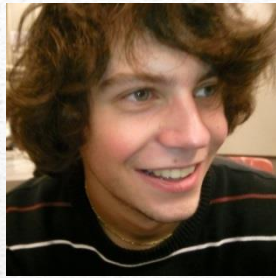# smargn

Reversing the n-grams searching process

1

Z. Gucevska   M. Monney   F. Junker   J. Gaspoz   J. Salathé

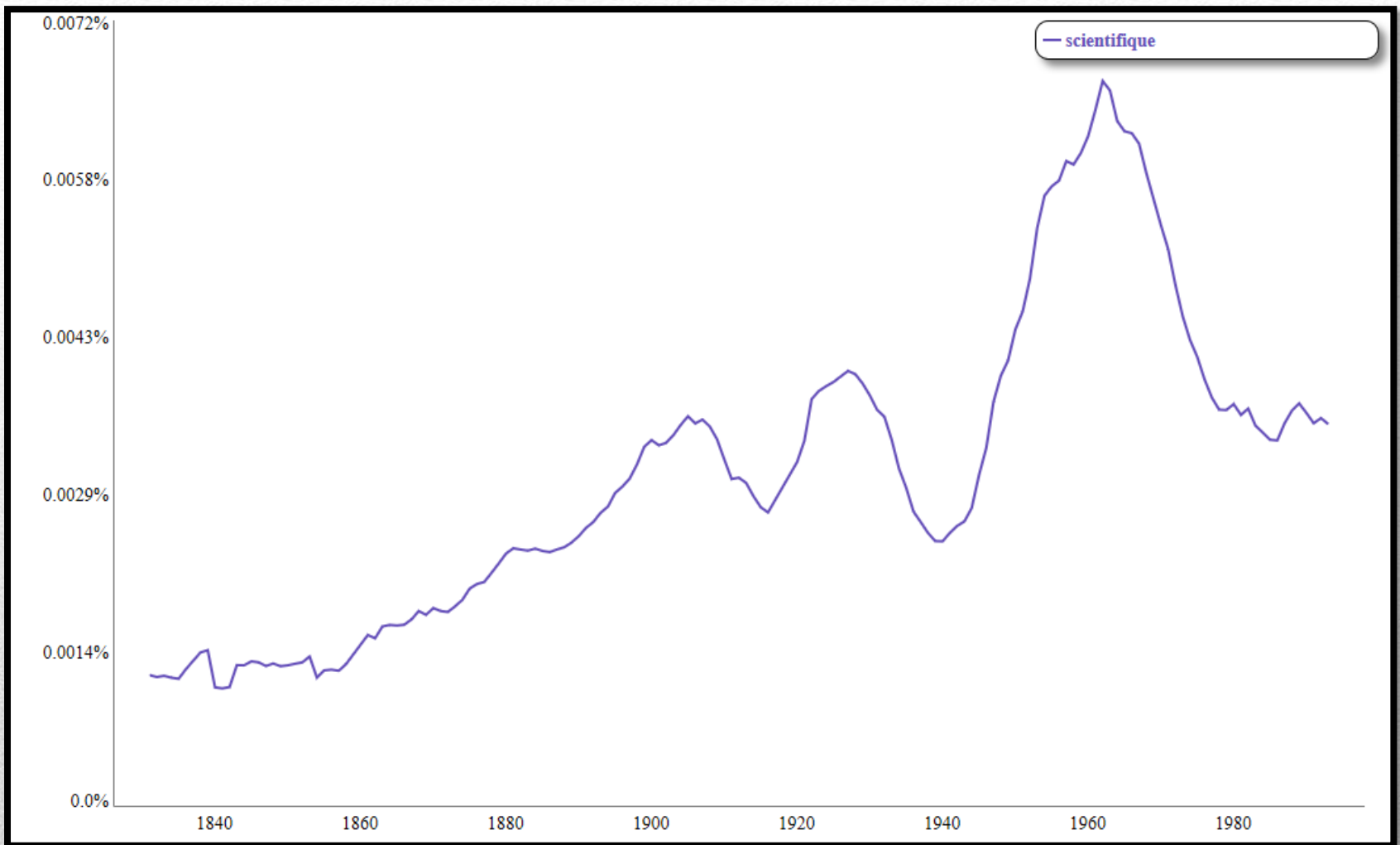A. Manasovska   F. Jolidon   V. Rutz   S. Bovet

