



CLARIN Annual Conference 2017

Budapest, 18-21 September 2017

Introducing HUNCLARIN

Tamás Váradi
Research Institute for Linguistics
Hungarian Academy of Sciences
varadi.tamas@nytud.mta.hu

Outline

- Background
- The Consortium
- Overview of Resources and Tools
- Reaching out
- Joining CLARIN ERIC
- Becoming CLARIN ERIC compliant

Background

- 2008-2010 Hungarian RI Roadmap (NEKIFUT) in active preparation
- HUNCLARIN branded a Strategic Research Infrastructure
- 2010 Government change
- NEKIFUT shelved for 6 years

- Work continued under different funding
 - META-NET (CESAR project)
 - Hungarian Academy of Sciences (MTA)

The HUNCLARIN Consortium

Research Institute of Linguistics, MTA (Tamás Váradi)

Budapest Technical University TMIT & MOKK (András Kornai, Géza Németh, Klára Vicsi)

Szeged University (Veronika Vincze, Richárd Farkas)

Pázmány Péter University (Gábor Prószéky)

Morphologic Plc (Gábor Prószéky, László Tihanyi)

Institute for Computer Science and Control, MTA (András Kornai)

Institute of Cognitive Neuroscience and Psychology MTA (Tibor , Bea Ehman)

Debrecen University (László Hunyadi)

Resources

Corpora:

Monolingual: Hungarian National Corpus, Hungarian Webcorpus, HunLearner, Sentiment Detection, MWE corpora etc.

Parallel: Hunglish, 1984 (Multext East)

Treebank: Szeged Treebank

Databases: Hungarian EuroWordNet, Inflectional Databases,

Tools

HUNxxx toolset (HUNalign, HUNtoken, HUNmorph, HUNtag, HUNner)

HUMOR morphological analyser

Magyarlánc, a complete toolchain

Mazsola, verb structure extraction and browser

MetaMorpho, parser

Integration of processing tools: e-magyar

Interoperable
Open
Modular
Free
Complete



Reaching Out

- Collaboration with narrative psychology research groups
- Digital Humanities Centre at ELTE University, Budapest
- MTA Centre for Social Sciences
 - Institute for Political Science

Joining CLARIN ERIC

- A breakthrough at last?
- Little interest by NKFIH
- Seems like we are left to our own resources ... ☹️

Becoming CLARIN ERIC compliant

- Hardware infrastructure is in place
- Single sign-on system available
- Running TLA, hosting
 - Budapest Sociolinguistics Interview,
 - BEA (huge annotated speech corpus)
 - HUCOMTECH multimedia corpus
- Final repository software not decided on yet
- Currently most resources and tools available through e-magyar.hu or Metashare
- Website at clarin.hu under development

Summary

- Underfunded but not underresourced
- We are slowly getting there ...

Thank you for your support!