

István Kenesei

## Hungarian as a Pluricentric Language\*

### 1. Overview

In this short essay I will summarise prevailing views concerning pluricentricity, point at the state of Hungarian as a pluricentric language, and review recent actions aiming at alleviating the problems or dangers of an asymmetric pluricentric situation.

### 2. Types of pluricentric languages

Following early attempts at classifying multilingual communities, which involved – in the terminology then suggested – the use of monocentric and polycentric languages, with “different sets of norms” existing simultaneously in the latter case (Stewart 1968: 534), current terminology has settled with the expression *pluricentric*.

According to Ulrich Ammon's definition, the term *pluricentric language* refers to the notion that “languages evolve around cultural or political centers (towns or states) whose varieties have higher prestige” (Ammon 2005: 1536). Such a language has two or more standard varieties in the traditional use of the term. For example, British English and American English are standard varieties of the English language. But in more modern parlance it can also be used to describe the case of languages without any territorial delimitation, such as Romani.

There are various cases of linguistic pluricentricity (cf. Clyne 1992), some of which are self-explanatory, such as plurinational (Spanish in Spain, Mexico, Argentina, etc.; French in France, Québec, etc., Portuguese in Portugal, and Brazil); or pluriregional (Northern and Southern Germany). We may speak of a pluristatal language, when a single nation has been politically divided into separate administrative units with different norms (Korean in North vs. South Korea, German in the former Bundesrepublik and GDR). Ammon also introduces the concept of divided language, as Serbian and Croatian, or we might add Romanian vs. Moldavian. In this case either a single language underlies two different languages, or, as was exemplified in pluristatal languages, speakers are/were separated by political boundaries.

Some pluricentric language communities are symmetric, as in American, Australian, British, etc. English(es), i.e., none has more prestige than any one of the others, but we often encounter pluricentric languages in which one variety has more power or prestige than the other(s), as in the French of the Île de France vs. Québécoise or at least in the perception of a large number of speakers regarding the differences between the German spoken in Switzerland vs. the German of Berlin. Such asymmetric linguistic situations can also be characterised by the term ‘unbalanced’, calling forth attributes as ‘dominating vs. non-dominating’ varieties (Muhr 2004), ‘centre vs. periphery’, or, in our practice below, making reference to ‘major and minor centres/varieties’.

---

\* I wish to thank Csilla Bartha and Tamás Váradi for their help in the preparations for my presentation at the EFNIL Conference in Madrid, November 20, 2006, which was the basis of the present paper. The use of maps from the Research Institute for National and Ethnic Minorities of the HAS is gratefully acknowledged.

### **3. Facts and dangers of pluricentricity**

The source of the differences between major and minor varieties of a pluricentric language may go back to intralinguistic or extralinguistic causes. Among the former, one can detect a) dialectal differences, including phonemic, morphological, lexical and syntactic ones, b) differences encoded by standards, distinguishing central and regional varieties, and c) speakers' attitudes, which may stigmatise one or more varieties as against others.

Extralinguistic causes include a) historical or political developments, such as new borders drawn within a single language community, b) institutional, i.e., the preference of one variety to another in a range of state, municipal, or local offices, and c) political, or the attitude of the body politic to language use.

It follows from the factors listed in the previous paragraphs that there may arise marked status differences between major and minor varieties of a pluricentric language, which may entail various dangers as to the state and future of these varieties. Whereas the variety of a major centre can attain the status of 'official language' due to its real or presumed historical origins and unbroken 'genealogy', its administrative use, its cultural significance and the number of speakers it can lay claim to, the varieties spoken in minor centres appear as the languages of minorities derivative as a sidetrack from the common historical source, usually suppressed in public life with only regional (as against national) significance, and distinctly less number of speakers. Consequently, the major variety has a full variety of styles and registers, including formal styles, a unified vocabulary, and, as a result, high prestige, while the minor ones have limited registers and styles, and often mixed vocabularies, so they have to put up with a much lower rung on the imaginary ladder of prestige hierarchy.

We will try to show here in a case study of Hungarian that the dangers inherent in a pluricentric situation can be reduced to some extent by applying state-of-the-art language technology.

### **4. Hungarian: The current situation**

Hungarian speaking language communities are currently scattered among eight countries in the Carpathian basin (see Map 1, p. 13).

With the notable exception of Transylvania in Romania, these language communities line up along the borders of the neighbouring countries, as is shown in the following map illustrating the concentration of speakers of Hungarian in varying shades of red (see Map 2, p. 14).

The following map shows a conservative estimate (based on censuses) of Hungarian speakers in the regions concerned (see Map 3, p. 15).

