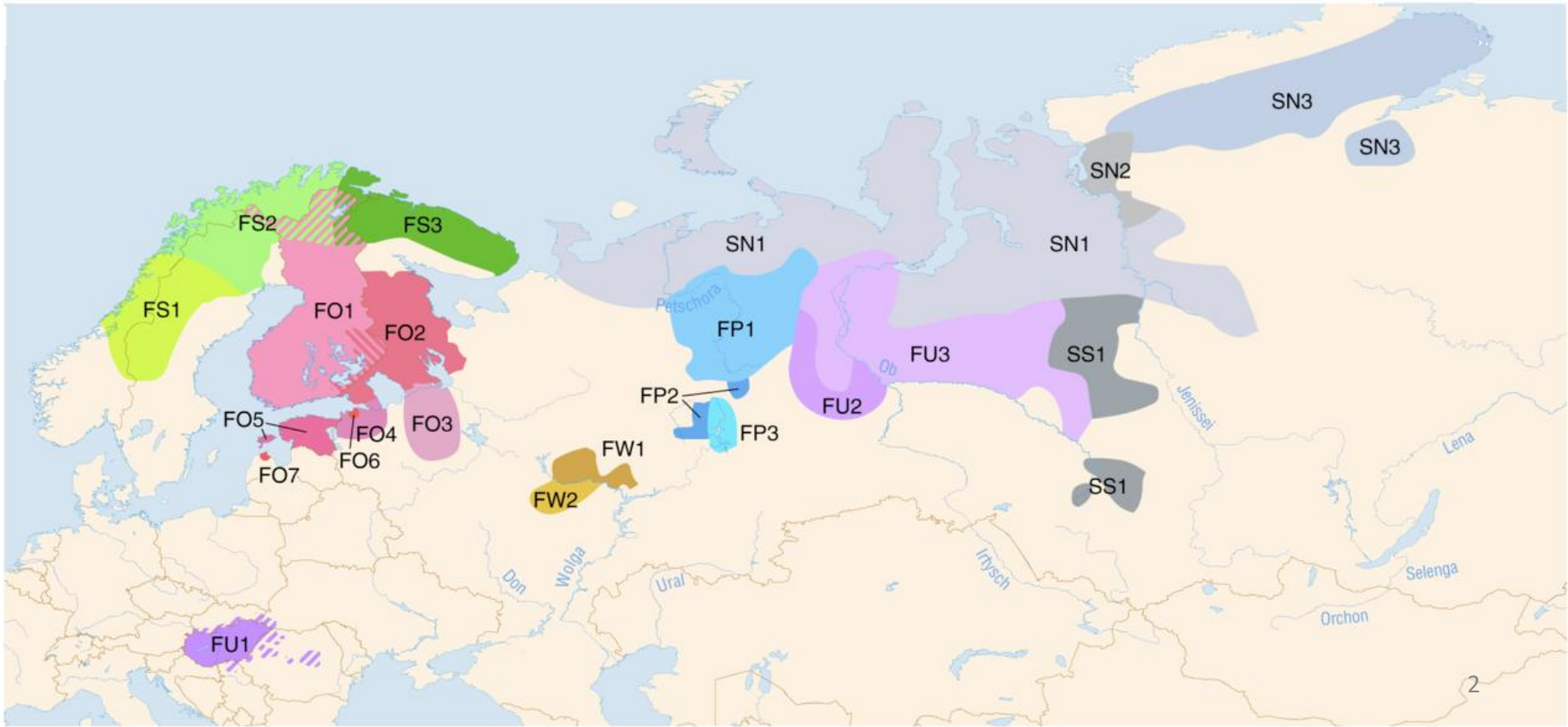
FP Finno-Permic
- FP1 Komi-Zyrian
- FP2 Komi-Permyak
- FP3 Udmurt

