



## The Project *Alpenwort* A corpus building bridges into different fields

Claudia Posch  
Gerhard Rampl

### *Alpenwort* - summiting strategy

'The main goal of this proposal is to make this source accessible for the scientific community enhanced by exhaustive linguistic annotation.'



## *Alpenwort – summiting strategy*

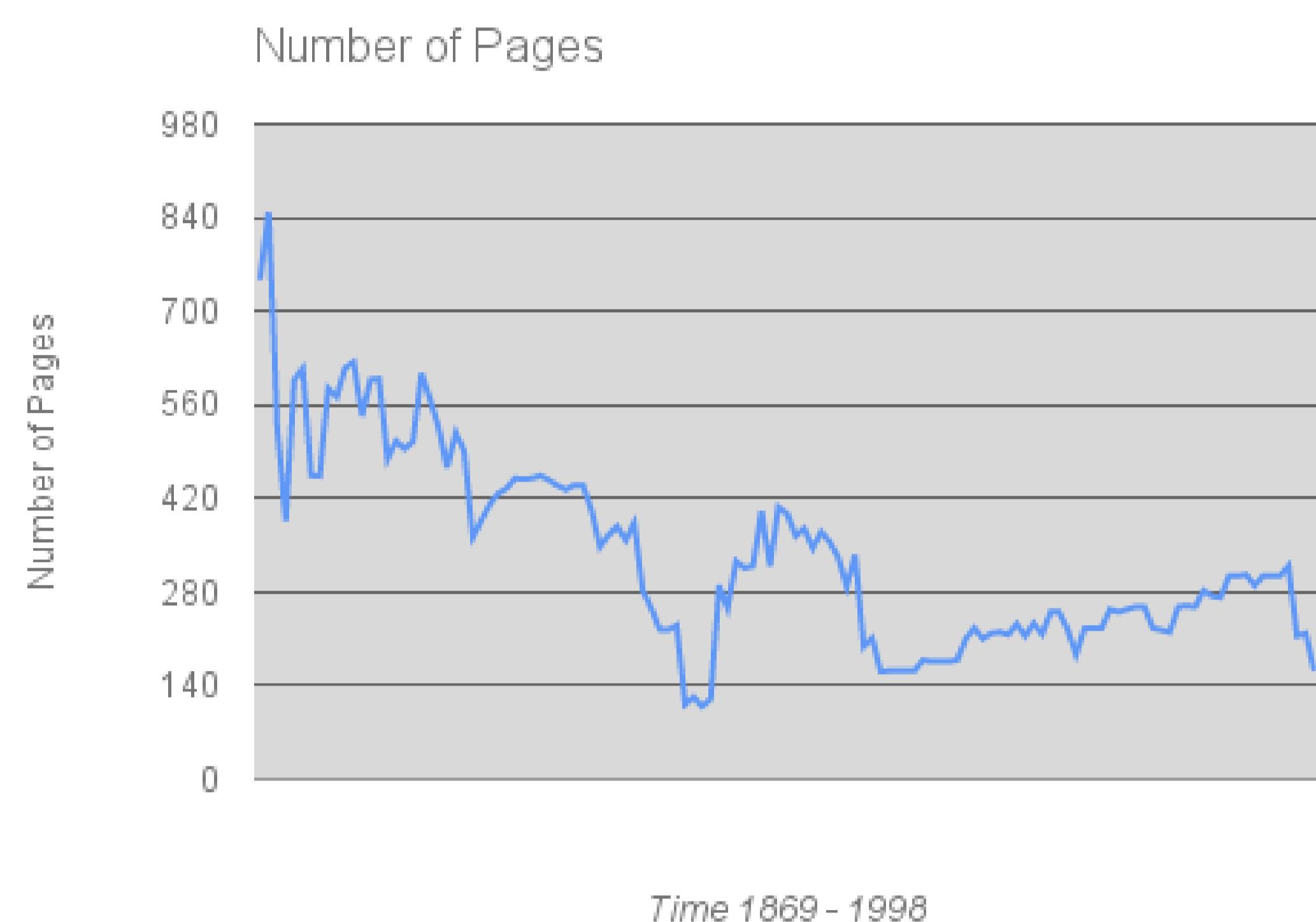
„The main goal of this proposal is to make this source accessible for the scientific community enhanced by exhaustive linguistic annotation.“

- » OCR and structure recognition (automated)
- » structure correction of OCR in FEP (semi-automated)
- » export of data in XML and version controlling (GITlab)
- » automated data annotation and enrichment with pipeline (automated)
- » pre-publication for testing (Hyperbase, CQP-Web)
- » OCR & text quality correction (semi-automatic)
- » re-tagging & publication

