

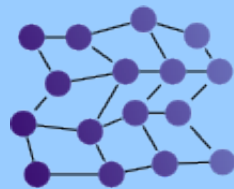
Concept mining with Self-Organizing Maps for the Semantic Web



Timo HONKELA and Matti PÖLLÄ



Helsinki University of Technology
Adaptive Informatics Research Center
Cognitive Systems Research Group



WSOM 2009

7th International Workshop
on Self-Organizing Maps

Traditional formalization of meaning



- Formalisms like first-order predicate logic have widely been used as a basis for theories of meaning (epistemology); consider also contemporary efforts such as Semantic Web
- These formalisms provide only limited means for creating in-depth theories of how language is understood

Traditional formalization of meaning, cont'd



- Traditional logic provides means e.g. for modeling quantification, connectives, analytical truths and conceptual hierarchies
- However, many semantic phenomena are matters of degree. Various proposals that apply Bayesian probability theory (e.g. Tenenbaum) or fuzzy sets deal with this.

Practical consequences



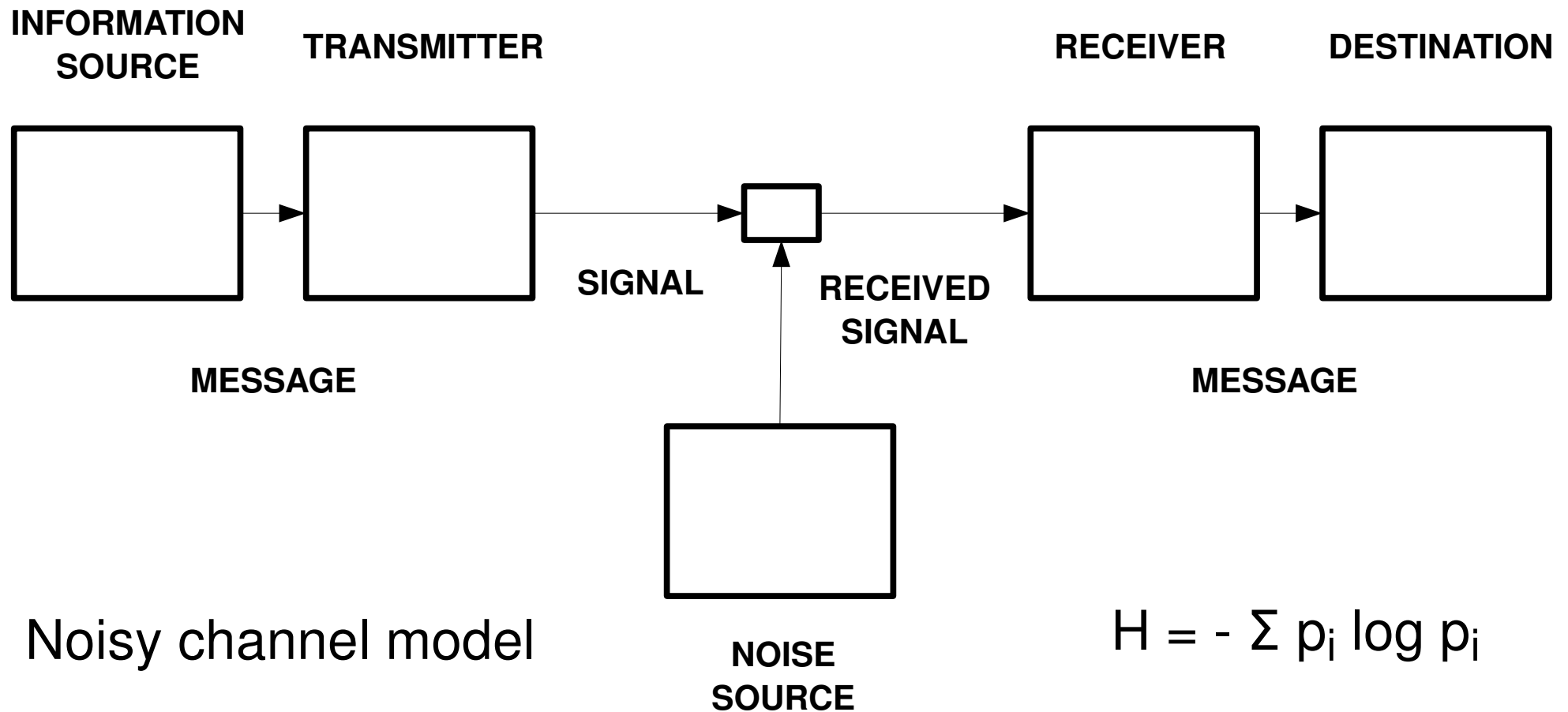
- The traditional notion of uncertainty in decision making does not cover the uncertainties caused by differences in conceptual systems of individual agents within a community
- In many transactions, including symbolic/linguistic communication, the differences in the underlying conceptual systems play an important role

