Romani dialectology and structure, sociolinguistic situation of Romani language speakers, Roma as multilingual speakers in monolingual states

Kimmo Granqvist
Charles University
March 17, 2021

Introduction: Roma and Romani language

Overview of Romani (1)

- Romani belongs to the Indo-European language branch of the Indo-Aryan subgroup of Indo-Iranian languages.

- one of the Indian languages, spoken outside of India by traditionally nomadic group of people, whose ancestors left India.

- since the Middle Ages it has been the only Indo-Aryan language spoken exclusively in Europe; Today it is spoken in the United States, South America and the Middle East.

- The name *rom* or *řom* is the counterpart of the names used by other Indian-speaking traditionally nomadic groups: *lom* people of the Caucasus and Anatolia, *dom* of the Middle East, and *dom* in northern Pakistan dom-nation (ROMANI Project Manchester, 2006a).

Overview of Romani (2)

- There is no definitive information on the number of speakers: a conservative estimate is 3.5 million speakers in Europe and 0.5 million in the rest of the world

- most speakers are in South-Eastern Europe: Romania, Bulgaria, Turkey, Macedonia, Serbia and Montenegro, but also in Greece, Slovakia, Moldavia and Hungary.

- Roma communities in Portugal, Spain, the UK and Finland have almost completely lost the Romani language and switched to the majority language. (with the exception of the old Scandinavian Roma)

- Nowadays, there are some estimates of about 50-60 dialects.

Speakers of Romani by country (Bakker & Kyuchukov 2000) (1)

country	speakers	%
Albania	90.000	95%
Austria	20.000	80%
Belarus	27.000	95%
Belgium	10.000	80%
Bosnia-Herzegowina	40.000	90%
Bulgaria	600.000	80%
Croatia	28.000	80%
Czech Republic	140.000	50%
Denmark	1.500	90%
Estonia	1.100	90%
Finland	3.000	90%
France	215.000	70%
Germany	85.000	70%
Greece	160.000	90%
Hungary	260.000	50%
Italy	42.000	90%
Latvia	18.500	90%
Lithuania	4.000	90%
Macedonia	215.000	90%
Moldova	56.000	90%
Netherlands	7.000	90%
Poland	4.000	90%
Romania	1.030.000	80%
Russia	405.000	80%
Serbia and Montenegro	380.000	90%
Slovakia	300.000	60%
Slovenia	8.000	90%
Spain	1.000	1%
Sweden	9.500	90%
Turkey	280.000	70%
Ukraine	113.000	90%
United Kingdom	1.000	0,50%

Romani dialects

Current dialects (1)

- Today in Europe there are at least 50-60 Romani dialects

- The dialect names are often derived from the self-appellation of the group
  - *lovarengi čhib* ’lovari language’ or Finnish *kaalengo tšimb* ’language of the Kaale',
  –
- The same process applies to adverbs: *sintikes* ’in the Sinti way'.

- The research literature on Romani dialects often refers to dialects by the plural forms of group endonyms or exonyms .
  - Plural *xaladitka* refers to the North Russian Romani dialect, from the name used by other Roma to refer to them (endonym is Russka Roma).

Current dialects (2)

- Some of group names used in the Balkans: *arli* 'settled person' < turk. *yerli*, *bugurdži* 'drill maker' < turk. *burgucu*, *čurari* 'sieve maker' < roman. *ciurar*, *kalderaš* 'cauldron maker' < roman. *căldărar*, *sepeči* 'basket wievers' < turk. *sepetçi* and *ursari* 'bear leaders' < roman. *ursar*

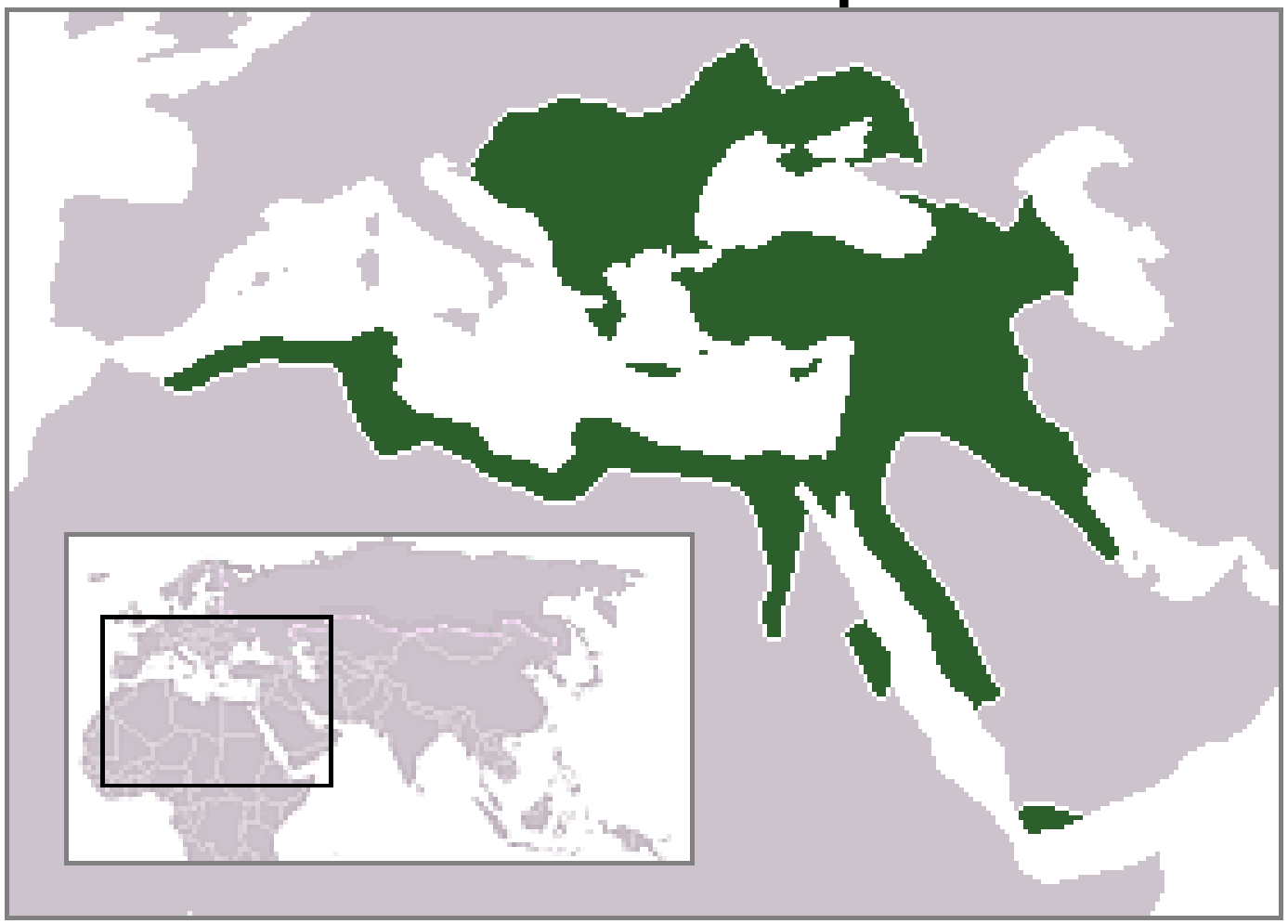
- The group names do not always match linguistic varieties one-to-one:
  - Groups using the same speech variety sometimes live as separate communities and use different names.
  - Sometimes groups speaking different dialects use the same self-appellation, based on common religious affiliation, trade, or region of origin

(Matras 2002: 5.)

Current Romani dialects – background (1)

- The Ottoman Empire expanded around 1300-1400 to include the formerly Byzantine areas of modern Turkey, Greece, the Balkans and the Crimea.
- The unrest in the region has driven some Roma to migrate from Byzantium to Central and Western Europe.
- In the 1500's the Ottoman Empire expanded along the Mediterranean coastline and included Tunisia, Egypt, Mesopotamia, Hungary and some regions of Europe all the way to Vienna.
- The Ottoman expansion in Europe stopped with the unsuccessful siege of Vienna in 1529.
- In 1571, during the battle of Lepanto, the Spanish fleet defeated the Ottomans.
- The Habsburg monarchy and the Ottoman Empire was then divided by a political boundary, preventing contacts between the different groups of the displaced Roma on either side of the border.

Ottoman Empire



(http://fi.wikipedia.org/wiki/Tiedosto:Ottoman_1683.png)

Great Divide (1)

- At this time the diffusion of linguistic features was inhibited, resulting in a dense and distinctive set of linguistic boundaries (isoglosses).

- Yaron Matras (2005) used this political frontier to explain the cluster of differences between the Northern and Southern European Romani dialects, terming it the Great Divide.

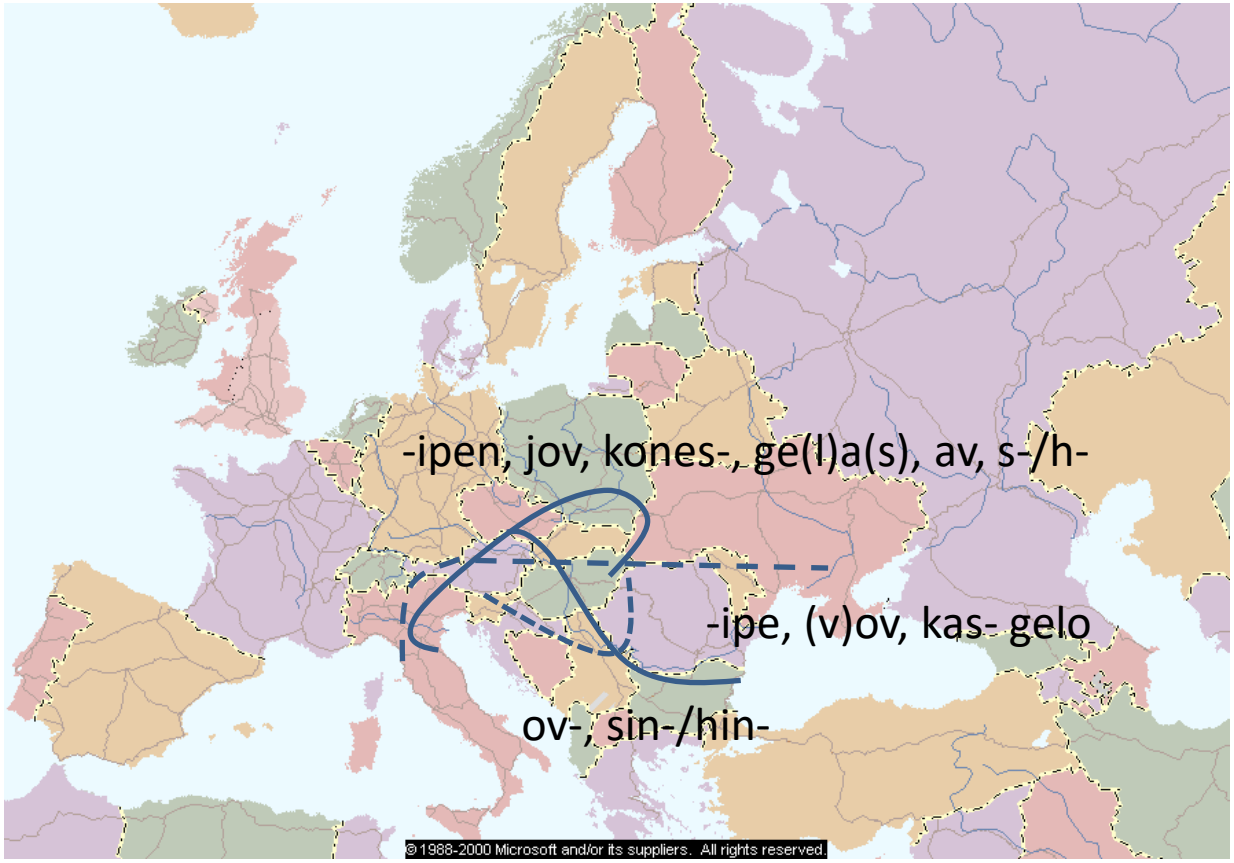
Great Divide (2)

- Both sides of the Habsburg-Ottoman border had separate centers of linguistic innovation of Romani.

- To the north of the border, there is the German-speaking northwestern region. Innovations in this zone have spread from the east and to the north, affecting British Romani, Finnish Romani and the Northeastern Romani dialects.

- To the south of the border, there is the South-Eastern zone of innovation. It has two distinct distribution patterns.
  - The first of these covers the entire Balkans
  - The second one is more specific to Transylvania/Wallachia, but often influences the Balkans, especially around the the Black Sea coast.
  - It sub-divides the southern Balkans into the eastern and western zones, as well as the Transcarpathian areas to the northwest of it. (Matras 2005: 13-14).

Great Divide (3)



Great Divide (4)

- Innovations spread on each of the two sides of the Great Divide, but each side's innovations are contained by the Divide and do not spread across it.

