



The comprehensive Slovenian- Hungarian dictionary: bilingual lexicography meets monolingual lexicography

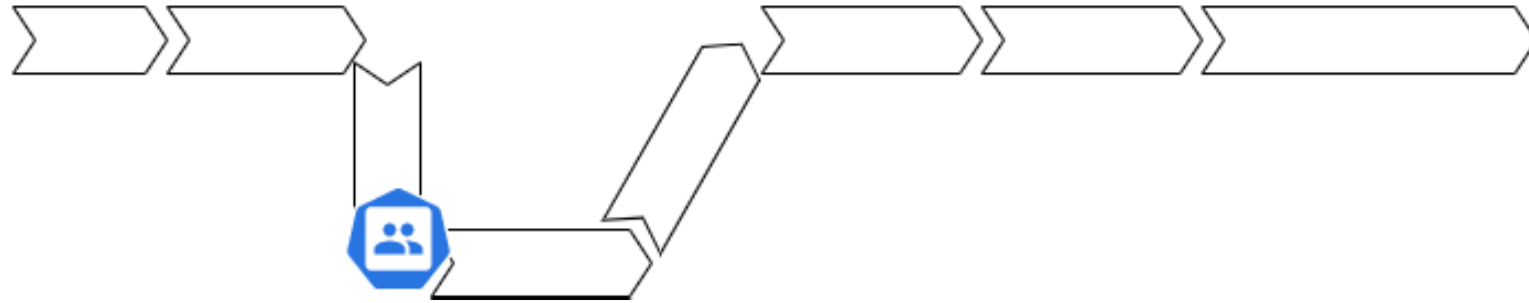
Iztok Kosem

iztok.kosem@cjvt.si

Centre for Language Resources and Technologies
Faculty of Arts, University of Ljubljana