To put it in a tabular format, cf.:

Transylvania (NE Romania)	1,500
Northern region (S Slovakia)	500
Southern region (Vojvodina, Serbia)	300
Eastern region (Sub-Carpathia, UA)	150
Minor communities:	
Osijek region (Croatia)	15
Burgenland (Austria)	<10
Mura region (Slovenia)	<10
<b>Regions</b>	<b>2,500</b>
<b>Hungary</b>	<b>10,000</b>

(in thousands of speakers)

This set of data clearly shows how overpowering the central variety is as compared to the regions, which have long been separated not only from the central variety/ies, but – and even more radically – from one another. With the fall of Communism and the (at least in some relations, relatively) free communication and travel between the respective countries, a new era of revitalisation of linguistic relations has come, among other things.

#### **4. Areas of influence in a new approach to pluricentricity**

With the decrease of the direct impact of highly centralised educational systems on language use it is now possible to attend to the needs of the minor varieties in all levels of education, from primary to tertiary. This includes systematic planning of language acquisition, taking into account problems of and solutions for bi- or multilingual communication. Language maintenance centres have been established to collect data on and promote the use of the local ‘minor’ varieties.

The domains in which these minor varieties need to be used increasingly or where their use must be (more and more) acknowledged comprise local governments, political institutions (parties, organisations). This can be achieved if the minor varieties acquire local prestige and if primarily linguists and teachers, and, in general, people with local prestige spread positive attitudes to their own language varieties.

In the four larger regions three-level bilingual educational systems provide the foundation for this work. They are complemented by the activities of the Catholic and Protestant churches, political organisations (parties, unions), the civil society (associations, societies, clubs, etc.), media and cultural channels (press, radio, television, theatres, publishing houses, libraries, etc.). In the regions where the relative weight of the Hungarian-speaking community is smaller, only the latter channels are available.

There are a number of actions linguists themselves are capable of carrying out in order to recognise pluricentricity and enhance the prestige of minor varieties. To list a few:

- Updating mono- and bilingual dictionaries
- Updating descriptive grammars

- Streamlining advisory services
- Setting up and supporting language maintenance centres in regions
- Developing new standards by
  - Corpus collection and implementation
  - Regular interaction and training

It is the last few items on this list that will be illustrated in more detail below.

## 5. Corpus linguistics: an aid to minor varieties

The Research Institute for Linguistics of the Hungarian Academy of Sciences (RIL), in particular the corpus linguistics team led by Tamás Váradi started work on the Hungarian National Corpus (HNC) in 1998. They set out to create a balanced reference corpus of 100 million words of present-day Hungarian. It collects data from the following genres: press, literature, official, scientific and personal. Its size is currently c. 160 million running words, annotated (tagged) automatically according to stem, word-class, and information on inflection class, an important record in this agglutinating language. The HNC is available on line following a routine registration procedure at [http://corpus.nytud.hu/mnsz/index\\_eng.html](http://corpus.nytud.hu/mnsz/index_eng.html).

In 2002 the data collection entered a different phase when a new project began under the title “Hungarian Language Corpus of the Carpathian Basin”. Its aim was to create a 15-million-word corpus of Hungarian language beyond the borders of Hungary. The truly all-Hungarian National Corpus, containing language variants from Slovakia, Subcarpathia, Transylvania and Vojvodina, was inaugurated in November 2005. It is the result of a collaboration of the Hungarian Language Offices and the researchers at RIL. The individual Hungarian Language Offices are available under the following links:

Gramma Language Office (Slovakia):

<http://www.gramma.sk/index.php?lang=en>

Szabó T Attila Linguistic Institute (Transylvania, Romania):

<http://www.lett.ubbcluj.ro/~tjuhasz/sztanyi/EnContent.html>

Hodinka Antal Institute (Subcarpathia, Ukraine):

<http://www.kmf.uz.ua/egysegek/kutatomuhelyek/magyarsagkutato/index.html>

Society for Hungarian Studies (Vojvodina, Serbia):

<http://www.korpusz.org.yu/nyelvikorpusz/index.htm>

Data from the four largest regional variants of Hungarian outside Hungary are collected under the same conditions as those from the major variant: words are annotated the same way and are grouped into five styles. In addition, spoken language is also recorded here. The amount of the data collected is shown here (see slide p 24).

The following table shows the ratio of the various styles according to the regions (see slide p. 25).

The next table adjusts percentages to the amount of data from the major region (see slide p. 26).

## **6. Conclusions and outlook**

Corpus collection and dissemination of linguistic findings relating to results of research gained from these efforts were coupled with consultations between the researchers in all the regions concerned. This was itself an educational experience and an occasion to evidence the importance of the minor varieties for speakers of the major ones. Due to these actions new (editions of) monolingual dictionaries now contain a large number of entries from the minor varieties of Hungarian, to the astonishment of some conservative linguists and thinkers from the major region. But the process of 'emancipation' has started and cannot be terminated.