- Northern Romani dialects show a preference for word-initial jotation (prothetic *j*-) rather than vowel-initial nominals, e.g.:
  - *aro* > *jaro* 'egg',
  - 3rd person pronouns *ov* > *jov*, *oj* > *joj, on* > *jon* 'he, she, they'

- With interrogative pronoun *kon* 'who', the conservative irregular oblique form *kas*- has been replace in the North by a regular form *kon-es,* transparently derivable from the nominative *kon*.

Great Divide (5)

- Another prominent feature to the north of the Great Divide is the disappearance of the gender-inflected perfective forms of verbs of motion, e.g. *gel-o* 'he went', *gel-i* 'she went'. These are replaced with *geljas/gejas/geja* 'he/she went' - forms found in the regular perfective paradigm, analogous to *kerdjas/kerdja* 'he/she did'

  - This replacement is a relatively late development, which has not had time to spread throughout the whole north.

  - Participial and verbal forms coexist, for example, in Austria, Hungary and parts of Romania. (Matras 2005: 14-15).

  - In Finnish Romani both of these forms are found

Great Divide (6)

- In addition, in the north the subjunctive and future copula (‘to be’, ‘to become’) relies on the root *av*- ‘to come’ . In the south, the original *ov*- is retained.

- Finnish Romani has documented irregular variant of *ov*- , *uljuv*- 'become something'.

- Dialects in the north tend to not have the extension -*in*- in the copula
  - exceptions include the transition zone around Slovakia, as well as Finnish Romani as a peripheral dialect.
  - In these dialects the copula form of the present tense 3rd person is *hin(o/i)* ’he/she is'.

- In southern Europe the -*in*- extension is more common and occurs sometimes throughout the whole paradigm (e.g. *sine* ’he/she is', *sinom* 'I am', *sinan* 'you are'). (Matras 2005: 15-16).

Great Divide (7)

- The isogloss bundle of the Great Divide is partly strengthened by a southeastern cluster
- To the East, this line connects the dialects of the Black Sea coast region with the Carpathian region.
- Some features stretch to Bulgaria, and as far north as Slovakia and southern Poland.
- In the center, this zone contains the northern Vlax dialects, which appear to be the source of many of the innovations on the eastern side of the line.
- These innovations include affrication of palatalized dentals **t'* and **d'*
  - E.g. *tikno > cikno* 'small', *dives* > **džives* > *dzes*, *zis* 'day' . *cikno* occurs from East Slovakia all the way to Epiros (Greece). Both of these extreme regions show de-palatalization in *dives*

Great Divide (8)

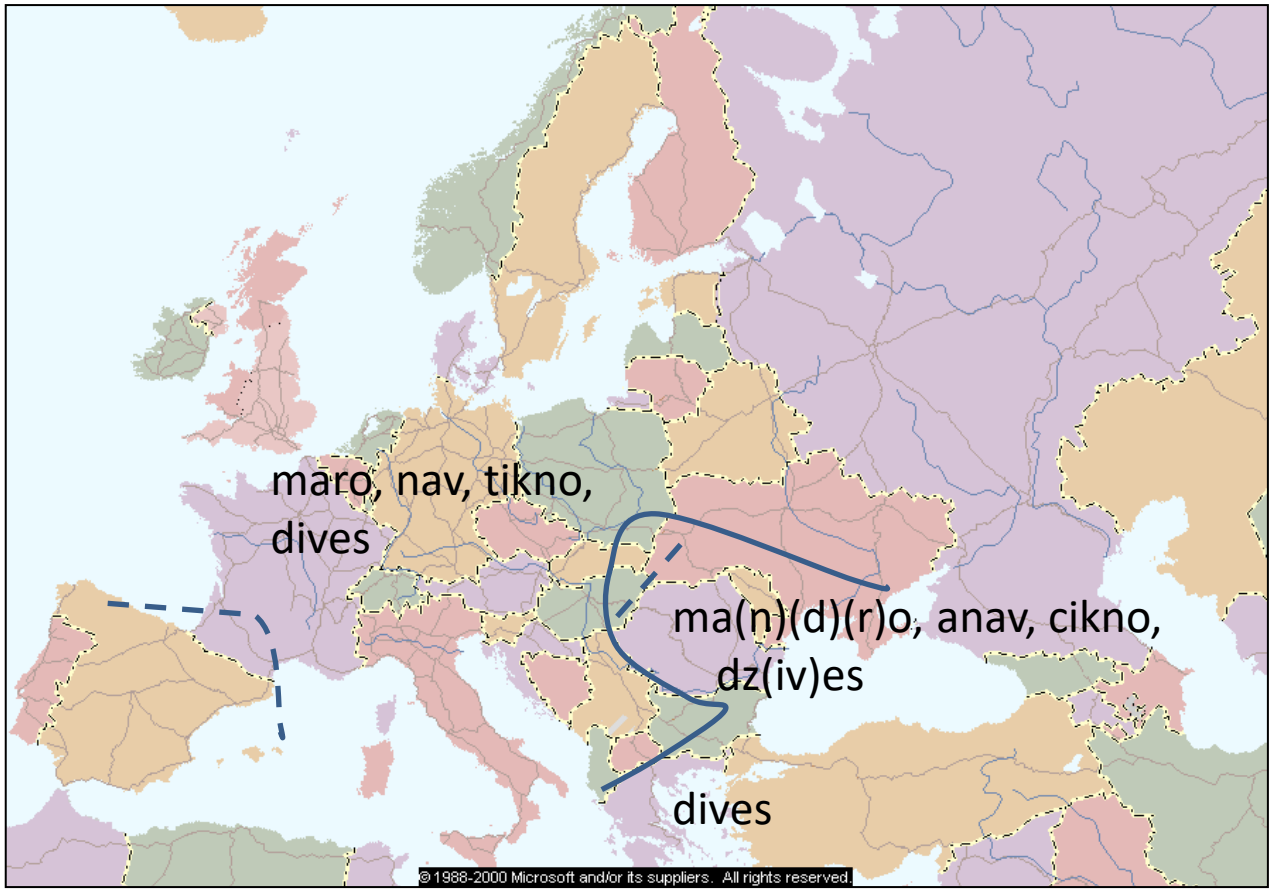
- Another eastern innovation is the prothetic *a-* in lexemes, e.g. *nav* > *anav* 'name', *bjav* > *abijav* 'wedding'. (Matras 2005: 16)

- Demonstrative pronoun stem *kad-* is common to northern Vlax dialects of Transylvania, East Slovak Romani, as well as Drindari and Burgudži varieties of Bulgaria

- Archaisms in the East:
  - The historical consonant cluster *-ndr-* is found mainly in the South-East Europe, where we find it continued in forms such as *mandro*, *manro* and *marno* 'bread'. Eastern Slovakia is a transitional area, where there is on the one hand *jandro* 'flour', but on the other hand *maro* 'bread'. (Matras 2005: 16)

  - Another archaism is the retention of Early Romani loan verb adaptation markers *-is-(ar -) /-iz- ,* found in Vlax group, Black Sea coast, but also in the periferal areas of Iberia and Wales. (Matras 2005: 16-17).

Great Divide (9)



South-East division (Matras 2005: 17.)

Early Romani option selections (1)

- We already mentioned some variation that existed already in Early Romani, and continued as options selection in the current dialects

- One of these concerns the *s~h* variation:
  - Copula verb forms in *s-/h-* 'to be',
  - Inter-vocalic position in grammatical markers (instrumental case, 2SG and 1PL present tense verb endings, imperfective marker)
  - Interrogative pronouns .

- The appearance of one of the forms over another is not a genetic trait that was passed to current dialects before leaving the Balkans

- The current variation rather reflects the selecetion of one option over another, which happened *in situ*

Early Romani option selections (2)

- Dialects with intervocalic and copula -h- are Sinti, Finnish, Central, Slovene, Gurbet (south Vlax), and some Arli varieties, transitional Vlax and Central varietis (such as Cerhari and Gurvari), and some fringe dialects, like in northern Greece.

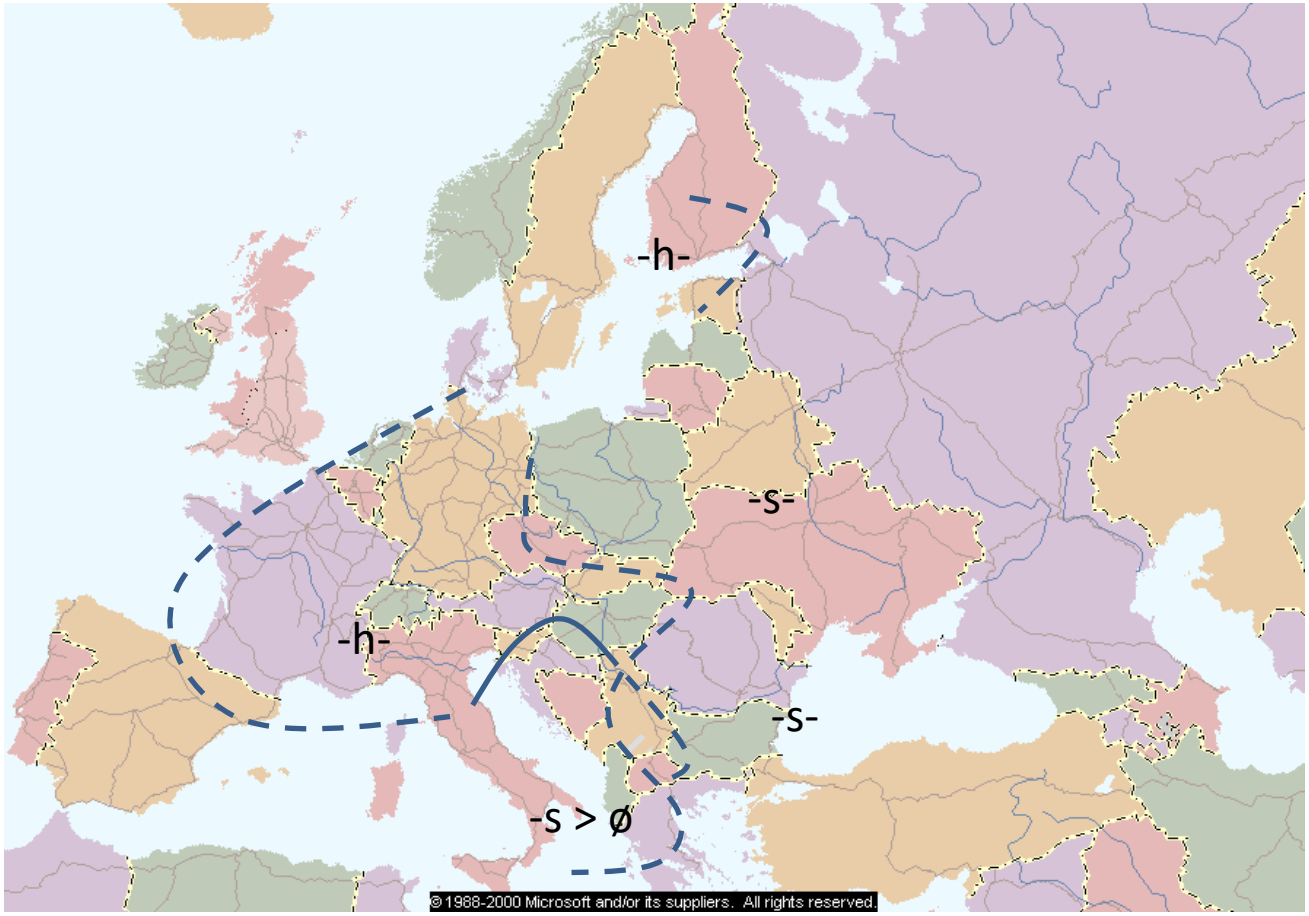
- The process affects partially some Kelderash varieties in Serbia and Transylvania.

- A more straightforward process is the loss of –s in final position in the zone of eastern Adriatic coast – Deljenski, Southern Central, southwestern Vlax and Arli.

- Remarkably similar to the isogloss representing -h- in internal position, is the isogloss representing differentiation b/w perfective concord markers of 2SG (-al) and 2PL (-an), both also used in present copula
  - Outside the zone both persons merge in-*an* by analogy. This is widespread and not contained in a single area (Matras 2005: 18–19.)