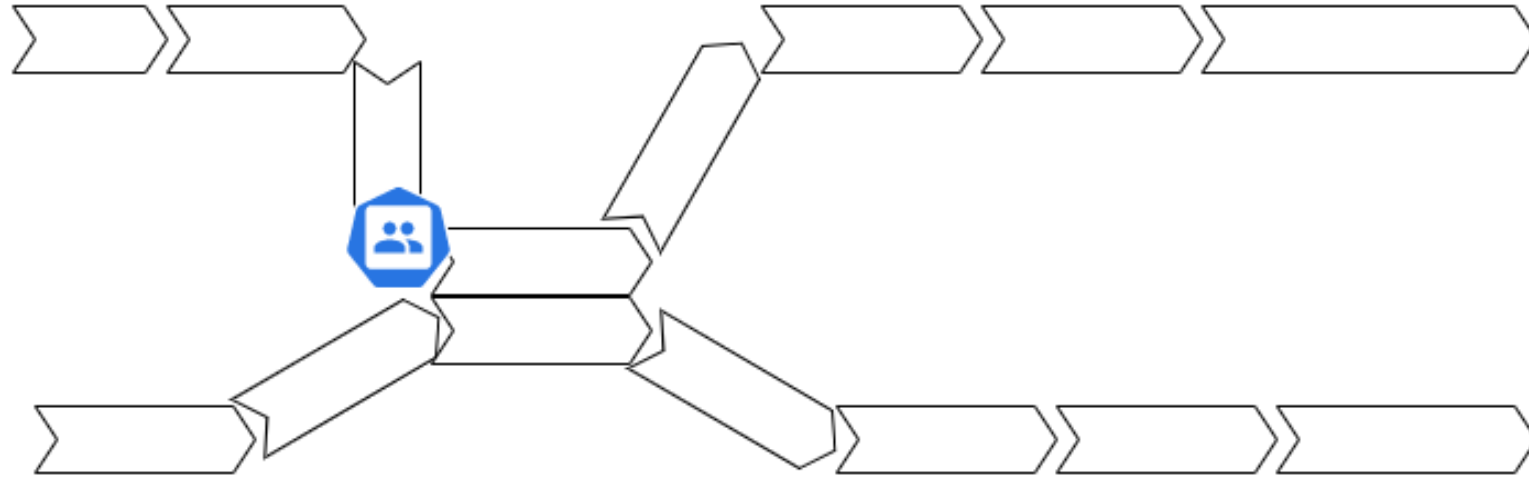
A

B



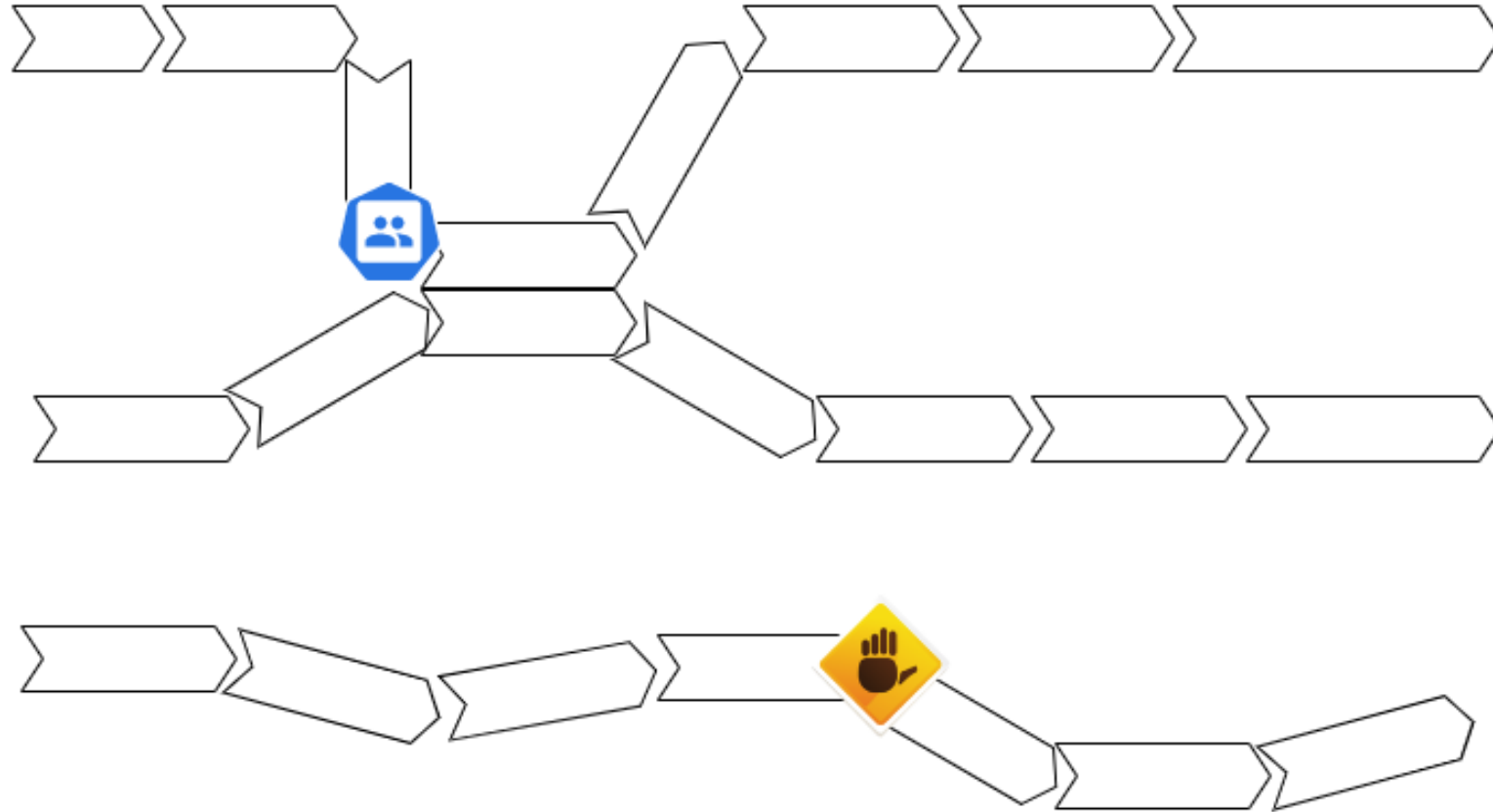