Practical consequences, 2



- Serious efforts have been made to harmonize or to standardize the classification systems or ontologies used by agents
- Even if standardization is conducted, there can not be any true guarantee that all participating agents would share the meaning of all the expressions used in the transactions in various context. Thus, the systems need to be able to conduct **meaning negotiations**

General communication system and measuring information (Shannon & Weaver)



Weaver on Shannon



- “Relative to the broad subject of communication, there seem to be problems at three levels. [...]
 - LEVEL A. How accurately can the symbols of communication be transmitted? (The technical problem)
 - LEVEL B. How precisely do the transmitted symbols convey the desired meaning? (The semantic problem)
 - LEVEL C. How effectively does the received meaning affect conduct in the desired way? (The effectiveness problem)”
- “The semantic problems are concerned with the identity, or satisfactorily close approximation, in the interpretation of meaning by the receiver, as compared with the intended meaning of the sender.” (1949, p. 4)

Neglected pattern recognition processes



- Even these methodological extensions do not suffice if the pattern recognition processes are not taken into account
- The world is not straightforwardly experienced as discrete objects and events but there are complex underlying cognitive processes involved

Some challenges in modeling linguistic phenomena



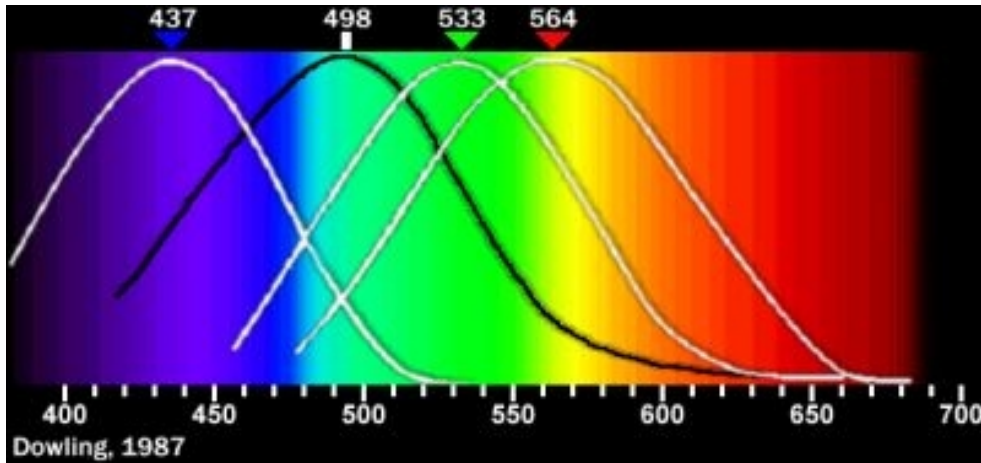
- A large proportion of modern human activity in its different forms (science, industry, society, culture, etc.) is based on the use of language
- Complexity of (natural) language
 - more than 6000 languages, many more dialects
 - Each language has the order of 10^5 to 10^{10} different word forms
 - Each word is understood differently by each speaker of a language at least to some degree
- Ambiguity at multiple levels