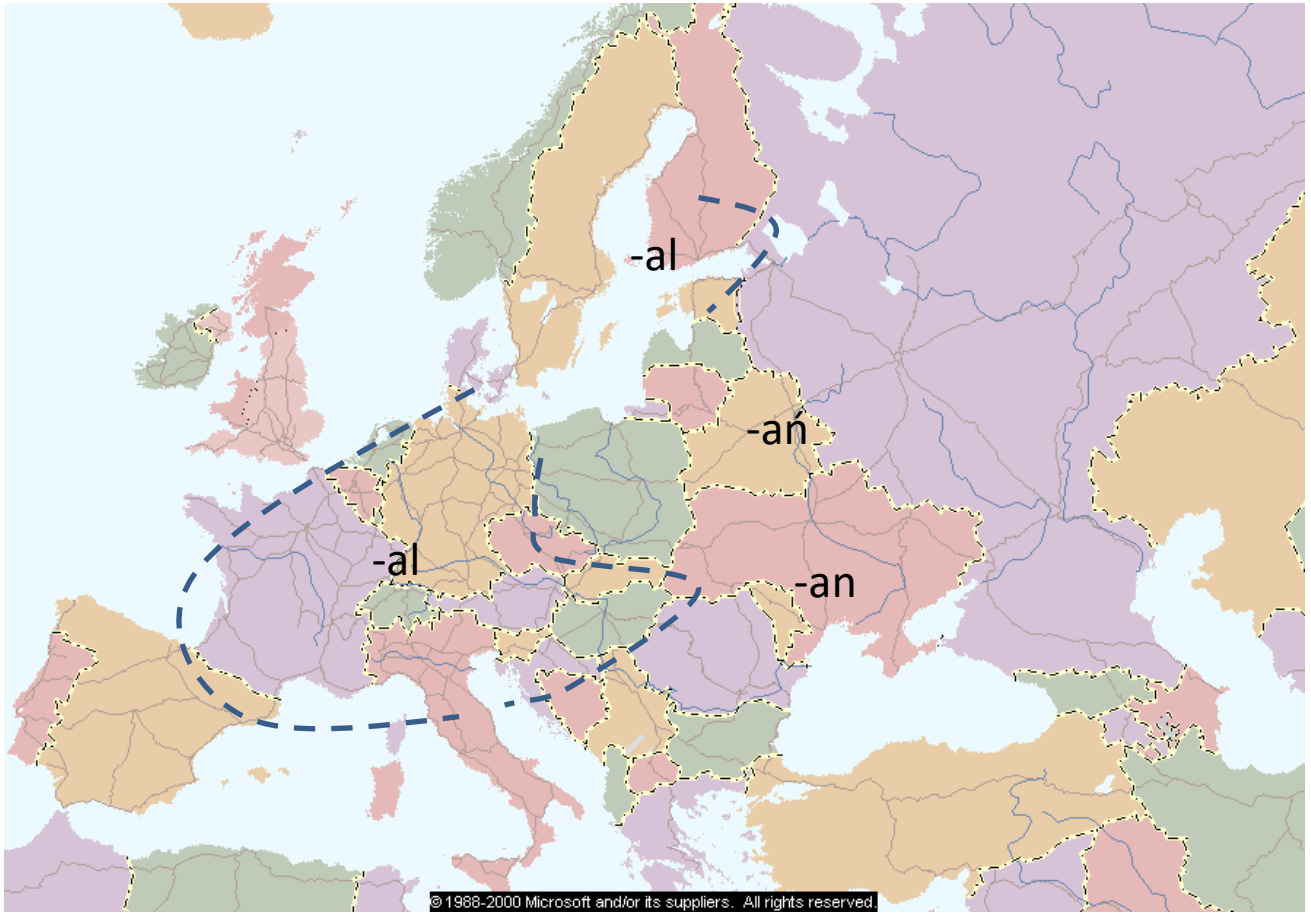
- This suggests that the –an/-al variation is also the result of option selection, with both options present already at the Early Romani stage

Early Romani option selections (3)



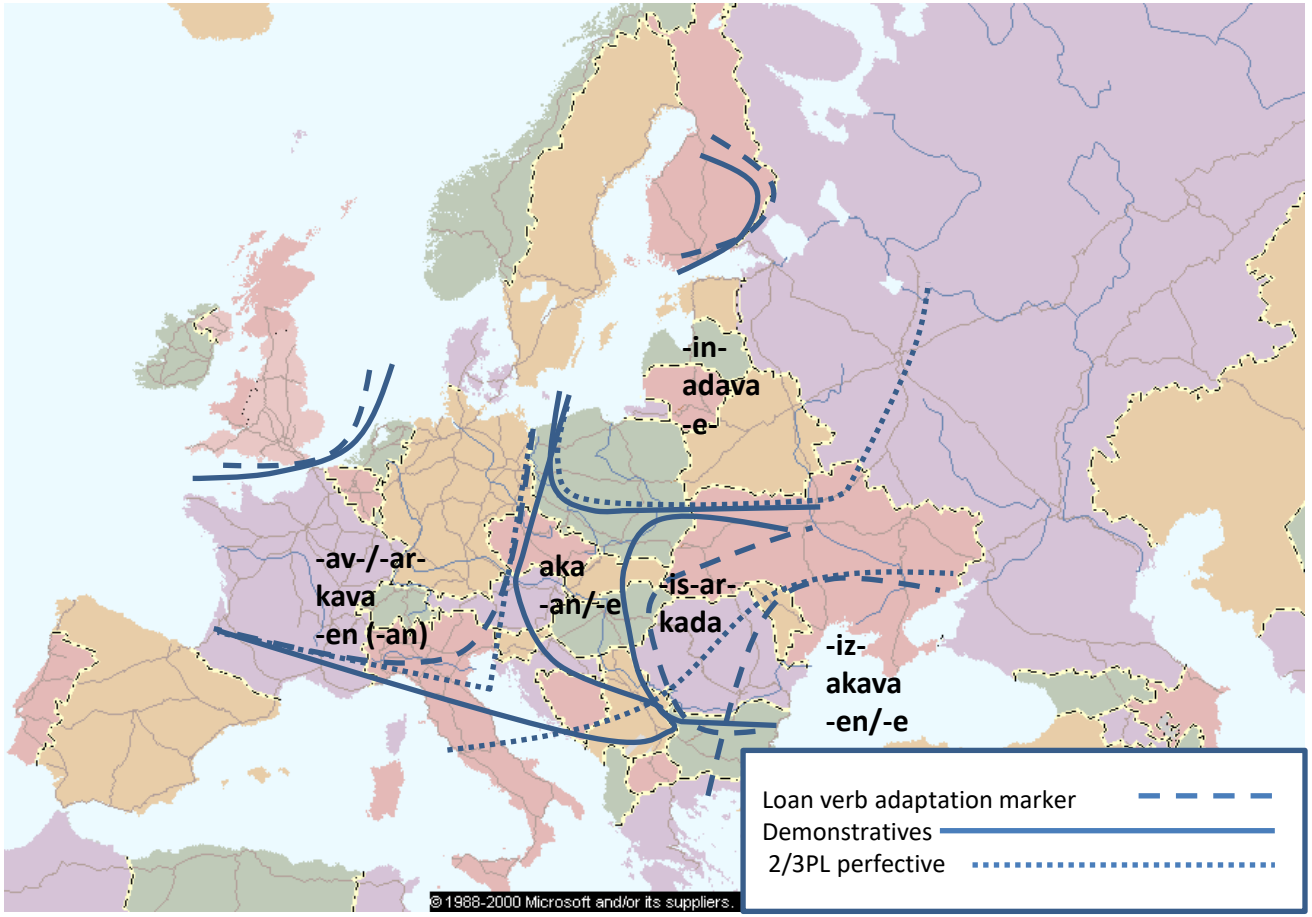
-h-option (Matras 2005: 18).

Early Romani option selections (4)



Preterit 2nd person forms(Matras 2005: 19).

Complex morphological paradigms



The consensus classification is more relevant with restructuring of complex morphological paradigms (Matras 2005: 24.)

Genetic classification of Romani (1)

- European Romani dialects have traditionally been classified using the genetic model; the same model is used when reconstructing language families. (Balto-Slavic -> Slavic -> east Slavic -> Ukrainian)

- During the Byzantine time, the Proto-Romani variety gave rise to several branches, that were further differentiated after moving to the Balkans.

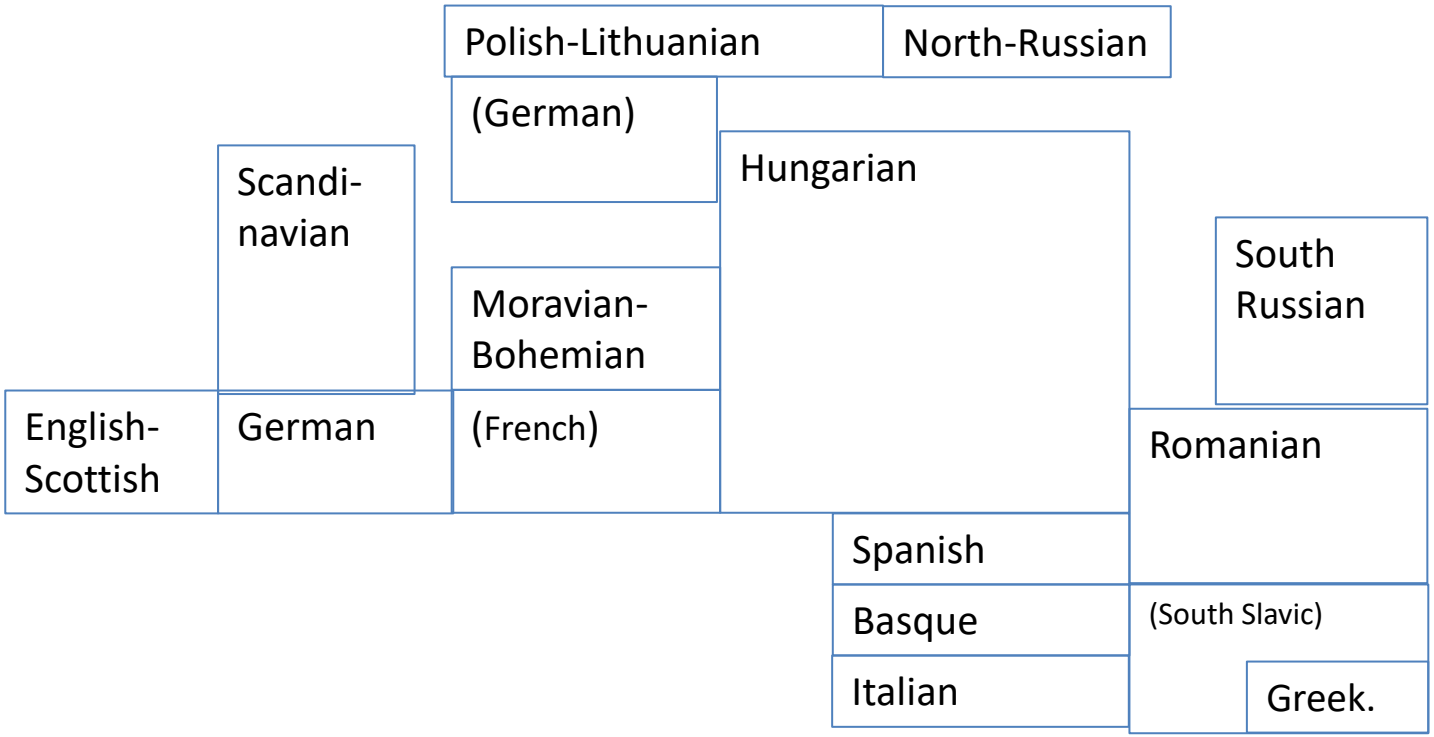
- Each current dialect is considered to belong to one or another genetic branch that came out of the Proto variety.

- The more two dialects resemble each other, the closer they are to each other genetically.