FW Finno-Volgaic
- FW1 Mari
- FW2 Mordvinic

S Samoyedic

SN Northern Samoyedic
- SN1 Nenets
- SN2 Enets
- SN3 Nganasan

SS Southern Samoyedic
- SS1 Selkup



Finno-Ugric (Uralic) studies

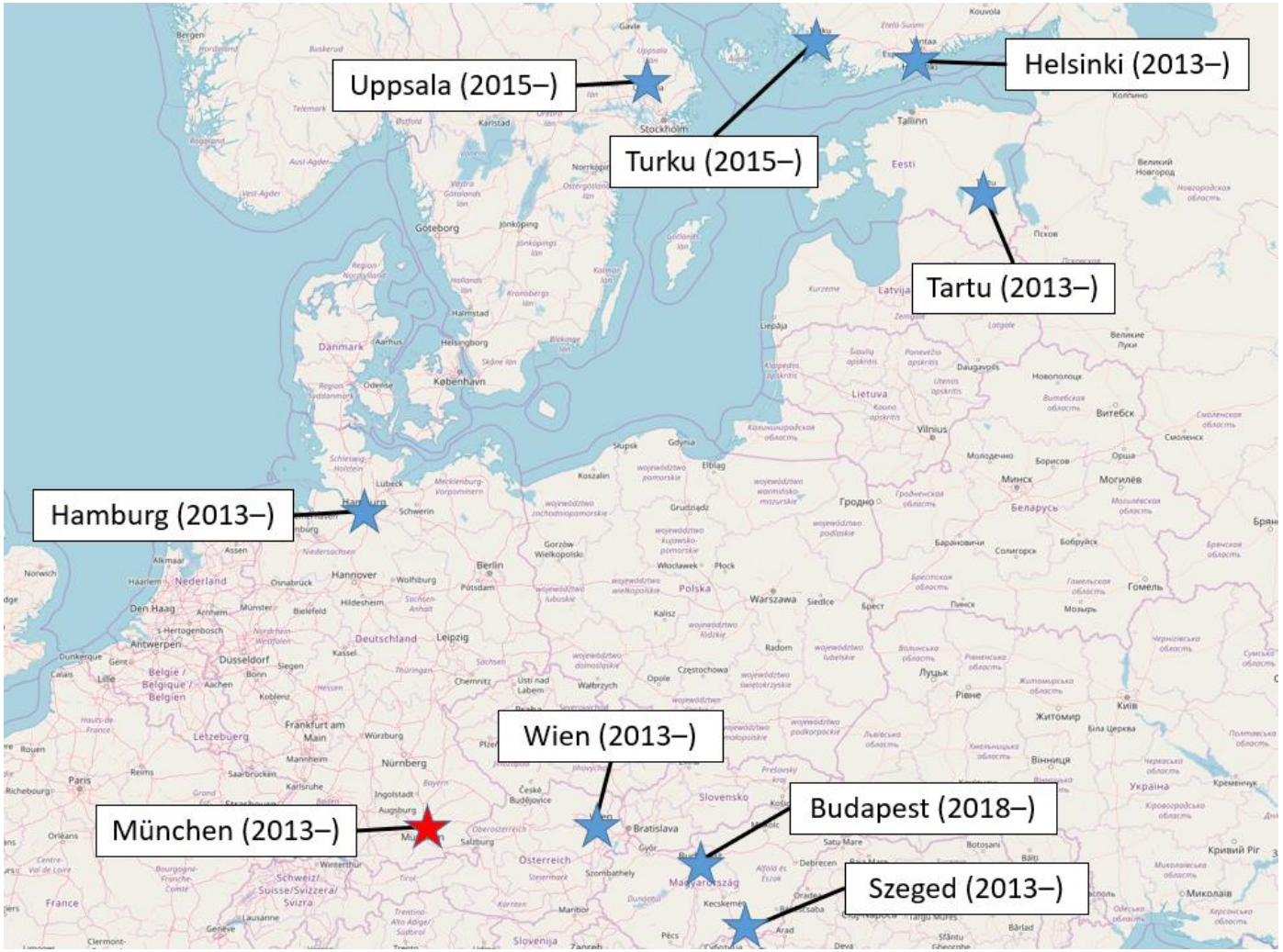
- In countries/regions with resident FU people, the focus of academic programmes is cultural/philological, otherwise (as in most of Western Europe) rather linguistic

- Students at LMU follow a linguistic curriculum with language specific modules (language, culture, literature): Finnish and Hungarian in the BA, Estonian and minor languages in the MA

- Strategic partnerships between institutions allow for pooling of competences and collaboration in the development of curricula

- The latest partnership COPIUS includes 9 universities and will shape the discipline for the next 3 years