Team

- What is a temporal profile?
- Motivation
- Corpus
- How smargn works
- Example I : "neige"
- Example II : "conflits"
- Conclusion

Plan

3

Understand a profile

- Studies based on n-grams have become popular

- Most of them relied on researcher's intuition

- We provide a way to reverse the process

  - Look for interesting words
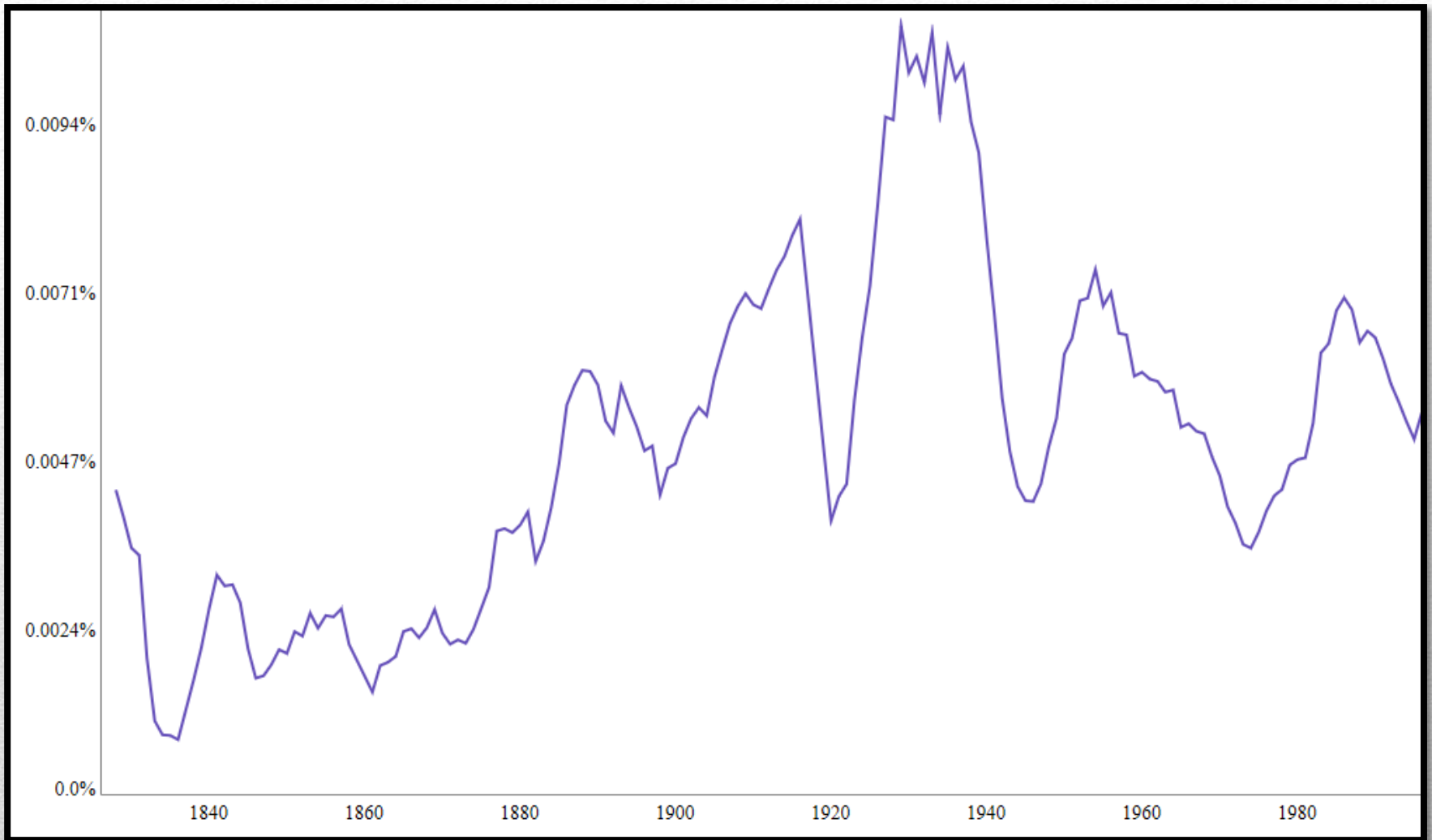
  - Find similar temporal profiles