## **7. References and further readings**

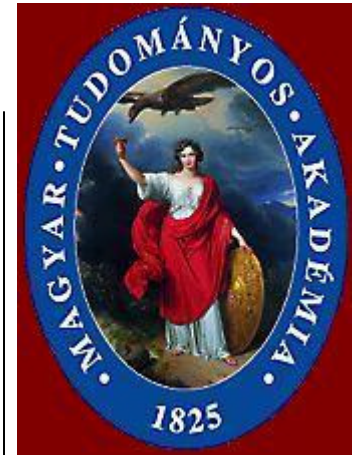
- Ammon, Ulrich (2005): Pluricentric and divided languages. In: Ammon et al. (eds.), pp. 1536-1543.
- Ammon, Ulrich/Dittmar, Norbert/Mattheier, Klaus J./Trudgill, Peter (eds.) (2005): Sociolinguistics: An international handbook of the science of language and society. Berlin/New York: Walter de Gruyter.
- Clyne, Michael (1992): Pluricentric languages: Differing norms in differing countries. New York: Mouton de Gruyter.
- Csernicskó, István (1998): A magyar nyelv Ukrajnában (Kárpátalján). [= The Hungarian language in Ukraine (Subcarpathia)]. Budapest: Osiris.
- Göncz, Lajos (1999): A magyar nyelv Jugoszláviában (Vajdaságban). [= The Hungarian language in Yugoslavia (Vojvodina)]. Budapest-Ujvidék: Osiris-Fórum.
- Kontra, Miklós (2005): Hungarian in- and outside Hungary. In: Ammon et al. (eds.), pp. 1811-1877.
- Lanstyák, István (1995): A magyar nyelv központjai. [= The centres of the Hungarian language]. In: Magyar Tudomány 1995/10, pp. 1170-1183.
- Lanstyák, István (1996): A magyar nyelv állami változatainak kodifikálásáról. [= On the codification of the national varieties of the Hungarian language]. <http://www.staryweb.fphil.uniba.sk/~kmjl/LI%20MNy%20allami%20valt%20kodif.htm>.
- Lanstyák, István (2000): A magyar nyelv Szlovákiában. [= The Hungarian language in Slovakia]. Budapest-Pozsony: Osiris-Kalligram.
- Muhr, Rudolph (2004): Language attitudes and language conceptions in non-dominating varieties of pluricentric languages. In: Internet-Zeitschrift für Kulturwissenschaften, Nr. 15. [http://www.inst.at/trans/15Nr/06\\_1/muhr15.htm](http://www.inst.at/trans/15Nr/06_1/muhr15.htm).
- Stewart, William A. (1968): A sociolinguistic typology for describing national multilingualism. In: Fishman, Joshua A. (ed.): Readings in the Sociology of language. The Hague: Mouton, pp. 531-553.

István Kenesei

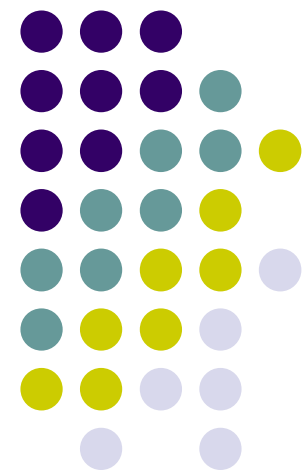
Director, Research Institute for Linguistics, Hungarian Academy of Sciences  
Budapest, Hungary

The following tables and maps represent the presentation given at the EFNIL conference.

# Hungarian as a Pluricentric Language



István Kenesei  
Research Institute for Linguistics  
Hungarian Academy of Sciences  
Budapest  
[www.nytud.hu](http://www.nytud.hu)



# The Research Institute for Linguistics

- One of 39 research institutes in HAS
- Ca. 100 research staff, 1/3 on grants
- Fields/teams/functions:
  - Uralic studies, history; phonetics; lexicography; psycho/neuro/cognitive, theoretical linguistics; advisory services; library; information centre
  - **Sociolinguistics** (Hungarian in and outside Hungary **and** the languages spoken in Hungary), **Language technology** →

# Overview

- Socio-political contexts
- Sociolinguistic contexts
- Hungarian in the larger region:
  - Speakers and institutions
  - Problems and objectives
- Linguistic devices
- Example: corpus building.



# Types of Pluricentric Languages

- Plurinational (Spanish, French, English ...)
- Pluristatal (North vs. South Korean)
- Divided (Romanian vs. Moldavian)
- Pluridialectal (North vs. South German)
  
- Asymmetric or **Unbalanced**:  
dominating vs. non-dominating; 'top-heavy';  
centre vs. periphery; **major vs. minor centres**:  
The case of Hungarian (German, French, ...)