## Gear list

- » duration: 10/15/2014 – 10/14/2017
- » funding: go!digital call of the Austrian Academy of Sciences 2013
- » staff: Daniela Feistmantl, Bettina Larl, Irina Windhaber
- » partners:
  - Austrian Alpine Club,
  - DEA – Digitisation and Electronic Archiving (@UIBK),
  - Institute of Computational Linguistics of the University of Zurich,
  - DBIS – Databases and Information Systems (@UIBK),
  - Unit for Surveying and Geoinformation (@UIBK)
- » linguistically annotated corpus of the ZAV (126 years, 123 volumes)

## Gear list



min. 109 – max. 848 pages per volume

total: 43.383 pages

average: 338 pages

1915 – 1961 gothic script

all others in antiqua script

1961/1962 in both

## Cruxes climbed

- » the corpus is completed as planned
- » annotation and correction were done automatically
- » semi-automated annotation and correction have been added
- » meta-data TEI conformant:
  - one XML-file per volume
  - TEI-header with project information similar to BNC
  - annotation of volumes: file name, volume, year, decade
  - annotation of articles: article number, author, title, pages,...
- » POS-tagging: automatically done by pre-trained TreeTagger + enhanced lexicon (historical spellings!) → output in XML format/CQP-web format

## Cruxes climbed

- » automated structure recognition had to be corrected manually
- » particularly problematic with two-column pages
- » particularly problematic with fancy designs
- » every volume is different has a certain degree of individuality!
- » OCR-correction of Gothic Script (1916-1961)
- » use of corpus tools to find errors (AntConc)
- » partly systematic errors – semi-automated with RegEx
- » some paragraphs were toggled

Das Pitzthal. 133



**Nasse und trockene Fahrten zwischen Ill und Inn**

author von  
**O. v. Pfister**  
in München.

paragraph

Es war ein wochenlanger Abschied zwischen Edderhucher, seinem alten, bewohnten Hause, und mir, als er, am Ende seiner Wünsche angelangt, das Watzenschenhaus als Hauswart bezog und wir damit für immer auf unsere gemeinsame Hochtour verzichten mussten. Für mich bedeutete dieser Abschied außerordentlich viel.

Bauer freut sich seines Sieges und baut seine Kartoffeln sowie seine Gerste an, während der Schnee rings um den freigegebenen Acker noch seine Dämme und Schutzwähren in nicht unbedeutlicher Höhe aufrecht stehen hat.

Sind alle Felder in der Thalsohle bestellt, — was immerhin 3 bis 4 Wochen später geschieht, als im unteren Pitzhale — ist das Heu eingehiemt, so geht es mit Sack und Pack, Gross und Klein, sogar mit dem Kleinvieh, auf die Berge, um an manchen Orten eine ganze Woche lang über Abgründen und Felsen zu heuen oder an steilen Abhängen mit Steigeisen das nötige Futter für den Winter zu sammeln. Bei hereinbrechender Nacht steigt man wieder ins Thal herunter. Es ist für den Wanderer so recht idyllisch, wenn dann aus einigen munteren jungen Kehlen ein Tiroler Volkslied in die Abendluft hinausklingt.

Das oben auf den Bergen gewonnene Heu wird auch hier in kleinen Holzstadeln bis zum Winter verwahrt, oder, wo solche fehlen, in grösseren Stössen aufgeschichtet und gegen Regen u. s. w. mit grossen Leintüchern überspannt. Ende September kann man viele solche Heuschober sehen. Im Winter wird das Bergheu zu Thal gebracht.

Da geht es mit einem Fuder auf Tannenästen über den Schnee hin, manchmal über Abgründe oder an Felswänden vorbei auf einem von der Natur freigelassenen Wege, so dass Ueingeweihte öfters die Frage hören lassen: wie kommen die Leute denn da mit dem Heu herunter?

Im oberen Pitzhale bestehen keine oder fast keine allgemeinen Alpenweiden für das Milchvieh, vielmehr bleibt dieses im Sommer daheim. Da findet man keine grossartige Alpen-Auffahrt und -Abfahrt mit allen möglichen Feierlichkeiten wie anderswo. Die Kühe werden daheim auf die Weide getrieben, welche für jeden Besitzer abgegrenzt ist. Wenn diese Weideplätze auch rauh und steinig sind, so müssen sie doch als vor treffliche angesehen werden nach dem Sprichworte: »Das Beste ist das Fleisch beim Bein und das Gras beim Stein«.