INFUSE/COPIUS



Our target audience

- Students in Finno-Ugric studies usually learn one or more of the languages
- Their research spans various topics: sociolinguistics, literature, pragmatics, folkloristics, or syntax (sample of BA & MA theses)
- Generally, they opt for a humanity on free choice because they do not want a tech-heavy curriculum
- Yet, there are many ways of using computers efficiently in their research, e.g. handling of corpus data, transcription tools, automated analysis, databases

Problems encountered

ńèβχa̭̮”nə̹ ŋàβχī̮̭” mā̹l̀̓eŋka̭̮”nə̹ jā̹n̹̮t̹́ś̹ērχa̭̮”nə̹ ŋōʙ” ŋārk̆kə̹ tā̹d̹'eb̬e ta̹̭n̆ńeββī̬
Nenets (Lehtisalo 1947: 30)

Stem alternation in Finnish (“gradation”: “**strong**” –**kk**– before open syllables, “weak” –k– before closed syllables)

*ku**kk**a*	*kuka-n*	*ku**kk**a-a*	*ku**kk**a-an*	*kuka-ssa*
flower	flower-GEN	flower-PART	flower-ILL	flower-INE
strong	weak	**strong**	**strong**	weak

Hungarian superessive case -*(V)n* (-*n* ~ -*on* ~ -*ön* ~ -*en*)

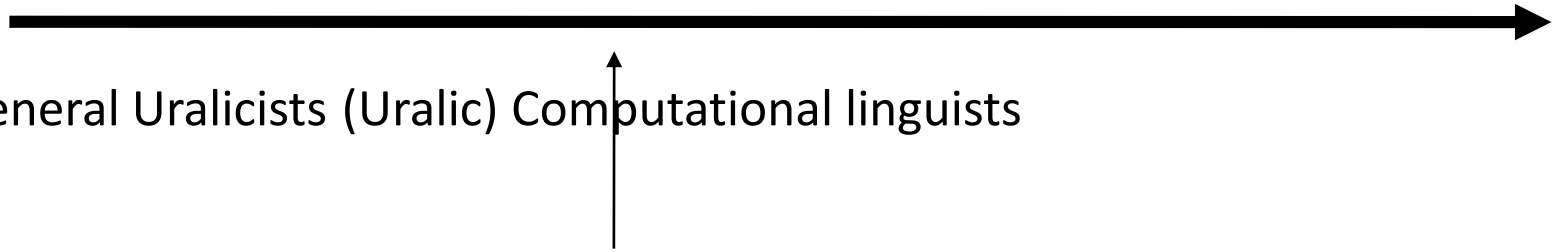
hajó-n	*ház-on*	*csütörtök-ön*	*szék-en*
ship-SUP	house-SUP	Thursday-SUP	chair-SUP

Finnish inessive case -*ssA* (-*ssa* ~ -*ssä*)

talo-ssa	*metsä-ssä*
house-INE	forest-INE

Conveying Uralic digital competence

Media competence



general Uralicists (Uralic) Computational linguists

bridging between Uralic linguistics and tech-savvy researchers (within or outside the discipline)

- Exploring new methodology for our research
- Contributing our knowledge to larger projects
- Enabling communication between both ends of the spectrum (CL learn how to work with our languages, FU learn to understand methods and research of CL)

Pilot project

- Explicit teaching part: basics of computing, encodings, types of programming languages and tasks
- Hands-on approach with example codes – learning basics of program design and restrictions of applications
- Discussing current topics in CL (machine translation, neural networks, big data) on a simple level
- Each example is related to our discipline and points at ways to handle issues, e.g. different orthographies or complex morphology
- Textbook for basic IT skills on the basis of Uralic linguistics

Outlook

- Ideally, including IT skills implicitly in all courses of the curriculum (see article for a rundown of options)
- Creating awareness of computational methods for handling Uralic languages
- Creating awareness of the issues in handling Uralic language data (alphabetical order, special characters, small corpora languages)
- Bridging the digital divide between pen-and-paper linguistics and computational linguistics
- Establishing courses through the strategic partnership for all students, thereby making these skills an integral part of the undergraduate training in Uralic linguistics

INFUSE/COPIUS

Hvala vam za pozornost

Weber.Tobias@campus.lmu.de

jeremy.moss.bradley@univie.ac.at