

General Online Dialect Atlas

Sheila Embleton,
Dorin Uritescu, and
Eric S. Wheeler

York University, Toronto, Canada

His voice was rather rich and dark; the accent was Middle Western, but underneath the nasalities there was something soft and furry that came from the South.

-- Mary McCarthy. 1942: *The Man in the Brooks Brothers Shirt. The Company She Keeps.* 67

Links

RODA

<http://pi.library.yorku.ca/dspace/>
(under the “dialectology” community, “RODA” collection)

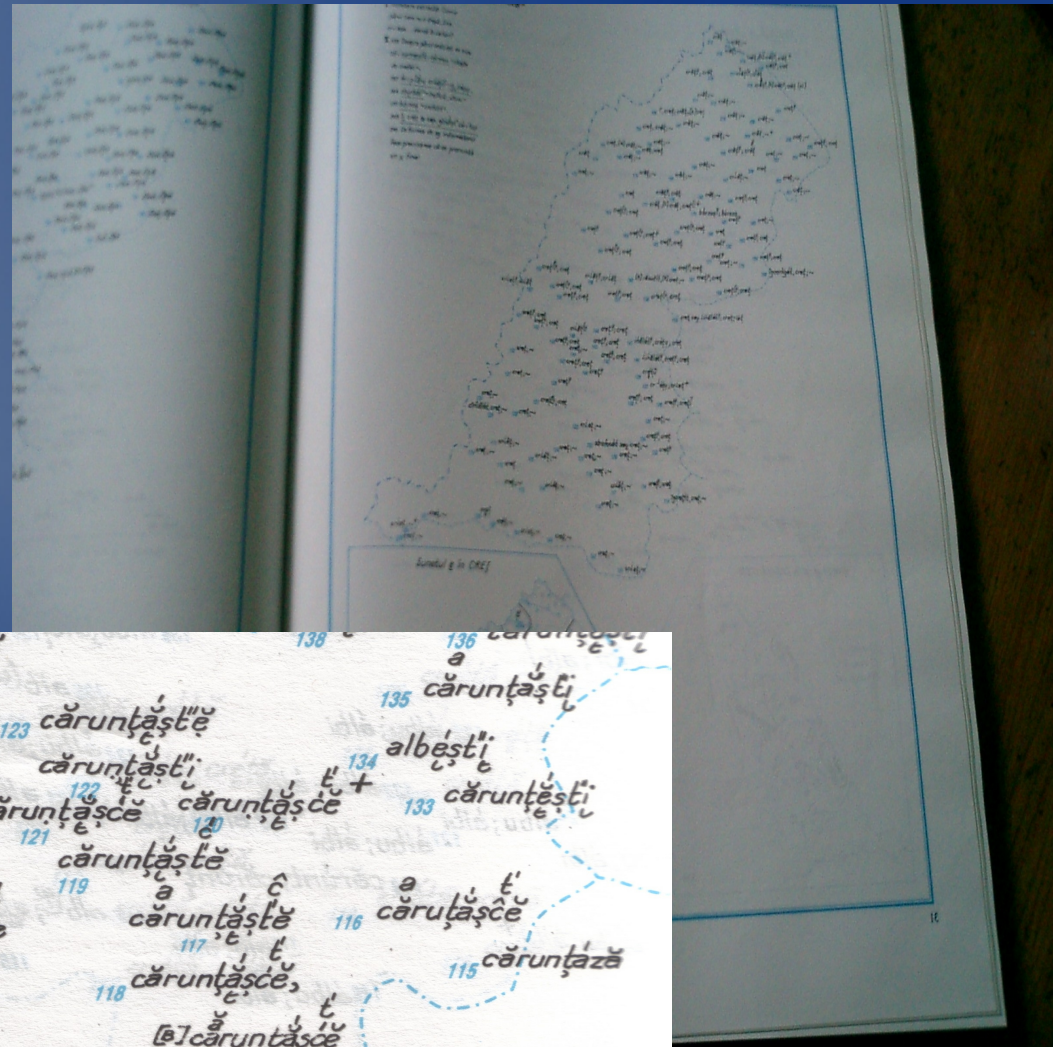
List of online dialect atlases:

<http://ericwheeler.ca/atlaslist>

Traditional Dialect Atlas

Set of maps

- *Prompt: 'Where do you keep the cows?'*
- *Data: "byre", "shed"*
- *Location: Eglington*