Das Erträgniss der Viehzucht ist zwar kein geringes, jedoch sind auch die Aus-



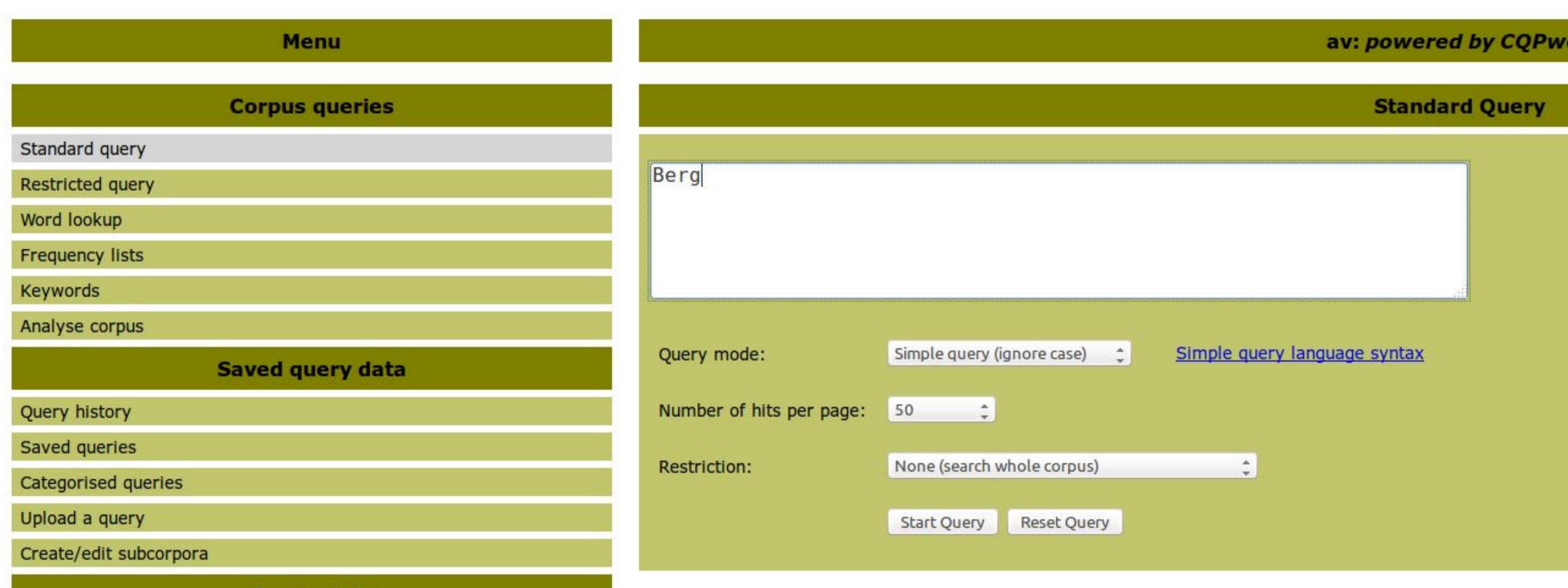
Bearbeitung eines Ackerfeldes im Pitzhale.

## Cruxes climbed

- » automated structure recognition had to be corrected manually
- » particularly problematic with two-column pages
- » particularly problematic with fancy designs
- » every volume is different has a certain degree of individuality!
- » OCR-correction of Gothic Script (1916-1961)
- » use of corpus tools to find errors (AntConc)
- » partly systematic errors – semi-automaged with RegEx
- » some pages were toggled

## North face

- » implementation into *CQP-Web* on a local server at DBIS
- » implementation into *Hyperbase*



The screenshot shows the CQP-Web interface. On the left is a sidebar menu with the following items:

- Menu
- Corpus queries
  - Standard query
  - Restricted query
  - Word lookup
  - Frequency lists
  - Keywords
  - Analyse corpus
- Saved query data
  - Query history
  - Saved queries
  - Categorised queries
  - Upload a query
  - Create/edit subcorpora
- Corpus info

The main area is titled "av: powered by CQPwe" and "Standard Query". It contains a search input field with "Berg" typed into it. Below the input field are several configuration options:

- Query mode: Simple query (ignore case) [Simple query language syntax](#)
- Number of hits per page: 50
- Restriction: None (search whole corpus)
- Buttons: Start Query, Reset Query

