Classification of the Belarusian Dialect Continuum by k-Nearest Neighbors

- Chris Marchant
- chris@vitba.org
- 435-213-2241

Timeline of Belarusian Lands

Polatsk, Hrodna, Brest, Vitsebsk, Minsk become vassals of Lithuania
1326
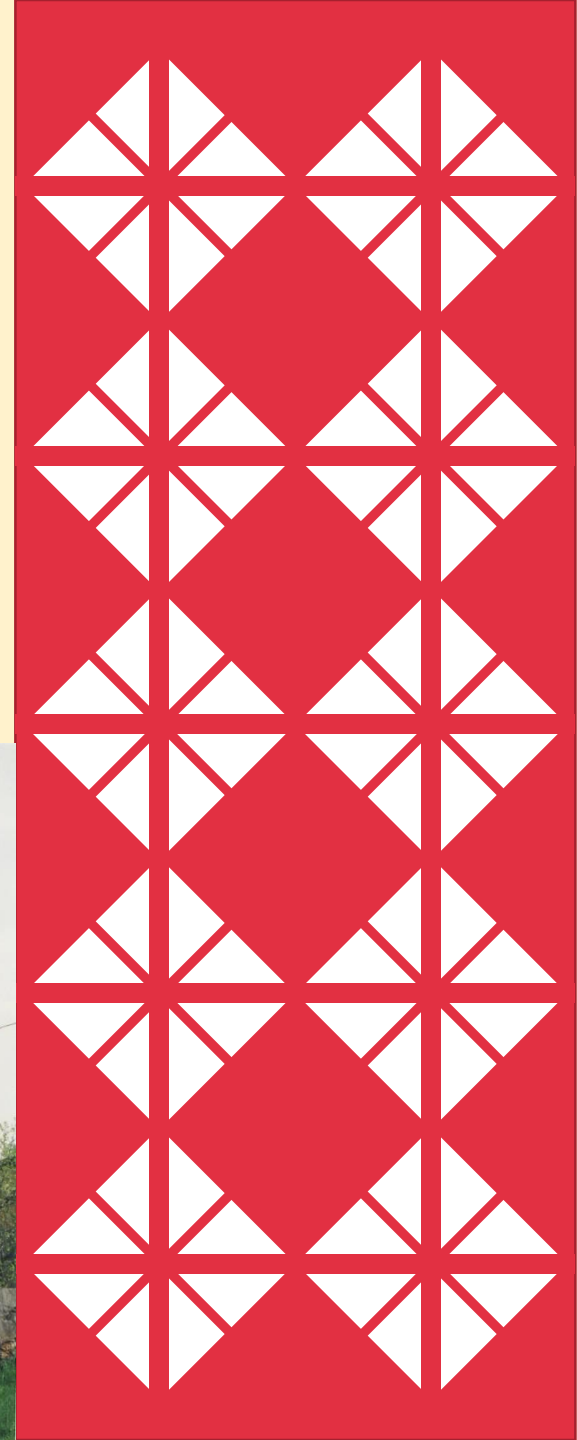
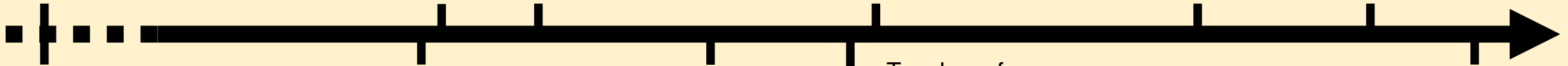Daniil becomes first prince of Moscow
1262

Treaty of Lublin creates Polish-Lithuanian Commonwealth
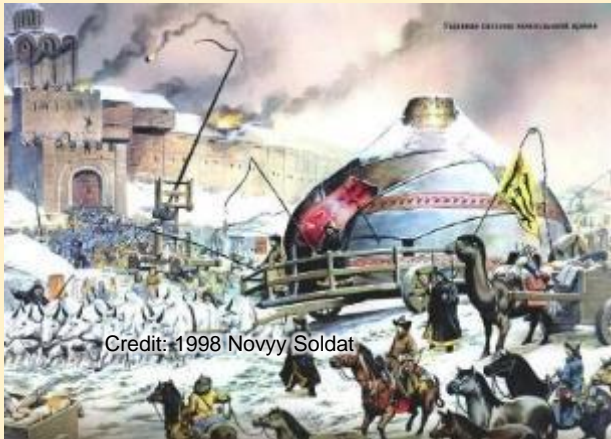**1569**

Russia annexes Lithuanian half of Commonwealth
1795

Declaration of Independence by Belarusian National Republic
1918

978

Vladimir I slays Rogvolod and conquers Principality of Polatsk

1240
Siege of Kiev

1410
Battle of Grunwald

1547 Tsardom of Russia declared

1991
Belovezha Accords dissolve Soviet Union

Credit: 1998 Novyy Soldat

Credit: Rada BNR

Timeline of Belarusian Lands

Credit: Koryakov Yuri, WikimediaCommons

Principality of
Polatsk
~1100

Credit: M.K., Wikimedia Commons

Grand Duchy of
Lithuania
~1430

Credit: Halibutt, Wikimedia Commons

Polish Lithuanian
Commonwealth
1772

Belarusian
National Republic
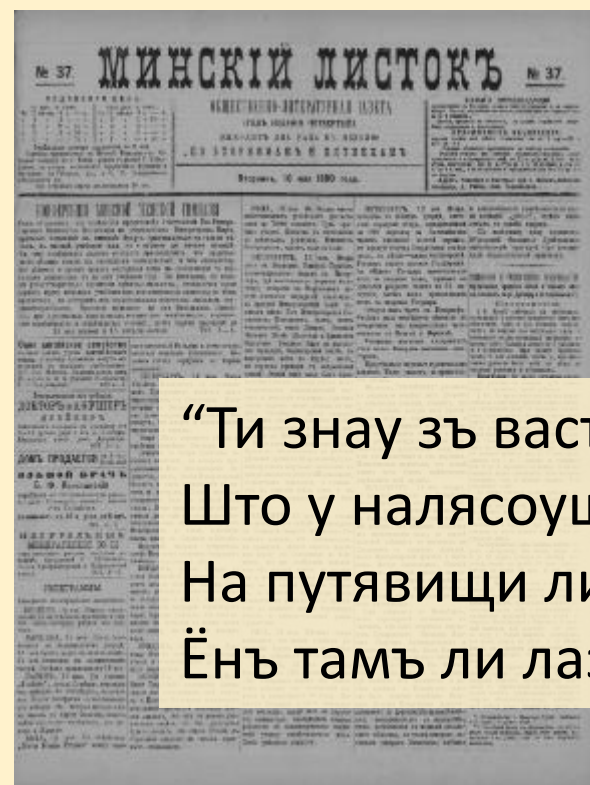1918

Republic of
Belarus
1991

First Printed Belarusian Literature

1517-1519, Francysk Skaryna publishes translation of Bible

1889, Minskiy Listok publishes *Taras*

Genesis 40:14-15
"Толико молютися воспомяни на мя егда добро тобе будеть, ивчини надо мною милость, ипонукни фараонови обо мне доповелить выпустити мене с темници сее"

"Ти знау зъ васъ, браццы, хто Тараса,
Што у налясоущикахъ ёнъ быу?
На путявищи ли Парнаса
Ёнъ тамъ ли лазни близка жиу."

Descriptions of Belarusian

1918, Taraškiewič publishes *Biełaruskaja hramatyka dla škoł*
- First comprehensive description of literary Belarusian
- Mentions variants of
  - Conjugation endings
  - Declension endings
 - Adjective endings
  - Numeral declension

1933, BSSR Counsel of National Commissars issues a description of the state-sanctioned standard of Belarusian, colloquially known as "Narkamaŭka".