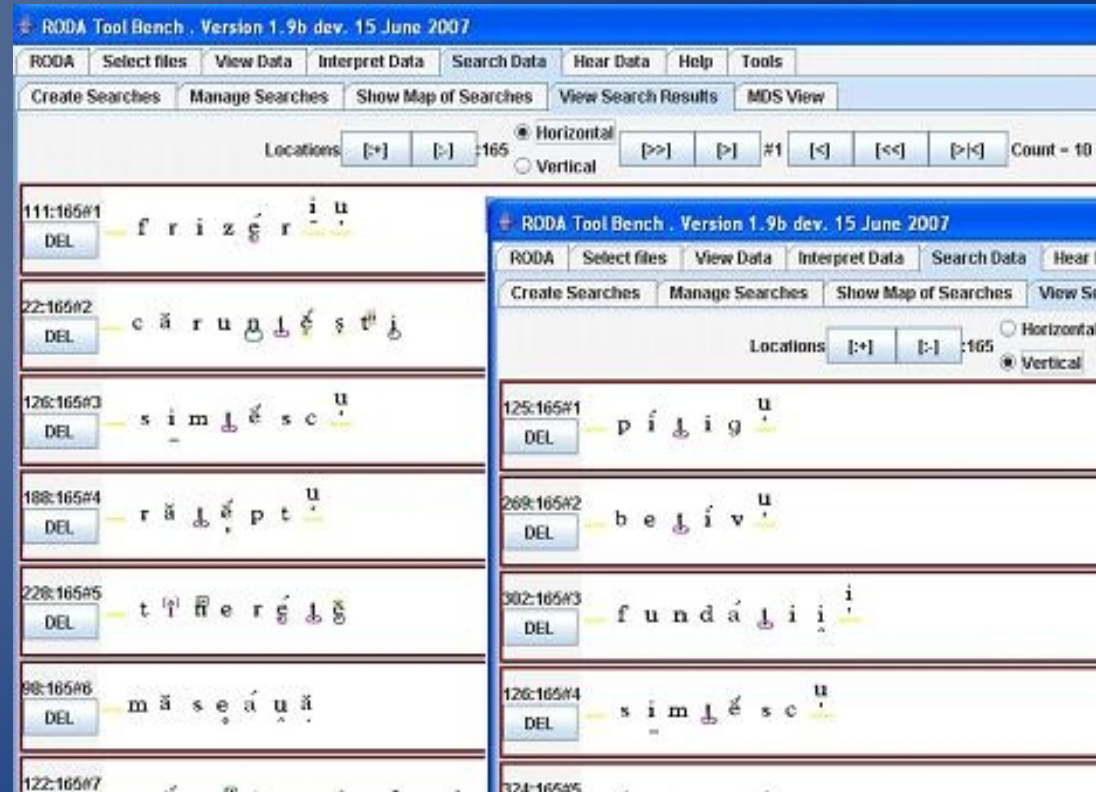


Raw data vs Interpretation

- Raw: response to a prompt.
 - “A byre or a cow shed”
- Interpretation:
 - “ /k/ vs /k'/ in word final position”
- Traditionally, there is good access to the editor's interpretation (maps, indices, etc.)
- Access to anything else is limited by the effort of getting it out of the raw data (if available)

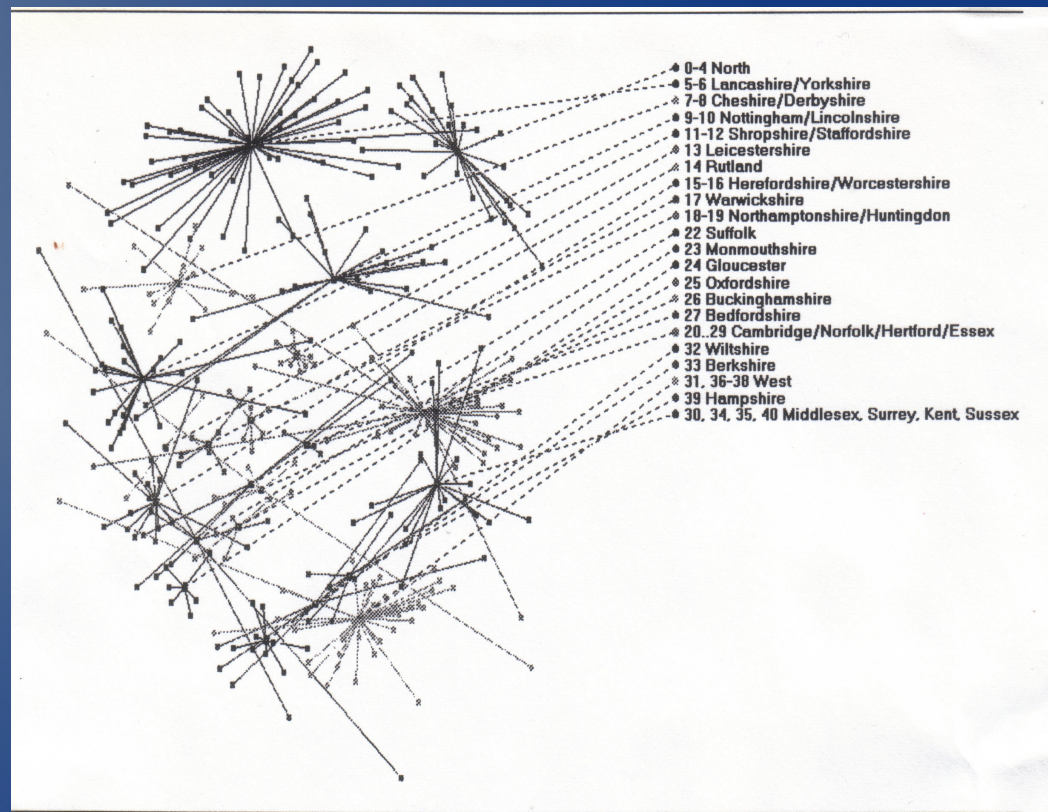
Information Technology

- digital data
- data bases and software applications
- graphics for maps, and multimedia
- internet for sharing
- very large amounts of (raw) data managed quickly and easily