Genetic classification of Romani (2)

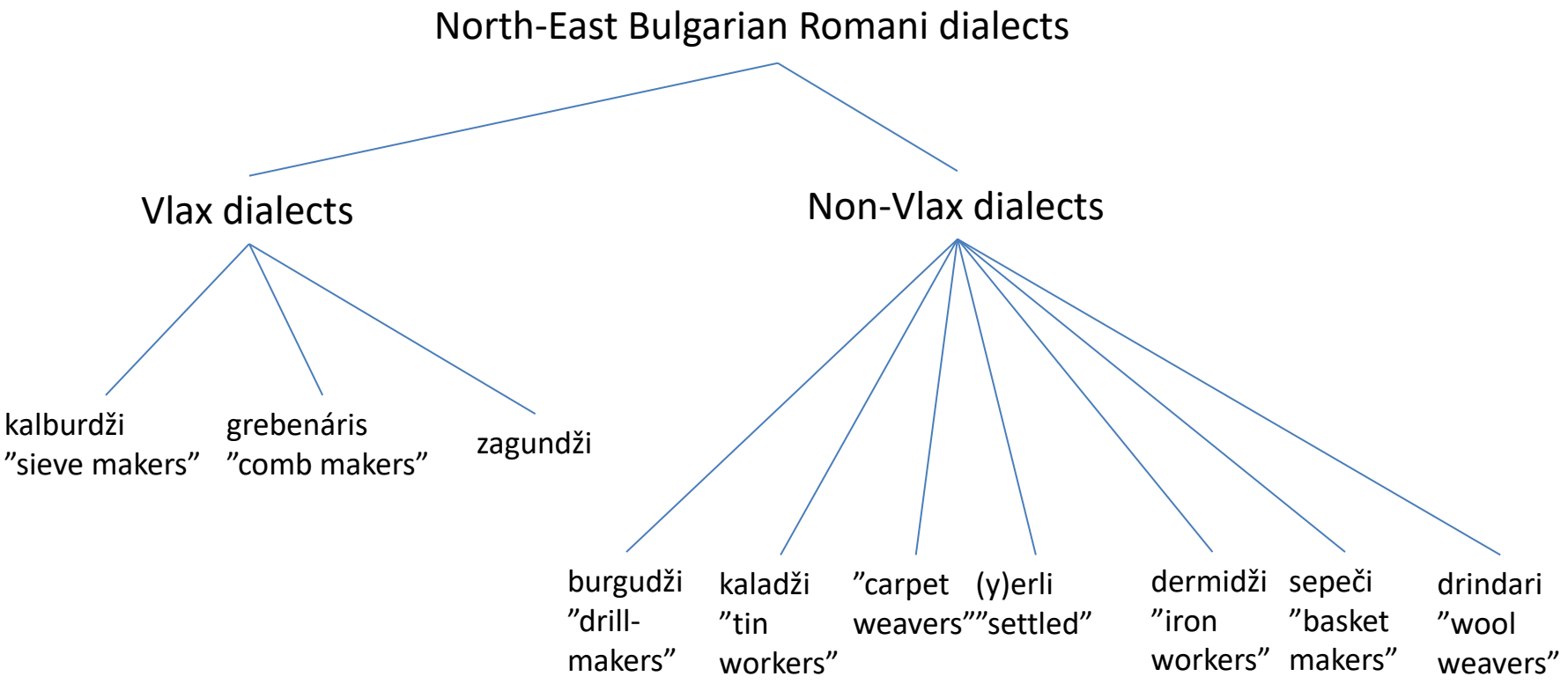
- Romani genetic dialect classification is based on the pioneering comparative and historical work of **Franz Miklosich** (1813-1891), a Slovenian linguist
- **Bernard Gilliat-Smith** (1883-1973) – Vlach- non-Vlach
- **Jan Kochanowski** (1920-2007) - attempted (1963-64) to synthesize the classifications by Gilliat-Smith and Miklosich
- **Tatjana Ventcel' and Lev Čerenkov** (1976) suggested a classification with 8 separate dialect groups
- **Terrence Kaufman** (1979: 134) considered about 60 dialects, and proposed a three-way division of the main dialects (Balkan, Northern, Vlach)
- **Norbert Boretzky** and **Birgit Igla's** (1991) – Vlach vs. non-Vlach

Genetic classification of Romani (3)

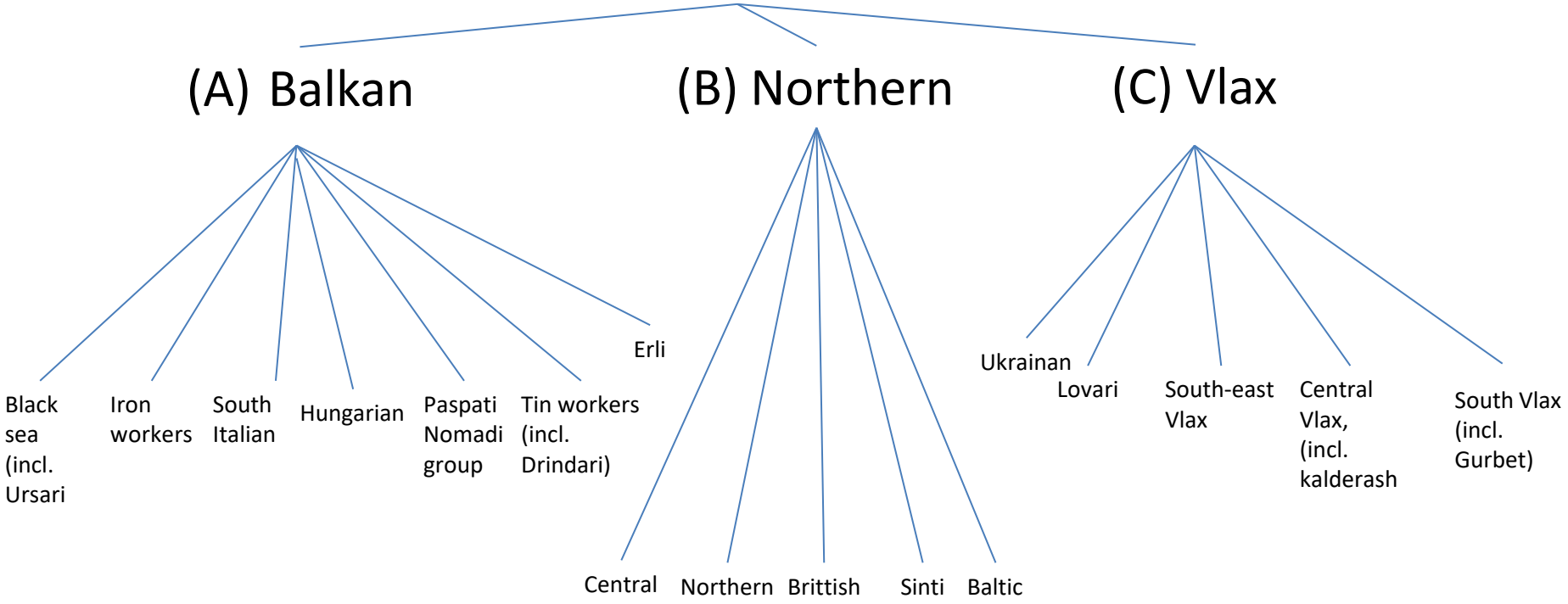


Miklosich (1872-1880, III). 13 dialects, based on contact influence

Genetic classification of Romani (4)



Genetic classification of Romani (5)



Kaufmann (1979) 3-way classification

Genetic classification of Romani (6)

- The division of dialects into Vlax vs. non-Vlax was finally dropped in the 1990's.

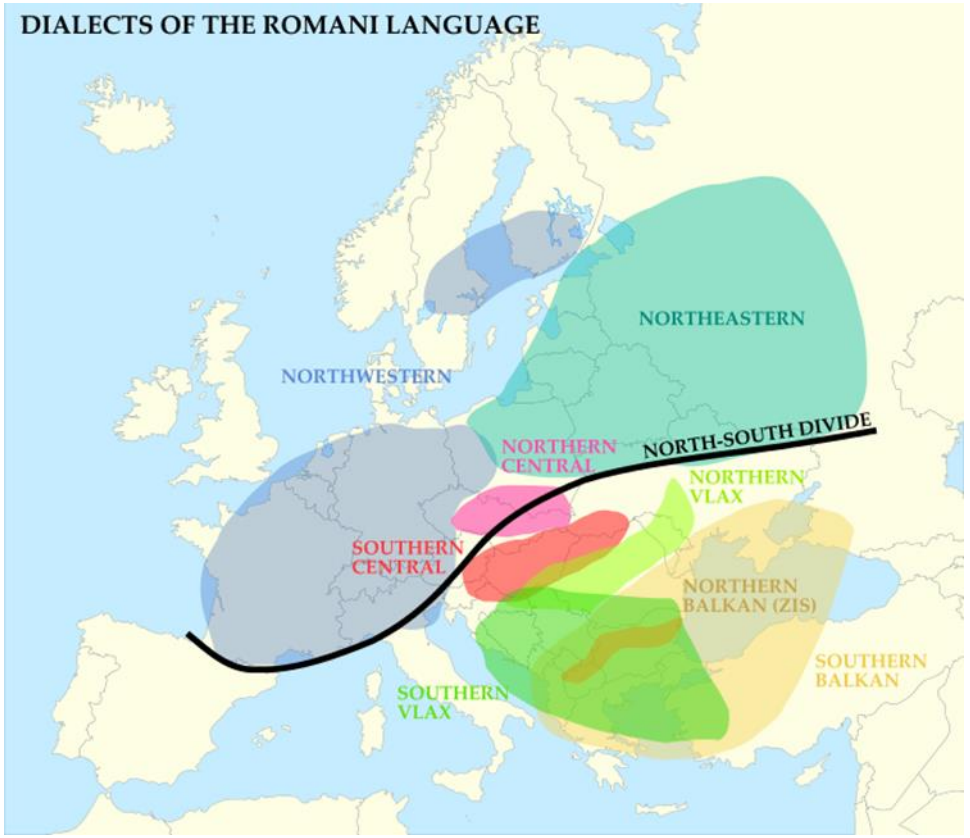
- It gave way to a 4-branch system, each with equal hierarchical status:
  - **Balkan,**
  - **Vlax,**
  - **Northern,**
  - **Central (Bakker & Matras 1997).**

- While this 4-way division is somewhat intuitive, there were later attempts to connect specific linguistic features to these groups.

- E.g., **Peter Bakker** (1999) listed a series of features, which he argues are genetic characteristics of the Northern branch.

Genetic classification of Romani (7)

- **A consensus classification grid**:
  - Differentiates around 4-5 principal divisions among dialect groups, with further sub-divisions:
    - 1) Balkan Romani;
    - 2) Vlach;
    - 3) Central Romani;
    - 4) Northern dialects.
    - Seen as hierarchically equal dialects groups (Bakker & Matras 1997).



http://en.wikipedia.org/wiki/File:Romany_dialects_Europe.svg.

Genetic classification of Romani (8)

- **Boretzky** (1999a, 1999b, 2000b, 2000c) took stock of various characteristics that are shared within each of the 4 branches

- He also further divides 3 of the groups into 2 subgroups each:
  - Central branch is divided into North Central and South Central,
    - The South Central is further divided into Vend-type and Romungro-type.
  - Vlax branch is divided into South Vlax and North Vlax
    - South Vlax is further divided into South-east and South-west.
  - Balkan branch is divided into South Balkan I and South Balkan II ('south', because Vlax is also spoken in the Balkans)
    - South Balkan I has Arli, Erli, Rumelina, Crimean etc.
    - South Balkan II has Drindari, Kalajdži and Burgudži).

Geographical diffusion model

- A competing, **geographical diffusion model** has been proposed by Matras (2002, 2005).
  - variation among dialects is subject to the geographical continuum and is best accounted for by assuming linguistic contact between the speakers of adjacent territories.
  - This model has been applied to the Northeastern group in Tenser (2008), and to Finnish Romani by Granqvist (2014).

- Centre-periphery dynamics - Northern Romani dialects

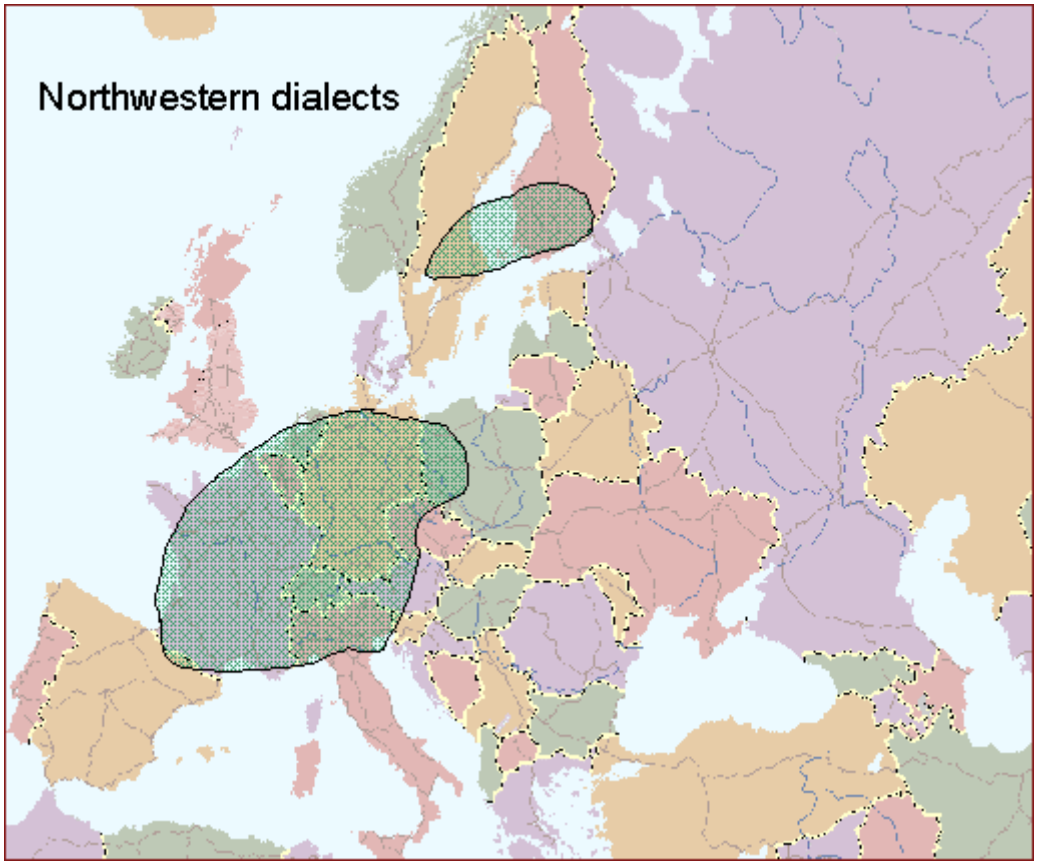
Centre-periphery dynamics - Northern Romani dialects

- **Bakker (1999):**
  - Proposed the term Northern metagroup, which was futher divided into Northwestern and Northeastern.
    - Northwestern subgroup included Finnish Romani, German Sinti, Spanish Calo, Welsh and English Romani, as well as some varieties spoken in Italy.
    - Northeastern included dialects of the Baltics, North Russia, some varieties of Poland
  - In this classification 'Northern' does not refer strictly to a geographical North
  - It is a reference to a genetic group, based on diagnostic linguistic features, which for the Northern group include:

Centre-periphery dynamics - Northern Romani dialects

1. Prothetic **j** in 3rd person pronouns **jov/joj/jon**- ’he/she/they’
2. Specific quantitative pronouns **čiči** ‘nothing’, **čimoni** ‘some’, **kuti** ‘a little’, **keti** ‘how much’ (Finn. rom,. Buutko? Sar buut?)
3. Location diectics **adaj/adoj/akaj** (vs. **kate/kote**, **katka/kotka**).
4. Demonstrative pronouns endings -**va**/-**ja**/-**la** (mask./fem./pl.) vs. -**o**/-**i**/-**la**.
5. Long forms of genitive marker: -**kir**- vs. -**k**-.
6. Negation adverbial **kek** ’no a single one’.
7. Prothetic **v** in **vaver** ‘other’.
8. **r** omission in **korkoro** > **kokoro** ‘alone’.
9. **vr** > **r** in **vraker**- > **raker**- ‘talk’.