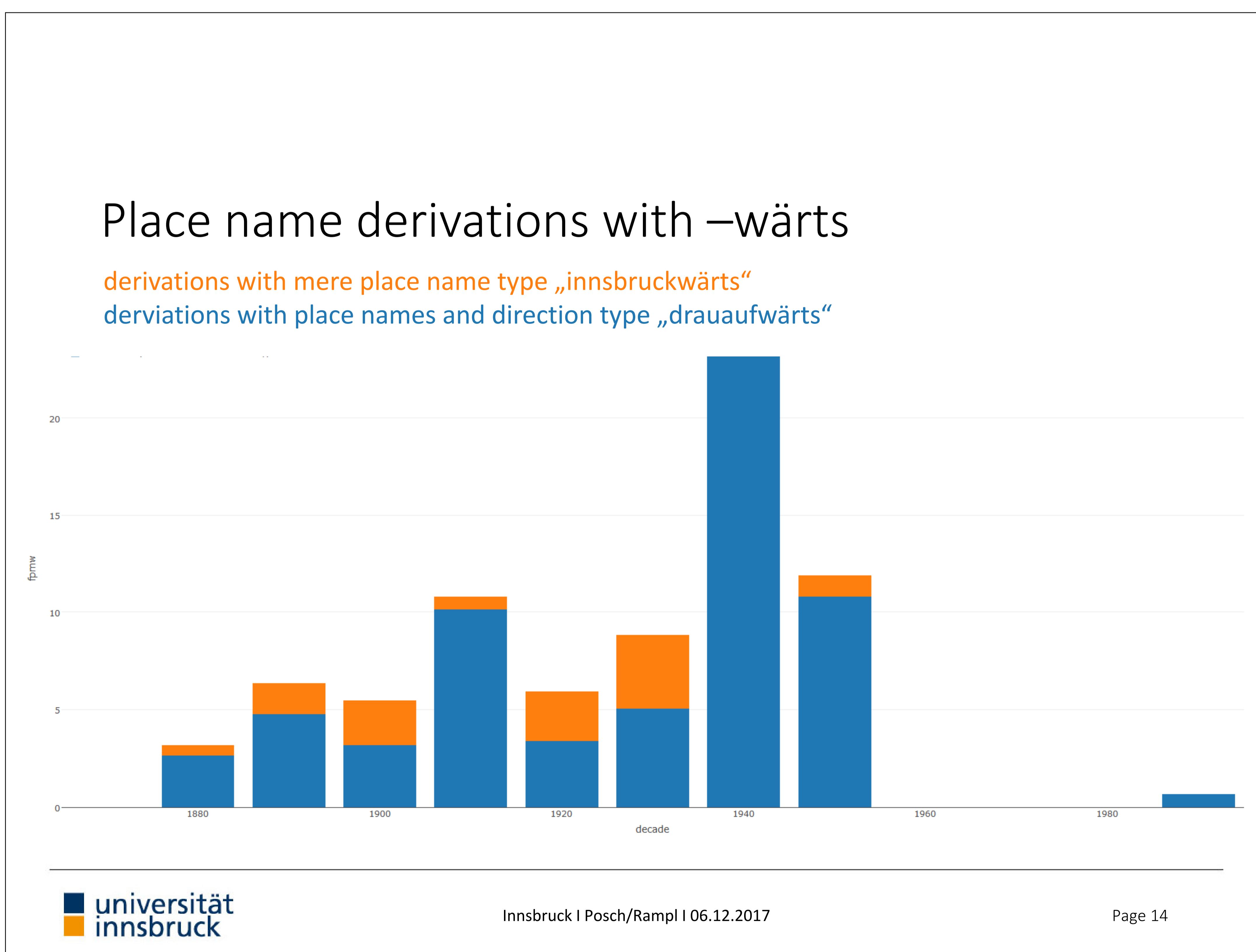
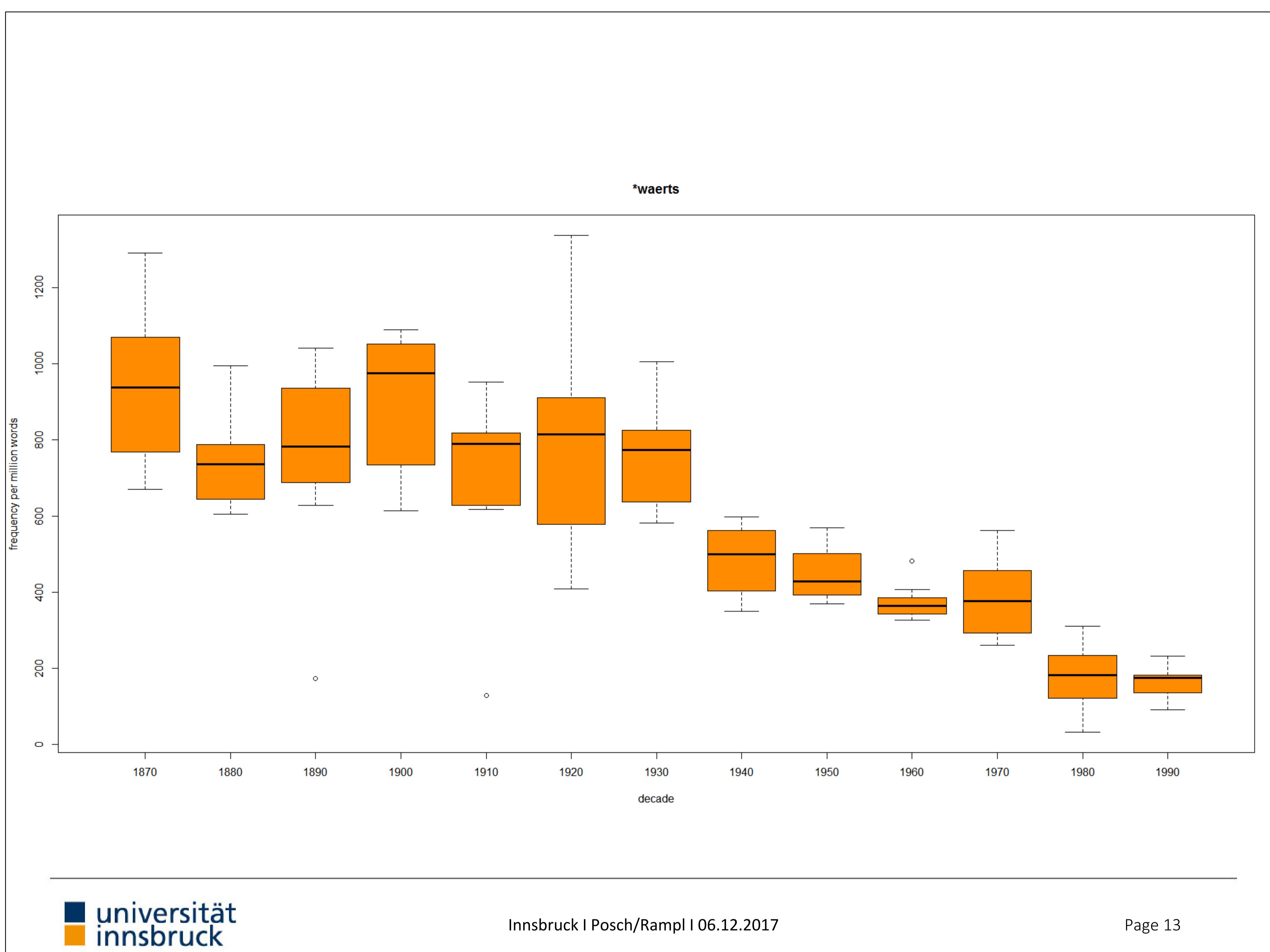
## South face