  - Understand why

Motivation

- Le Temps newspaper
  - La Gazette de Lausanne
  - Le Journal de Genève

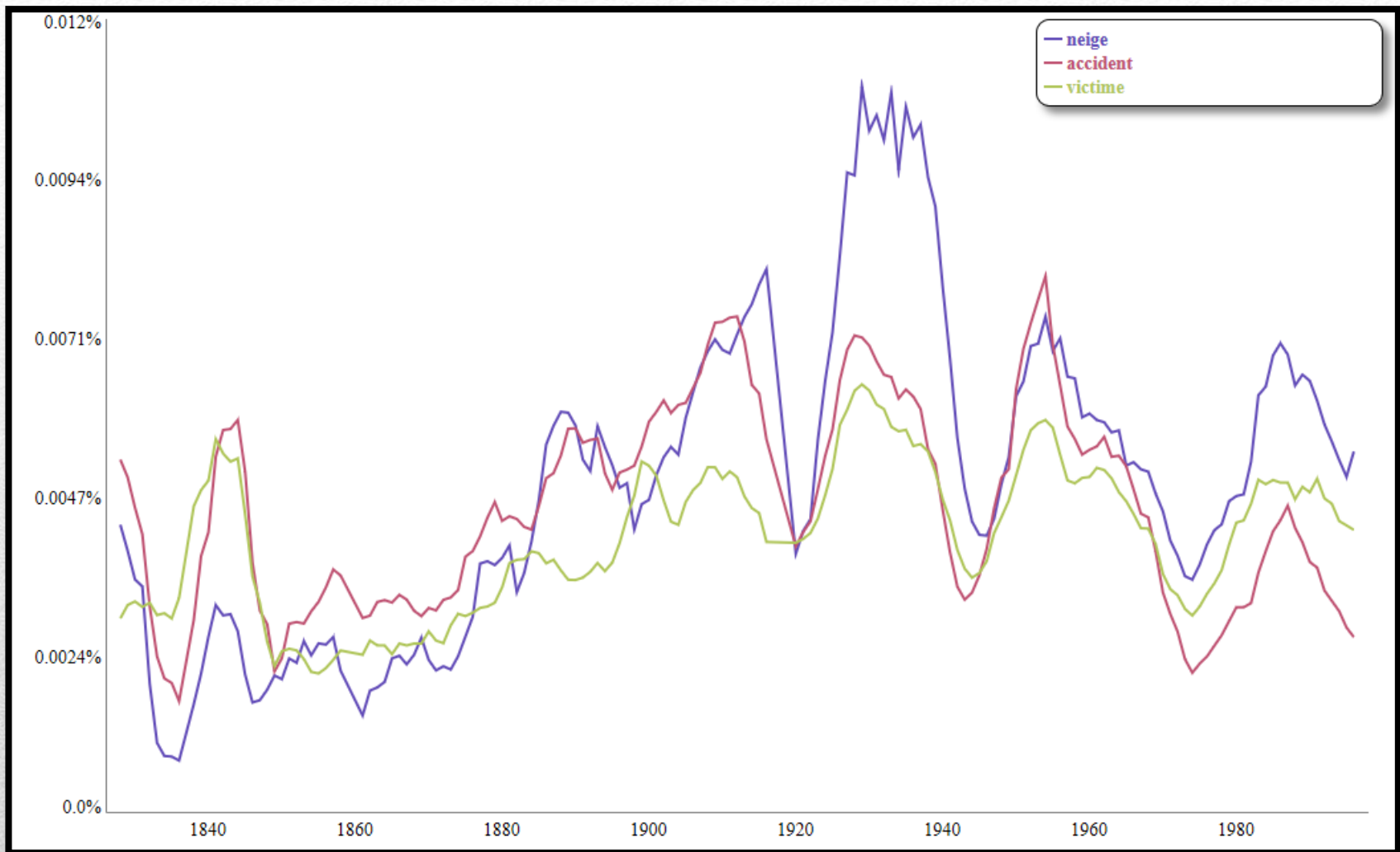- OCR from 1840 to 1998
→ Cleaning of the data required