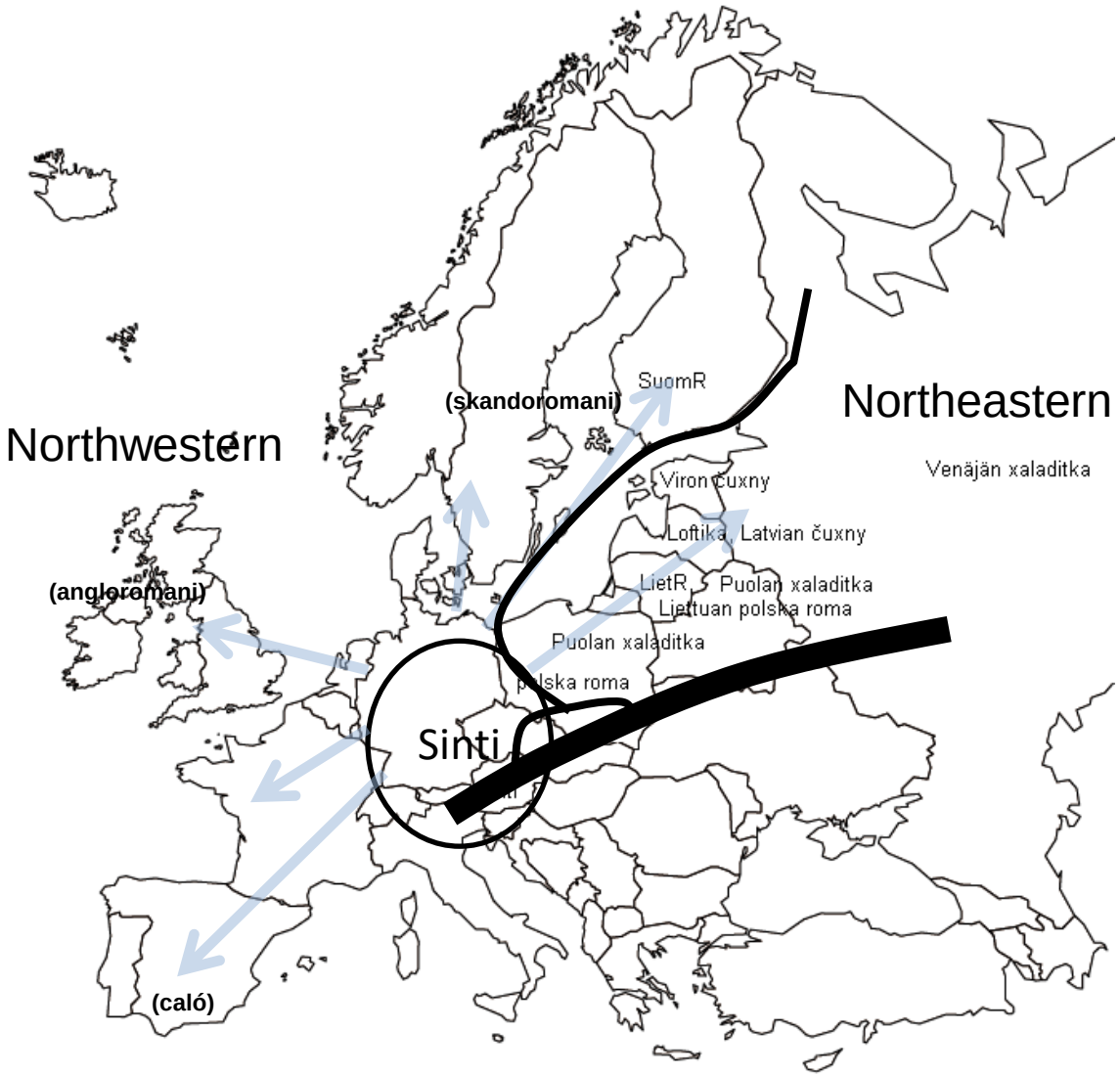
Northwestern



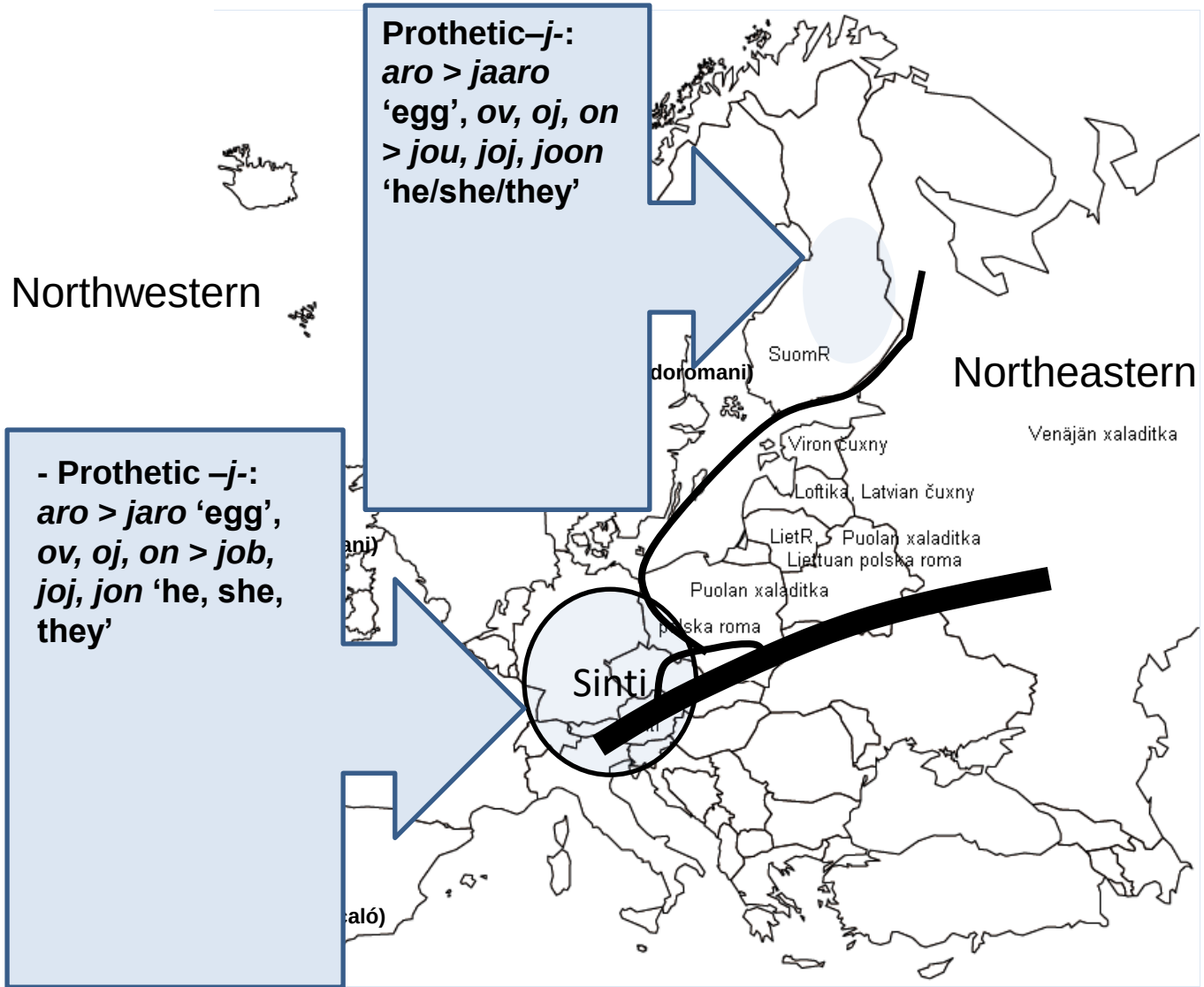
(http://romani.humanities.manchester.ac.uk/wh
atis/classification/dialect_classify.shtml)