- » implementation into *CQP-Web* on a local server at DBIS
- » implementation into *Hyperbase*

partie gauche	pivot	partie droite
	Berg als Hochgrindl , während es zweifellos der Botzer ist .	
	Berg auf dem Ferner bis zur Südseite zu umgehen .	
	Berg von allen Seiten umgebenden Felsenmassen .	
	Berg erstieg , sprach im Fremdenbuche zu Heiligenblut seine Vermuthung dahin aus , dass der	
	Berg nordöstlich davon , zwischen Hochgruber- und Bockkahrgletscher gelegen , der Kleine , der Gipfel	
	Berg der Ostalpen .	
	Berg u.	
	Berg . — wenn ihm unbekannt , dass Gader- , Enne-berger- , Manubio- ,	
	Berg = Berg , Km .	
	Berg mit dem Kleinen Bärenkopf , der sich als langgestreckter Eisrücken über ein morschem Schiefergerüste	
	Berg zwischen dem Schwarzen	
	Berg Wilden Kreuzjoch , der Burgum .	
	Berg , ist der einzige Fundort , der den Wissens Zirkon , mit Iridolith ,	
	Berg Rundschau aufnahm , weil mir die Gewisheit mangelte , ob es auch wirklich der Berg ist .	
	Berg wie Thal mit ihren auserlesenen Lieblingen .	
	Berg Angesichts der Alpen sich geltend macht , ist wohl die nach der Bildung von Berg und Thal , nach der Form der Gesteinsmassen im Grossen , nach ihren Lagerungsverhältnissen	
	Berg die Bildung von Berg und Thal eine nähere Betrachtung .	
	Berg hohe Berg ( Fig .	
	Berg dieses Namens .	
	Berg „Fluchttog! ? Ausser Nicodem Klotz kannte Niemand aus der jetzigen Generation einen Berg dem Reize des Panoramas bietet unser Berg dem Geologen reichlich Gelegenheit zu interessanten Studien .	
	Berg die Centralalpen vertasse , verweise ich noch auf den nunmehr erschienenen zweiten Theil der Berg - und Gleitscherreisen des Herrn Dr.	
	Berg Über diesen oft genannten Berg sei mir hier eine kurze Notiz gestattet : Der Gebirgsstock des Steinernen Meeres findet	
	Berg verblichen und bald sassen wir vereint am flackernden Herdfeuer eines rauchigen Alpenkässes , während Berg wird zu erkennen vermögen .	
	Berg Die Bahn ersteigt den Berg auf Luzerner Gebiet vom Vierwaldstätter See aus , hat also , da der See	