Belarusian Dialect Maps

1903 • Karskij publishes map of Belarusian dialects, based on key features
1927 • Buzuk publishes overview of the south-western Belarusian dialect
1963 • Avanesaŭ publishes *Dyjaliektalahičny Atlas Bielaruskaj Movy*
1993 • National Academy of Sciences of Belarus publishes follow-on work *Lieksičny Atlas Bielaruskich Narodnych Havorak*
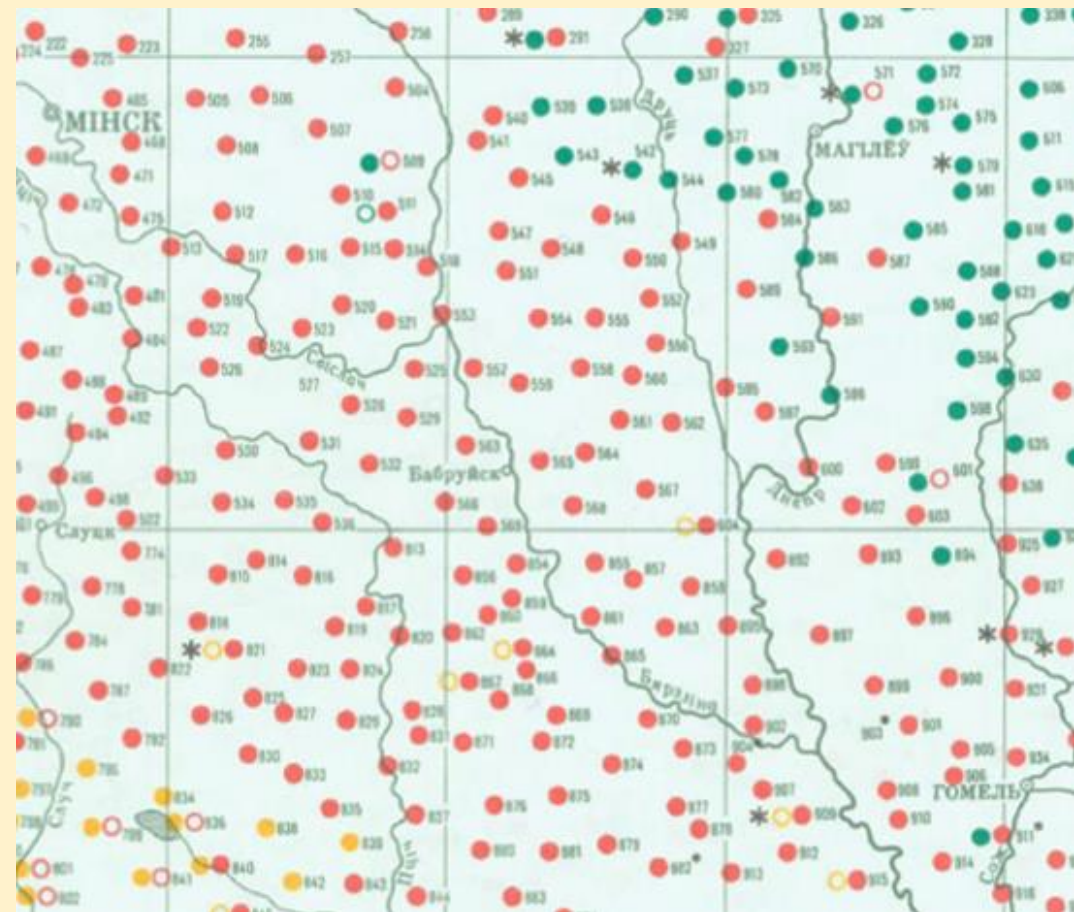
Map 34
Pronunciation of accented "e"
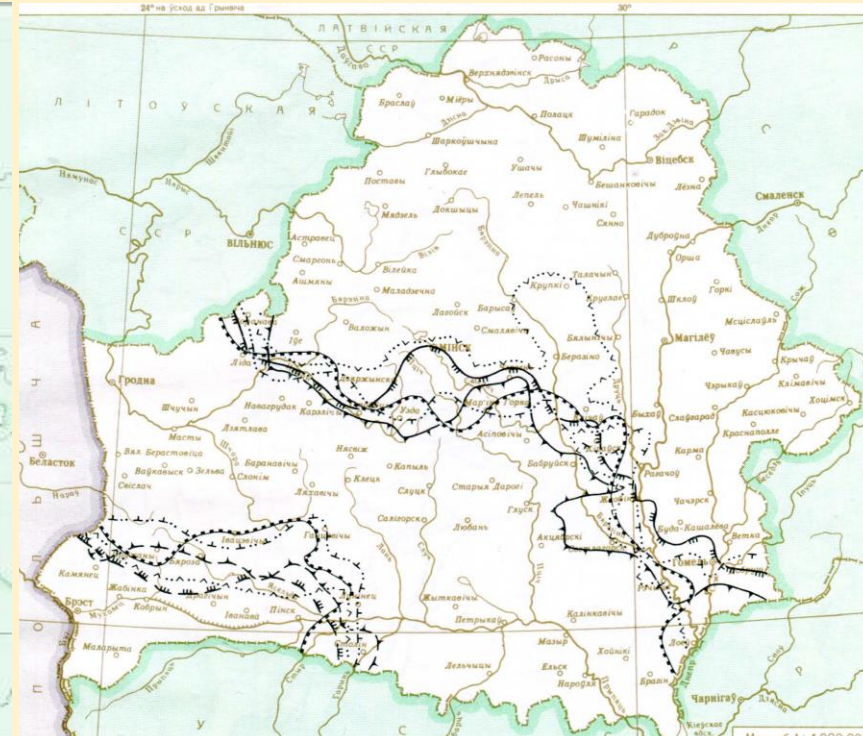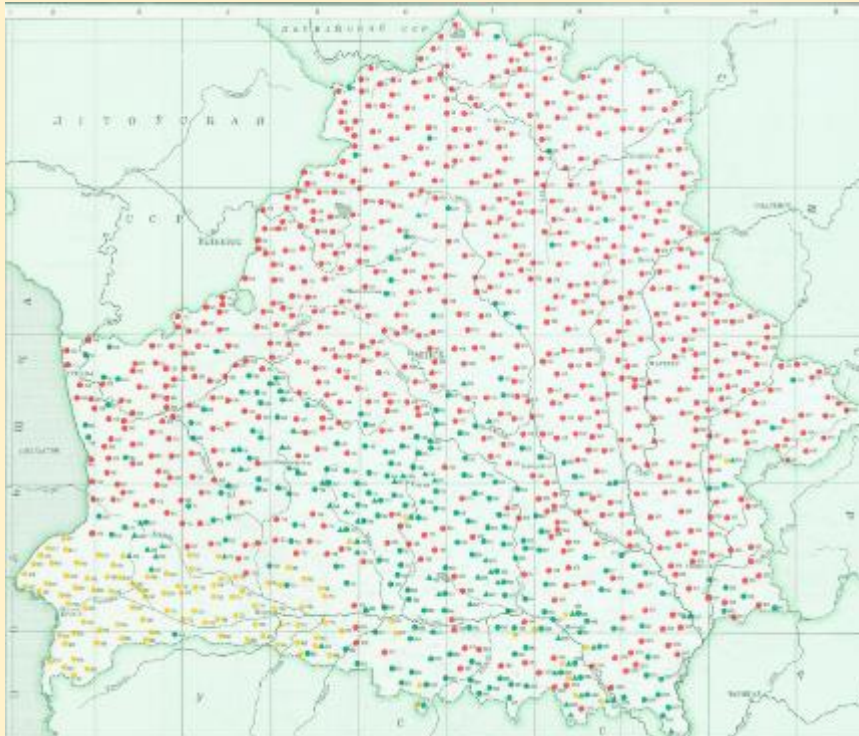
Isoglosses derived from maps 34, 35, 84, 86, 87, 88, 117, 135, and 142
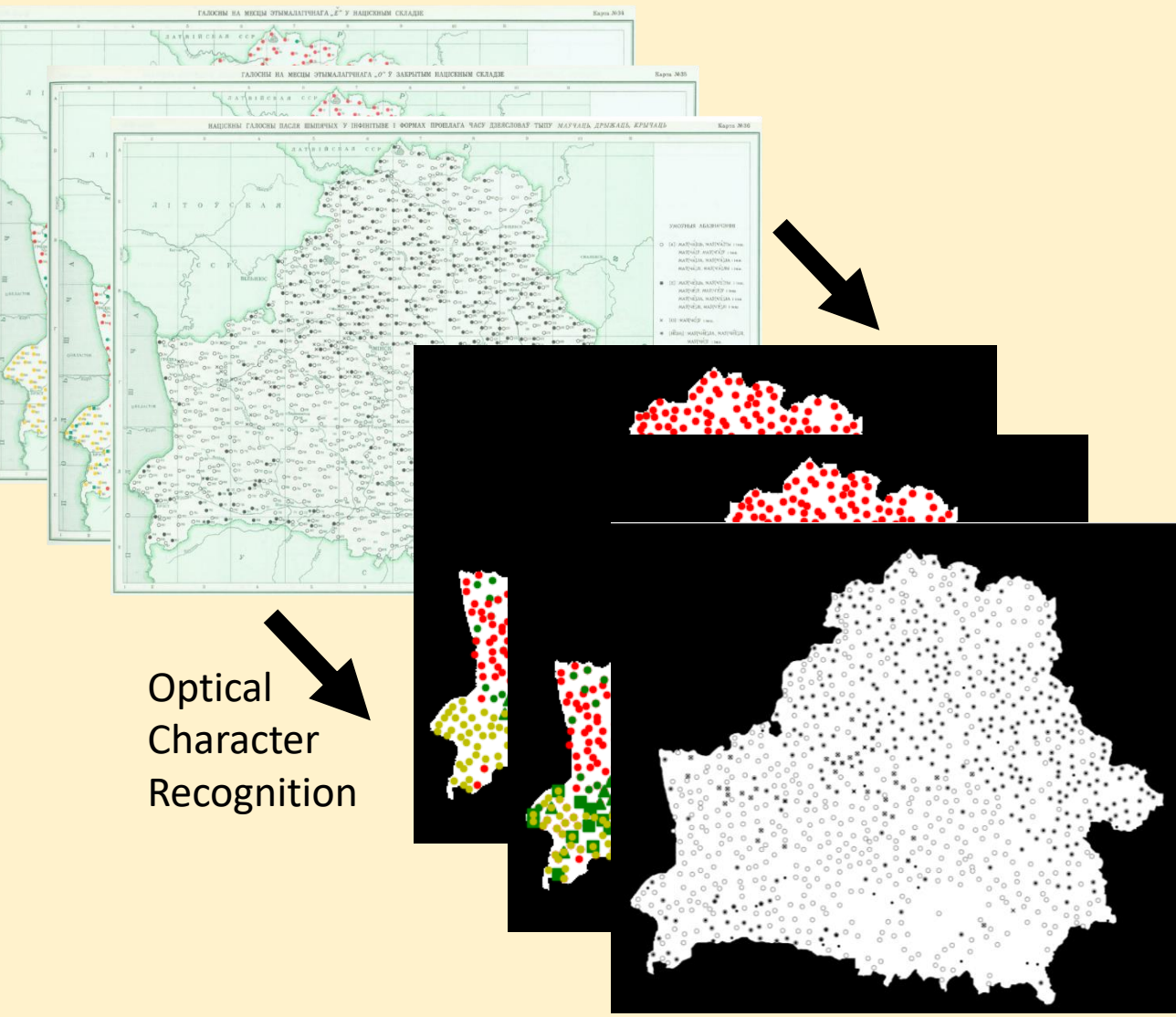
Dialectal Atlas of the Belarusian Language