Northern Dialects

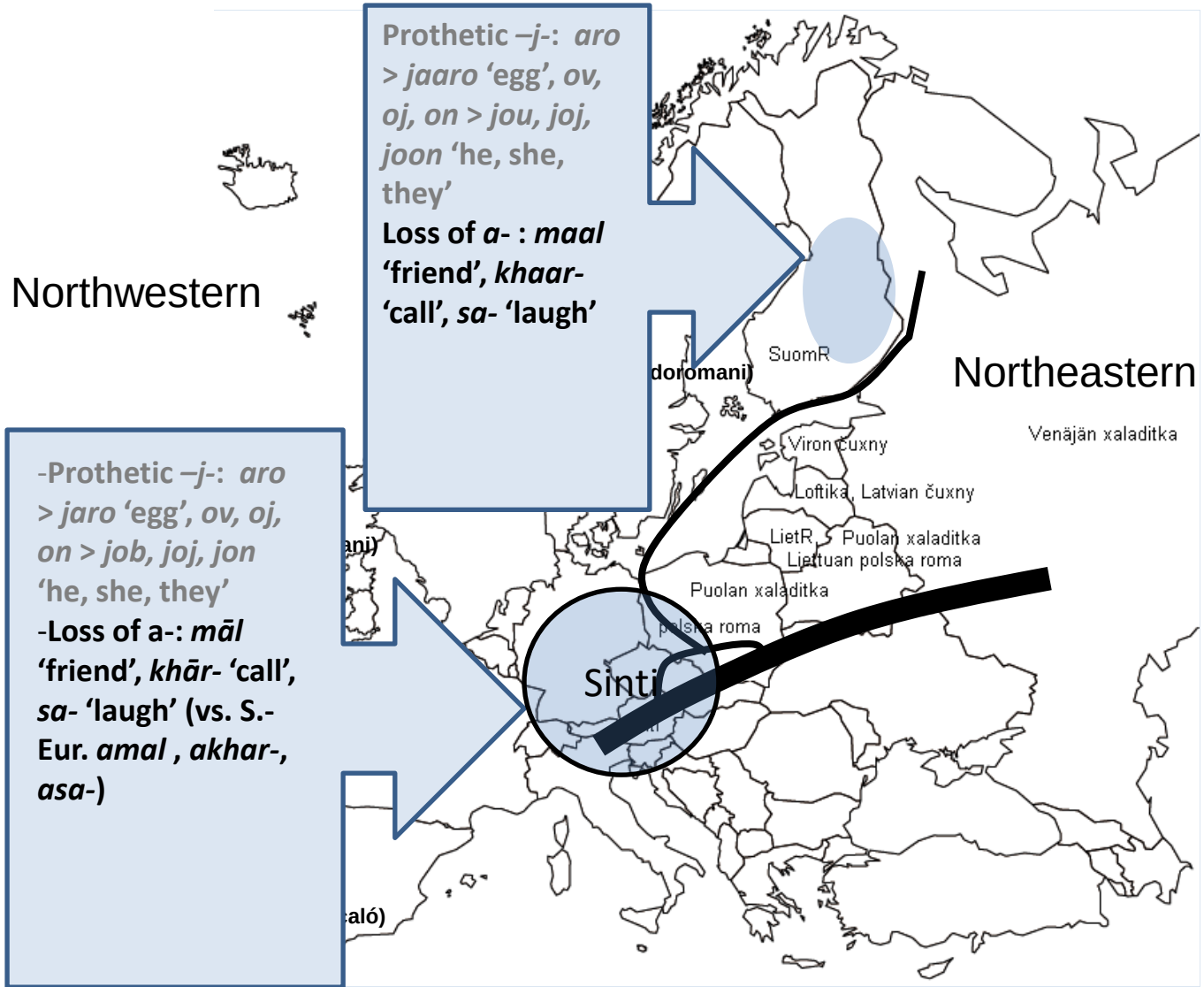
German Sinti Center of Innovation



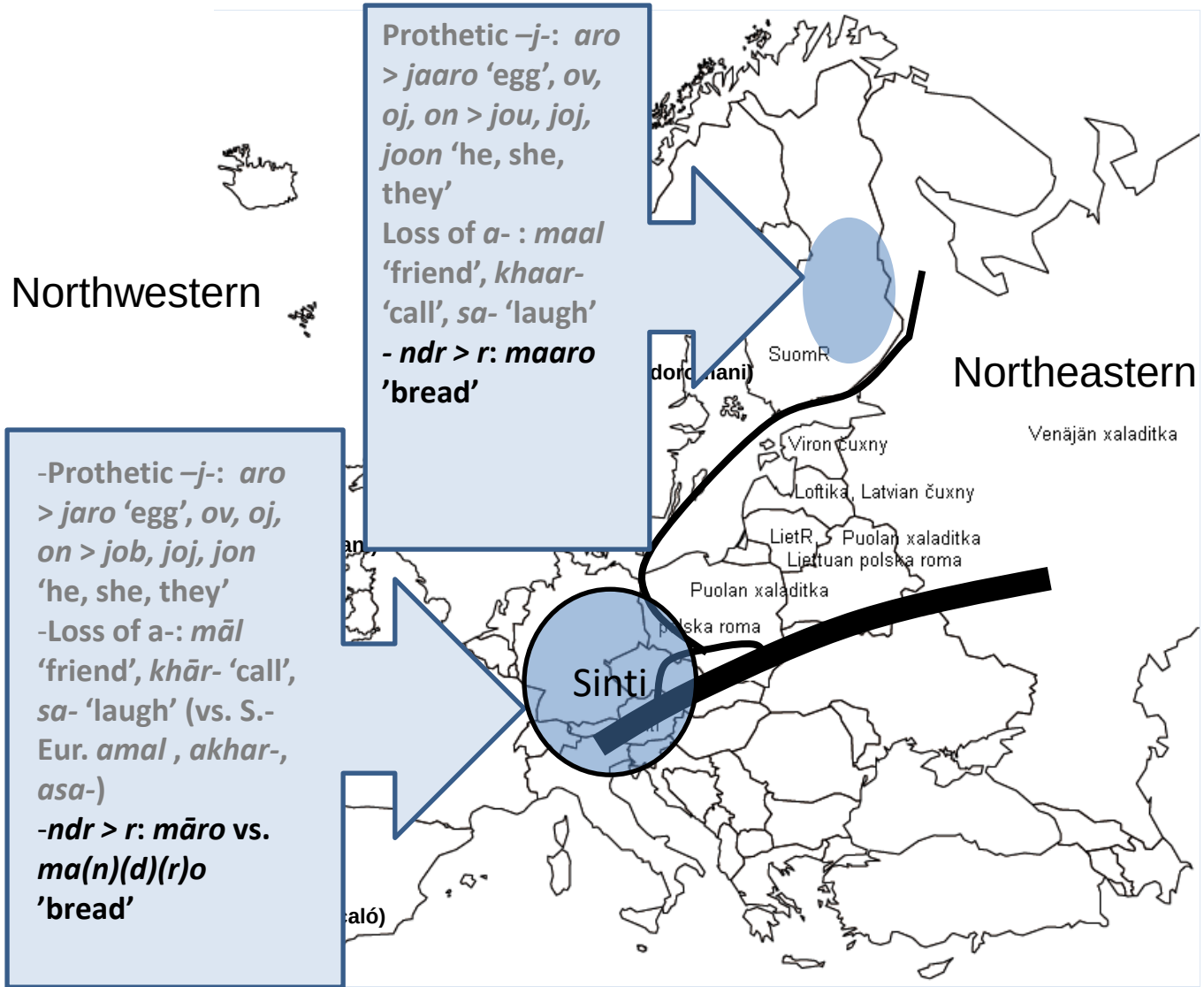
German Sinti Innovation Center– Finnish Romani



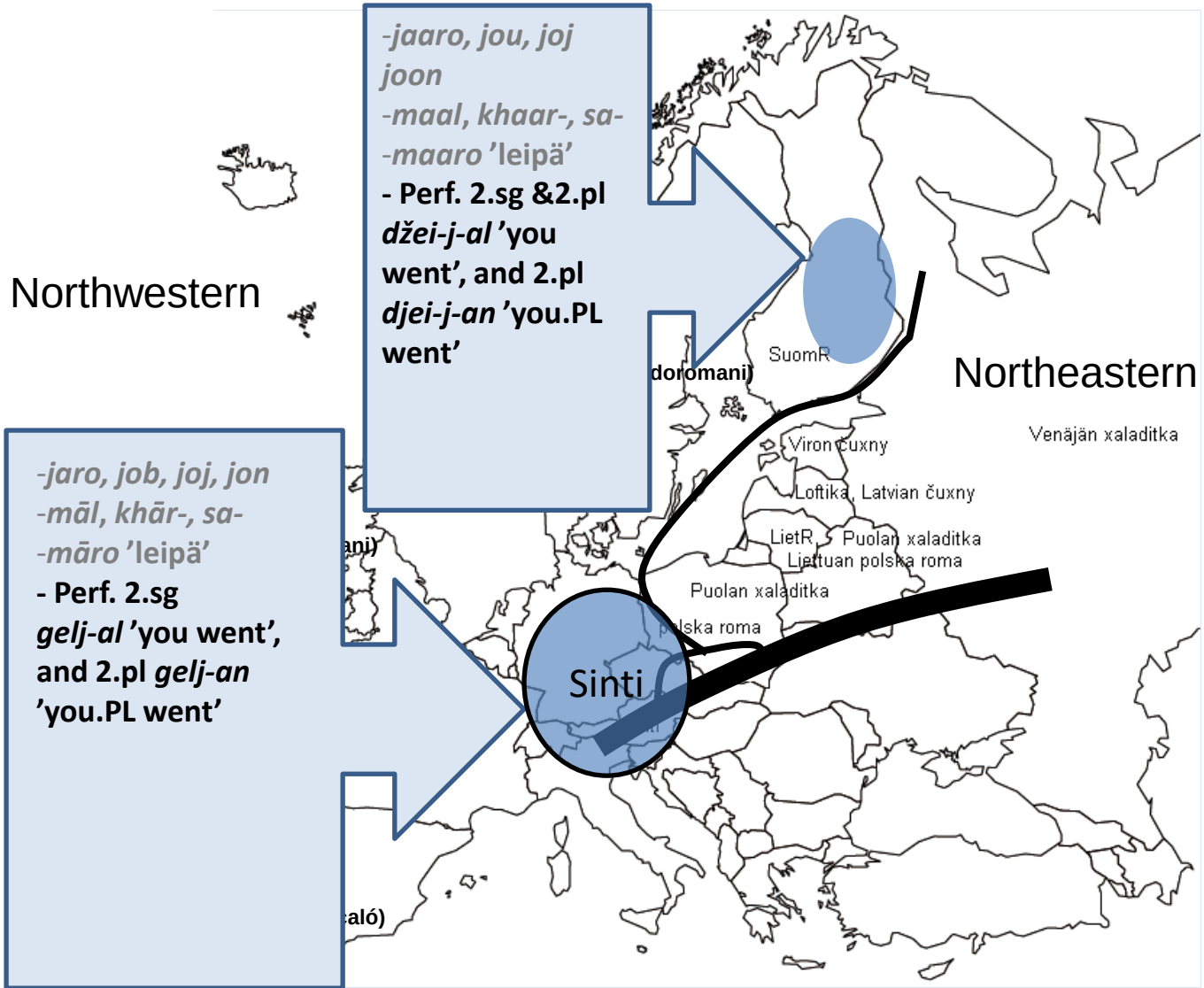
German Sinti Innovation Center– Finnish Romani



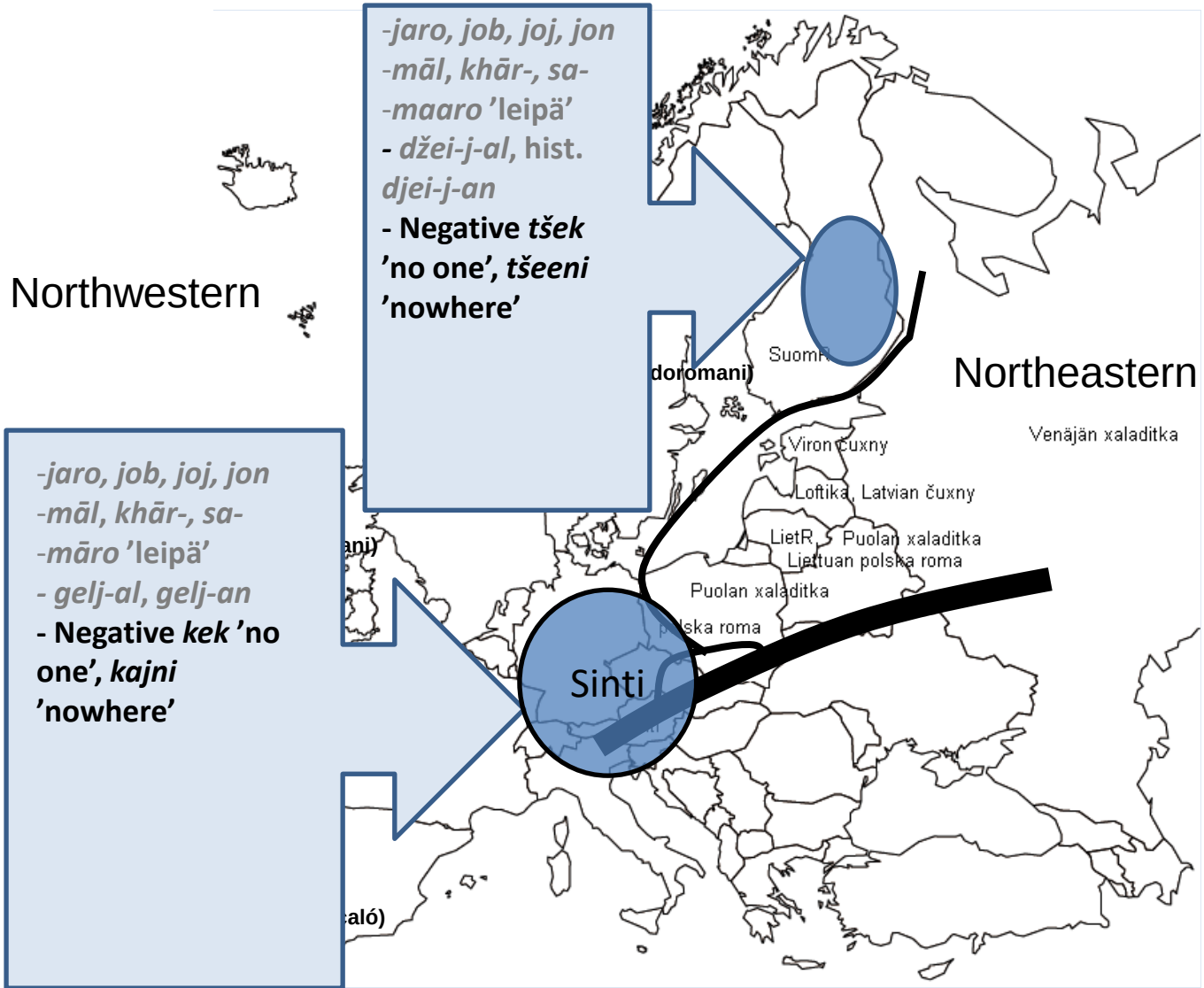
German Sinti Innovation Center– Finnish Romani