# Sources & types of differences

- Intralinguistic
  - Dialectal  
(phonemic, morphologic, lexical, syntactic)
  - Standards (central vs. regional – at best)
  - Attitudes (of speakers to own language)
- Extralinguistic
  - Historical (new borders)
  - Institutional (range of municipal/state offices)
  - Political (attitudes to language use).

# Status differences: Facts and/or dangers

## Major

- Official language
- Historical source
- Administrative use
- Cultural import
- More speakers
- High prestige

**HIGH**

## Minor

- Language of minority
- Derivative
- Suppressed in public life
- Regional significance
- Less speakers
- Low prestige

**LOW**

# **(Socio)Linguistic distinctions: More (potential) dangers**

## **Major**

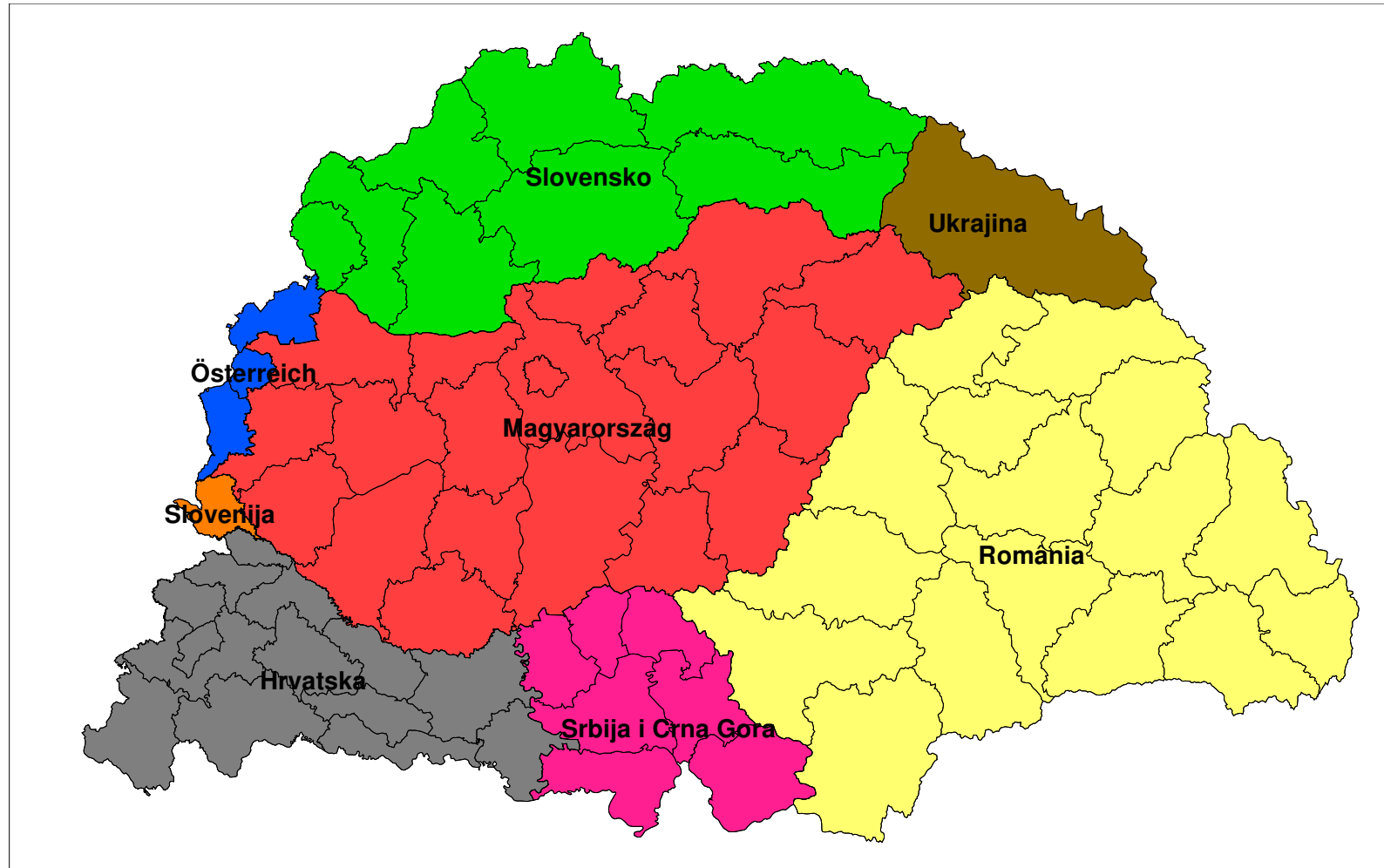
- Formal styles
- Full scale of registers
- Full variety of styles
- Unified vocabularies
- Positive attitudes