A

B



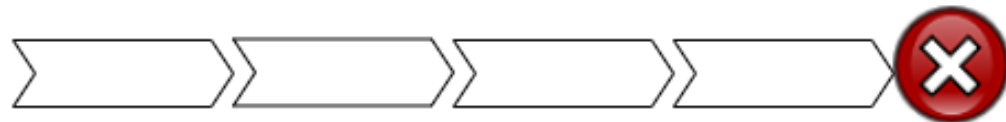
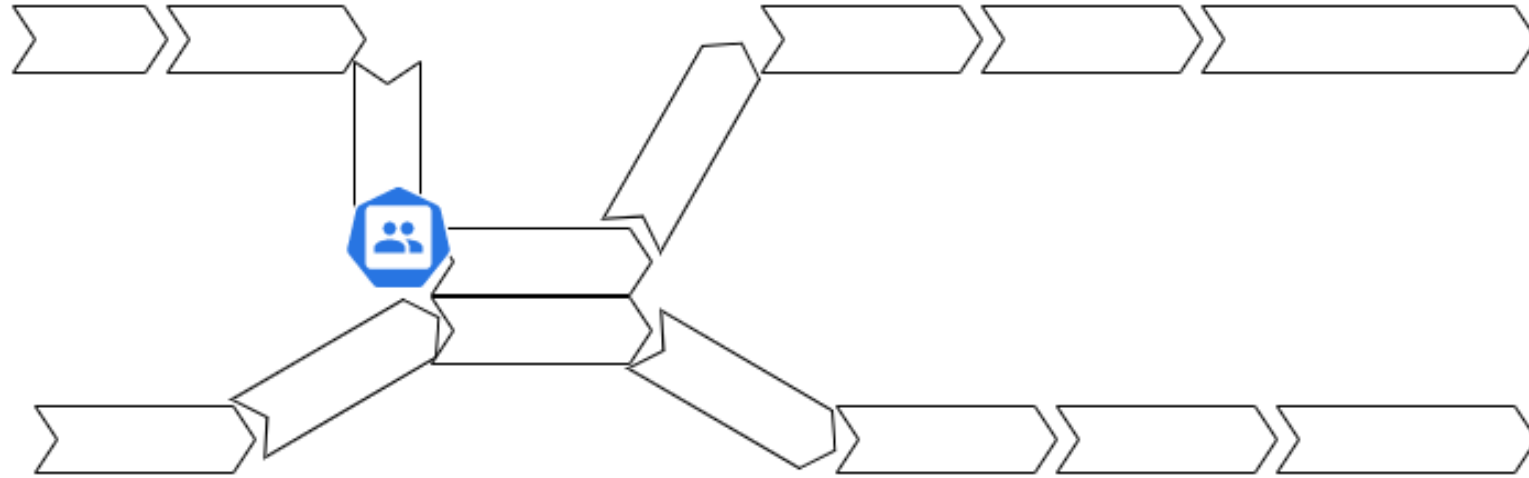
A