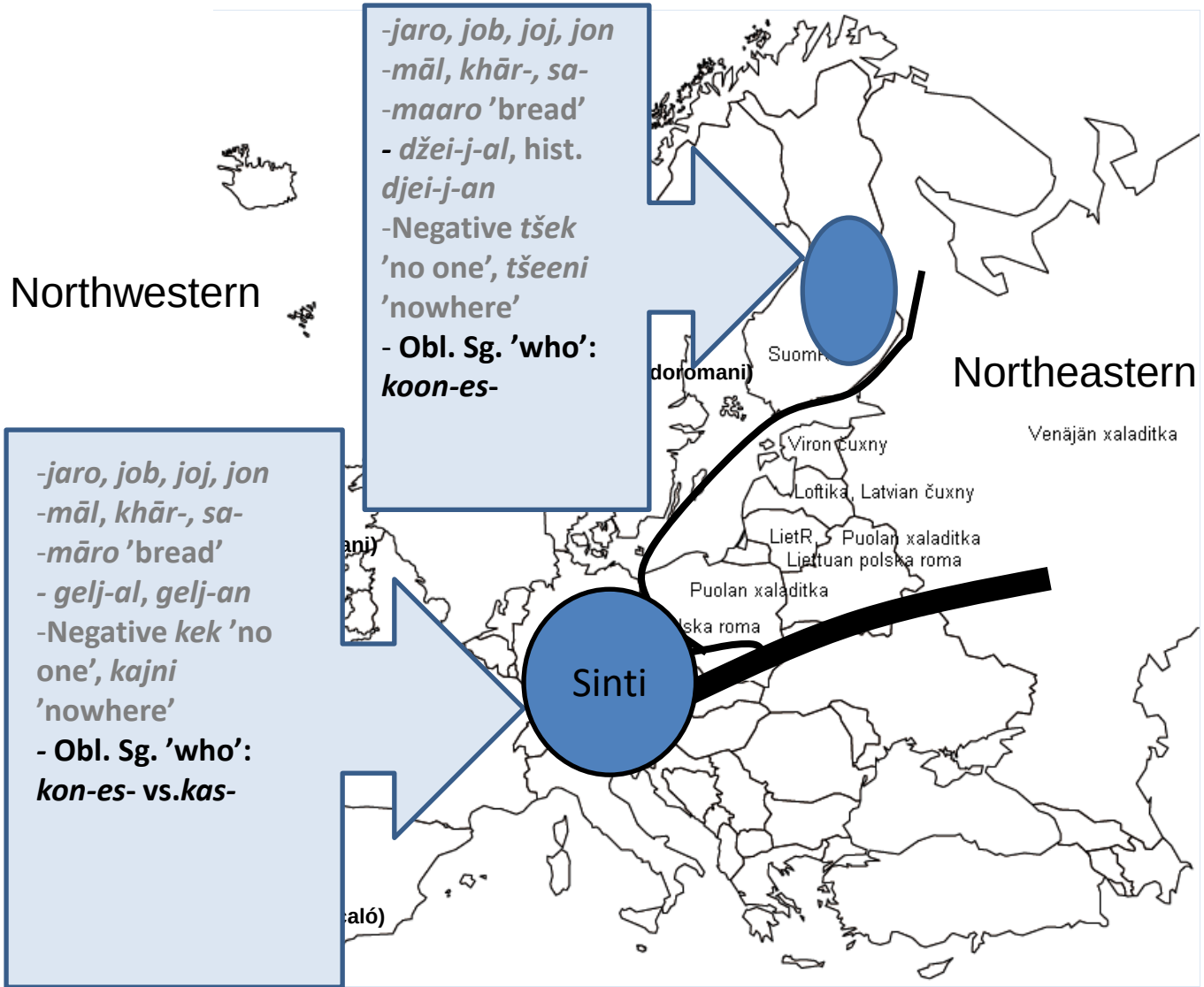
German Sinti Innovation Center– Finnish Romani



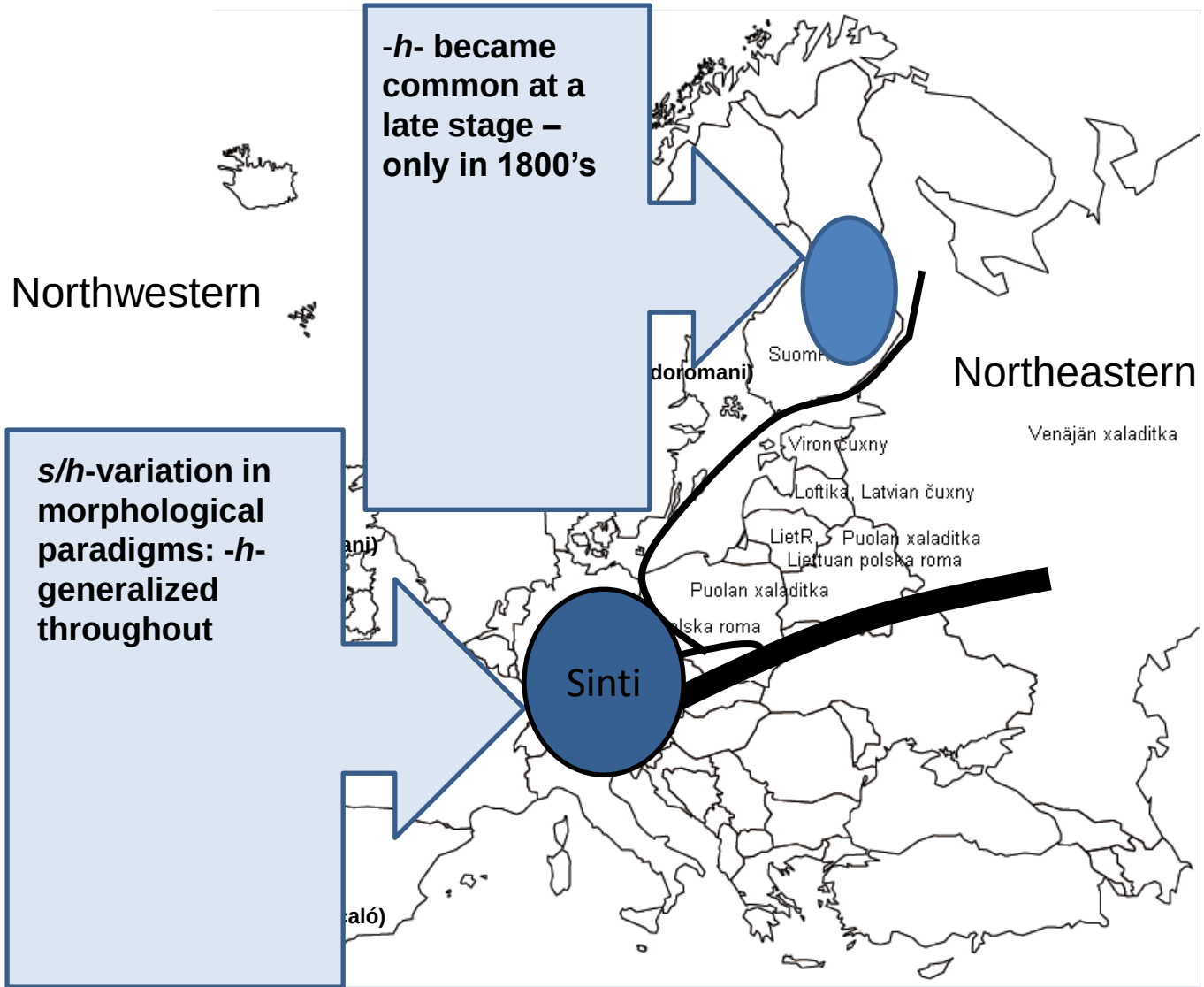
German Sinti Innovation Center– Finnish Romani



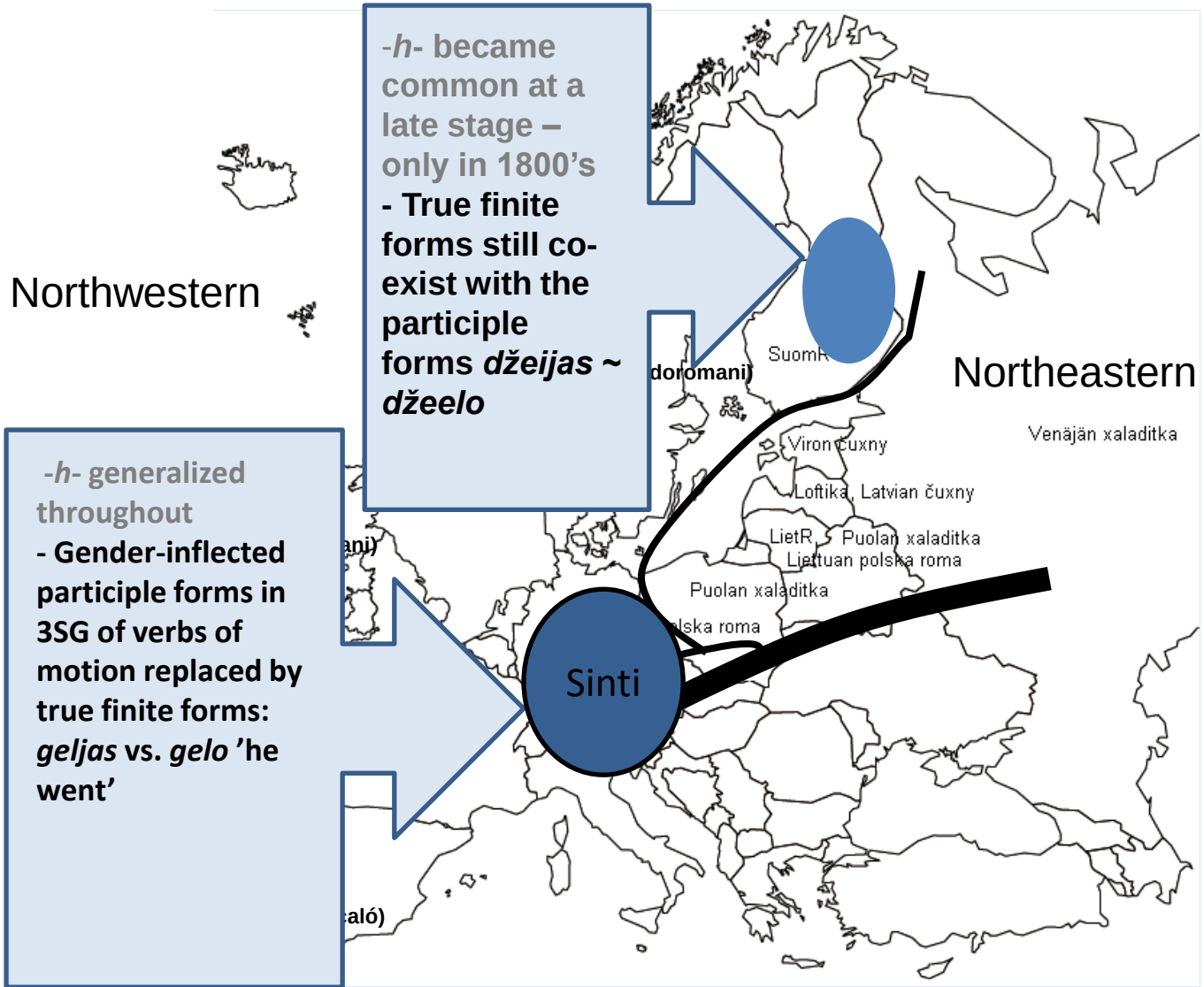
German Sinti Innovation Center– Finnish Romani