Relationship between language and thought

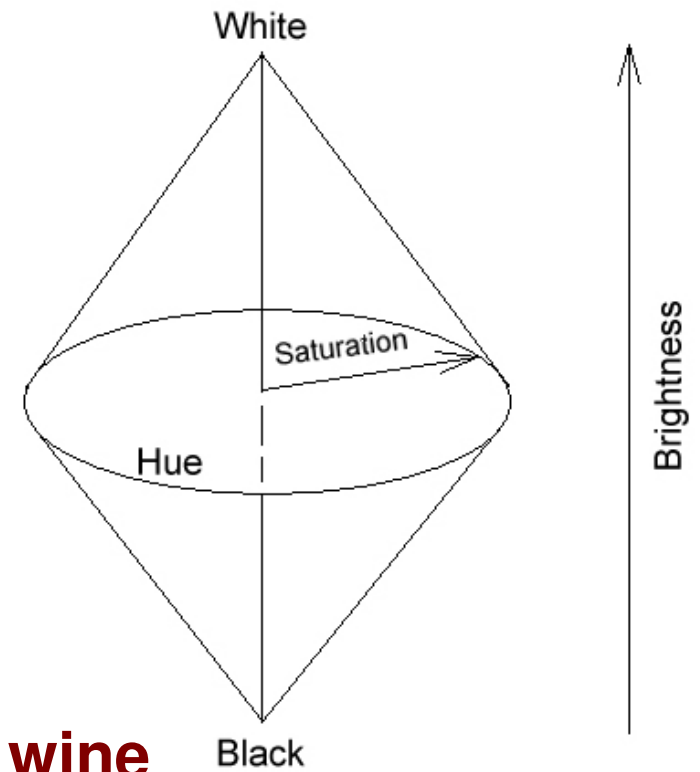


- Consider e.g. Vygotsky (1934): “The relation of thought to word is not a thing but a process, a continual movement back and forth from thought to word and from word to thought. In that process the relation of thought to word undergoes changes which themselves may be regarded as development in the functional sense.”
- >> Building ontologies does not help much in developing systems that understand language.

Simple challenge: Color naming



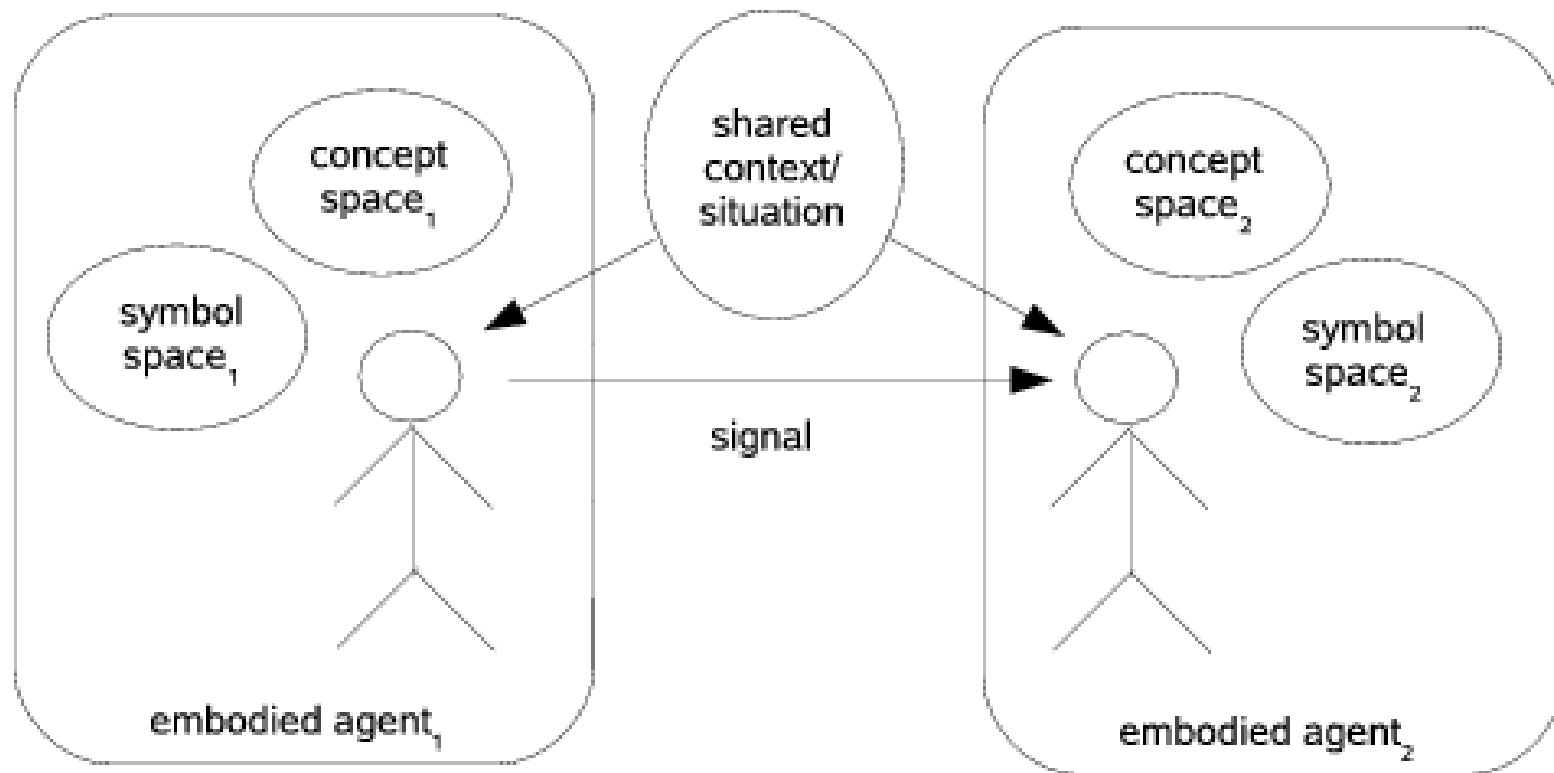
Human vision: rods, cones,...
Physical reasons for color
Contextuality of naming