ДЫЯЛЕКТАЛАГІЧНЫ
АТЛАС
БЕЛАРУСКАЙ МОВЫ

- 338 maps
- 1027 villages
- 2279 unique features
- 1948-1959

- 16 contributing institutions
- >65 years out of date

Map Digitization

Optical Character Recognition

- 1027 villages
  - 26 no longer exist

Map Digitization

1027 villages

Mathematical linear matrix $M_{2280 \times 1027}$ represents the features of the atlas, with 2280 feature rows and 1027 feature columns

338 maps, 2280 features

Each column is the measurement vector $\mathbf{m}_i$ of an individual village.

Principle Components

Singular value decomposition of feature matrix M:

$$MM^H = UDU^H$$

Columns of *U* are the principle components of *M*.

Each village measurement vector can be approximated as a linear combination of the first four principle components.

The principle component coefficients map the villages onto an abstract space, revealing relationships of statistical similarity, independent of spatial location.

k-Means Clustering

Iteration #0
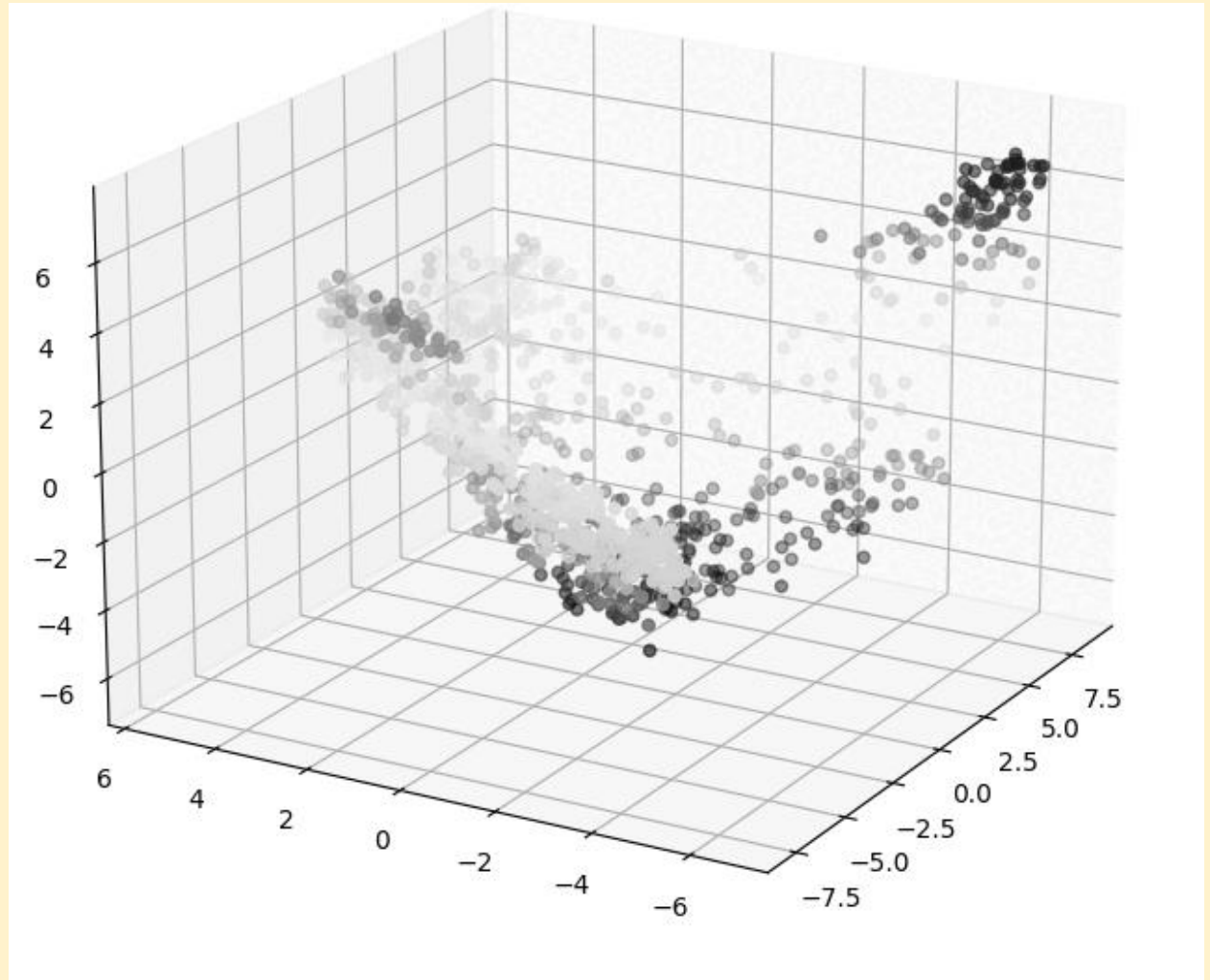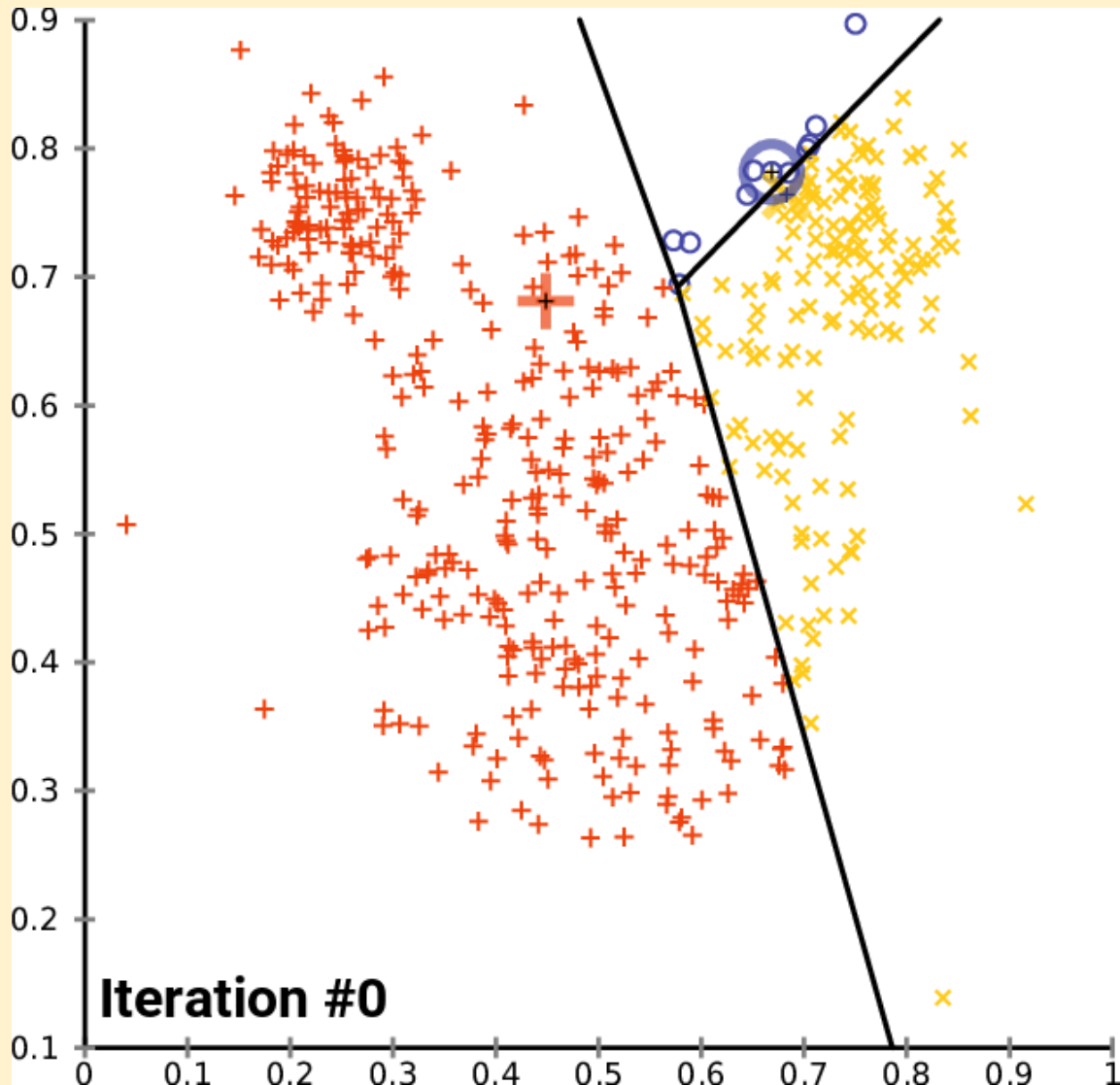
The similarity between two measurement vectors m_i and m_j can be directly compared using the Euclidean norm.

$$e = (\mathbf{m}_i - \mathbf{m}_j)^T (\mathbf{m}_i - \mathbf{m}_j)$$

The algorithm operates by classifying the points into classes such that the points of each class are more similar to the average of that class than to the average of any other class.

In the case of village measurement vectors, each village is more similar to the average of its own group's dialect than to the average of any other group.

