Text Meets Space: Geography and text
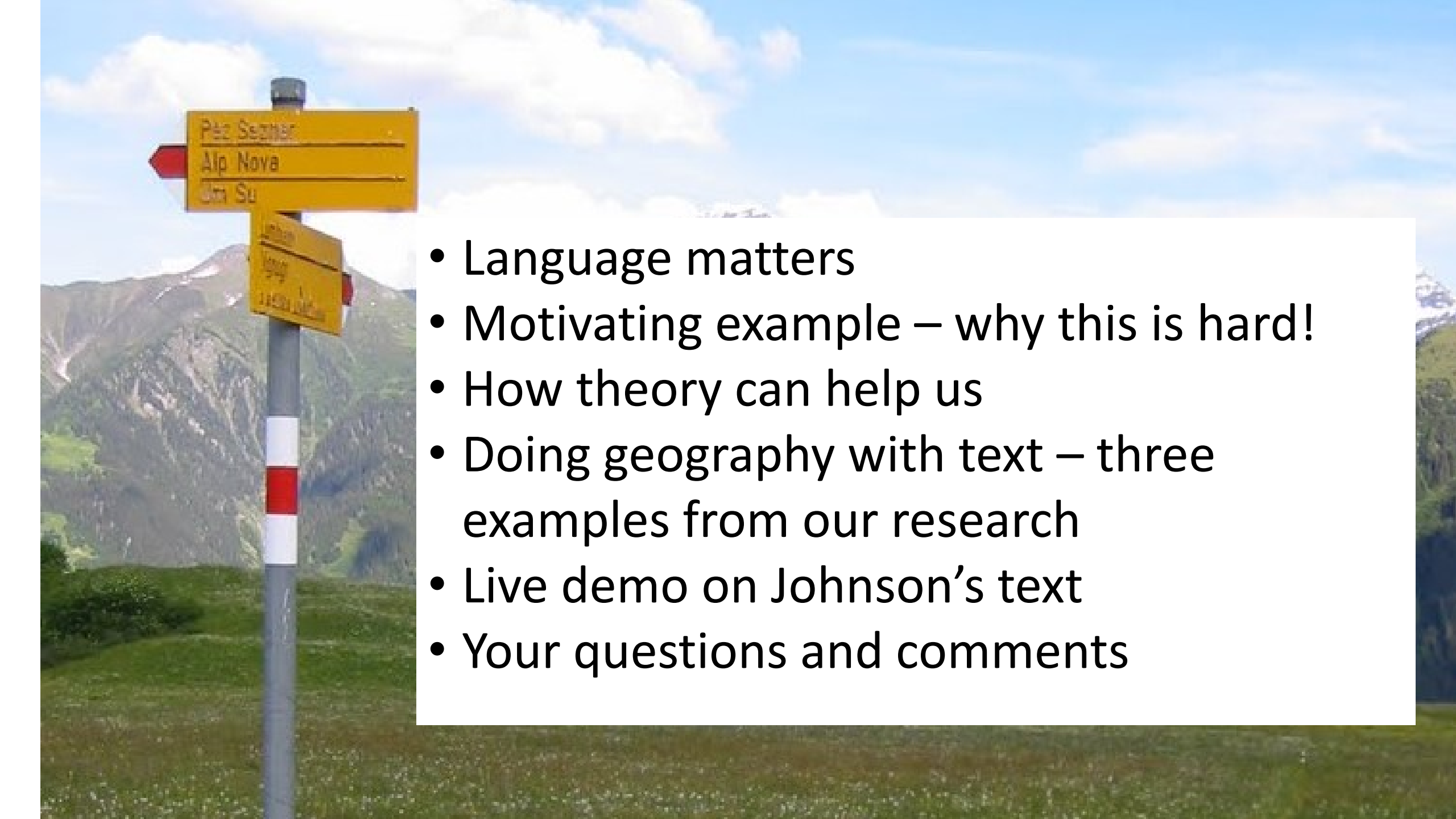
Ross Purves

- Language matters
- Motivating example – why this is hard!
- How theory can help us
- Doing geography with text – three examples from our research
- Live demo on Johnson's text
- Your questions and comments

A comment on language

"**Do** state the name of the language that is being studied, even if it's English. Acknowledging that we are working on a particular language foregrounds the possibility that the techniques may in fact be language specific. Conversely, neglecting to state that the particular data used were in, say, English, gives [a] false veneer of language-independence to the work. (Bender 2011:18)"

My examples are in **English** and **German**

I want to go further than Bender – there are **multiple ways** of thinking about and **categorising** the world (see e.g. Burenhult et al. 2017)

Let's start with a motivating example

Imagine a query of the form…

<beautiful mountains> <near> <Zermatt>

Requires methods to deal with queries of form

<theme><spatial relationship> <location>

Purves et al. 2018

Multiple research challenges include…

<beautiful mountains><near><Zermatt>

# Capturing multiple perspectives on concepts: what is a <beautiful> mountain?

"...von einer geplanten **künstlichen Erhöhung des Fletschhorns** zum Viertausender, einem Unternehmen, das heftige Diskussion ausgelöst hat. Während der **Gemeindepräsident für eine Erhöhung des Gipfels** und damit auch für einen **Zuwachs an Bedeutung plädierte**, zumal der Berg nur durch Erosion die Fehlenden sieben Meter von seiner ehemaligen Höhe verloren habe, **wandten sich die Umweltschützer** strikt dagegen, da sie die **"unberührte Natur"** gefährdet sahen."

