

Basics of subproject 4.8 – Mental Maps
A first unspecified version

Text Mining for Classical Studies
Leipzig, 2009/10/01

Marco Büchler
Natural Language Processing Group
Department of Computer Science
University of Leipzig

- Research question
- Approaches from NLP
 - Co-occurrence analysis
 - Trend Mining
- Main objectives
- Challenges

Mental Maps: Research question

- Research question:
 - How are concepts changing over time?
- Remember: **Psycholinguistic experiments**

Stimulus	Antwort VP	Anz. VPs	Kollokation	Signifikanz
Butter	Brot	60	Brot	51
	weich	40	Käse	49
	Milch	32	Zucker	29
	Margarine	27	Milch	23
	Käse	20	Margarine	22
	Fett(e)	16	Mehl	18
	gelb	14	Eier	16
	Butterbrot	8	Pfund	14
	Dose	6	zerlassener	13
	essen	6	Fleisch	13

- Expectation: Different persons answer with different words
- During their life time every person learns a conceptual map connecting different words.
- Means: Every person has an individual word map - **mental map**.
- However: On Ancient Greek texts we can't compute a mental map for every author (data sparseness) but for **centuries, locations** or in combination **cities and locations**.

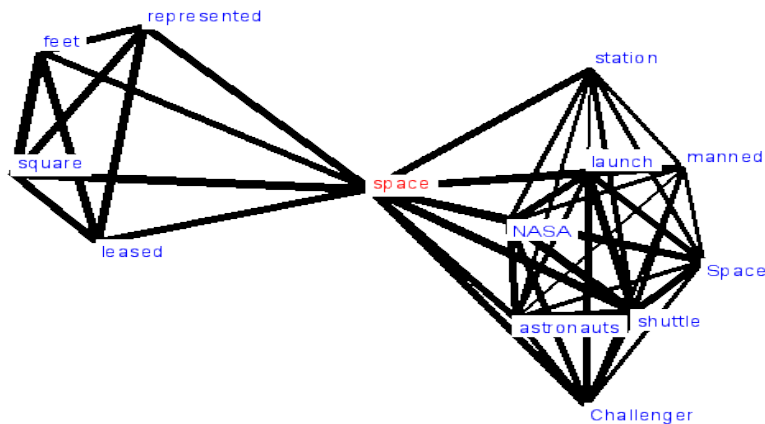
Mental Maps: Approaches from NLP

- Basics (semantic level):
 - Distributional Hypothesis (Harris 1953)
 - The underlying idea that "a word is characterized by the company it keeps" was popularized by Firth (1957).
- Why Distributional Hypothesis?:
 - Investigation of different locations or/and centuries can cause other associations.
 - Cause:
Existing mental map M_i
+ new information (influence, event) I_j

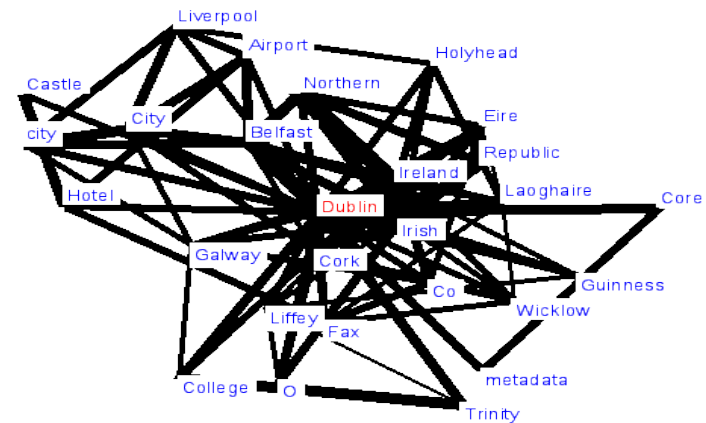
= modified mental map M_{i+1}
 - Motivation:
Geographical distances in the Ancient world encapsulate events or information stronger than nowadays.

- How does the number of meanings change?