k-Means Clustering

k=2   k=3   k=4   k=5

k=6   k=7   k=8   k=9
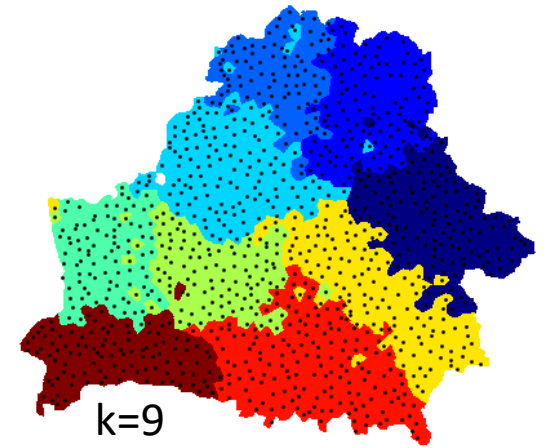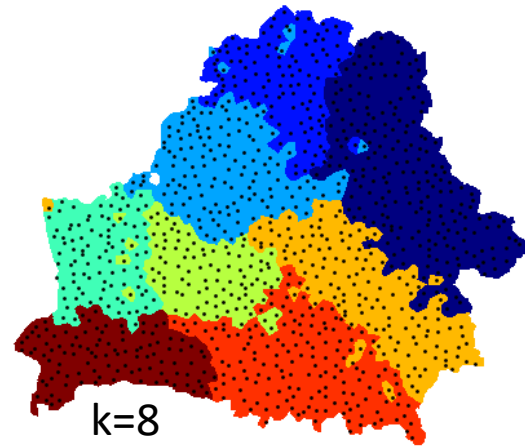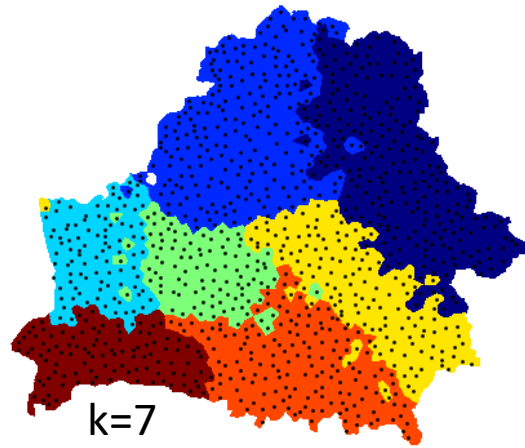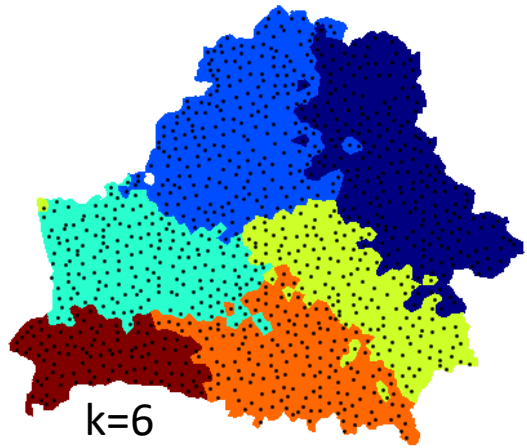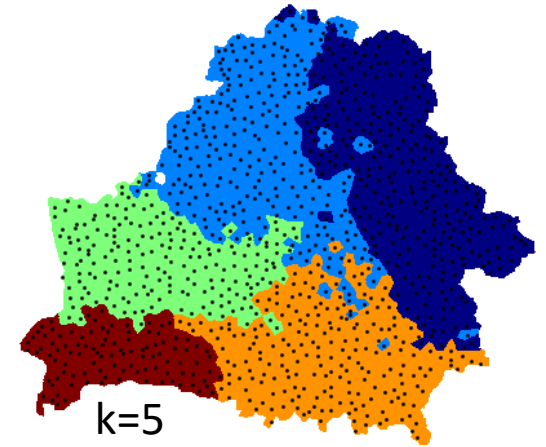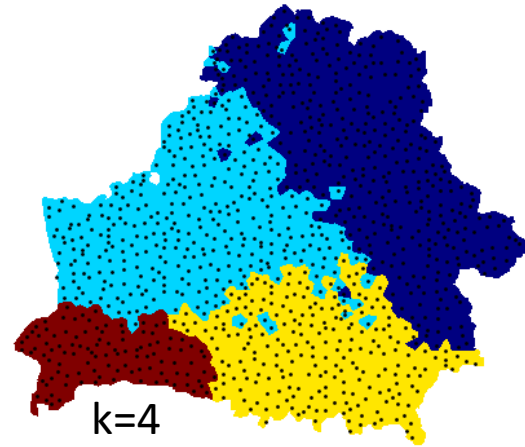