Munz, 2001, pp. 19
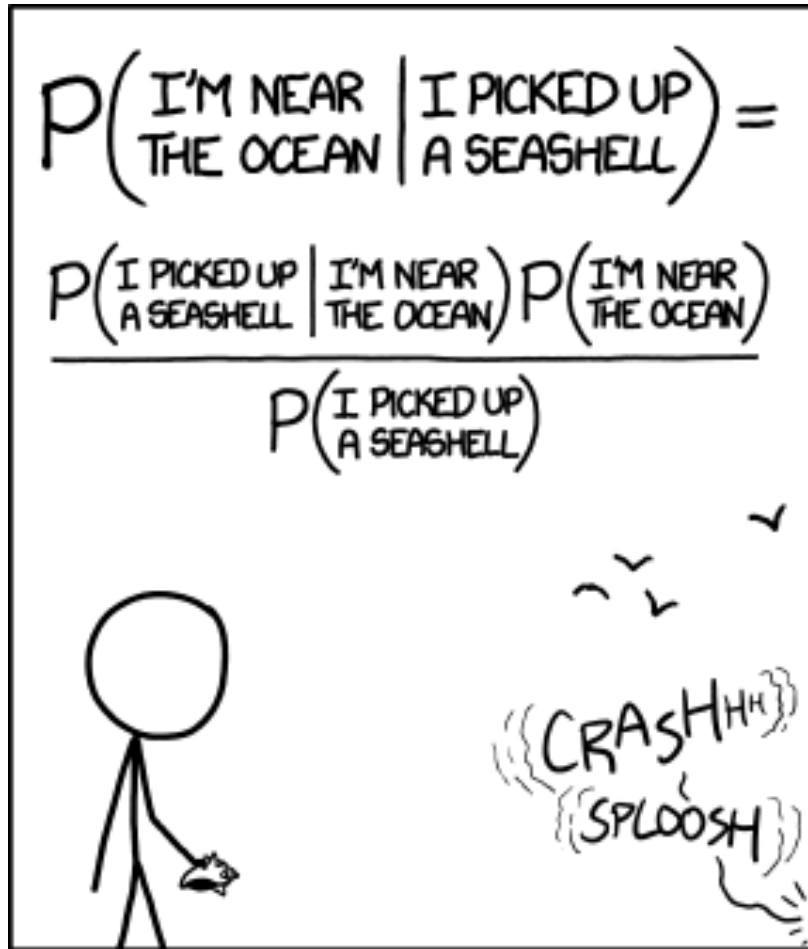
<beautiful mountains><near><Zermatt>

Spatial language

Cartoon from https://xkcd.com/1236/

Context dependent nearness to ocean defined here by **availability of shells** (and hearing the ocean)
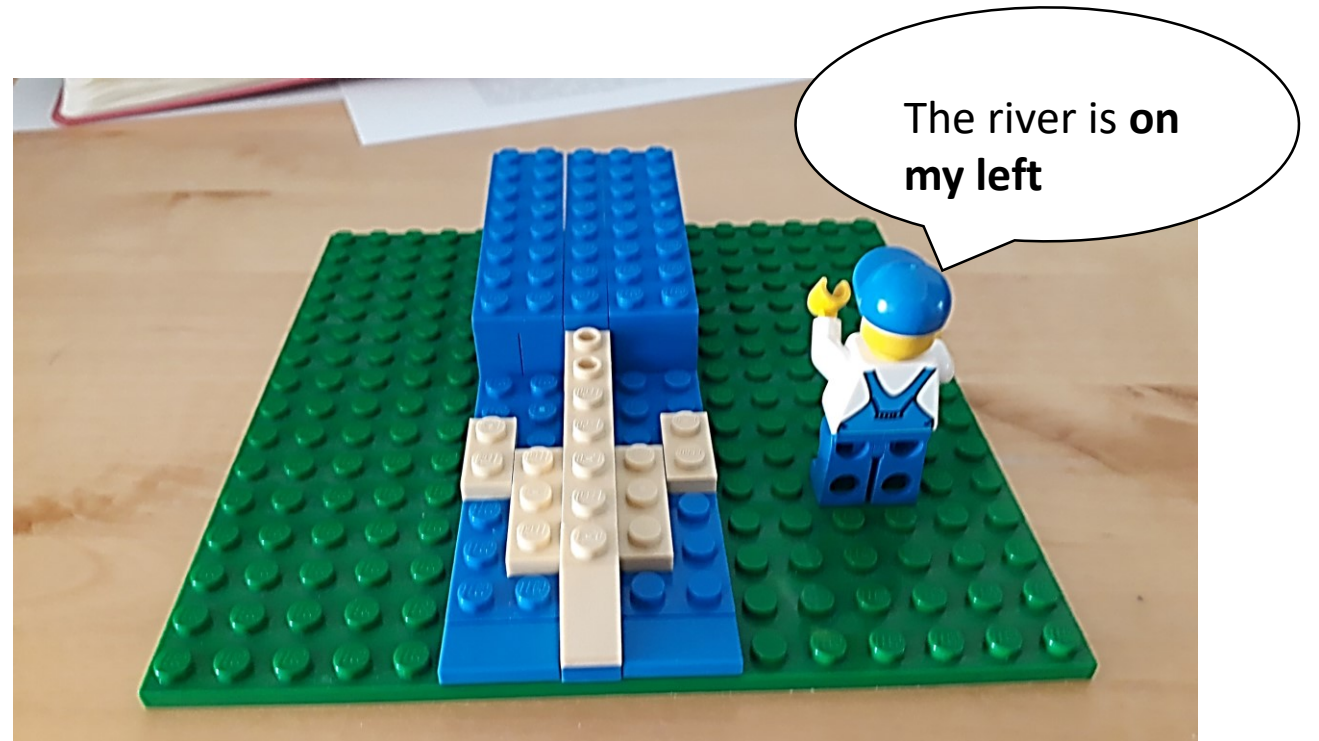
Spatial language