red wine
red skin
red shirt

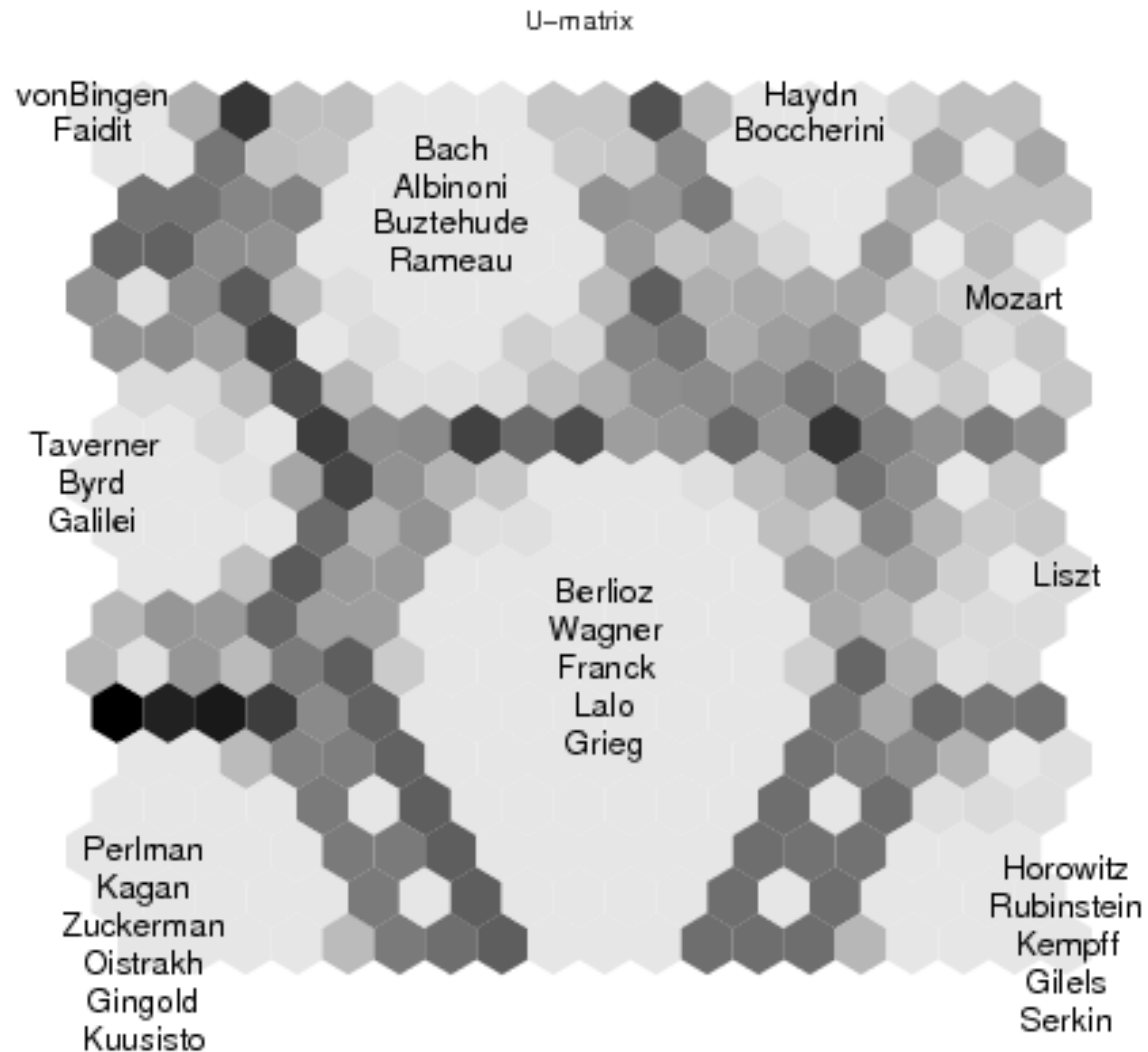
→ Hardin
→ Gärdenfors

Theory of Intersubjective Meaning Spaces (IMS)