German Sinti Innovation Center– Finnish Romani as periphery



German Sinti Innovation Center– Finnish Romani as periphery



Participle vs. Regular Perfective 3SG

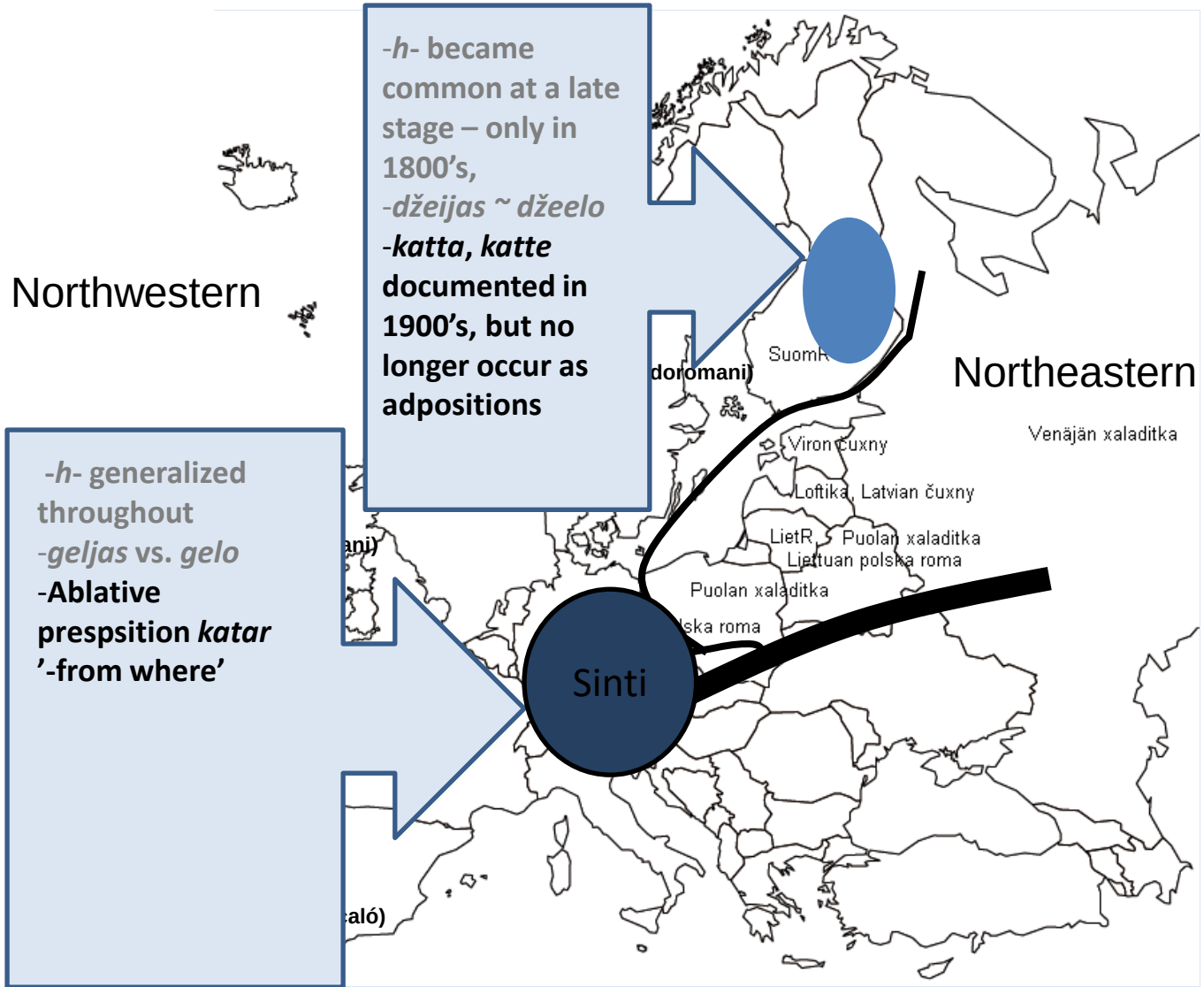
- Finite form in **–as** always possible.
- Hierarchy: motion verbs> **meer-** ‘die’ > other verbs
- Sources:
  - Spoken (2000-2001)
  - Written sources (1960s-**2001**)

Verb	Past participle	Non-participle
(a)v- come’	345 (**aulo**: 330, **auli**: 2, **veelo** 13)	9 (**aujas**: 6, **avjas** 1, **veijas** 1)
dža- ’go’	217 (**džeelo**: 216, **džeeli**: 2)	1 (**veijas**)
meer- ’die’	87 (**muulo**: 85, **muuli**: 2)	72 (**muulidas**: 13, **multas**: 1, **merdas**: 58)
d- ’give’	5 (**diilo**)	138 (**diijas**)
l- ’take’	2 (**liilo**)	112 (**liijas**)
pi- ’drink’	1 (**piilo**)	6 (**piijas**)

Participle vs. Regular Perfective 3SG

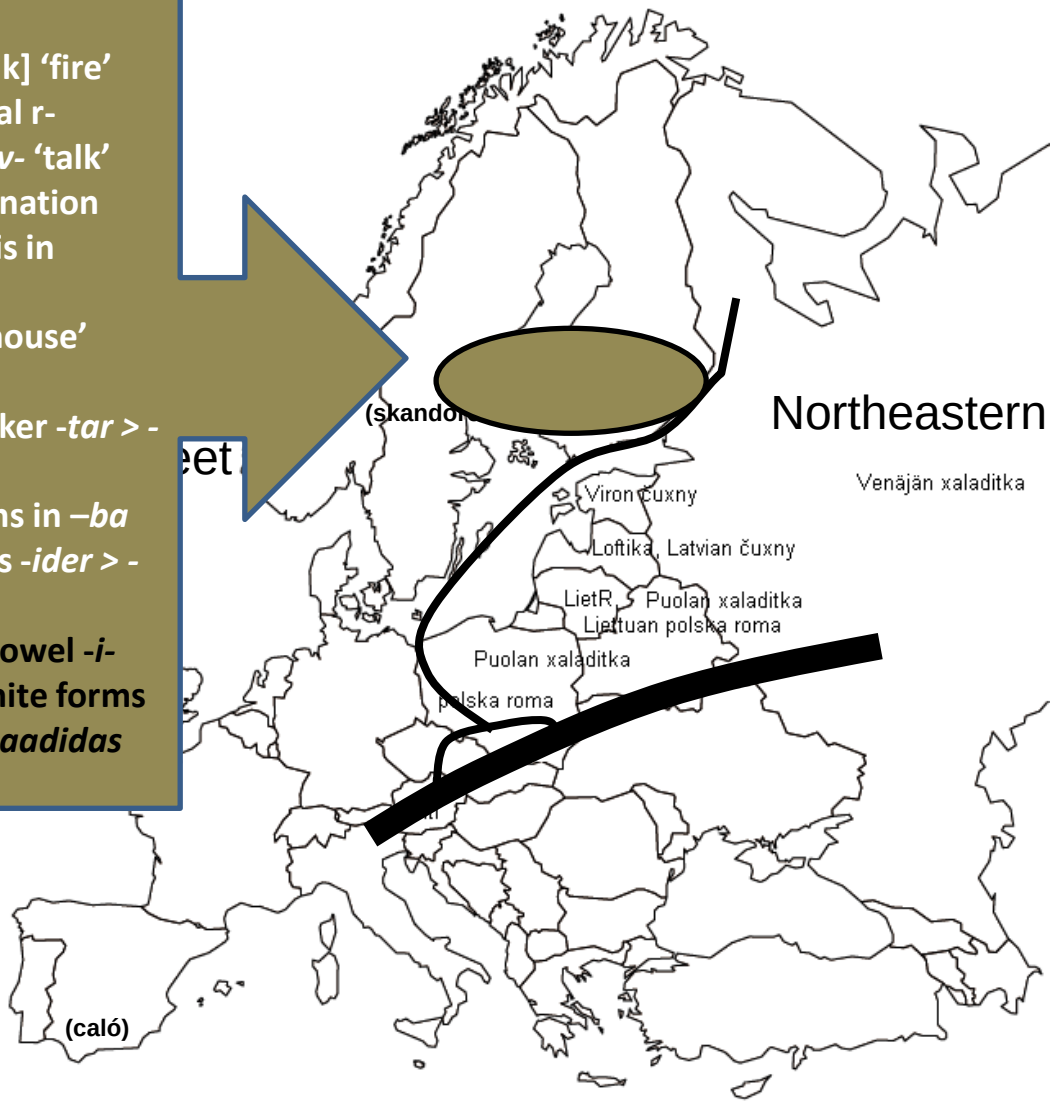
- It is unusual for Northern dialects to have participle forms for 3SG perfective verbs (e.g. **gelo/geli** ‘he/she went’) (Matras 2005: 15)
- In Finnish Romani this exists as an option:
  - Verbs of motion (**aulo** ‘(he) came’ (sometimes **veelo**) and **džeelo** ‘(he) went’)
  - A number of other verbs (**diilo** ‘gave’, **liilo** ‘received’, **piilo** ’drunk’, **muulo** ‘died’),
  - Forms **aahto** ‘was’, **behto** ‘sat’ ja **nahto** ‘fled’ are only used as adnominals, and not perfective verbs
  - **pelo** ‘fallen’, **rundlo** ‘cried’ and **suto** ‘slept’ are not used
  - **Hierarchy: verbs of motion > meer- ‘die’ > other verbs**

German Sinti Innovation Center– Finnish Romani as periphery



Innovations in Finnish Romani

- Pre-nasalisation of voiced stops
jag > jang [jaŋk] ‘fire’
- Word initial r-
rakkav-/akkav- ‘talk’
- *kh*:n affrikatination and metathesis in ‘house’
kher > škeer ‘house’ idiolects.
- Ablative marker *-tar > -ta*
-Abstract nouns in *–ba*
- Comparatives *-ider > -ide*
- Connecting vowel *-i-* occurs with finite forms in preterite: *traadidas*



- Sociolinguistics

Sociolinguistics

- Typically and historically a means of oral communication and interaction (Matras 1999: 482).
- Romani is mainly used for interaction with family members, friends and acquaintances Identity or boundary marker, emblem
- Secret language

Sociolinguistics

- Most Roma are multilingual
- Many migrant Roma also have an extensive linguistic repertoire and are proficient not only in various Roma dialects, but also in national and minority languages in their countries of origin
  - such as Albanian, Czech, Finnish, Polish, Serbian, Slovak, Turkish, and in the various variants spoken in the host country.
  - Czech Lovari have said that in their everyday life in Sweden they use Czech, Slovak, Hungarian, Romani, English and Swedish;
  - Arli-Roma from Macedonia again Macedonian, Serbian, Romani, English and Swedish;
  - Ursari från Rumänien använder rumänska, romani, engelska och svenska.

Sociolinguistics

- Swedish example:
  - the language choice within the family seems to vary depending on the generation and nature of kinship:
  - most Roma from the Arli, Lovari and Kalderash groups have stated that they speak Romani and sometimes Swedish with their parents,
  - slightly more Swedish with their siblings and partners,
  - either Romani or Swedish with their children and either Romani and / or Swedish with other relatives,
  - National languages of the Roma's countries of origin such as Serbian and Romanian are in use in parallel with Romani and Swedish.

Sociolinguistics

- Not all Romani groups speak Romani:
  - Beaš/ Rudari
  - Aškali
  - Egyptians
- Pararomani:
  - Angloromani
  - Scandoromani
  - Caló, Calão, Erromantxela
- Artificial Romani:
  - Romanó-Caló
  - Kalderaš-mixed Finnish Romani in Sweden

Sociolinguistics

- In recent decades, however, the language has undergone a transformation and also become a written language.
- Roma migration and transnational networks - not least social media - have opened up opportunities to interact with other Roma speakers throughout virtually the world.
- Despite the fact that few Roma read literary texts published in Romani, a large number of novels and poetry have been published in Romani in recent decades
- for example, in Italy, Poland, Romania, Slovakia, the Czech Republic, Hungary, and Russia

Sociolinguistics

- Printed collections of oral folklore have been published in Albania, Austria, Bulgaria, the Czech Republic, Hungary, Macedonia, Norway, Serbia and Ukraine
- Pushkin, Shakespeare, and Tolstoy have been translated from world literature into Romani
- New Testament Romani texts in several dialects of the Romani language

Sociolinguistics

- The Roma greatly influenced the development of the Karagöz Shadow Theater
- There have been Romani-language theater groups, especially in the Soviet Union and the Balkans
- Many professional Roma groups and choirs sing both in Romani and in the languages of their countries of residence
- Developing movie industry
- Romani-language radio and television broadcasts are increasingly available in various countries

Sociolinguistics

- The teaching of the Romani language in many places in Europe has been sporadic and experimental
- In Czech Republic and Slovakia, the Romani language is taught regularly in a few schools
- In Finland, Roma culture and language have been taught more widely in some primary schools since 1989
- University teaching subject: Finland, Sweden, Czech Republic, Romania etc.

Sociolinguistics

- Used in some international political and cultural meetings, sometimes with interpretation
- But: rarely a language of administration

- Constitutional status in several European countries
  - In Finland, the 1995 fundamental rights reform safeguarded
  - In 1999, Sweden recognized it as one of the country's five national minorities and Romani Chib as one of its national minority languages

- Standardisation and codification

Standardisation and codification

- IRU: 1990, ”Cortiade alphabet”- did not succeed to gain ground except Romania and Spain
- Standardisation has failed:
  - Romani society has traditionally lacked a fixed hierarchical, social structure
  - There is therefore no dominant class that has been able to decide for the whole group
  - Relations between different Roma groups are also not the best possible
  - There is no form of spoken language that is acceptable to everyone
  - The geographical spread of Roma groups = > the spread of a standard cannot be controlled by the imposition of sanctions or exclusion measures on those who do not use the correct language.

Standardisation and codification

- The Roma prefer pluralism, that is, written Romani that shows regional codification with a certain international focus.
- A new generation of Roma intellectuals today uses various forms of language, both orally and in writing.
- Meetings with Roma around the world at international conferences, seminars and the like, regular e-mail communication and text messages:
- Roma today encounter a number of different forms of their language.

Standardisation and codification

- The semi-formal, oral, conference communication between Roma has contributed to a semi-formal written communication
- strengthened in digital communication
- Romani is widely used in digital communication, such as in various mailing lists or in social media
- The written exchange of thoughts in Romani is spontaneous and not uncontrolled by spelling rules or other norms or grammatical constraints.

Standardisation and codification

- More successful local standards
  - First attempts in Soviet uninion during Stalin's period
  - Czech Republic, Slovakia, Macedonia, Hungary, Austria, Finland

- Resources

Resources

- RomLex, dictionaries of various dialects:

http://romani.uni-graz.at/romlex/lex.xml

- RMS – morphosyntax, maps

http://romani.humanities.manchester.ac.uk/rms/

- Fact sheets on Roma – language and culture

http://romafacts.uni-graz.at/

Thank you!