B



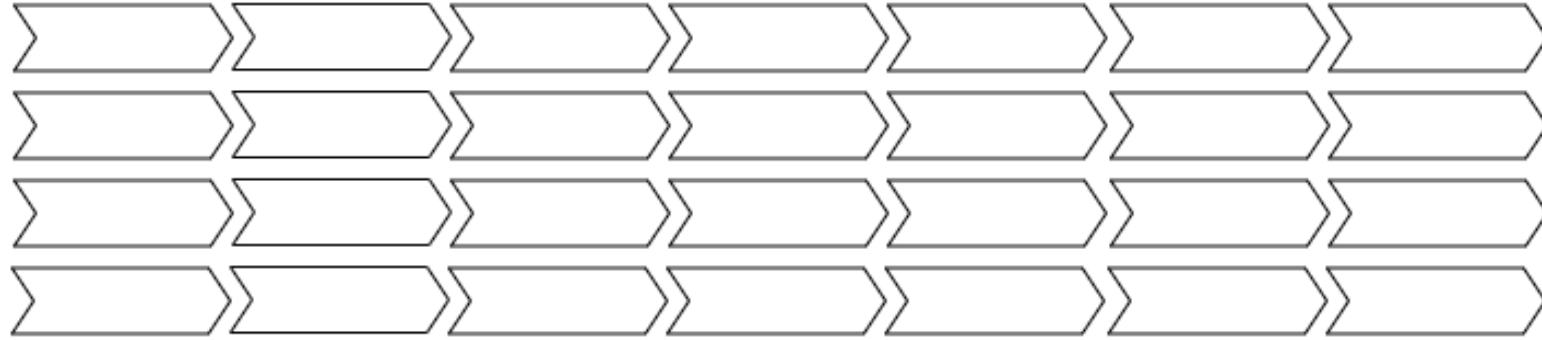
A

B



A

B



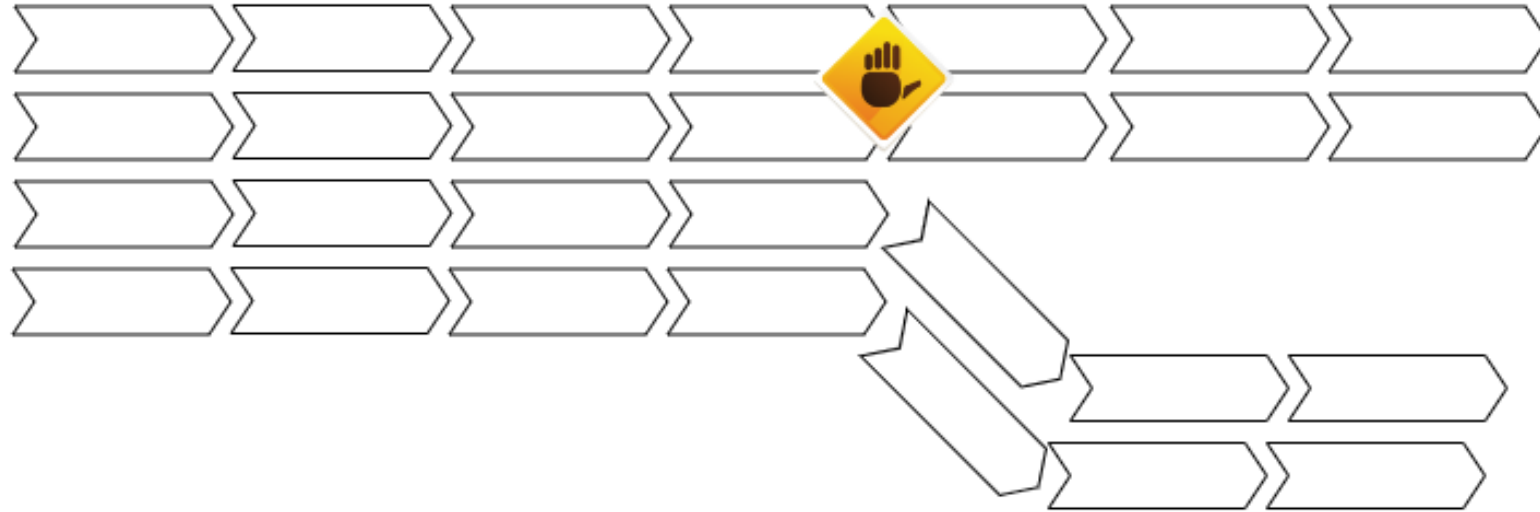
A

B



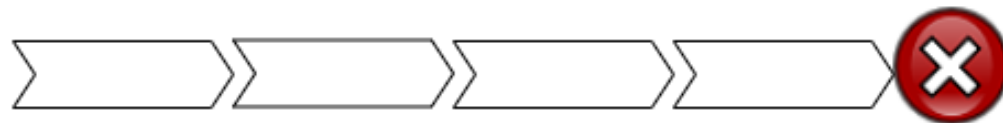
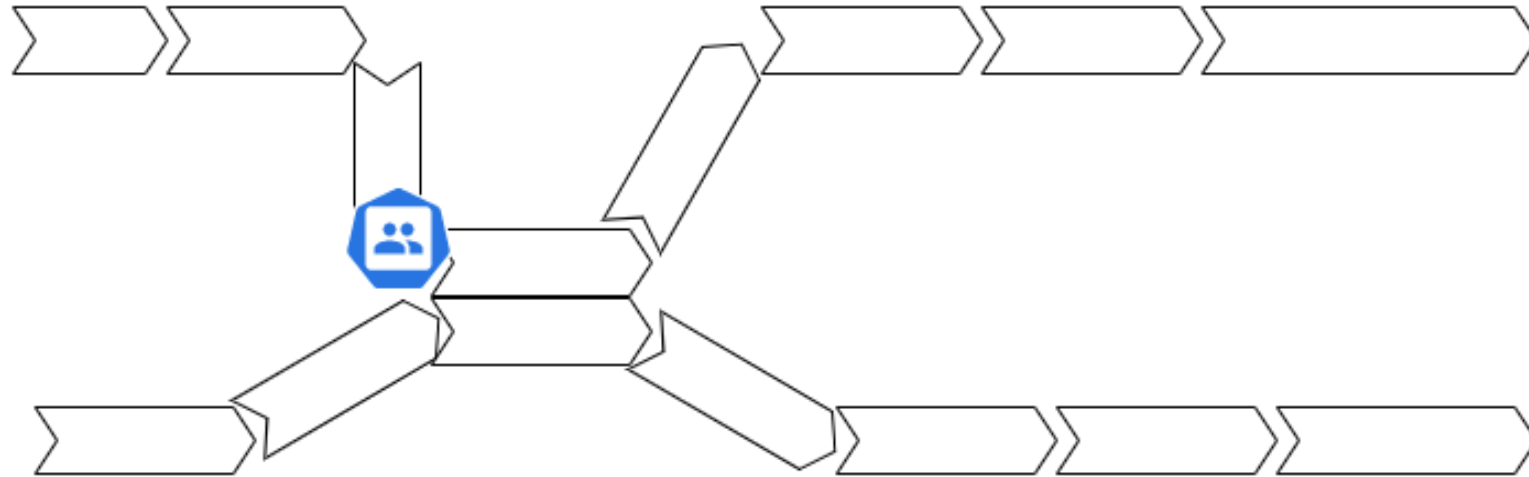
A