I'm on the **right bank** of the river

Left bank   Right bank

True left and true right bank defined by direction of flow

Cartoon from https://xkcd.com/1236/

Spatial language

The river is **on my left**

Cartoon from https://xkcd.com/1236/

Spatial language

<beautiful mountains> <near> **<Zermatt>**
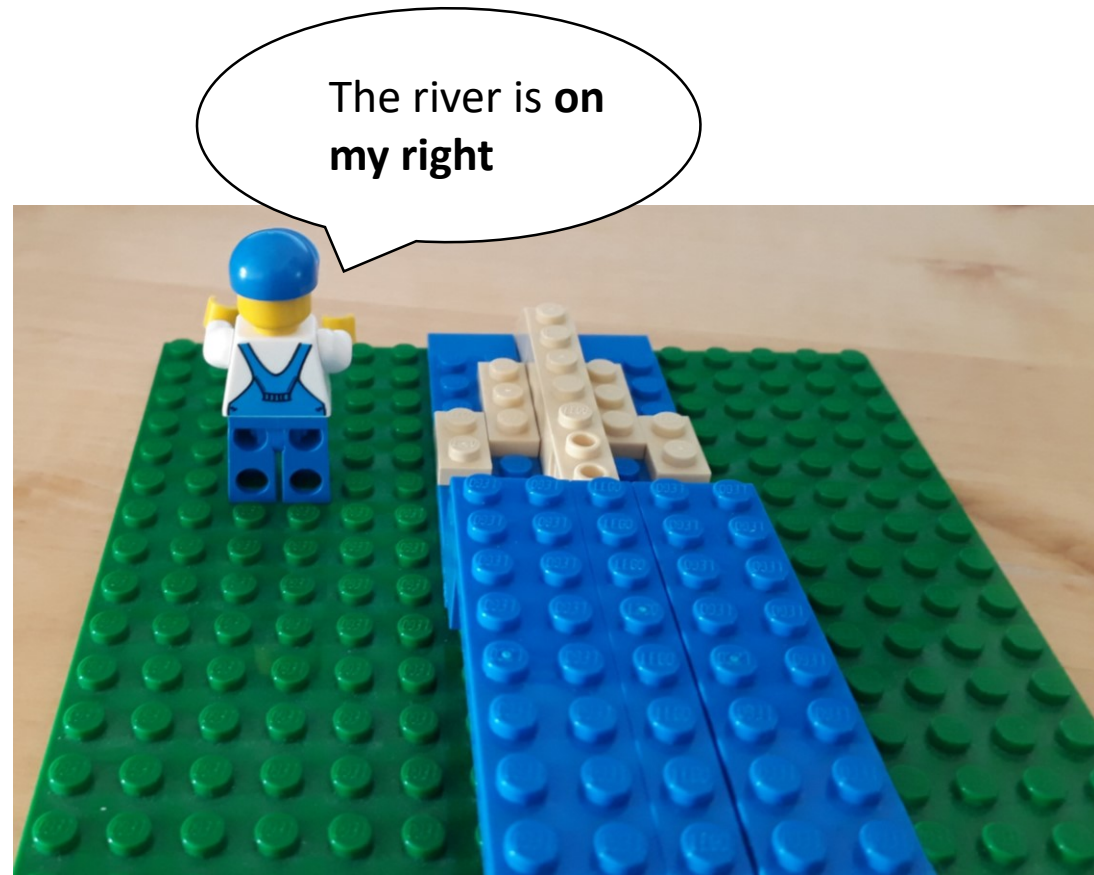
Identifying toponyms

Was steht nun auf den Thurgauer Landkarten?

Von Christof Widmer. Aktualisiert am 24.09.2009

Im Streit um die Schreibweise der Thurgauer Orts- und Flurnamen wird es eng. Die neuen Karten gehen bald in den Druck.

(Bild: Tom Werner)

Artikel zum Thema

▶ Wenn die Karte Velofahrer total verwirrt

Kleinere Auflage als Lösung

Da die Neuauflage der Kartenblätter kaum aufgehalten werden kann, steht eine pragmatische Lösung im Raum: Es könnte

Die Vorbereitungen im Bundesamt für Landestopografie (Swisstopo) sind schon weit gediehen. Die Luftaufnahmen vom Thurgau sind gemacht. Die neusten baulichen Veränderungen können so für die Neuausgabe der Thurgauer Kartenblätter der Landeskarte berücksichtigt werden. Die elf Blätter der Karte im Massstab 1:25'000, die den Kanton abdecken, werden nächstes Jahr neu gedruckt. Ein Jahr später folgen die Karten im Massstab 1:50'000. Die in der Karte verzeichneten Orts- und Flurnamen übernimmt das Bundesamt vom Thurgauer Amt für Geoinformation.

Hier liegt das Problem: Wie die Lokalnamen im Thurgau geschrieben werden sollen, ist inzwischen

1. Same referents can be referred to by **multiple toponyms**
2. Referents can be **spatially vague**
3. Toponyms may be used **vernacularly**
4. Toponyms **are not constant** in time or space

From www.tagesanzeiger.ch and www.guardian.co.uk

Resolving toponyms

Land down blunder: teen heading to Australia lands in Sydney, Nova Scotia

Milan Schipper ended up in a snowy city 10,000 miles from his intended destination – and he's not the first tourist to make the mistake, even on his flight

Referents may **share** the **same toponym**

Some useful concepts

Naive geography

"Field of study ... concerned with formal models of the common-sense geographic world" (Egenhofer & Mark 1995)

Some useful concepts

Naive geography

"Field of study ... concerned with formal models of the common-sense geographic world" (Egenhofer & Mark 1995)

Basic levels

Useful abstractions in explaining the world to one another (e.g. chair and table rather than furniture or desk chair) (Tversky & Hemenway 1983)

Some useful concepts

Naive geography

"Field of study … concerned with formal models of the common-sense geographic world" (Egenhofer & Mark 1995)

Basic levels

Useful abstractions in explaining the world to one another (e.g. chair and table rather than furniture or desk chair) (Tversky & Hemenway 1983)