## **Minor**

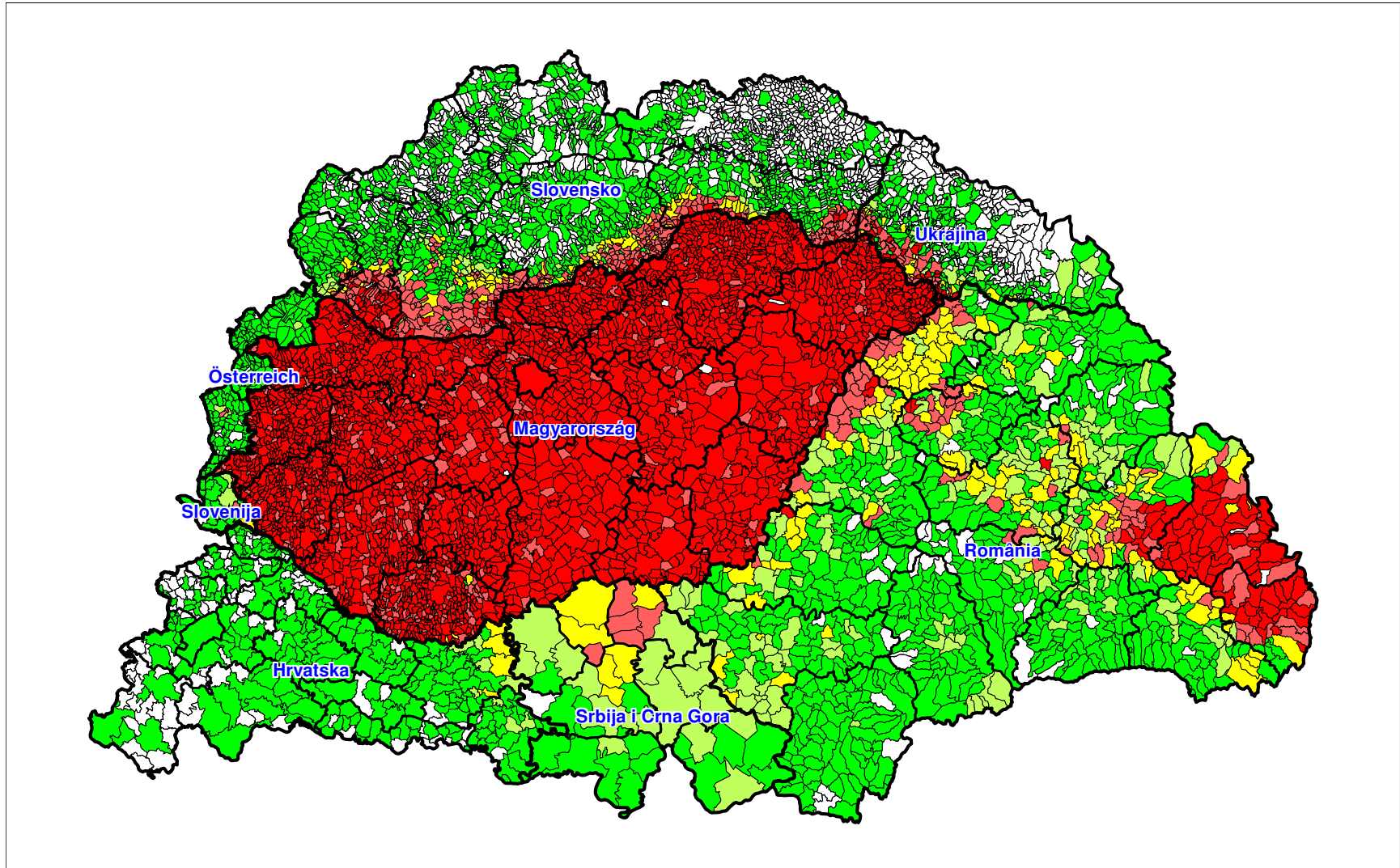
- Vernacular only
- Limited registers
- Limited styles
- Mixed vocabularies
- Negative attitudes