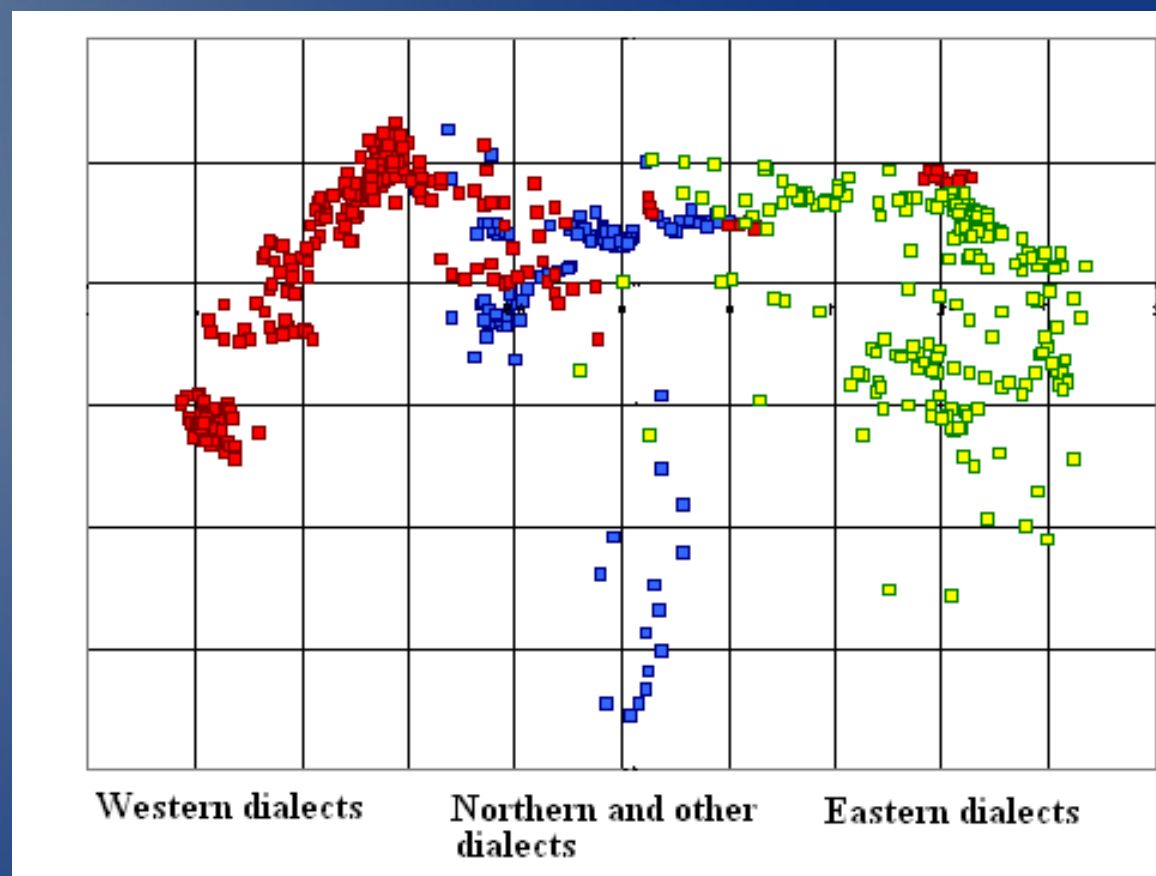
CLAE: English

- Computer Developed Linguistic Atlas of English (Viereck and Ramisch. 1997)
- MDS analysis of the English dialects in England



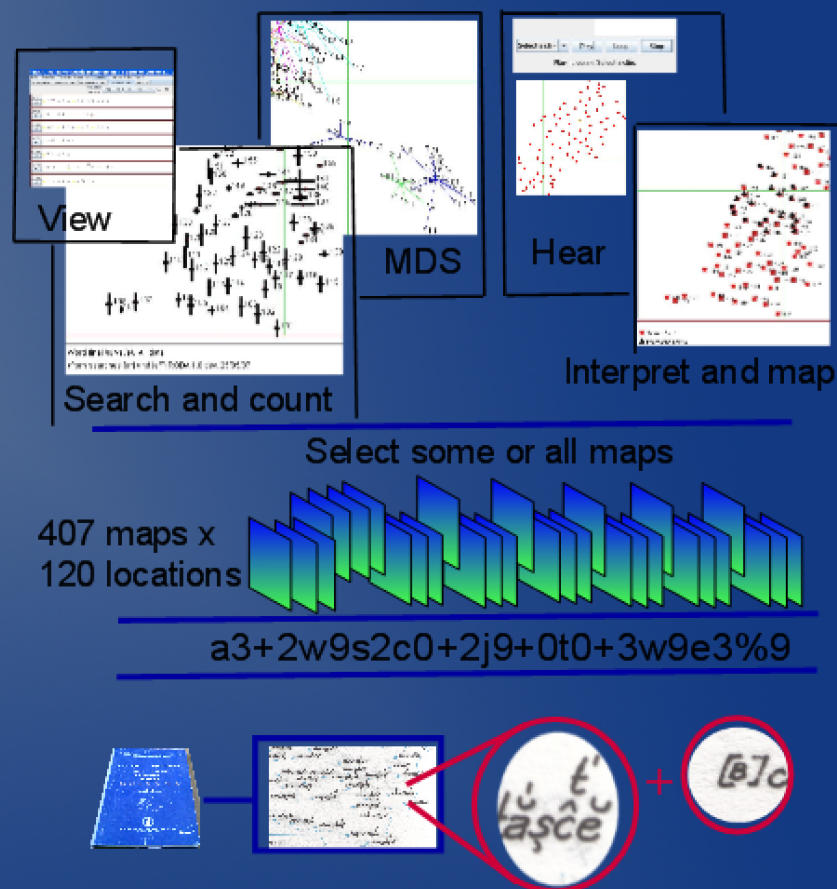
Finnish Online Dialect Atlas

- Digitization of a classic atlas (Kettunen 1940)
- MDS analysis
- Data available for other scholarly use



Romanian Online Data Atlas: Crișana

- Digitize an existing atlas
- Data key to Romance linguistics
- Built an application to give users advanced access to the data