Corpus

- Different comparison techniques implemented
  - Naïve comparison
  - Peak detection
  - Dynamic Time Wrapping
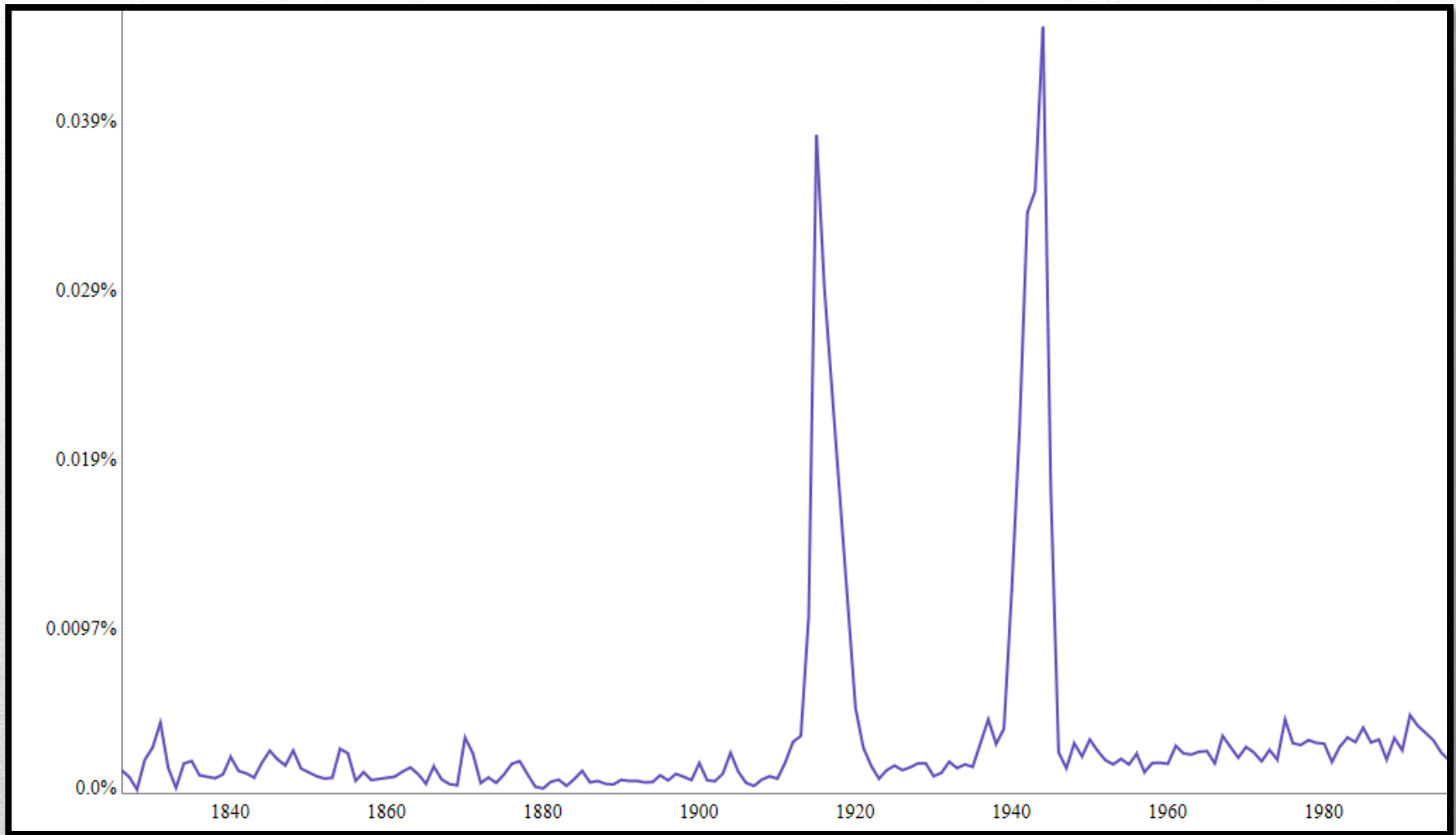 - …
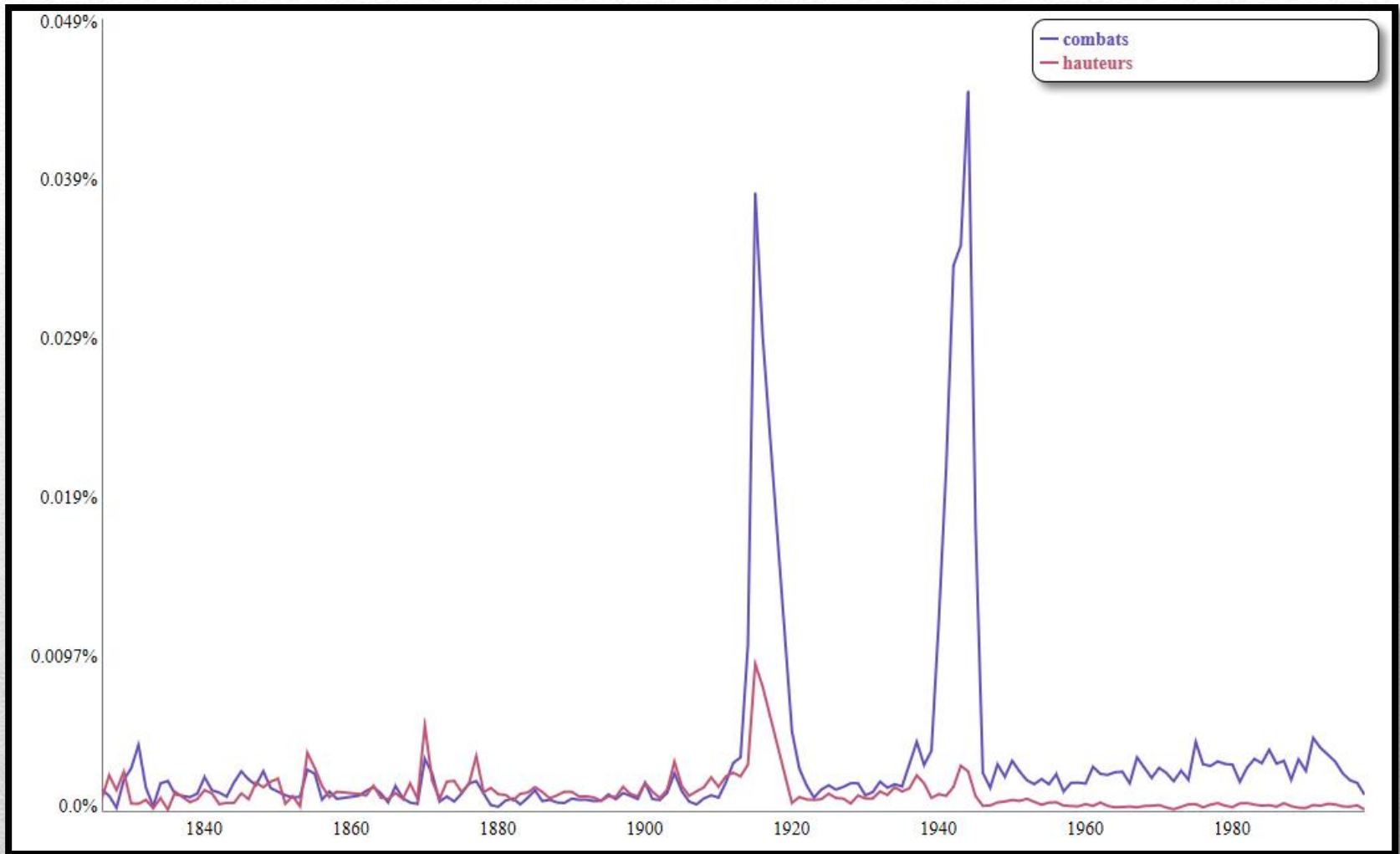- Check a word's profile against all the other