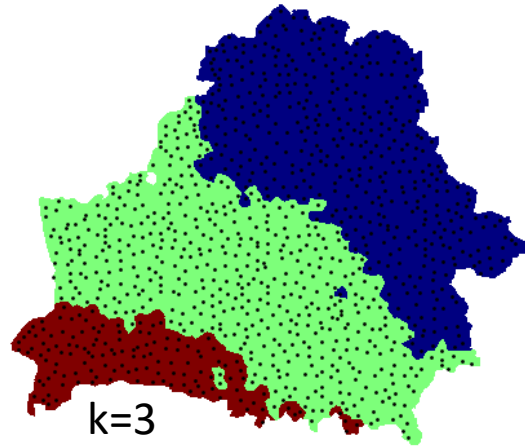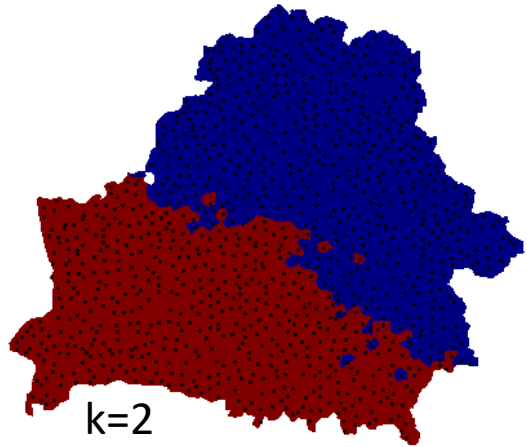
Class Boundaries vs Watersheds

k=10
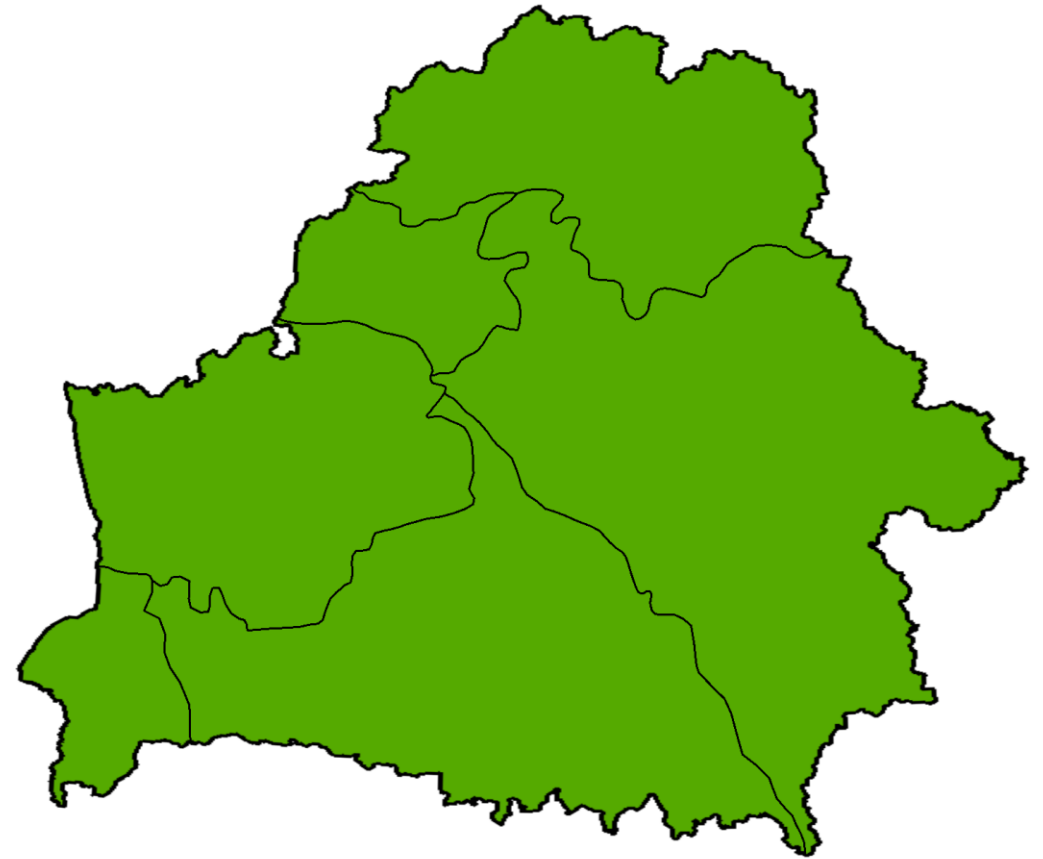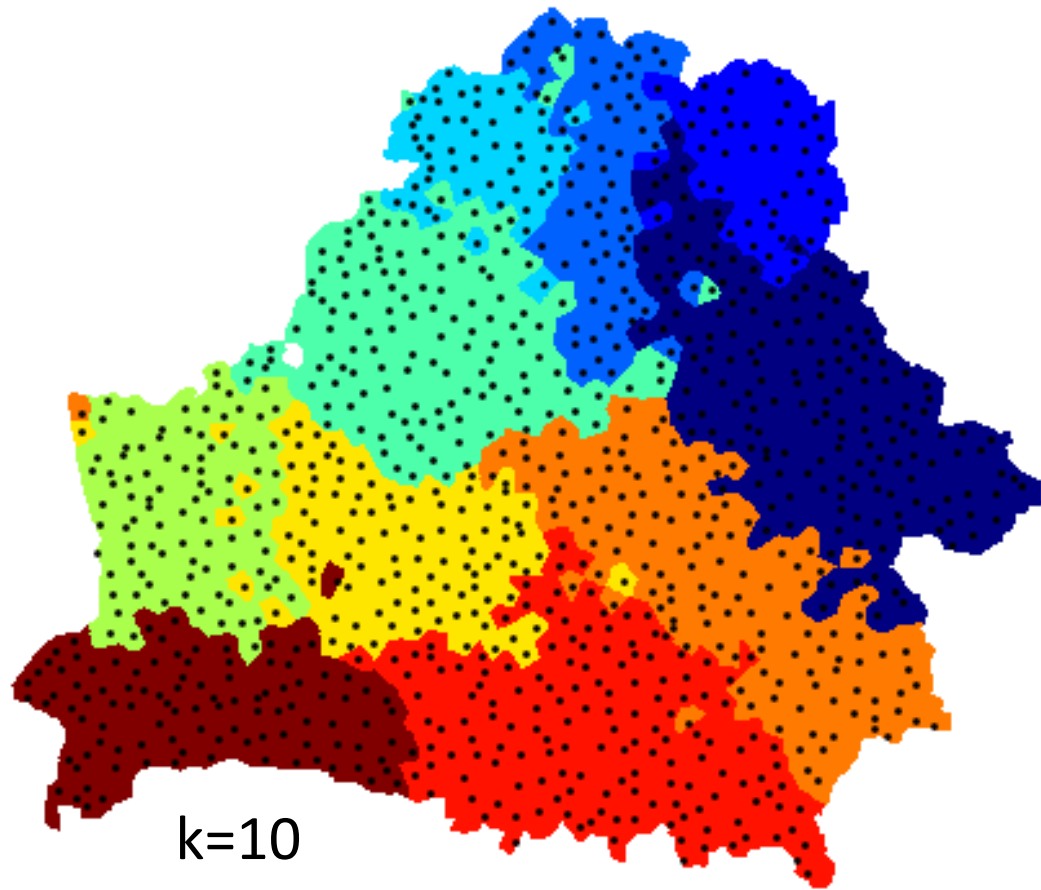
Dialect boundaries appear to superficially match with the watershed boundaries of catchment basins.