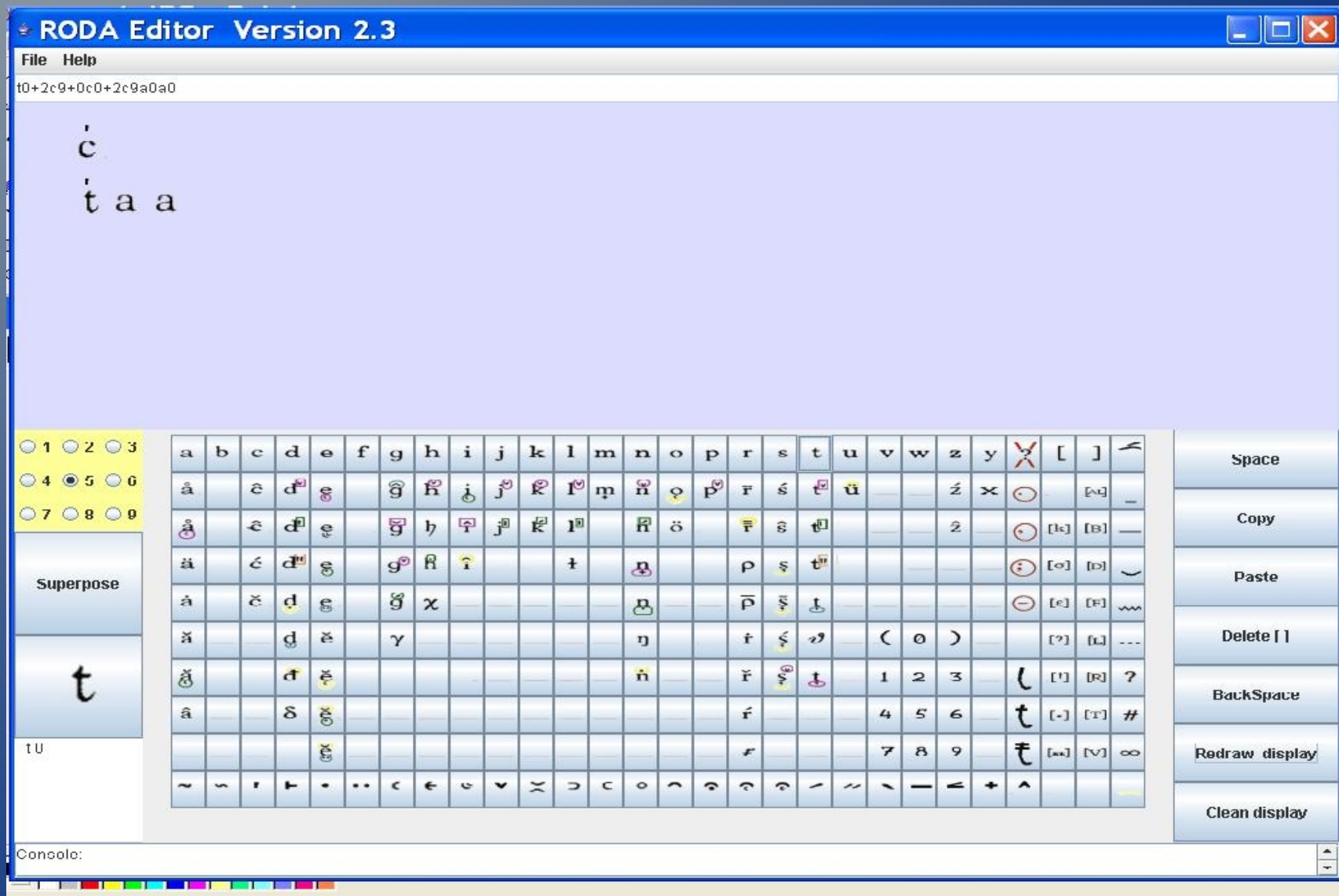
Progression of Requirements - 1

- Scanned version of hard copy book



Progression of Requirements – 2

Digital data



Progression of Requirements – 2b

Custom encoding

a5+2c9s2c0+2x9+0t0+3c9e5%9

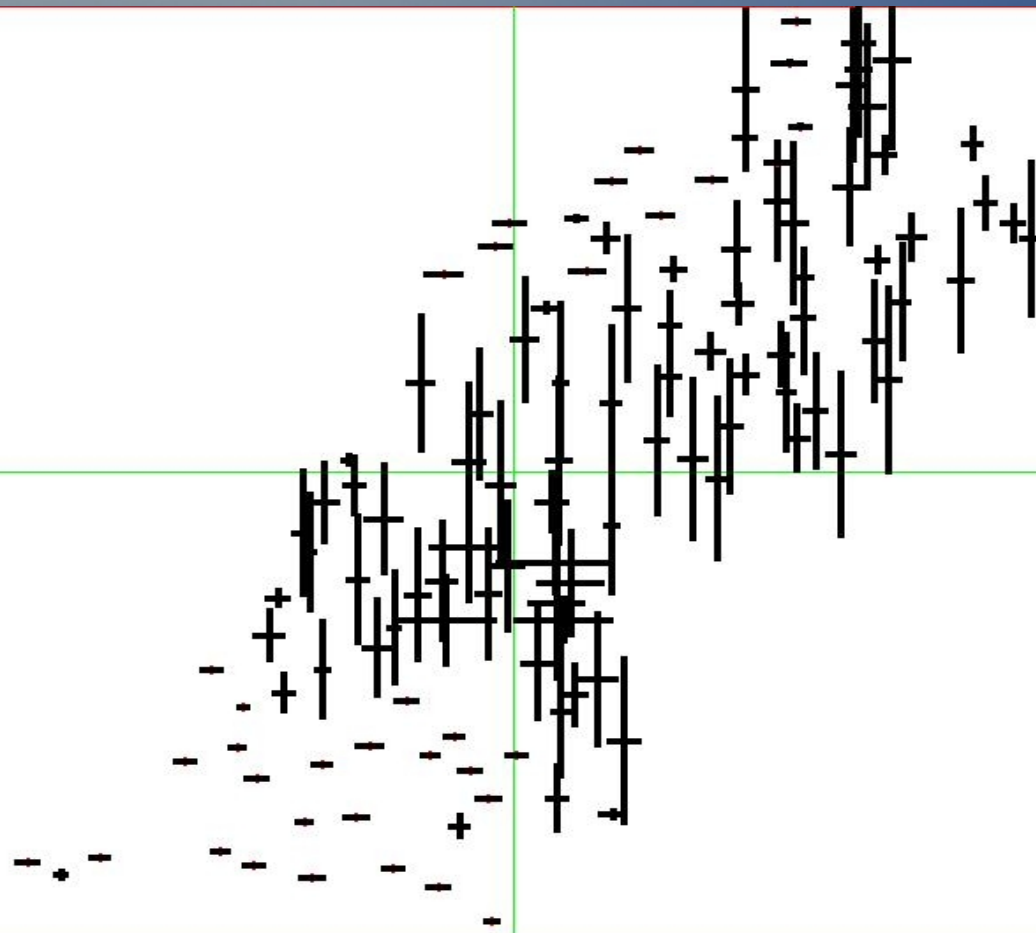


+

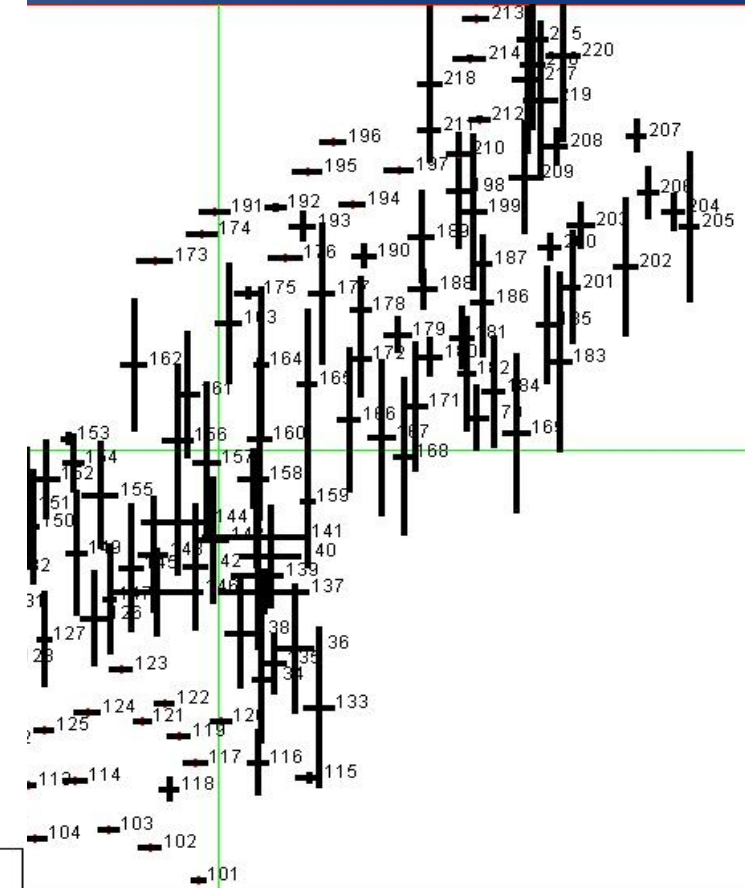


Progression of Requirements – 3

Dynamic Maps