B



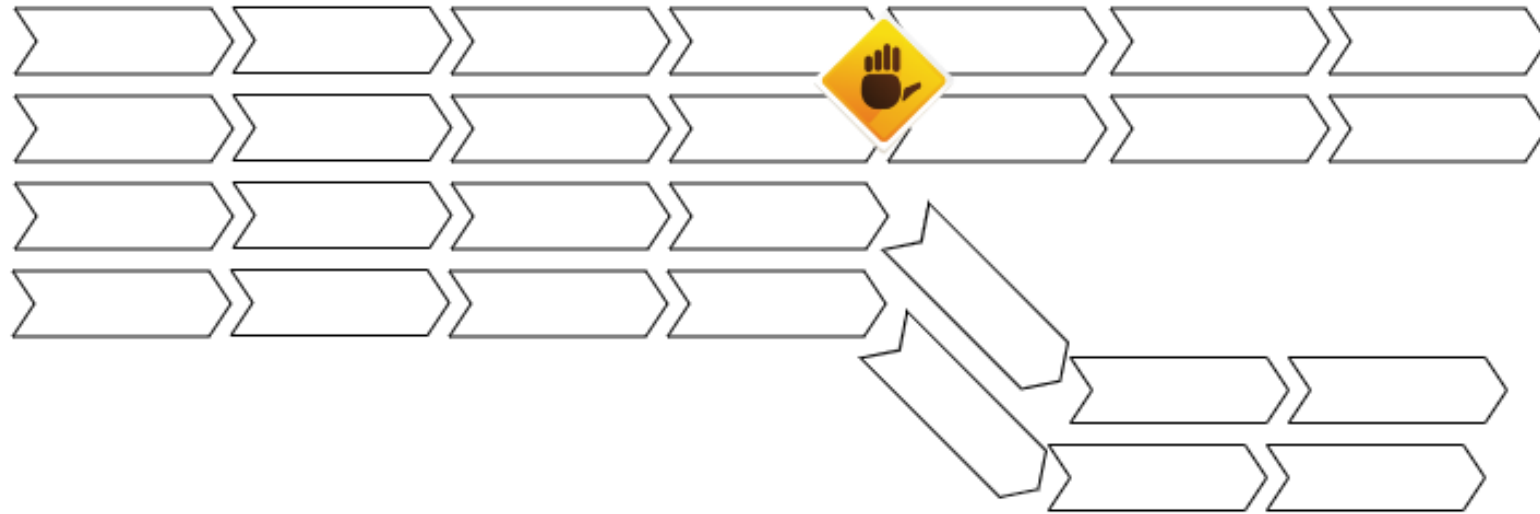
A

B

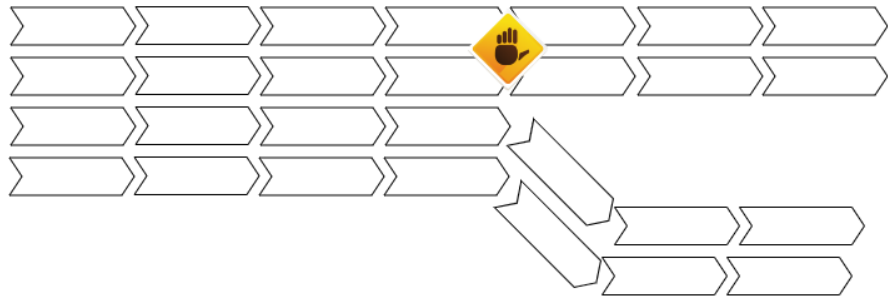


A

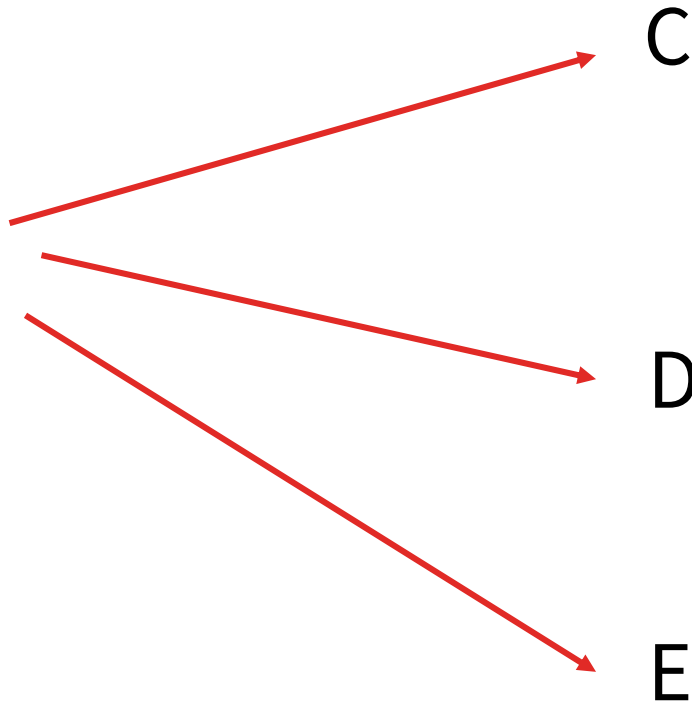
B



A



B



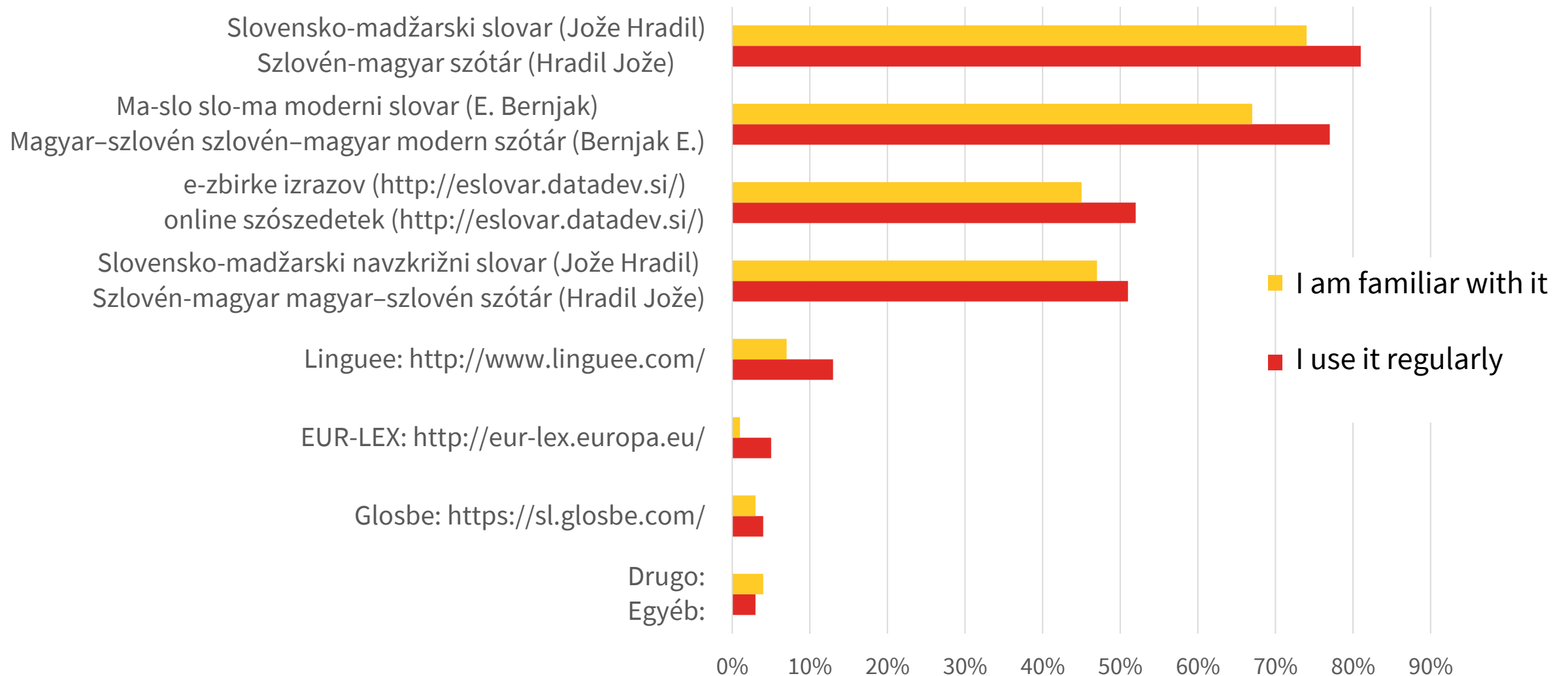
NO DICTIONARY IS AN ISLAND



Dictionary situation in Slovenia

- Slovenian-Hungarian situation:
 - existing dictionaries very outdated
 - not corpus-based
 - so far mainly compiled by individuals
 - few available in electronic form (desktop application only)
 - all under copyright
 - but still regularly used