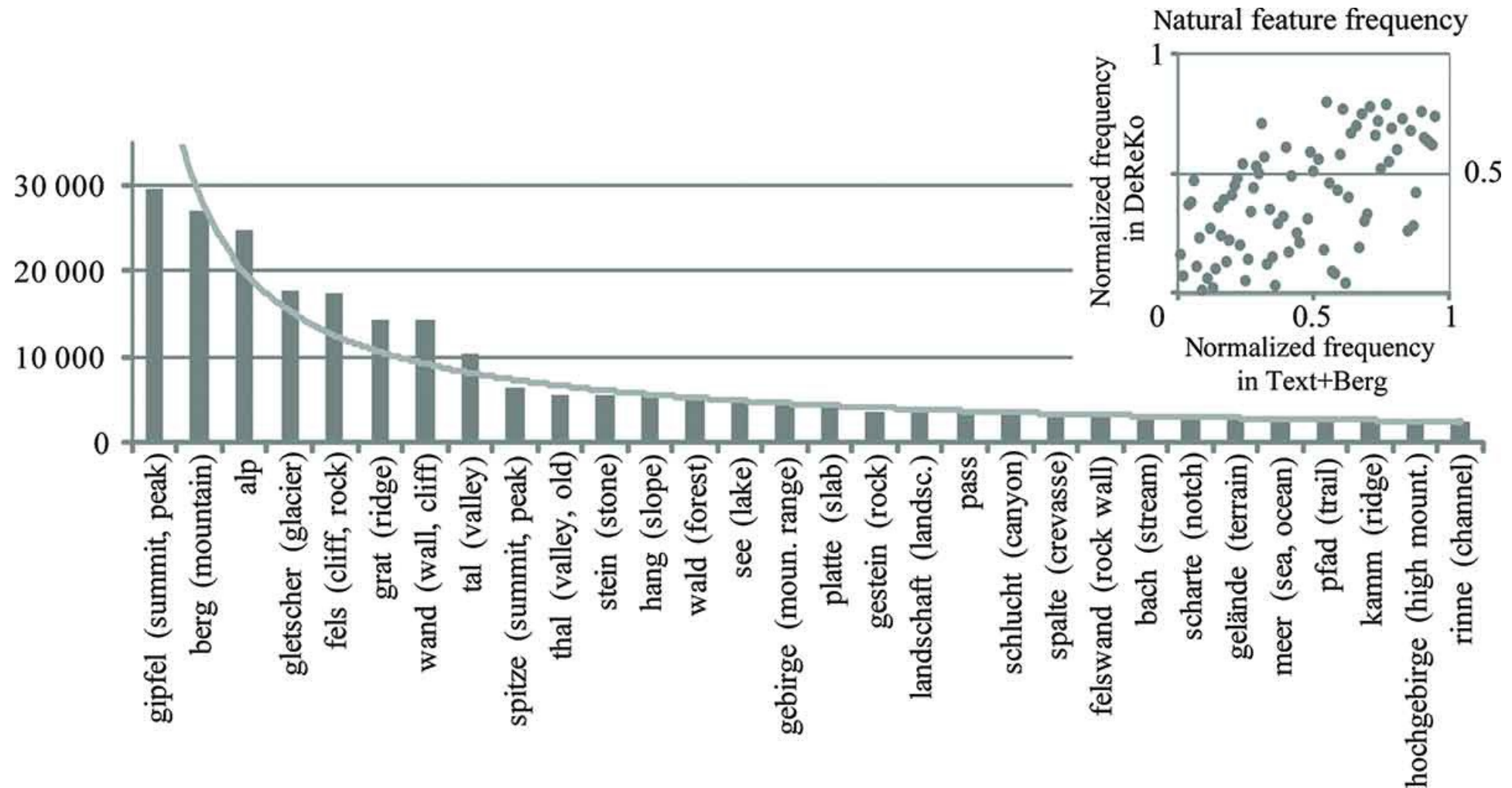**Qualitative spatial reasoning**

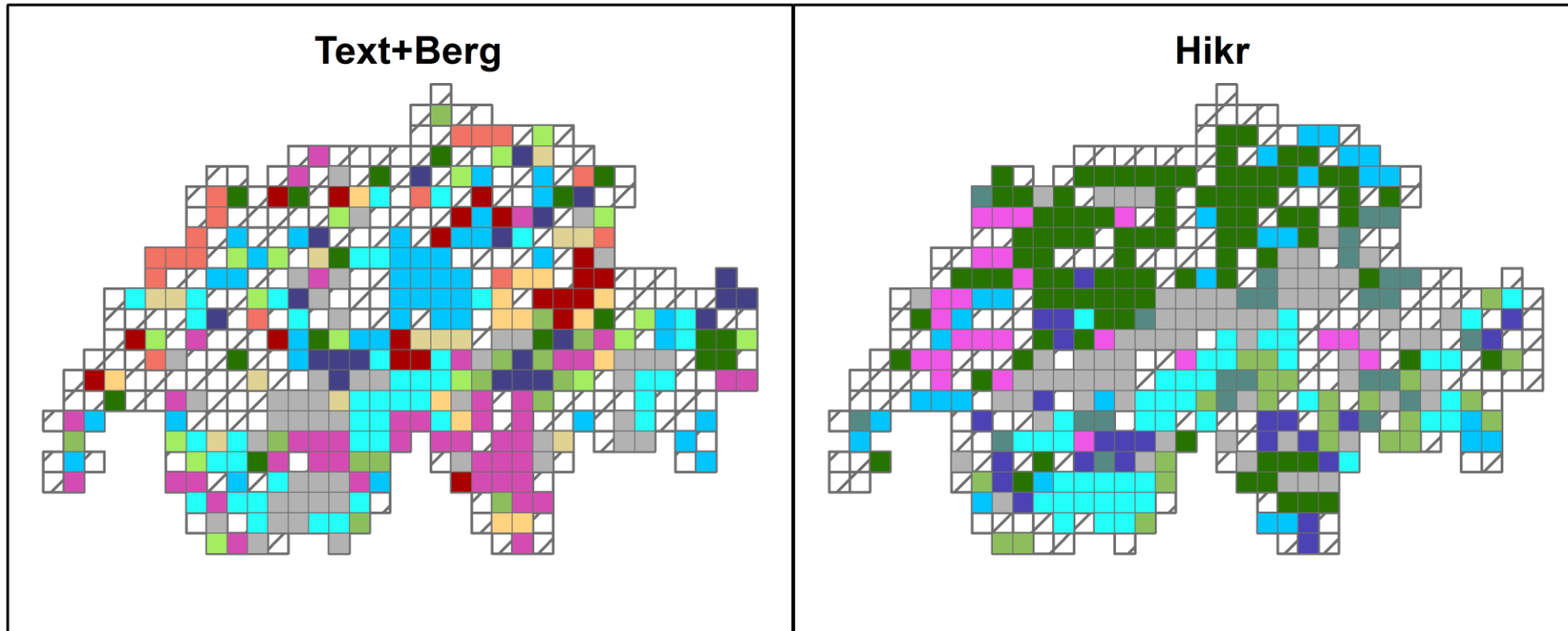Allow reasoning about space in human-like ways (Moratz 2017)