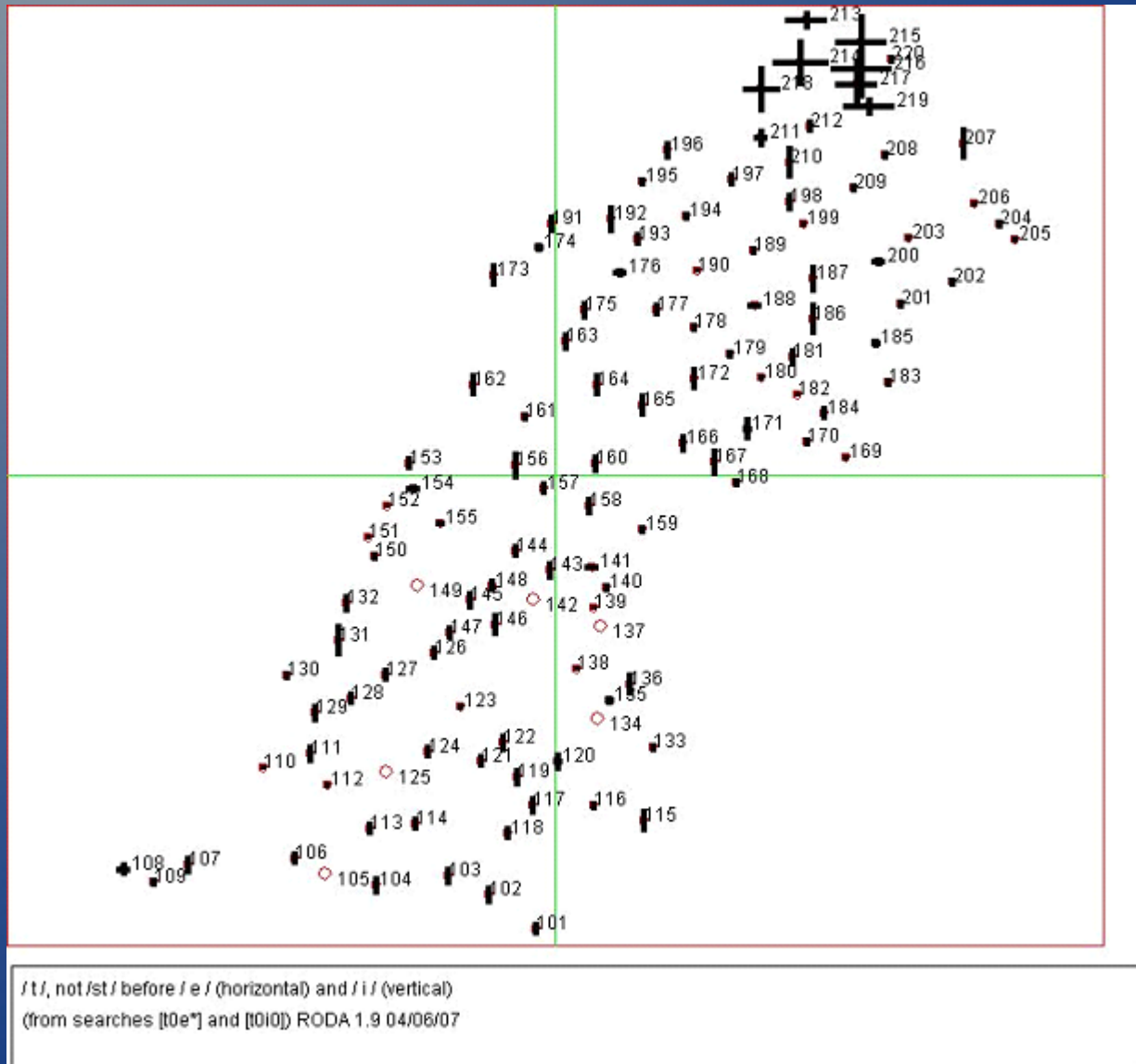
/u/ word final. Syllabic (horizontal) and non-syllabic (vertical)
(from searches [u*/ syllabic word final] and [u*/ non-syllabic word final]) RODA 2.01 2008/03/26



on-syllabic (vertical)

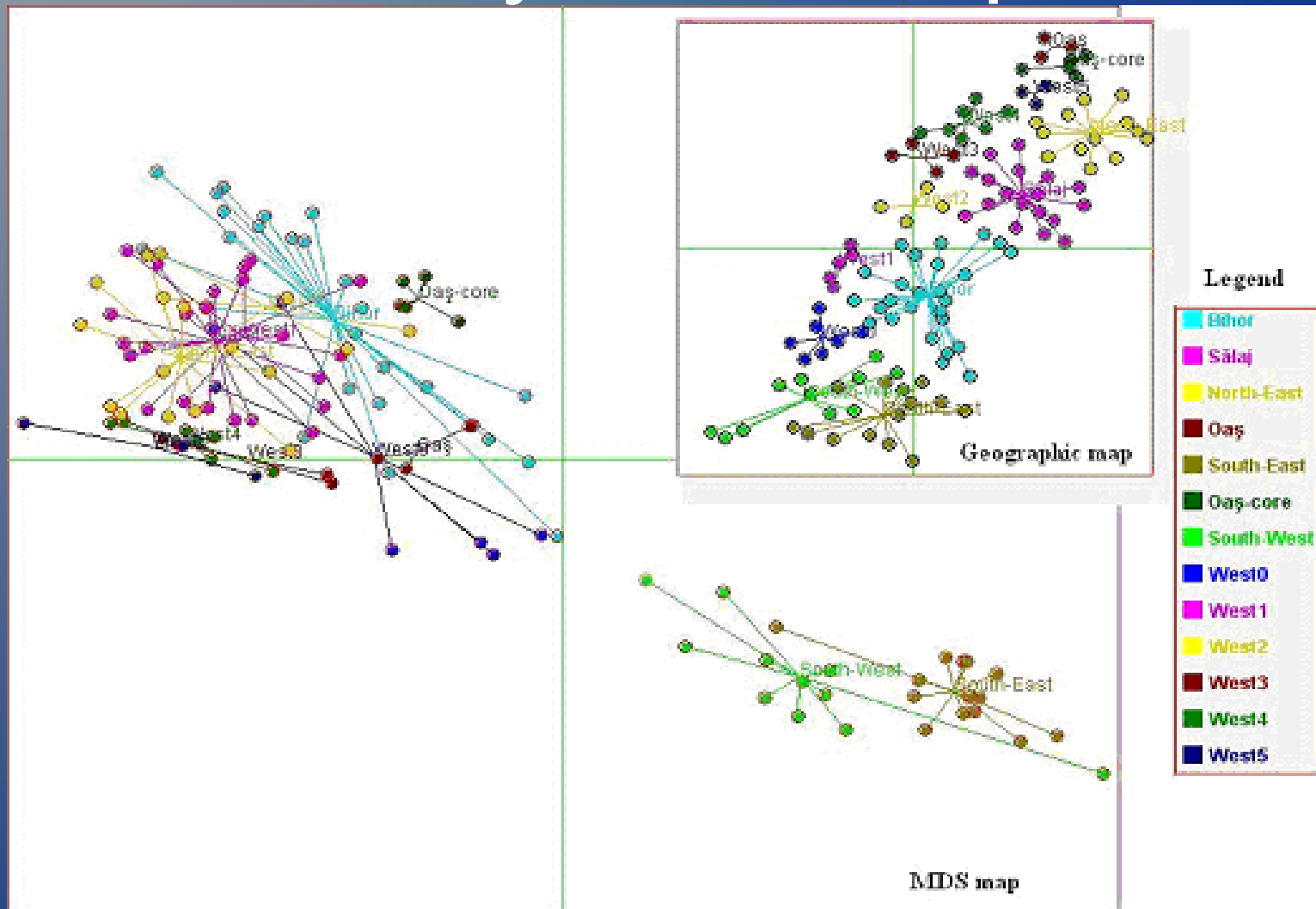
(from searches [u*/ syllabic word final] and [u*/ non-syllabic word final]) RODA 2.01 2008/03/26