## **Case study: Hungarian**

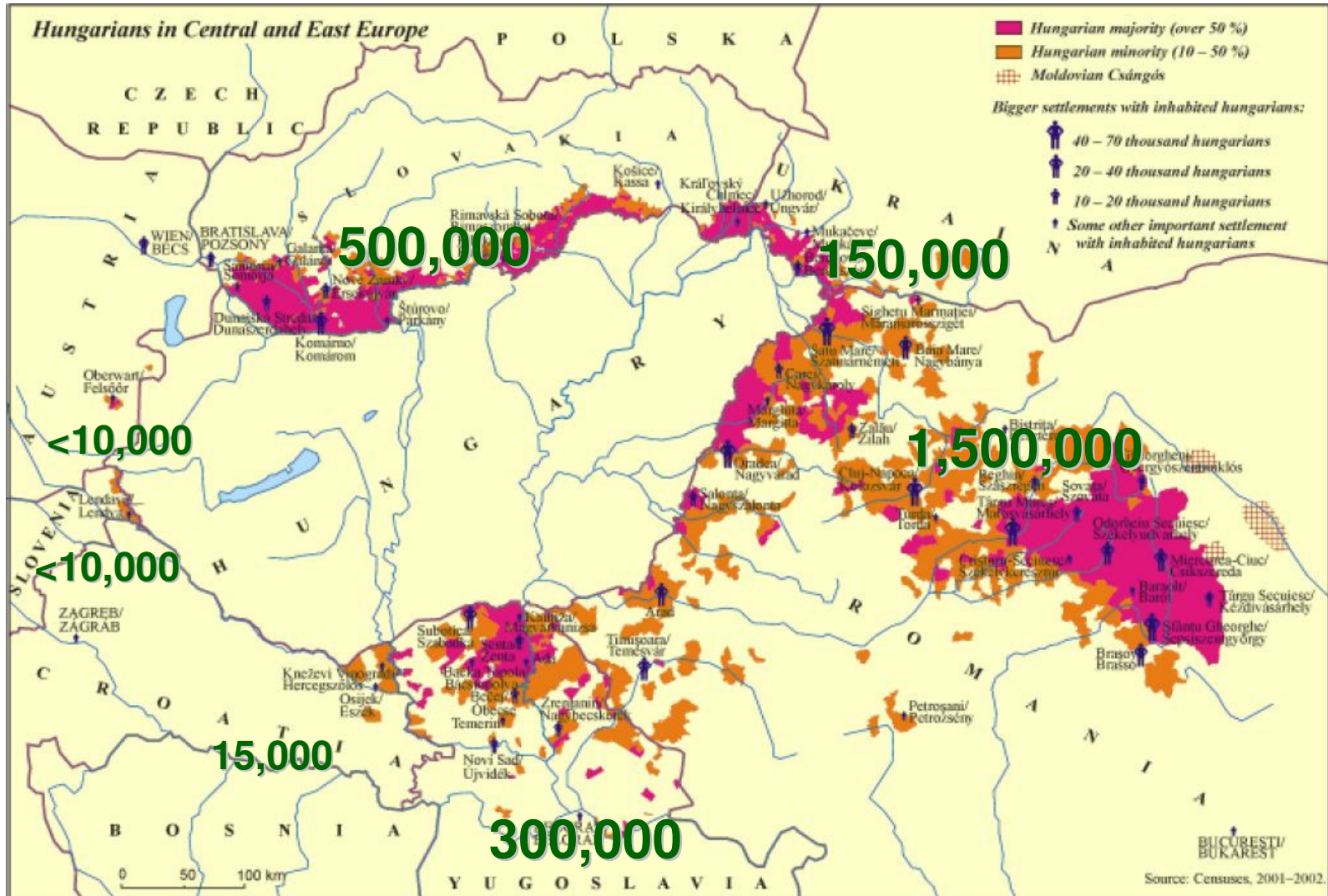
# Map 1: Countries in the Carpathian Basin



# Map 2: Hungarians in the region



# Map 3: Hungarians in the region





# Number of speakers in regions

- Transylvania (NE Romania): 1.5 M
- Northern region (S Slovakia): 500 K
- Southern region (Vojvodina, Serbia): 300 K
- Eastern region (Sub-Carpathia, UA): 150 K
- Minor communities:
  - Osijek region (Croatia): 15 K
  - Burgenland (Austria): <10 K
  - Mura region (Slovenia): <10 K
- **Regions:** **2.5M**
- **Hungary:** **10.0M**