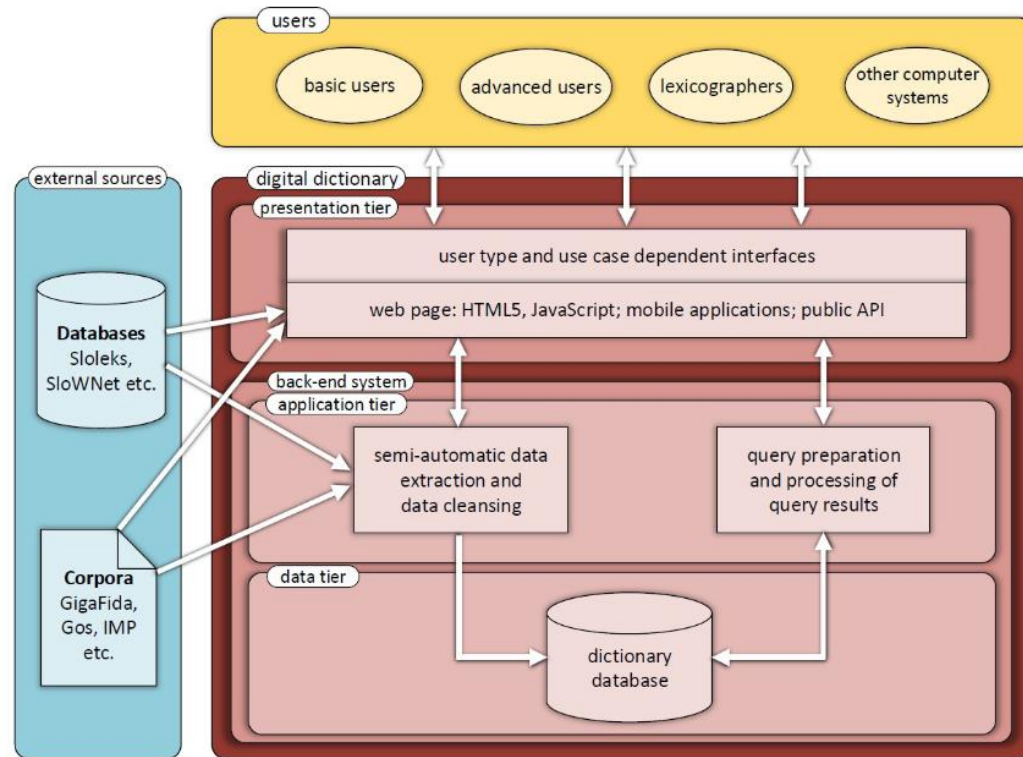
Slovenian-Hungarian language resources (N=142)



Dictionary situation in Slovenia

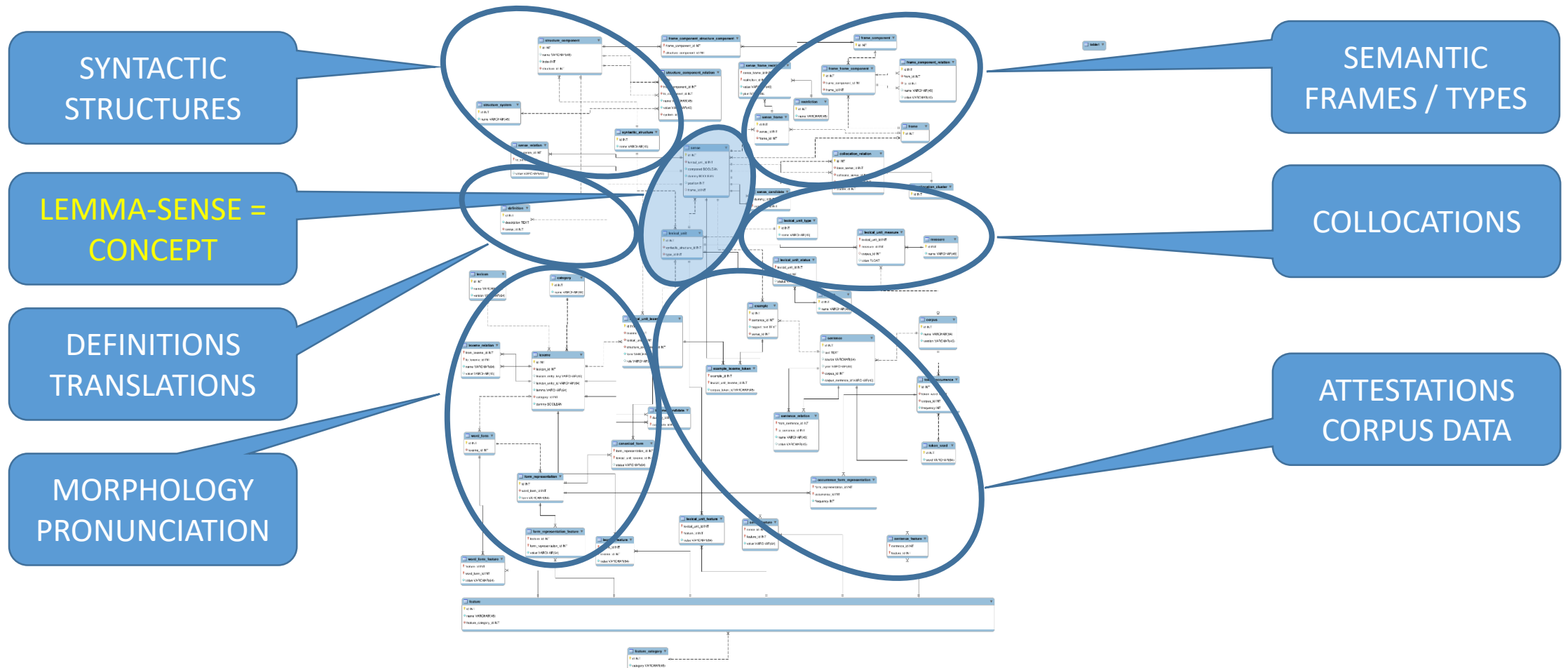
- Bilingual dictionary situation
 - Most dictionaries outdated
 - Dictionaries under copyright (some are being bought out as part of the Development of the Slovene Language in Digital Environment project)
- Monolingual dictionary situation
 - More up-to-date (*)
 - Corpus-based
 - Good variety of types (general, specialized etc.)
 - Some available under open licence (but not the main ones)