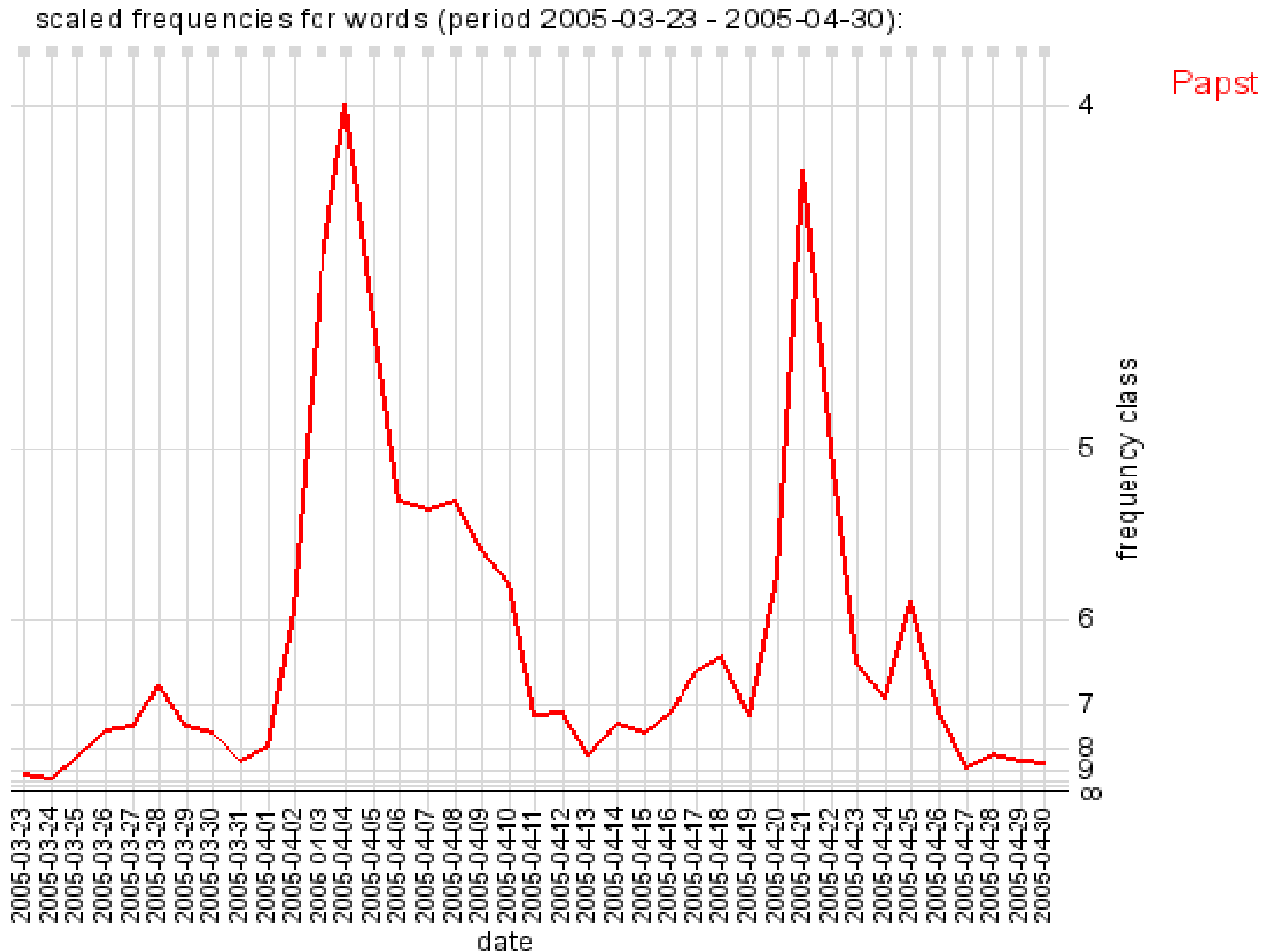
Graph v. 1.5 für space



Graph v. 1.5 für Dublin



- Working hypothesis 1:
 - Words tend to have several meanings in earlier stages (lifetime of concept): The number of meanings is decreasing over time.
 - Motivation: If a new concept (word) occurs, every person (or location) tries to put it into the mental map. Implicitly, several properties would be inherited.
- Working hypothesis 2:
 - Words tend to have several meanings in later stages (lifetime of concept): The number of meanings is decreasing over time.
 - Motivation: Human beings try to use learned knowledge (least effort approach). Example: Usage of term „computer“



Mental Maps: Main objectives

- Creation of a time dependent graph
- Creation of a city dependent graph
- ??? Needs to be done

Mental Maps: Challenges

- Degree of normalisation
- Critical mass of text per slice (century, location)
- Co-occurrence statistics (special data sparseness)
 - How to handle very low frequent co-occurrences?
 - Is a co-occurrence of frequency 1 significant? Mostly not.
 - Is a co-occurrence of frequency 2 significant? Mostly not, too.
 - However: Current research topic in my PHD.
- Good visualisation of a mental map
 - Expected number of datasets: 100 Mio – 500 Mio