How smargn works

neige

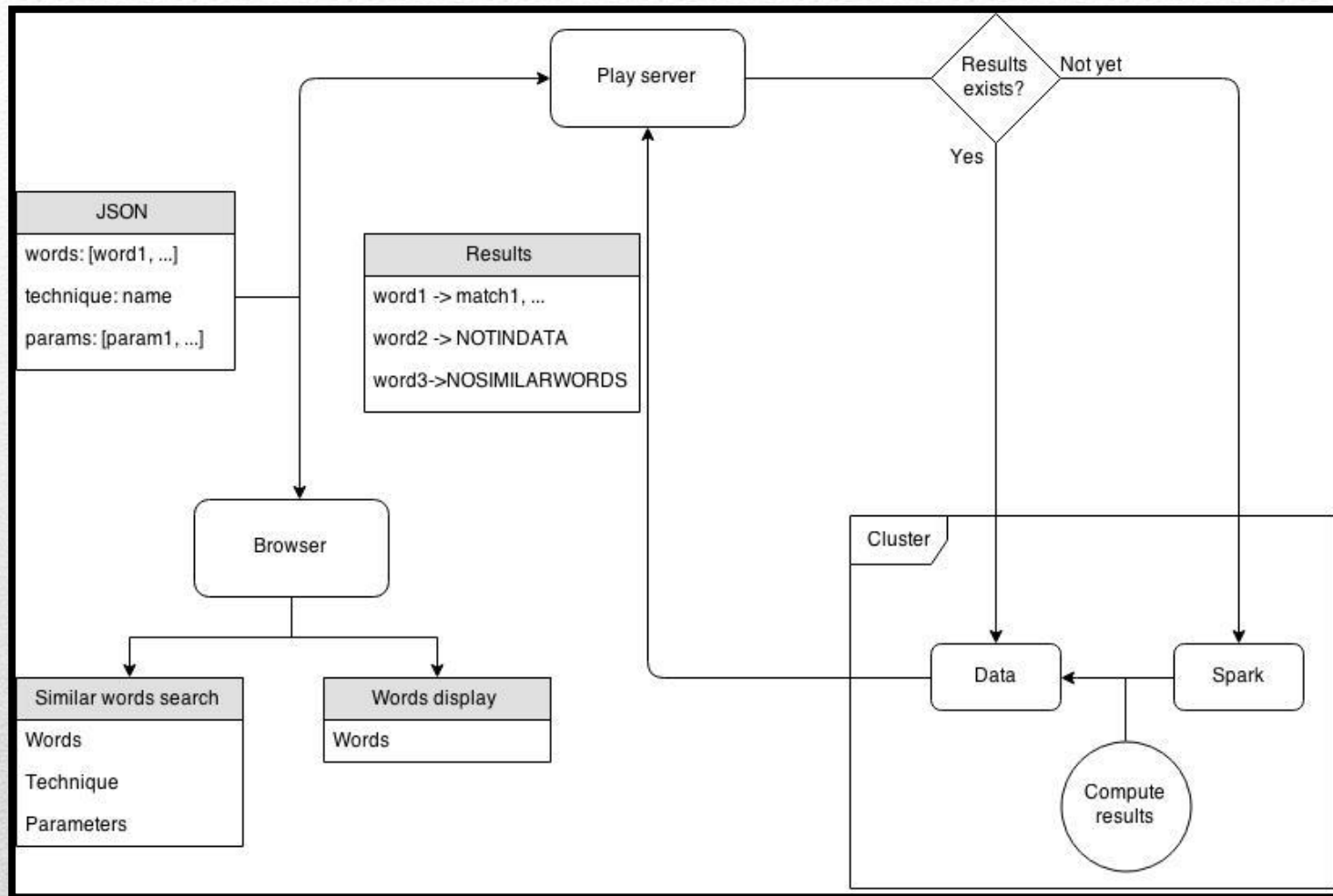# neige, accident & victime   9

combats

Big Data 2015

combats & hauteurs

- Can help linguistic studies