Dialect Continuum
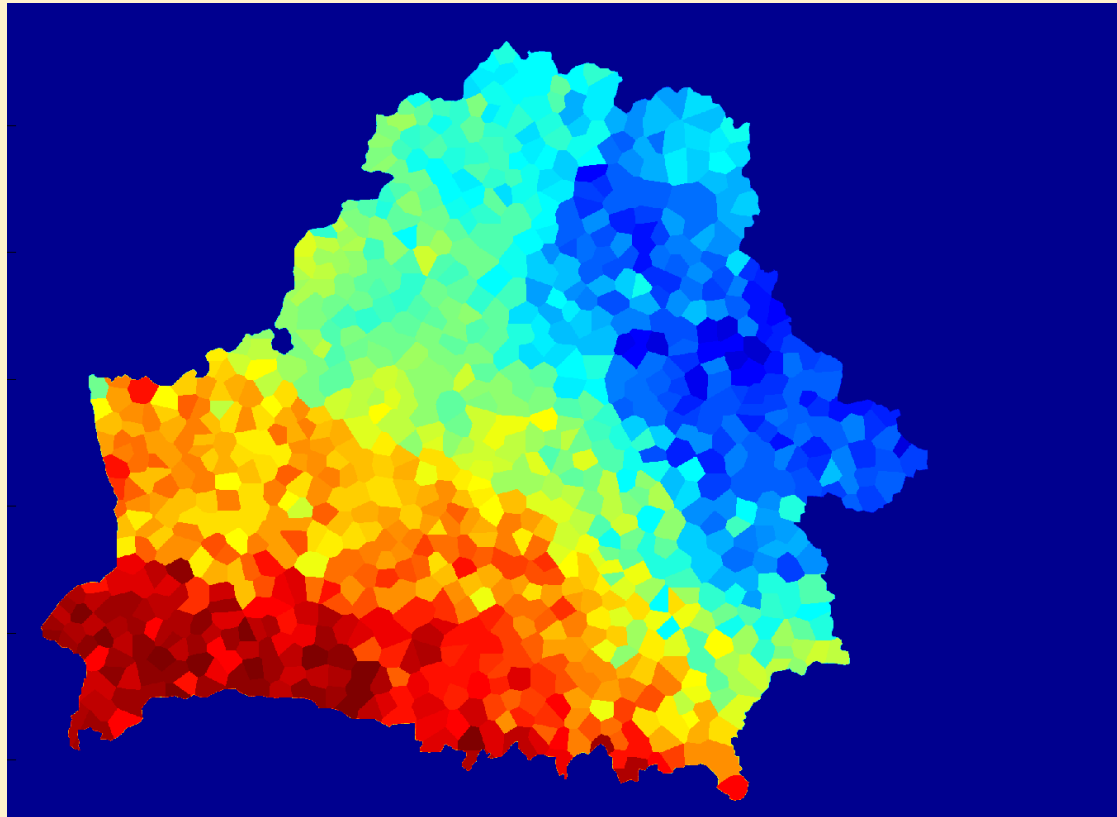
Difference values map compared to south-west corner (red = most similar, blue = least similar)

Dialect Continuum

Create a vector gradient to visualize direction and rate of spatial change.
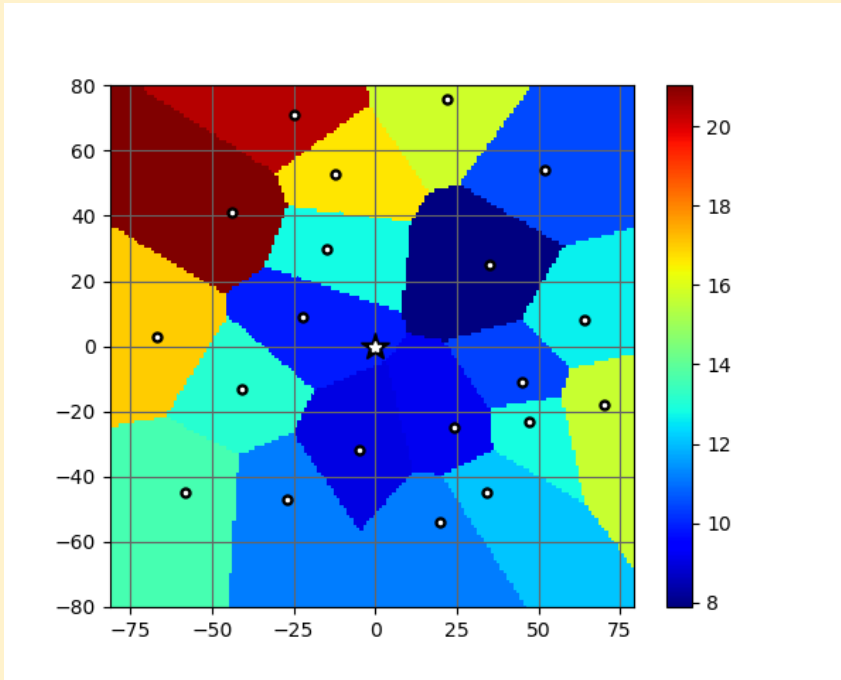
Calculate the difference between a measurement vector and each of its closest 20 spatial neighbors.
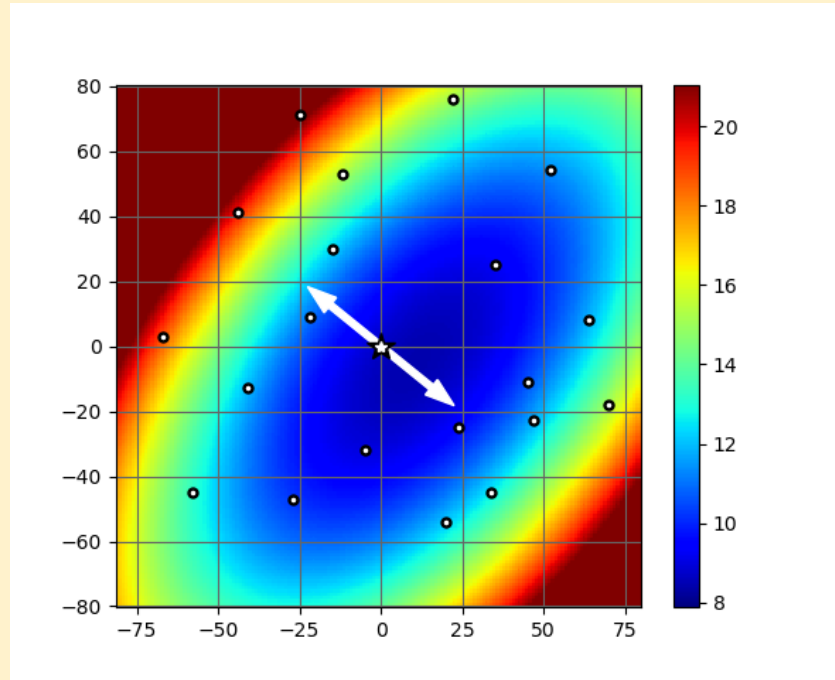
Least-squares fit a polynomial to the error function. Direction and slope of steepest change points in the direction of the gradient.
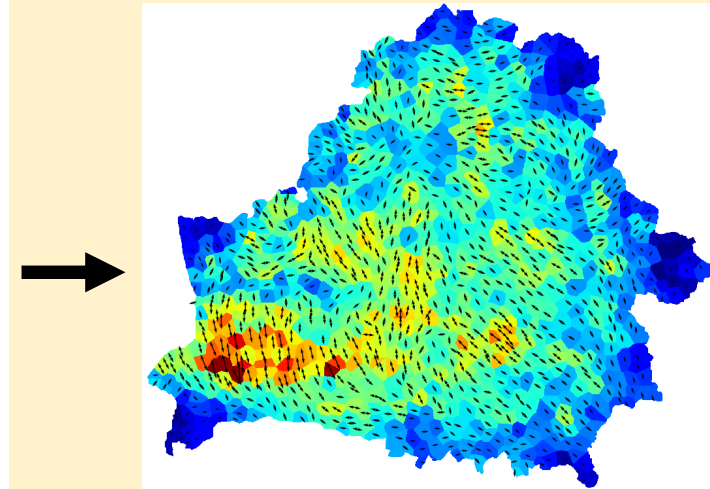
Dialect Continuum Boundaries