# Areas of influence:

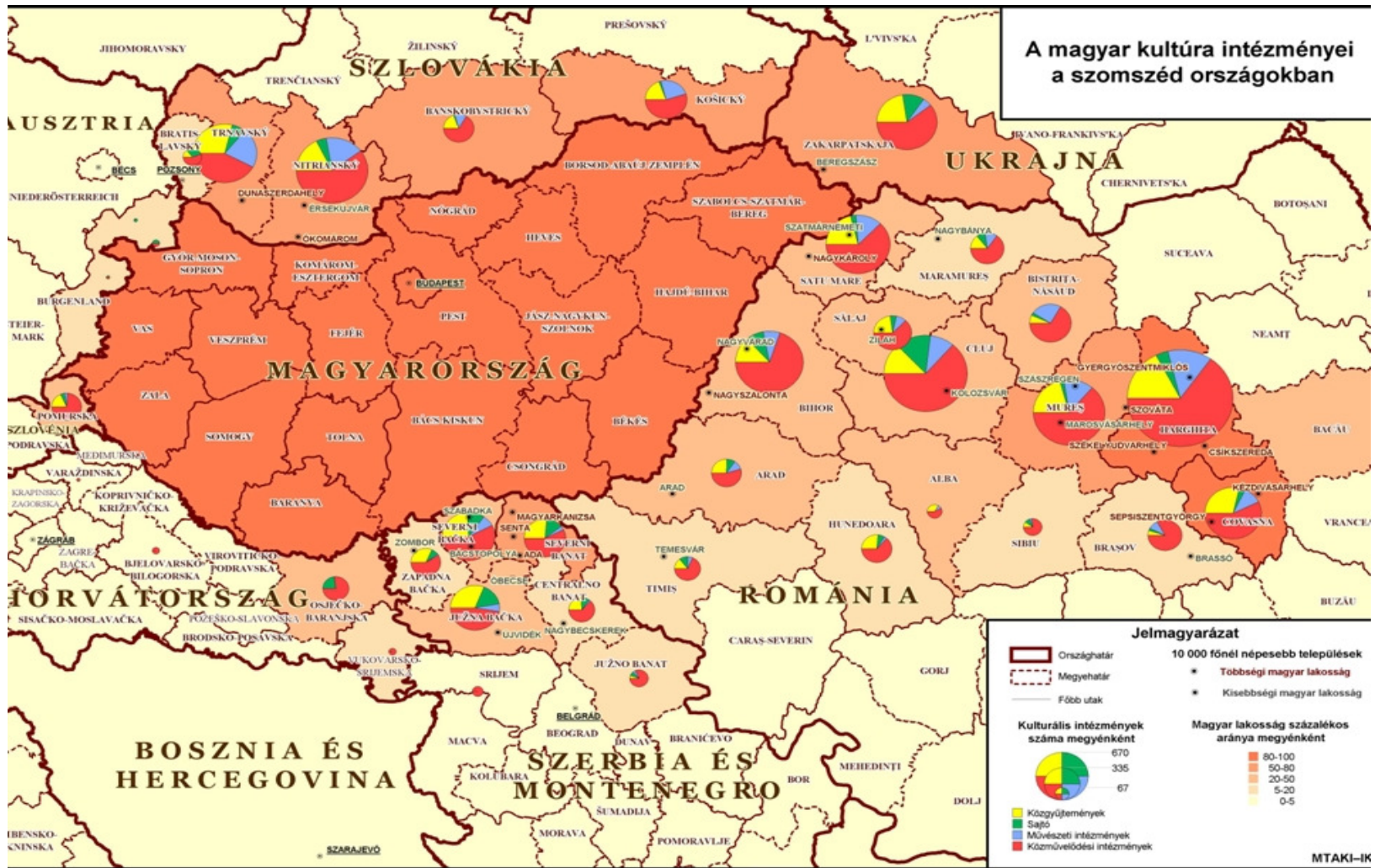
## Maintenance and revitalization

- Language maintenance centers
- Educational system
  - domains (pre-primary to tertiary)
  - acquisition planning
- Local government
- Political decisions
- Creating local prestige
- Spreading positive attitudes to local varieties.

# Language maintenance

- Romania, Slovakia, Vojvodina, Subcarpathia:
  - state and private general & higher education
  - churches (Catholic and Protestant)
  - political institutions (parties)
  - civil society (associations, organisations)
  - press, media, theaters
  - publishing houses.
- Elsewhere:
  - mostly cultural institutions: libraries, cultural centres, press, museums, etc.

# Map 3: Cultural institutions



# Toward a more balanced picture

- Recognizing pluricentricity
  - Updating mono- and bilingual dictionaries
  - Updating descriptive grammars
  - Streamlining advisory services
- Support for regional centers
  - Language maintenance centers in regions
- Developing new standards →
  - Corpus collection and implementation →
  - Regular interaction and training →

# Hungarian National Corpus (HNC)

- General reference corpus of present-day written Hungarian used inside Hungary
- 165 million running words
- Balanced selection in uniform annotation
  - treasure house of data attesting actual use
- Intelligent corpus available at [www.nytud.hu](http://www.nytud.hu)
  - every word linguistically annotated
  - query by linguistic features as well

# Hungarian Minority Language Corpus (HMLC)

- 4 regional minority language variants
  - Transylvania, Slovakia, Subcarpathia, Vojvodina
- 5 styles
  - same structure as in HNC
- spoken language material as well
- uniform coding, same query system as in HNC
- affords easy comparison with dominant language documented in HNC

# Objectives of HMLC

- preserve and disseminate
- provide solid empirical evidence for status of language (dispelling prejudices)
- provide solid data for comparative analysis
- spread the use of language technology in the areas concerned
- foster collaboration between research centres in the regions