- Gives hints on where to look at
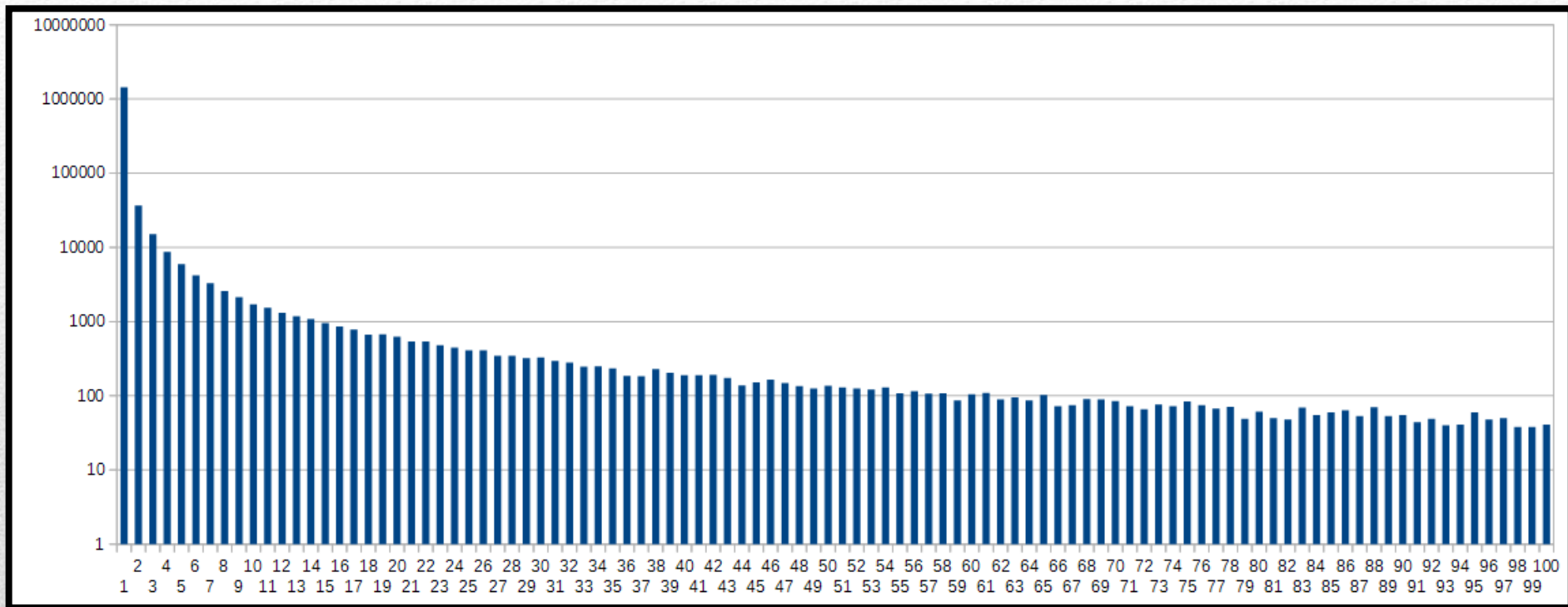- Still needs some global intuition

Conclusion

Thank you!

13

Application flowchart

Data histogram