Vector gradient reveals directions of most rapid change.

Areas of rapid change (red color, long arrows) are natural boundaries between dialects.

Dialect Features for the Five Dialect Configuration

For each feature, for each dialect, identify the variant with highest rate of occurance

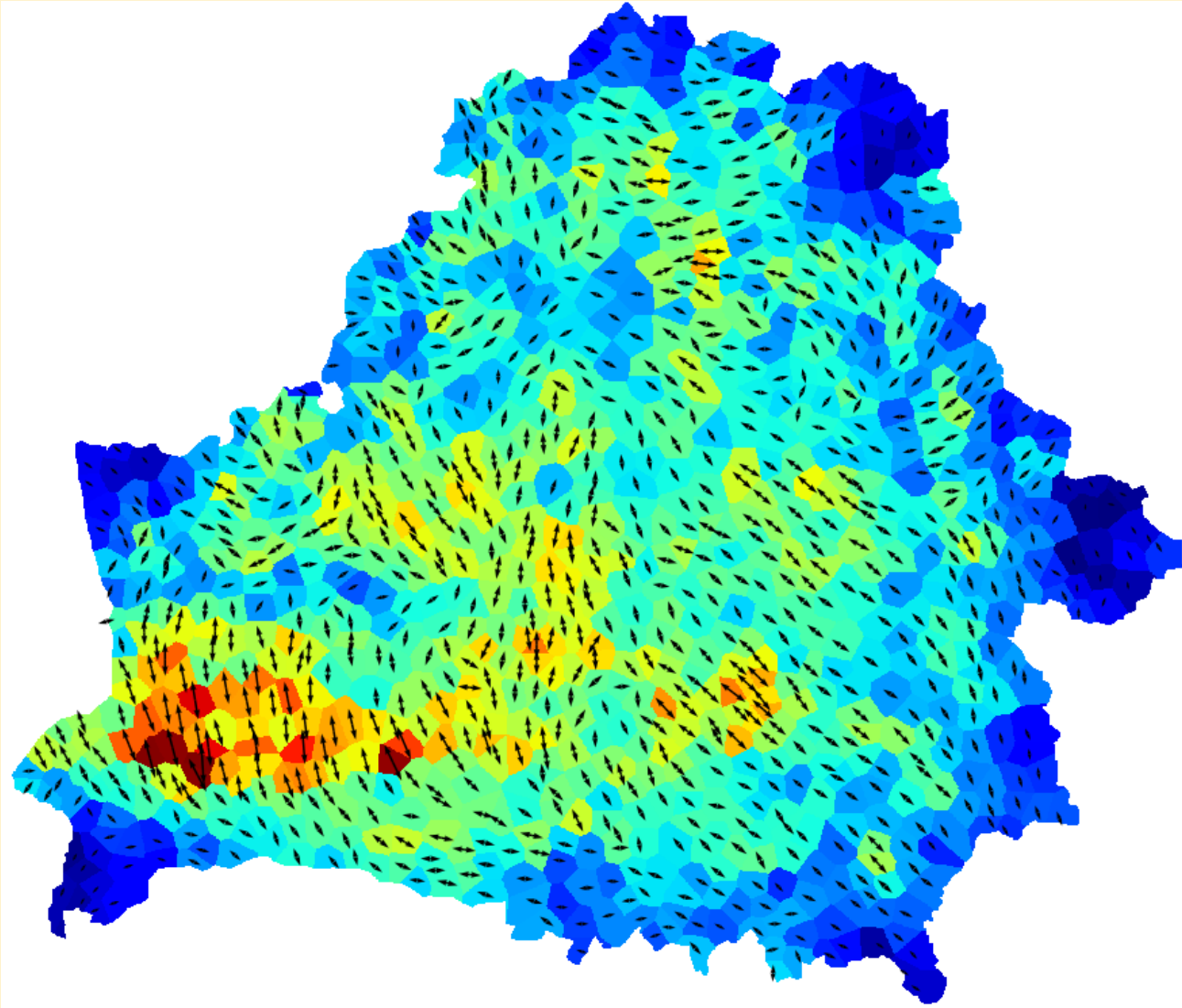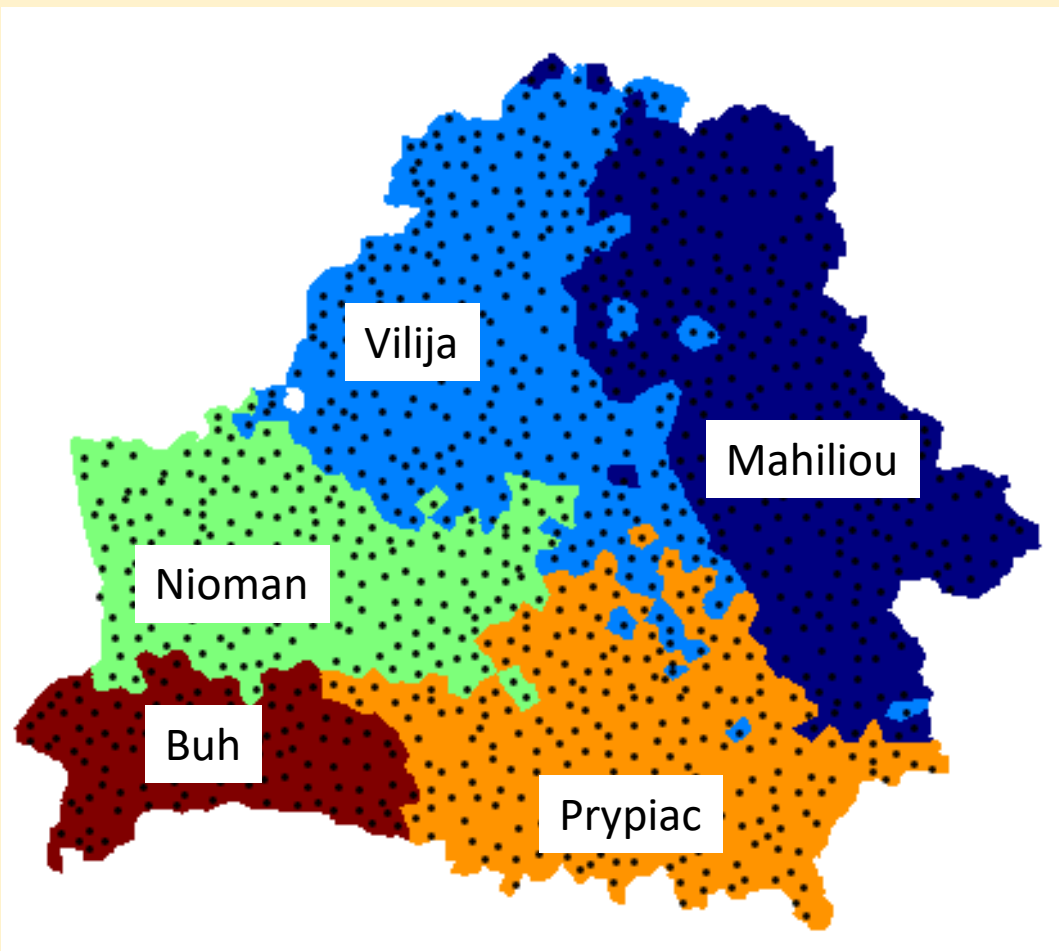| Mahiliou dial.	Viliya dial.	Nioman dial.	Prypiac dial.	Buh dial.	
57%	93%	94%	98%	41%	новы
55%	14%	2%	10%	78%	новый
3%	0%	0%	0%	0%	other
94%	91%	65%	72%	38%	новае (-ая) вядро
2%	2%	22%	0%	0%	новаё вядро
0%	0%	2%	8%	18%	новэе вядро
0%	1%	12%	13%	8%	нова вядро
0%	0%	17%	8%	15%	ново вядро
0%	0%	6%	24%	58%	новэ вядро
1%	0%	0%	0%	0%	other
97%	98%	87%	74%	55%	новая
0%	2%	58%	67%	75%	нова
2%	0%	1%	4%	1%	новаа
83%	96%	78%	68%	33%	новай хаты
3%	4%	43%	52%	23%	новае хаты
0%	0%	9%	5%	69%	новаі хаты
34%	0%	0%	3%	0%	новый хаты
4%	0%	0%	0%	0%	новэй хаты
0%	0%	2%	2%	1%	other