Exploring our challenges using text

1. **Landscape concepts** in a text corpus

2. **Spatial relationships** derived from online advertising

3. **Delineating vernacular toponyms usage** using tags in user generated content

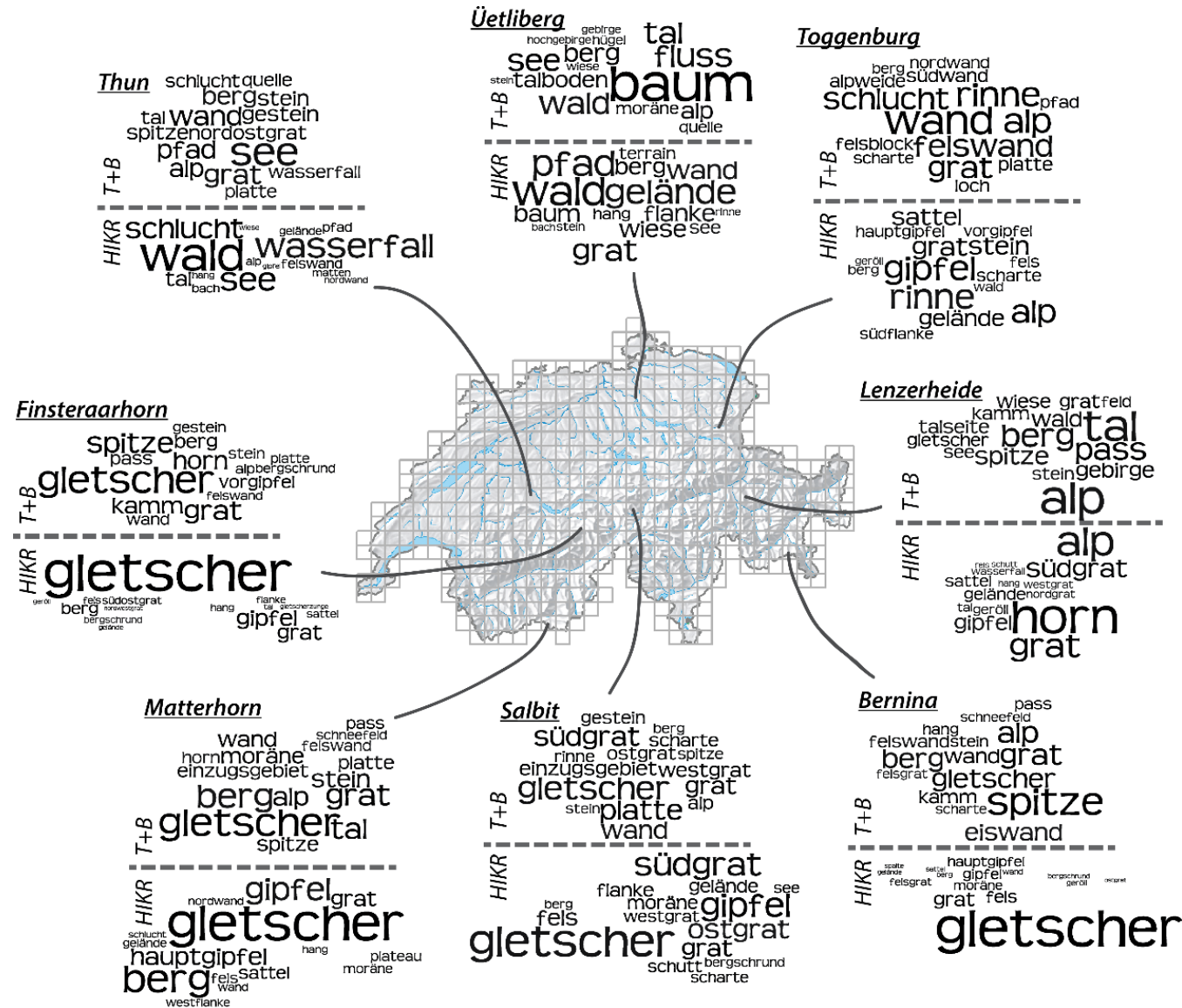# Landforms in Swiss Alpine Club Yearbooks

Derungs & Purves 2014

Most prominent natural feature (tf-idf)

Text+Berg

Hikr

Pixel's most prominent natural feature

- bach (T+B)
- berg (T+B)
- fluss (T+B)
- gestein (T+B)
- gipfel (Hikr)
- **gletscher**
- **grat**
- horn (T+B)
- **pass**
- pfad (Hikr)
- schlucht (Hikr)
- **see**
- tal (T+B)
- **wald**
- wand (T+B)