# Data in Hungarian Minority Language Corpus (HMLC)

	Text (wordcount)	Speech	
		(wordcount)	(duration)
Slovakia	<b>9529021</b>	<b>75376</b>	<b>13:57:35</b>
Transylvania	<b>8856014</b>	<b>54272</b>	<b>7:56:14</b>
Subcarpathia	<b>2506004</b>	<b>141254</b>	<b>20:44:23</b>
Vojvodina	<b>2043650</b>	-	-
Total	<b>22934689</b>	<b>270902</b>	<b>42:38:12</b>



# Structure of HMLC by region and genre in comparison with HNC

	<i>press</i>	<i>literature</i>	<i>science</i>	<i>personal</i>	<i>official</i>	<i>Total</i>
<b>Slovakia</b>	5737022 60.2%	1353586 14.2%	2286044 24%	- 0%	155369 3.6%	9532021 100%
<b>Transylvania</b>	5505248 62.2%	760499 8.6%	1603377 18.1%	380238 4.3%	606652 6.8%	8856014 100%
<b>Subcarpathia</b>	739869 29.5%	379896 15.2%	745900 29.8%	374492 14.9%	265847 10.6%	2506004 100%
<b>Vojvodina</b>	1460102 71.4%	216677 10.6%	337325 16.5%	26239 1.3%	3307 0.2%	2043650 100%
<b>HNC</b>	71012200 43%	3546912 21.5%	2053555 13%	17838598 11%	19854717 11.5%	164710196 100%
<b>Total</b>	84454446 45%	3817978 20.3%	25508196 13.6%	18619567 9.9%	20885892 11.2%	187647885 100%

# Structure of HMLC and HNC by genre and region

	<i>press</i>	<i>literature</i>	<i>science</i>	<i>personal</i>	<i>official</i>	<i>Total</i>
<b>Slovakia</b>	5737022 6.8%	1353586 3.5%	2286044 9%	- 0%	155369 0.7%	9532021 5%
<b>Transylvania</b>	5505248 6.5%	760499 2%	1603377 6.3%	380238 2%	606652 2.9%	8856014 4.7%
<b>Subcarpathia</b>	739869 0.9%	379896 1%	745900 2.9%	374492 2%	265847 1.3%	2506004 1.3%
<b>Vojvodina</b>	1460102 1.7%	216677 0.6%	337325 1.3%	26239 0.1%	3307 0.01%	2043650 1%
<b>HNC</b>	71012200 84.1%	3546912 93%	2053555 81.5%	17838598 95.9%	19854717 95%	164710196 88%
<b>Total</b>	84454446 100%	3817978 100%	25508196 100%	18619567 100%	20885892 100%	187647885 100%

# Hungarian Minority Language Corpus in the Hungarian National Corpus

	<i>press</i>	<i>literature</i>	<i>science</i>	<i>personal</i>	<i>official</i>	<i>Total</i>
<b>Slovakia</b>	5737022 6.8% 60.2%	1353586 3.5% 14.2%	2286044 9% 24%	- 0% 0%	155369 0.7% 3.6%	9532021 5% 100%
<b>Transylvania</b>	5505248 6.5% 62.2%	760499 2% 8.6%	1603377 6.3% 18.1%	380238 2% 4.3%	606652 2.9% 6.8%	8856014 4.7% 100%
<b>Subcarpathia</b>	739869 0.9% 29.5%	379896 1% 15.2%	745900 2.9% 29.8%	374492 2% 14.9%	265847 1.3% 10.6%	2506004 1.3% 100%
<b>Vojvodina</b>	1460102 1.7% 71.4%	216677 0.6% 10.6%	337325 1.3% 16.5%	26239 0.1% 1.3%	3307 0.01% 0.2%	2043650 1% 100%
<b>HNC</b>	71012200 84.1% 43%	3546912 93% 21.5%	2053555 81.5% 13%	17838598 95.9% 11%	19854717 95% 11.5%	164710196 88% 100%
<b>Total</b>	84454446 100% 45%	3817978 100%20.3%	25508196 100%13.6%	18619567 100% 9.9%	20885892 100%11.2%	187647885 100% 100%

# Summary

- Unbalanced pluricentric language
- Determining status problems
- Registering sociolinguistic features
- Determining current and future objectives
- Realising (some) objectives
  - Example: status enhancement through corpus building.

# Sources:

- General:  
U. Ammon, M. Clyne, H. Kloss, W. Stewart,  
R. Muhr
- Hungarian:  
I. Lanstyák, I. Csernicuskó, L. Göncz, S. Gal,  
Cs. Bartha, K. Kocsis, M. Kontra, T. Váradi
- Maps:  
Institute for National and Ethnic Minorities of  
HAS, HNC/HMLC Project (RIL, Budapest)