Progression of Reuirements – 4 Counting



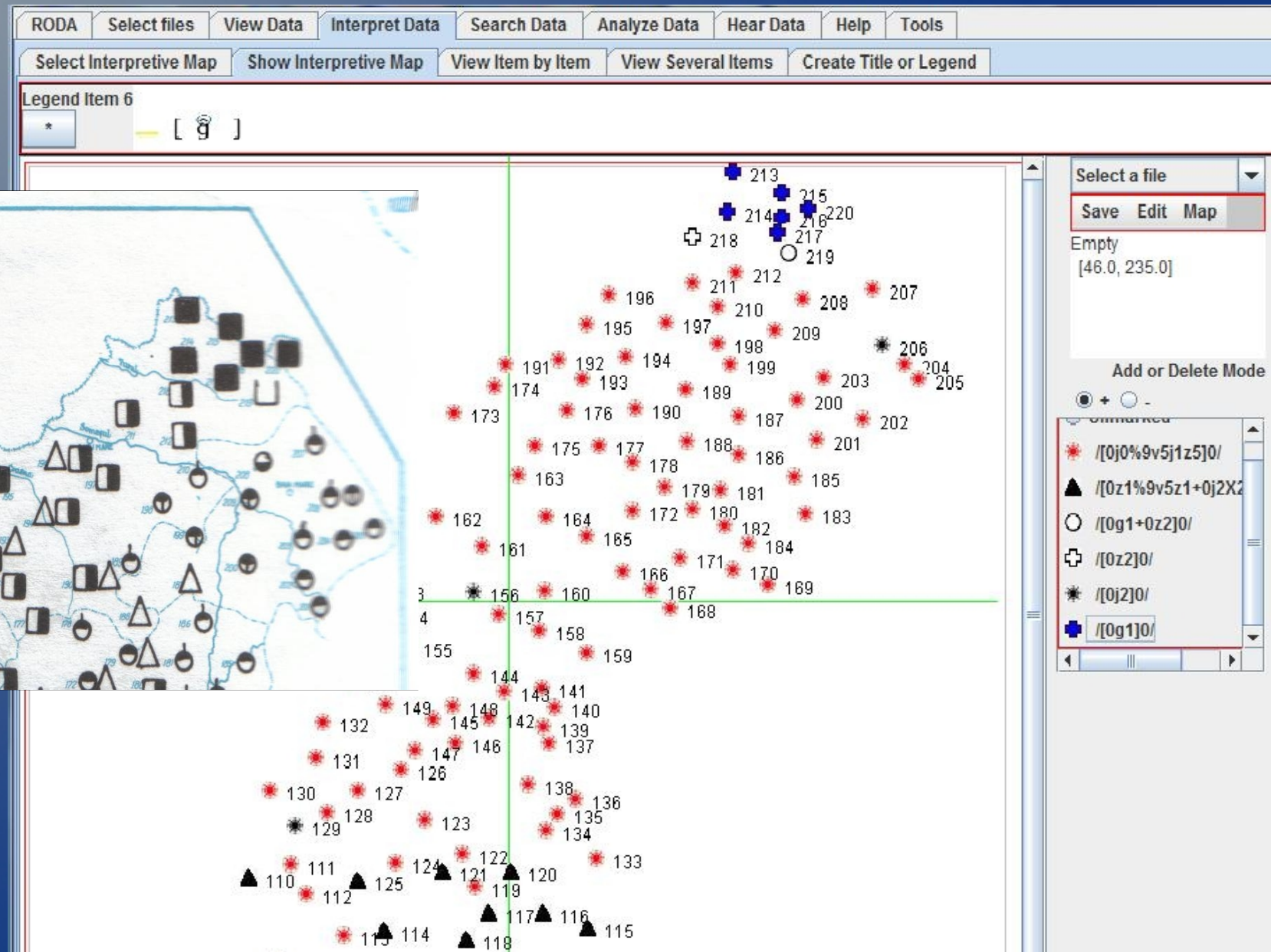
Progression of Requirements- 5

Analytic techniques



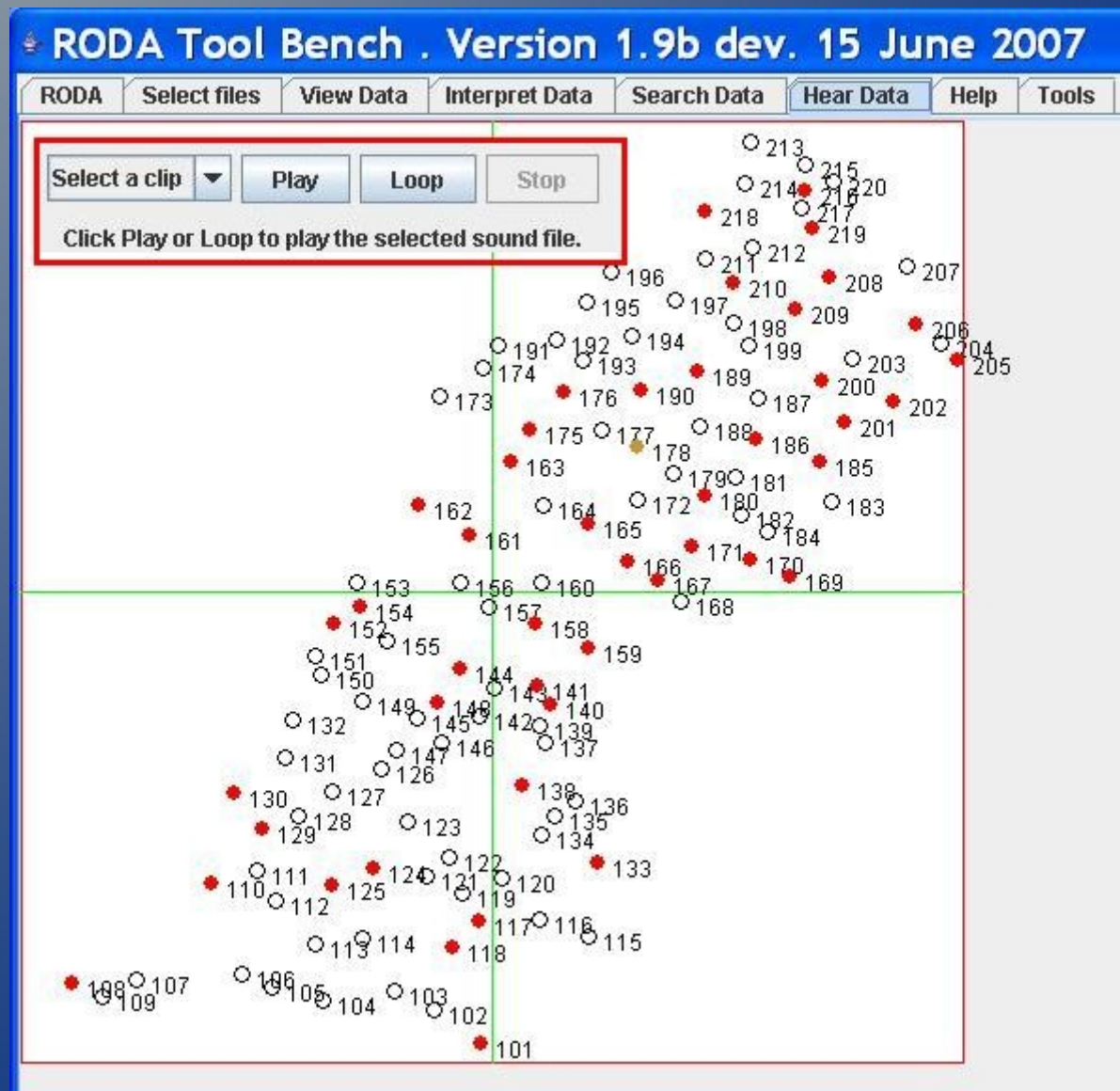
MDS of All Data. Standard geographic groups

Progression of Requirements – 6 Interpretations



Additional Requirements

Sound samples



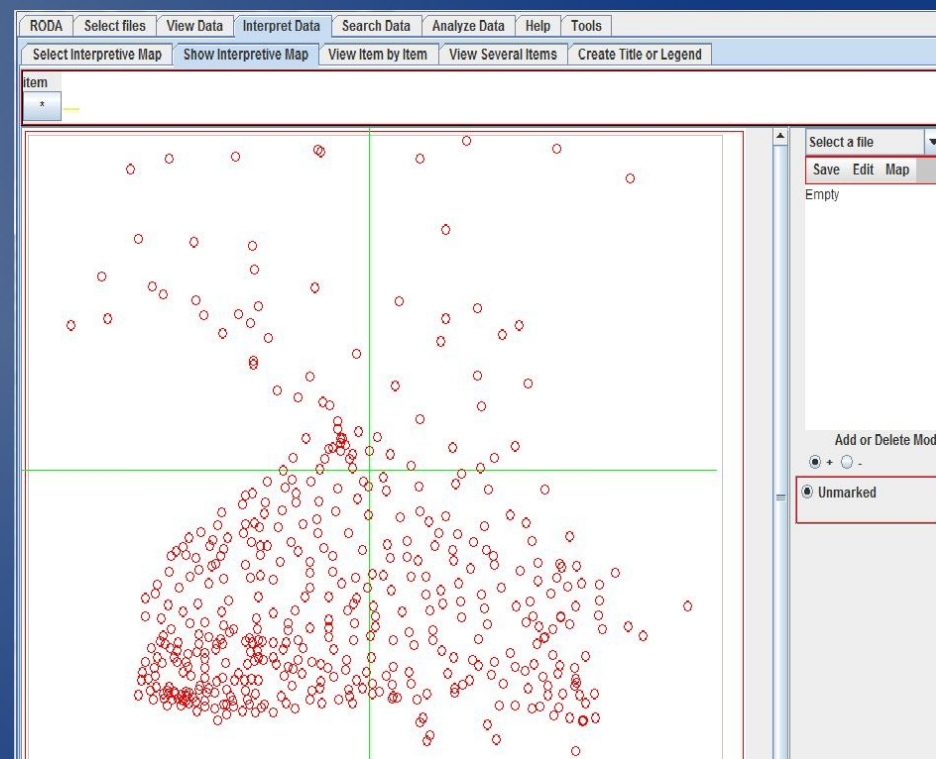
Prototyping

Repeated cycle of requirements, design, build, test and assess

- Each cycle leads to something “better”
- A good approach to System Development when the end point is unclear
- Drawback: leads to commitments that may not be optimal later on
 - Custom encoding works for Romanian
 - May be cumbersome for other data

Finnish Online Dialect Atlas

- Source data is a set of maps showing features
- Customize RODA to have correct parameters
- Display Finnish data as “Interpretation” maps



Customizing RODA code

- Finnish data is not “raw” data
- Therefore, some RODA functions are not usable; function needs to be disabled
- Code can be reprogrammed (ad hoc)
 - prototype code was not set up to be turned on and off easily

Prototype works, but it is not the optimal general application.