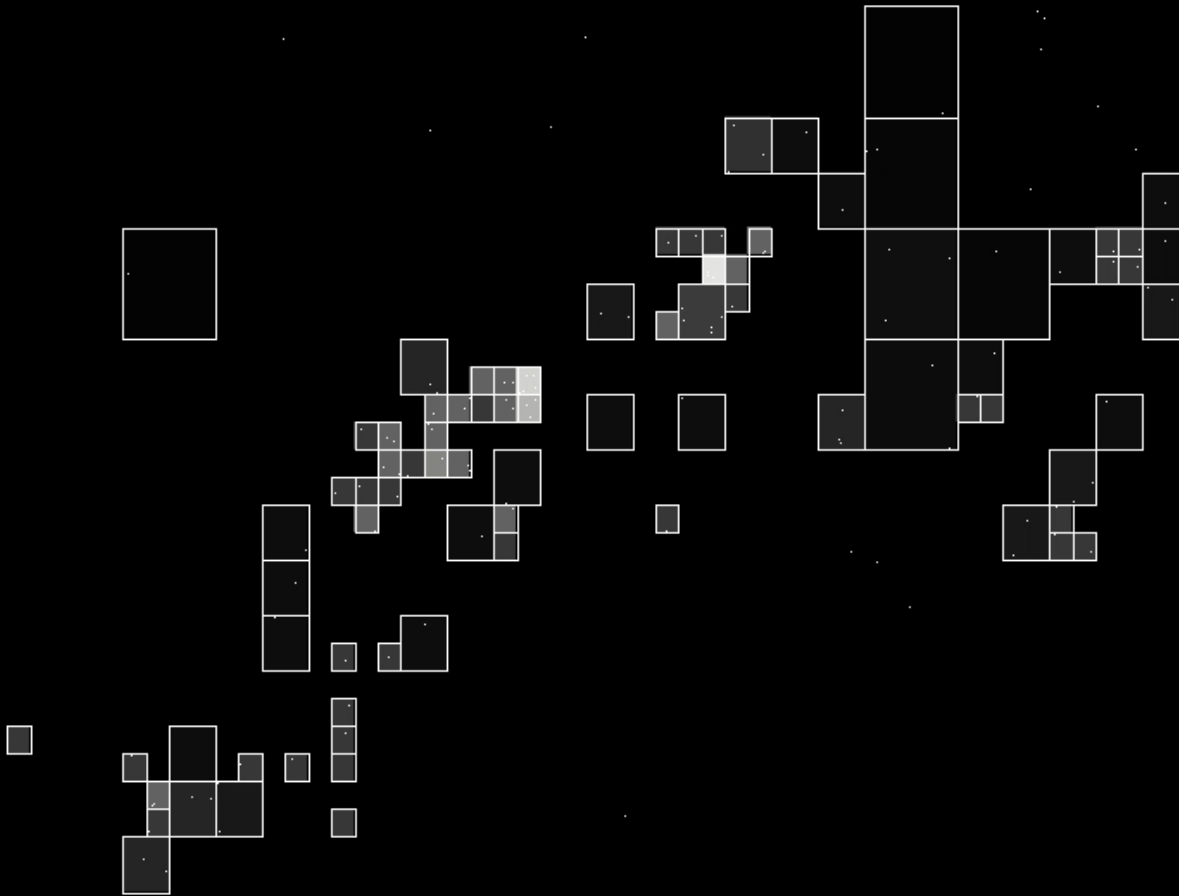Term occurs in less than 10 pixels

TF-IDF descriptions of grid cells

Derungs & Purves 2016

Up to 1868 a total of 190 documents

steil x 76
tief x 42
klein x 41
groß x 37
weit x 37
lang x 35
südlich x 34
östlich x 33
hoch x 32
bedeutend x 31
westlich x 29
nördlich x 26
eben x 26
mächtig x 26
jäh x 26
ausgedehnt x 24
felsig x 22
breit x 21

Adjectives
related to *Gletscher*
(glacier) in time and
space in Text+Berg

Aside on research questions

Research questions **matter** – here we posed the following

- How can **natural landscape descriptions** … be **linked** to **footprints**?
  (methodological – **toponym recognition & resolution**)
- **What natural features occur** in Swiss alpine descriptions?

  (descriptive - **ranking**)

- How are **different alpine landscapes** described by natural features?
  (comparative - **similarity**)

- Our underlying aim was to use text to **"do" geography**
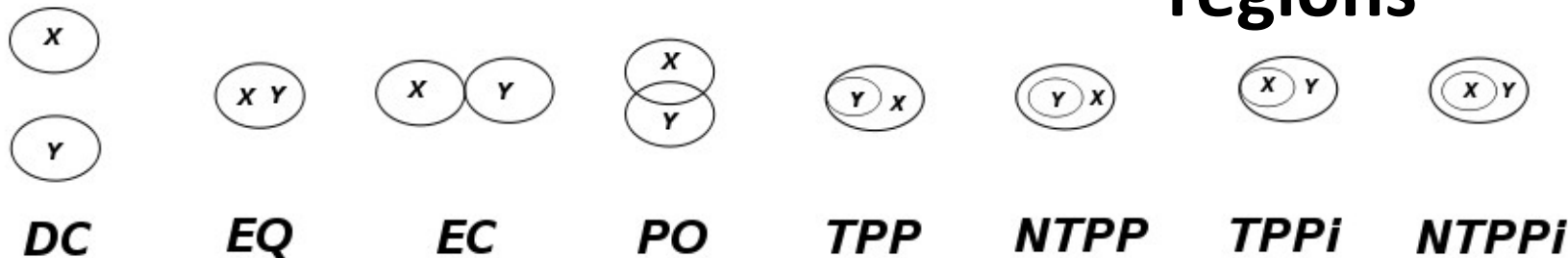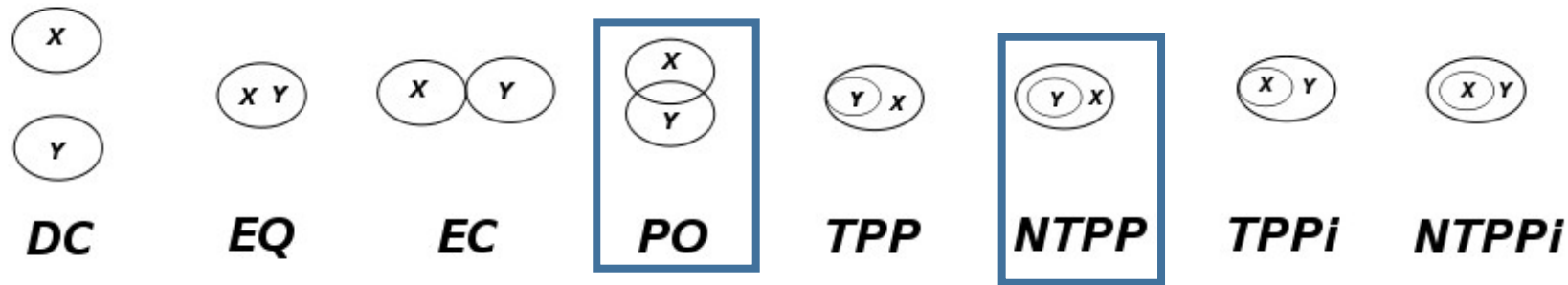
Spatial relationships mined from GumTree
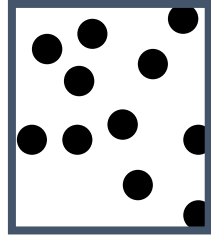
Regions derived as bounding boxes for toponyms in Cardiff
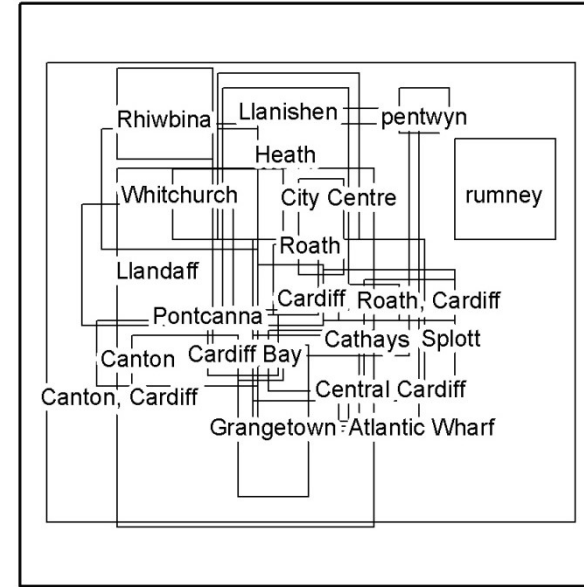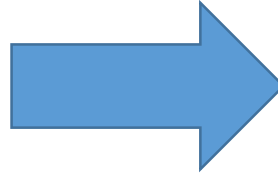
Multiple overlapping regions