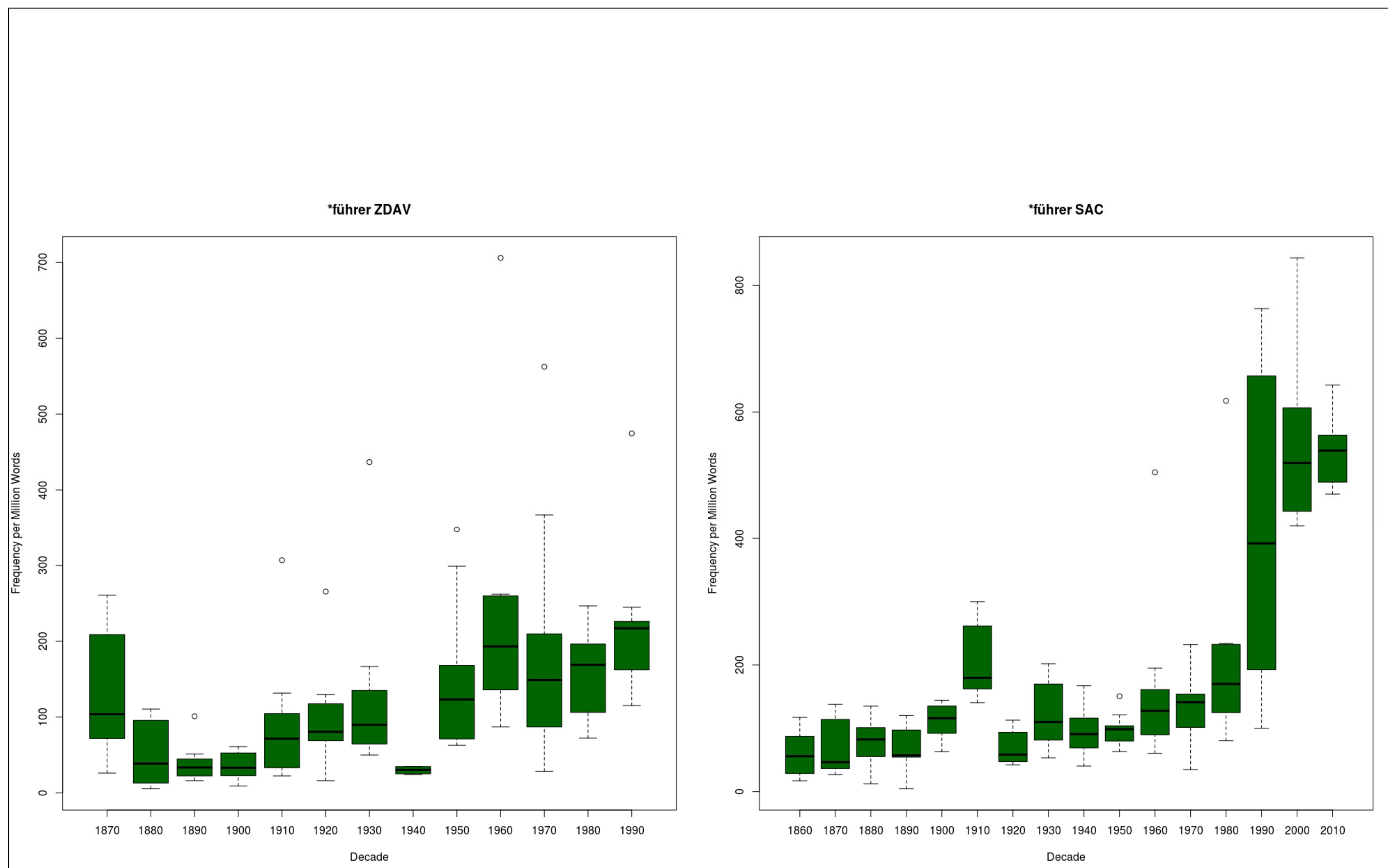
## Bridges and knots

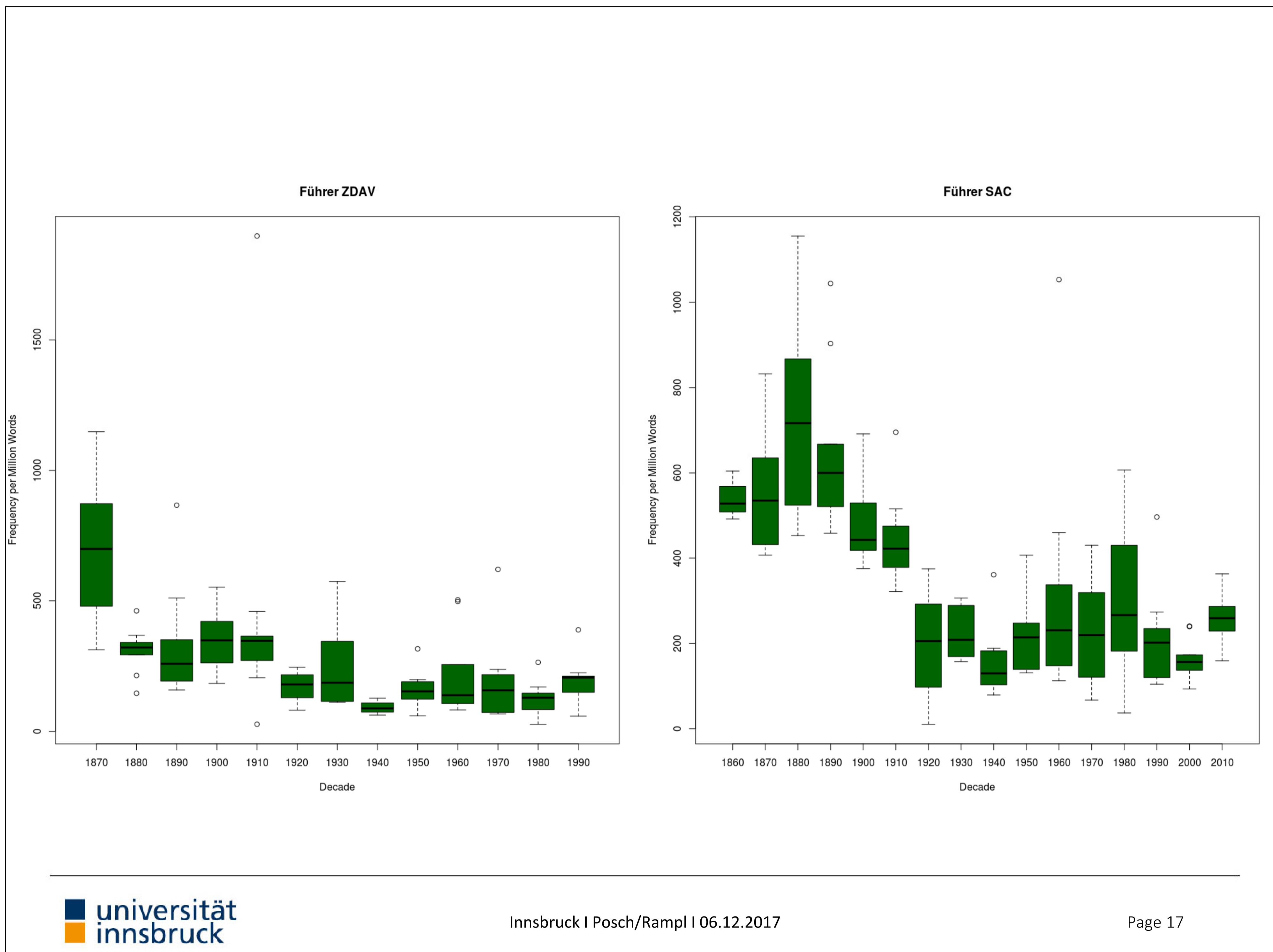
- » uses for other fields within linguistics
- » historical linguistics
- » discourse studies
- » onomastics



## Bridges and knots

- » Alpine Club
- » history of the Alpine Club (including world wars and Nazi-regime)
- » first ascents
- » exploration of unknown areas
- » re-publication of historic articles (Powder GUIDE)
- » search for people/places/events
- » links to other data possible (other corpora, archives, library, art depot, ...)