Data model (for all types of lexicographic data)

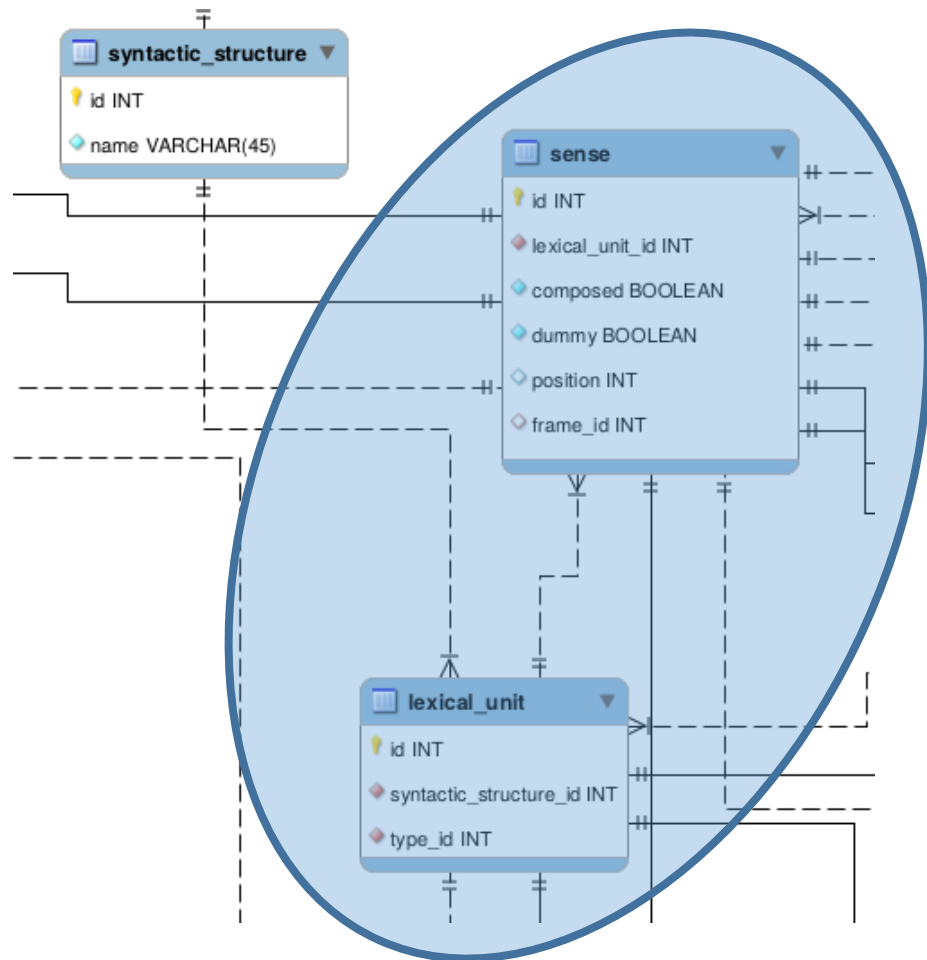


- Dictionary of Modern Slovene: Problems and Solutions (e-monograph, 2017)
- Bojan Klemenc, Marko Robnik-Šikonja, Luka Fürst, Ciril Bohak and Simon Krek: Technological Design of a State-of-the-art Digital Dictionary

Digital Dictionary Database (for Slovene)



Lexical unit + Sense = Concept (stable id)

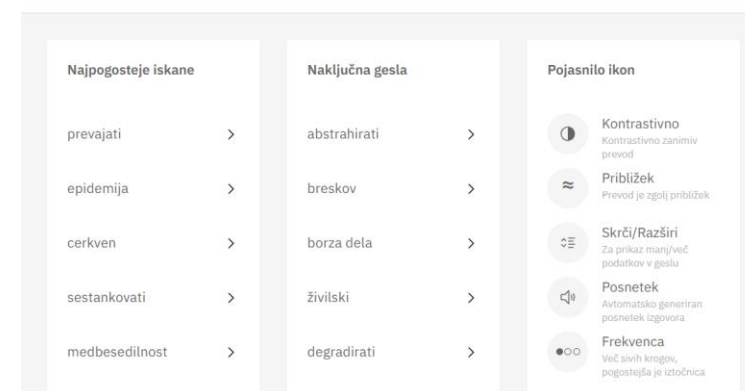


- Lexicographic Data as a Service (LDaaS)
- Available through API service
- Open Access (CC BY-SA)
- In CLARIN.SI repository