Qualitative spatial relationships

Cohn et al., 1997; Twaroch et al., 2009

Qualitative spatial relationships

Toponym	Cardiff Bay	Canton	Cathays	Roath	Cardiff	Rumney
Cardiff Bay	EQ	PO	PO	PO	PO	DC
Canton	PO	EQ	PO	PO	NTPP	DC
Cathays	PO	PO	EQ	PO	NTPP	DC
Roath	PO	PO	PO	EQ	NTPP	DC
Cardiff	PO	NTPPi	NTPPi	NTPPi	EQ	NTTPi
Rumney	DC	DC	DC	DC	NTPP	EQ

Relations for one of the previous geometries:

Rumney is in Cardiff   Cardiff Bay is near Canton

Twaroch et al., 2009

Delineating vague and vernacular toponyms with Flickr tags

The Loop   Downtown   City

Hollenstein and Purves, 2010
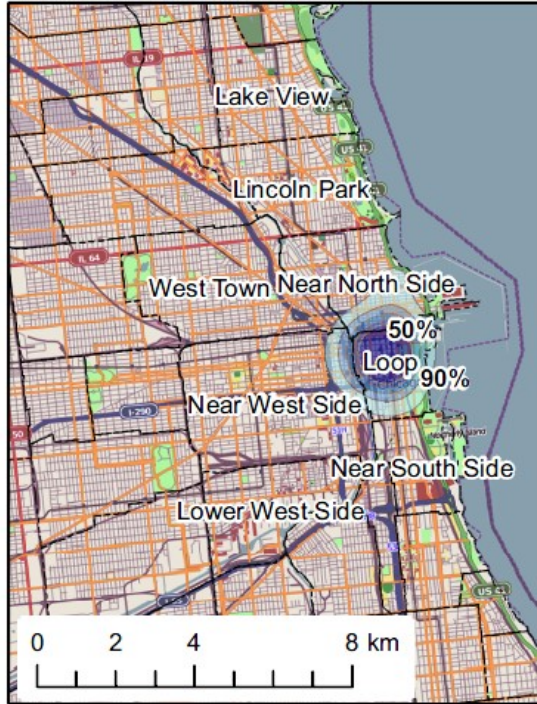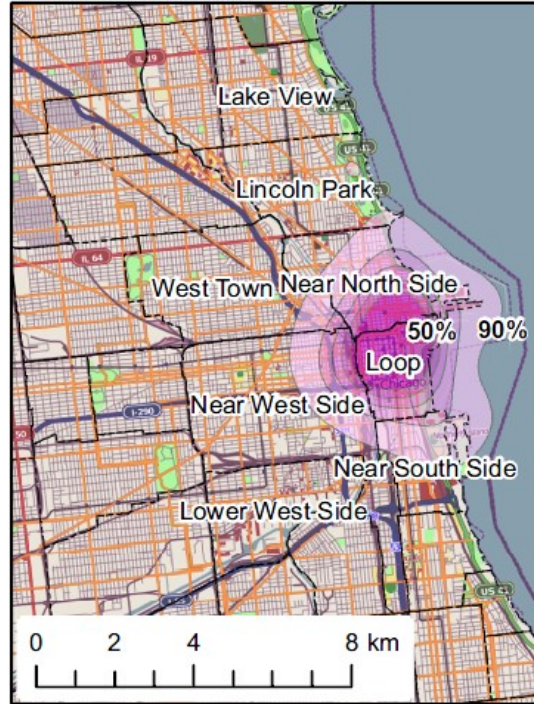
Let's introduce a new text
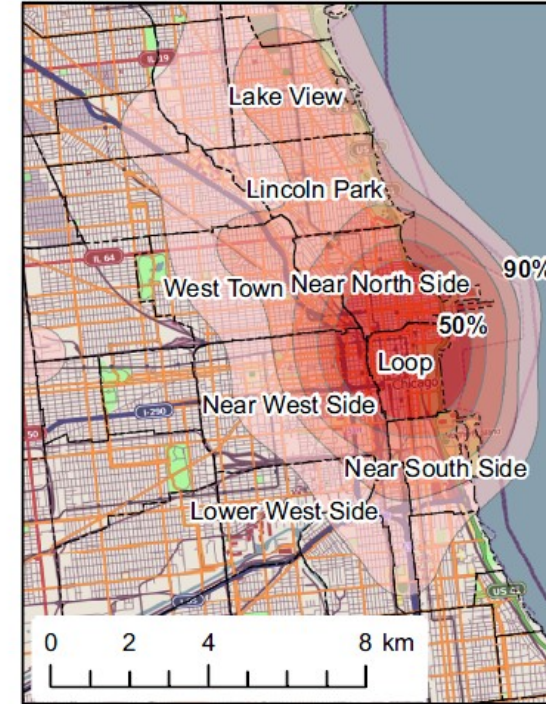
"Till the Union made them acquainted with English manners, the culture of their lands was unskilful, and their domestic life unformed; their tables were coarse as the feasts of Eskimeaux, and their houses filthy as the cottages of Hottentots."