List of Online Dialect Atlases

List of online dialect atlases:

<http://ericwheeler.ca/atlaslist>

You are invited to contribute

- List shows there is no consensus on what an online dialect atlas should be

General Online Dialect Atlas

Here are some design features we think should be in any online dialect atlas

- Based on our experiences
- Not “ultimate”
- Not exclusive of other approaches

Feature: General data

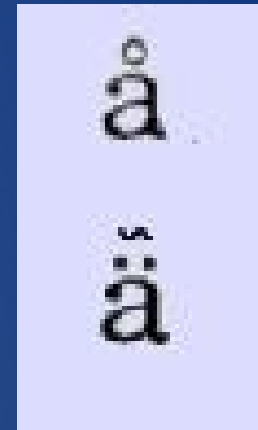
- Handle a wide range of data
- Existing data may have to be transformed
- New data can be created to a standard format
 - XML description for external use
 - internal use can be simpler

Key: transformation of data from one format to another, and tools to do the transformation

XML

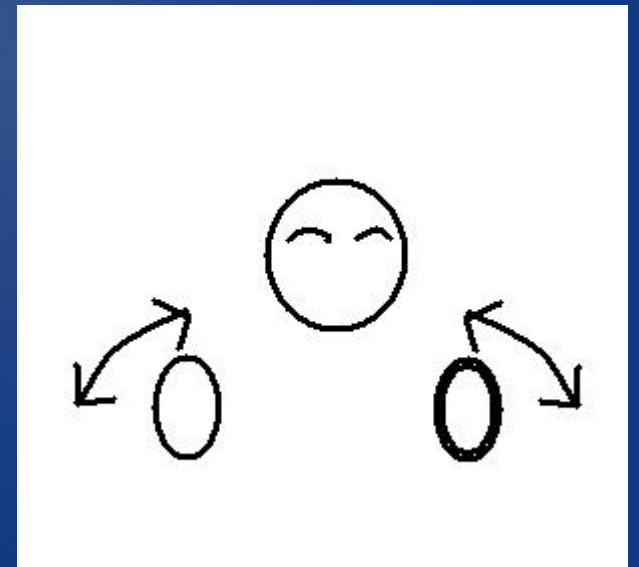
a3+2b9+0a1

```
<character>
  <glyph>a3
    <accent position=2>b9
  </accent>
</glyph>
<superposition>
  <glyph>a1</glyph>
</superposition>
</character>
```



Content

- Some data is simple text
 - ASCII or Unicode
 - left-to-right; right-to-left; r-l embedded in l-r, etc.
- Some data is more complicated
 - Field notes with idiosyncratic notation, direction
 - Photos, drawings, figures



Two Approaches

- See all data as a sequence of complex symbols
 - A complex symbol can have special representation, direction or presentation
 - Application allows plug-in processing for non-standard symbols
- Multimedia
 - links to source (images of the field notes or videos of the interview, etc.)
 - underlying is a plain text description of the data

Feature: Online Atlas

- Online application
 - Accessed via Internet from a central server
 - Stand-alone downloaded from a site
- “Best” solution changes with new technology and internet economics
- Advice: (if possible) don't make an application that is dependent on any particular technology.

Feature: Functions

- Atlas should provide a set of common functions
 - Select and Search
 - Count and Analyze
 - Interpret and Present

Function: select

- View prompts (“1. What is the opposite of 'yes' ?”)
- Select data by prompts
 - List of prompts (“Prompts 2, 4,7”)
 - Defined, labeled set (“Phonology”)
 - Boolean combination (“Phonology less 7”)
- Display the selected data, by location
 - on a map
 - in a table
 - in a file

Function: search

Search for a pattern in selected data

- One or more symbols (e.g. *find /ti/ or /te/*)
- Symbols defined by characteristics (e.g. *find /t+Highvowel/*)
- Symbols in context
 - with or without characteristics
 - in contexts with or without characteristics
 - (e.g. */t/ with or without accents, at word-end and in a word marked /noun/*)

Function: Count

- Count the occurrences of a search pattern over a set of selected data
- Display the counts
- Allow for comparison of different counts
- Allow the user to review the data behind a count, and revise the displayed count (e.g. by deleting known exceptions)

Function: Analyze

- Data can be exported in a file to other applications
- At a minimum, the application can calculate the linguistic similarity of locations
 - user definition of “similar”
 - selected set of data
 - display of similarity (e.g. in MDS map)

Function: Interpretation

- Allow users to create an “interpretation”
 - display as a map
 - save as an object that can be reworked
- Allow “automatic” creation of interpretations
 - based on searches
 - basis for further manual rework

Function: presentation

Display as maps

- Customize maps
 - titles, descriptions
 - data and location labels
 - map keys
- Make maps “zoomable”
- Save maps as images for use in documents
 - colour, resolution, format

Interface

Use maps as an interface

- See data by location: click on the place
 - raw data
 - interpretation
 - sound and multimedia associated data

Data Entry

Digitization Projects

- Data Entry tools can make the work
 - more efficient
 - more accurate
 - less error prone

The screenshot shows a software window titled "FINNISH Project: Edit Data". The interface includes a menu bar with "File", a toolbar with "Save", "Initial", and "Exit" buttons, and a status bar showing "67" and "Map-9". The main area is divided into several sections. On the left, there is a vertical list of "Next Question" items, with labels A, B, and C pointing to the first three. In the center, there is a list of "Next Location" items, with labels D, E, F, G, H, and I pointing to the first six. On the right, there is a list of "Next Location" items, with labels D, E, F, G, H, and I pointing to the first six. The "Next Location" list includes the following items: (tarvitse-, keritse-jne.), (tarvitse-, keritse-jne.), ts: ts, 't's: t's, 't's: t's', ss: ss, ££: ££, tt: tt, tt: t, ht: ht, and ht: h.

A

B

C

D

E

F

G

H

I

FINNISH Project: Edit Data

File

Save Initial Exit

67 Map-9

Next Question 1 (tarvitse-, keritse-jne.)

(tarvitse-, keritse-jne.)

ts: ts

't's: t's, 't's: t's'

ss: ss

££: ££

tt: tt

tt: t

ht: ht

ht: h

Next Location

1 Ahlaine

114

General Online Dialect Atlas

- Use RODA and adapt to its forms and methods
(You are invited to do this)
- Create your own, but build on the design features prototyped by RODA
- Build a new, broad use application, based on the prototype
 - funding
 - time and effort

Links

RODA

<http://pi.library.yorku.ca/dspace/>
(under the “dialectology” community, “RODA” collection)

List of online dialect atlases:

<http://ericwheeler.ca/atlaslist>