- Development of Slovene Language in Digital Environment (2020-2022)

Comprehensive Slovenian-Hungarian Dictionary

- Funded by Slovenian Research Agency (SRA) since 2018
- Based on a concept (the project also funded by SRA)
- Entries compiled completely from scratch
- version 1.0 at <https://viri.cjvt.si/slovensko-madzarski/eng/>:
 - 10,946 entries
 - 15,265 collocations
 - 2,416 examples
 - 33,298 translations
- Date of publication: 20. 10. 2021
- Database available in CLARIN.SI repository under 4.0 CC-BY-SA license (<http://hdl.handle.net/11356/1453>)



Team

- Over 20 colleagues
 - lexicographers
 - lexicographers-translators
 - proofreaders
 - terminologists
 - linguists
 - computational linguists, IT staff
- International advisors:
 - dr. Jelena Kallas (Institute for the Estonian Language)
 - dr. Katalin P. Márkus (Károli Gáspár University of the Reformed Church, Budapest)
 - dr. Attila Mártonfi
 - Michael Rundell (Lexicography Masterclass)
 - dr. Carole Tiberius (Dutch Language Institute)
 - dr. Tamás Váradi (Research Institute for Linguistics, Hungarian Academy of Sciences)

cerkven pridevnik

Variante Homonomije Povezano Oznake Opombe

Show examples

Show clusters

4-L **cerkven** pridevnik

Variante Homonomije Povezano Oznake Opombe

Show examples

Show clusters

SKE

collocations ▼ simple ▼

examples

collocations

Vstavi v: ▼

OK Prekliči

³cerkvna ladja^{506 6.44} templomhajó^E +

³cerkvna klop templompad +

New ID Save Cancel Clone Delete

SKE

examples ▼ simple ▼

Isci

- Ob 15-letnici delovanja bodo posebno priznanje podelili tudi **cerkvenemu** pevskeemu zboru Ignacij Hladnik.
- Po navadi na tako velike praznike poje na koru **cerkveni** zbor.
- Petje je lepo oblikoval mladinski **cerkveni** zbor iz Škocjana.
- In **cerkveni** dostojanstveniki si lahko mirno predstavljajo, da so prav oni voditelji vsega tega.
- Njeni predniki so bili črnski najemniki na polju, oče pa je bil **cerkveni** dostojanstvenik.
- To je ena od najglobljih misli, ki jih je **cerkveni** zbor povedal o Cerkvi.
- Predstavili so se mladi talenti iz Šentjerneja (na sliki otroški **cerkveni** pevski zbor) in Kamniški koledniki.
- Umetnostna **cerkvna** glasba srednjega veka je izključno vokalna.
- Ob tej priložnosti bo blagoslov novega parkirišča, nato pa srečanje na **cerkvenem** dvorišču.
- Za zraven nje stoječo **cerkvno** hišico so iztržili 300 goldinarjev.
- Pri tem se smiselno uporabljajo vsa določila civilne in tudi **cerkvene** poroke.
- Sedanje **cerkvno** vodstvo je zaradi papeževe lastne izkušnje temu še posebej naklonjeno.
- Neposredno po koncu vojne je oblast dovolila celo odpiranje **cerkvenih** šol.
- Ta logika deluje praktično pri vseh projektih, tudi **cerkveni** niso nobena izjema.
- Nemogoče se zdi, da bi spregledal zasuk **cerkvenih** krogov k novim zavezništvom.

²⁶cerkvna zgodovina^{355 4.69} egyháztörténet +

Dictionary content

- Novelties:
 - Slovene-centred sense division
 - Multiword units as entries
 - New system of labels
 - Speech synthesis for all parts of the dictionary (Slovenian and Hungarian)
 - Each sense has an indicator
 - Each sense attributed (at least one) semantic type

šah (samostalnik)



1 AKTIVNOST-DEJAVNOST-negospodarska-igra (namizna igra) sakk

igrati šah sakkozik
turnir v šahu sakktorna
prvak v šahu sakkbajnok
ženski šah női sakk
moški šah férfi sakk
aktivni šah
turnirski šah

All the content from the database available, but not all is translated

3 AKTIVNOST-DEJANJE-stanje (položaj v igri) sakk

napovedati šah sakkot ad

- Tudi če s skakačem **napoveste šah** , zgubite in ste po devetih do desetih potezah poraženi.
- Figura, ki prestreže **šah** na lastnega kralja, lahko sama s šahom napade nasprotnikovega kralja. **Az a sakkfigura, amelyik elhárítja a király elleni sakkot, az ellenfél királyának sakkot adhat.**

Conclusions and plans

- Bilingual and monolingual lexicography exploiting mutual interests
- Dictionaries and other projects “helping” each other
 - A project of updating Thesaurus and Collocations dictionary underway
 - Semantic data for/from ELEXIS Word Sense Disambiguation task
 - New corpora, e.g. monitor corpus (*presentation tomorrow at 9am!*)
- Future
 - More bilingual data to be added (recently approved project, Ministry of Culture)
 - Customized editor for the Digital dictionary database
 - Including state-of-the-art technologies (definition extraction etc.)
 - Linking dictionaries (Dictionary Matrix, ELEXIS)

Acknowledgments

This project was funded by ARRS (Slovenian Research Agency) through the research programme P6-0411 "Language Resources and Technologies for Slovene", and the infrastructure programme I0-0022 "Network of Research Infrastructure Centres (MRIC)" at the University of Ljubljana.