Samuel Johnson's *A Journey to the Western Islands of Scotland*, published in 1775, is both an early example of **travel writing** and a remarkably detailed piece of **cultural anthropology.** It records a journey made by Johnson and his friend and biographer, James Boswell, in the summer and autumn of 1773. Johnson was then 63 years old, and had rarely travelled outside London: the Scotland he describes is a **remote and alien place** in which everything he experiences, from the **lack of trees to the windows of Scottish houses**, is seen as evidence of the country's uncivilised state. His descriptions of a country still adjusting to the effects of the Highland clearances are **vivid, opinionated and wide-ranging**, encompassing subjects such as the Gaelic language, the primitive nature of Scottish shoes, and the consumption of whisky before breakfast.

Experiment 1

How does Johnson's use of concepts compare with modern experiments where participants were asked to list "natural earth formations" (Battig & Montague 1969[1])?

We'll count and explore context using

AntConc

```
mountain 36
  hill 32
valley 10
 river 8
 rock 58
 lake 6
canyon 0
 cliff 0
ocean 5
 cave 21
```

[1]Unfortunately this paper is not freely available

Corpus Files

johnson_pure_t

Concordance   Concordance Plot   File View   Clusters/N-Grams   Collocates   Word List   Keyword List

Concordance Hits 36

Hit	KWIC	File
1	alley. Such are the effects of habitation among mountains, and such were the qualities of the Highlanders,	johnson_p
2	outhern inhabitants of Scotland, the state of the mountains and the islands is equally unknown with that	johnson_p
3	for themselves : Hither the young nymphs of the mountains and valleys are sent for edu- cation, and	johnson_p
4	, and found a wider basis of analogy. Regions mountainous and wild, thinly inhabited, and little cultivated	johnson_p
5	these any aquatick plants. The vallies and the mountains are alike darkened with heath. Some grass, how-	johnson_p
6	desire me to suppose, for the manners of mountaineers are com- monly savage, but they are rather	johnson_p
7	it into small cheeses. The stags of the mountains are less than those of our parks, or	johnson_p
8	ing in security the more fruitful provinces. As mountains are long before they are conquered, they are	johnson_p
9	carried their weapons to visits, and to church. Mountaineers are thievish, because they are poor, and having	johnson_p
10	he and his adherents were suffocated together. Mountaineers are warlike, because by their feuds and competiti	johnson_p
11	f adventures, stratagems,surprises and escapes. Mountainous countries are not passed but with difficulty, not	johnson_p
12	as these before us are generally distinguished. Mountainous countries commonly con- tain the original, at lea	johnson_p
13	solitude. The way makes a flexure, and the mountains, covered with trees, rise at once on the	johnson_p
14	skip over the bog, or clam- ber the mountain. For a campaign in the wastes of America,	johnson_p
15	, ex- cept against the king. The inhabitants of mountains form di- stinct races, and are careful to	johnson_p
16	bog has firmness to sustain them: besides that, mountaineers have an agility in climbing and descending distin	johnson_p
17	life in very diffe- rent climates; and the mountains have no greater examples of age and health	johnson_p
18	there were goats feed- ing or playing. The mountains have red deer, but they came not within	johnson_p
19	pecu- liar form of the country, broken by mountains into many subdivisions scarcely accessible but to	johnson_p
20	effect of a thousand streams poured from the mountains into one channel, struggling for expansion in a	johnson_p
21	always necessary : but because that which is not mountain is commonly bog, through which the way must	johnson_p
22	a, the representation will be fallacious. These mountains may be properly enough measured from the inland	johnson_p
23	Trent. A tract intersected by many ridges of mountains, naturally divides its inhabit- ants into petty n	johnson_p
24	ssia.) This young gentleman was sporting in the mountains of Sky, and when he was weary with	johnson_p
25	pri- mitive manners, whether they live among the mountains or in the islands; and in that sense	johnson_p
26	ein- forced by winter. The height of mountains philosophically considered is properly computed f	johnson_p
27	sure to happen before the company dis- persed. Mountainous regions are sometimes so remote from the seat	johnson_p
28	to con- template the appearance and properties of mountainous regions, such as have been, in many countries,	johnson_p

Search Term ☑ Words ☐ Case ☐ Regex   **Search Window Size** 50

mountain*   Advanced

Start   Stop   Sort   Show Every Nth Row 1

Total No.
1

Files Processed

Kwic Sort
☑ Level 1 1R   ☑ Level 2 2R   ☑ Level 3 3R

Clone Results

Experiment 2

How does Johnson use the spatial prepositions "in", "near" and "through"?

Rounding up

- I listed **three challenges** in dealing with geographic text
- I **presented some useful theory** from my field

  N.B. there is lots of other work to look e.g. linguistics, psychology, computer science – making progress requires meaningful **interdisciplinary**

- I gave **three examples** where text helped address the challenges
- We started to **(naively) explore** a text using AntConc