Dialect Features for the Five Dialect Configuration
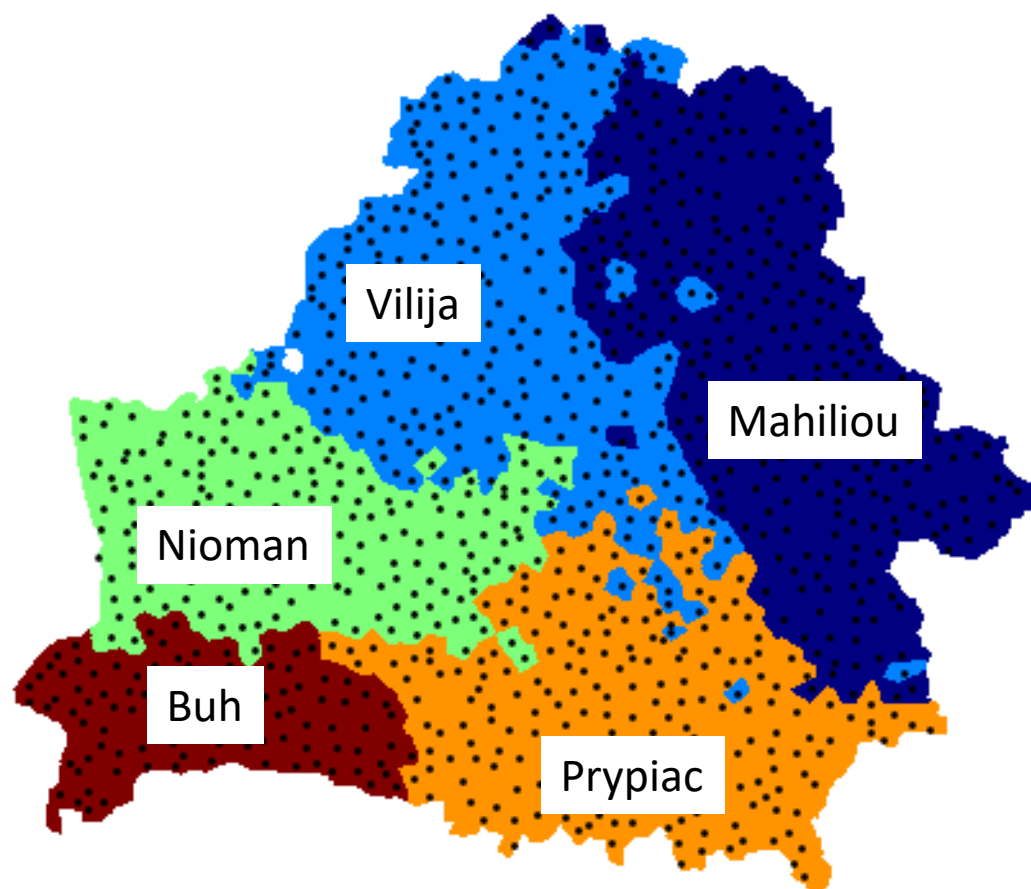
	Mahilioŭ	Vilija	Nioman	Prypiać	Buh
Nom. masc. adj. ending (unstressed)	новы	новы	новы	новы	новый
Nom. neut. adj. ending (unstressed)	новае	новае	новае	новае	новэ
Nom. fem. adj. ending (unstressed)	новая	новая	новая	новая	нова
Gen. fem. adj. ending (unstressed)	новай	новай	новай	новай	новаі
Instr. fem. adj. ending (unstressed)	новай	новай	новаю	новаю	новаю
Nom. plur. adj. ending (unstressed)	новыя	новыя	новыя	новыя	новы
Nom. masc. adj. ending (stressed)	малады	малады	малады	малады	маладый
Prep. Masc. adj. ending (stressed)	аб маладым	аб маладым	аб маладом	аб маладом	аб маладому
Gen. fem. adj. ending (stressed)	маладой	маладой	маладой	маладой	маладоі
Instr. fem. adj. ending (stressed)	маладой	маладой	маладою	маладою	маладою
Nom. plur. adj. ending (stressed)	маладыя	маладыя	маладыя	маладыя	маладыі
Nom. masc. 3rd pers. pronoun	ён	ён	ён	ён	вон
Fem. instr. 3rd pers. pronoun	ёй	ей	ею	ею	ею
Nom. masc. pronoun *that*	тэй	той	той	той	той
Gen. fem. pronoun *that*	тые	тэй	тае	тае	тэе
Instr. fem. pronoun *that*	той	той	тою	той	тэю
Nom. plur. pronoun *those*	тыя	тыя	тыя	тыя	тый
3rd pers. plur. conj. стрыгчы *to clip*	стрыжом	стрыжом	стрыжом	стрыгом	стріжем
Infinitive end. verbs like пячы *to bake*	печ	пячы	пячы	пекці	пячы
Masc. past tense of verb быць *to be*	быў	быў	быў	буў	був
Fem. past tense of verb быць *to be*	была	была	была	була	була
Short form adjective *distant*	далёка	далёка	далёка	далека	далека
Ending of noun дзіця *child*	дзіцёнак	дзіця	дзіця	дзіця	дзіця
Masc. instr. ending of цяля *calf*	цялёнкам	цялём	целям	цялям	цялятам
Plural ending of noun сяло *village*	сёлы	сёлы	сёла	сёла	сёла
Gender of noun яблыка *apple*	яблыка	яблык	яблыка	яблыка	яблыка
Noun meaning *potato*	бульба	бульба	картофля	картофля	картофля
Adverb meaning *very*	дужа	надта	вельмі	вельмі	вельмі

Chris Marchant

www.vitba.org

chris@vitba.org