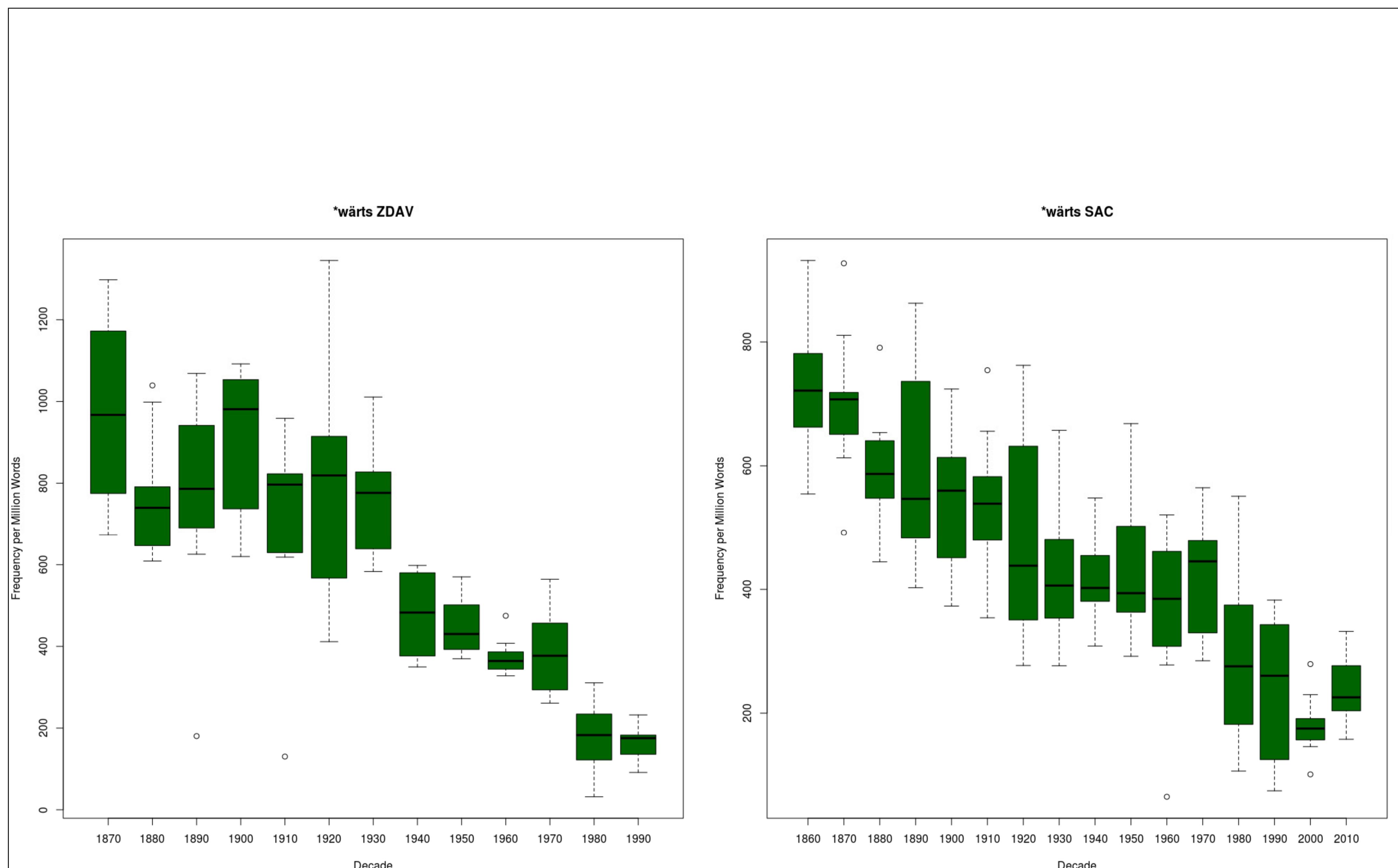


## Bridges and knots

- » Informatics and Geography
- » excellent training corpus for NER and language processing (including 19<sup>th</sup> century)
- » pipeline development for language processing
- » semantic tagging, geo-tagging
- » data linking to other data types (e.g. CIDOC, gazetteers)

## Panoramic view

- » corpus analyses
- » follow-up project SEMOHI – Semantics of Mountaineering History
- » follow-up project KEA (Keywords of the Alps)
- » comparison with existing corpora like text&berg
- » enrich annotation
- » link to other data (Lichtbilderverzeichnis, Autorenkatalog, biogr. Lexicon)
- » subsequent new corpora, e.g. Mitteilungen des Alpenvereins, sources from the Alpine archive, other languages
- » use of corpora in university teaching



## Panoramic view

- » publication of corpus on different platforms → end of 2017 (no pressure felt)
- » text-only XMLs (no POS) with TEI-conformant meta-data → on  zenodo
- » publication of annotated and ready-to-search corpus on  Hyperbase  
Web Edition
- » information on release will be available via:  
<http://sprawi.at>  
on Twitter: @SpraWiTwit  
newsletters
- » project website (under construction): <http://alpenwort.at>