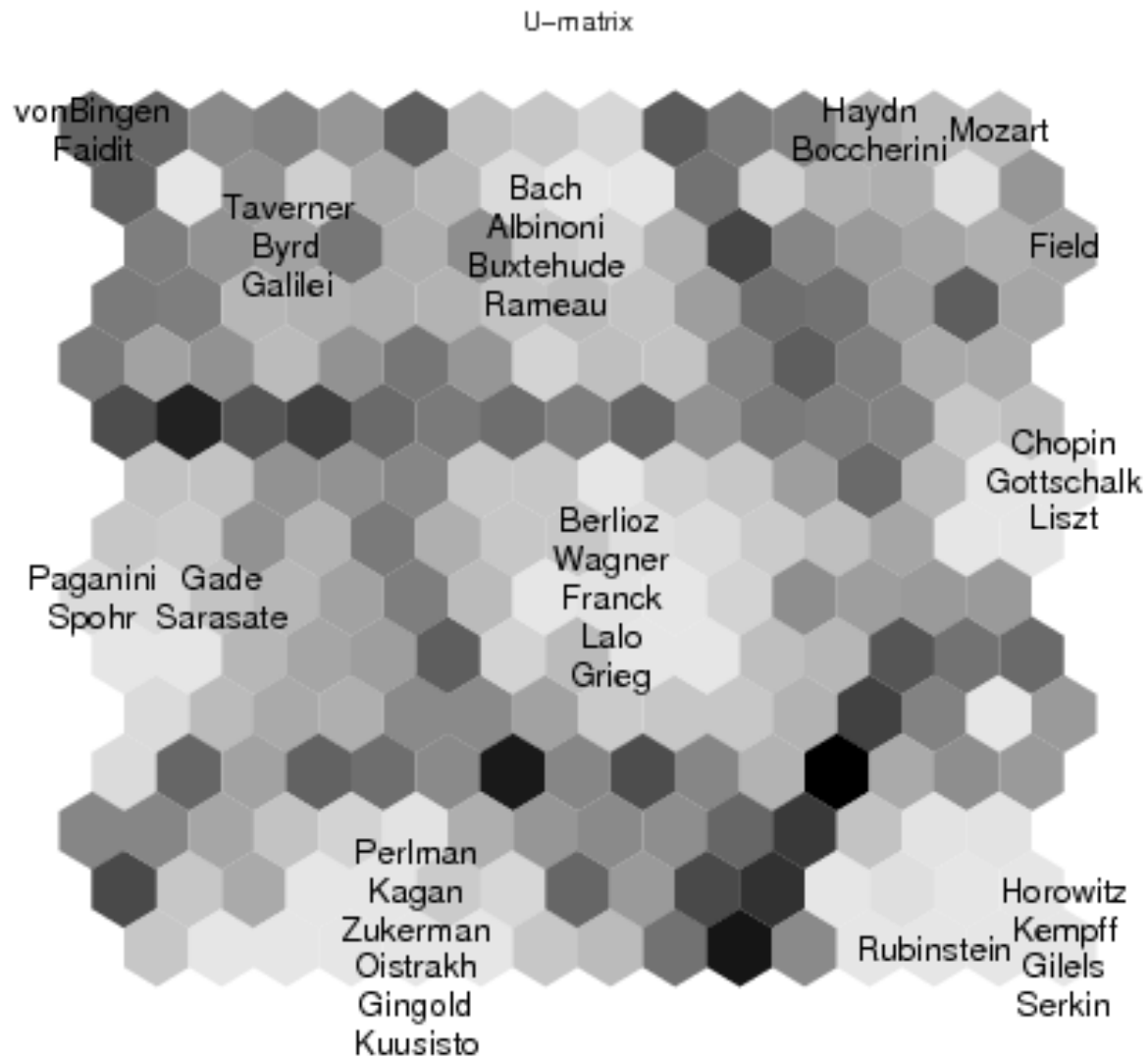


(Honkela, Könönen, Lindh-Knuutila & Paukkeri 2008)

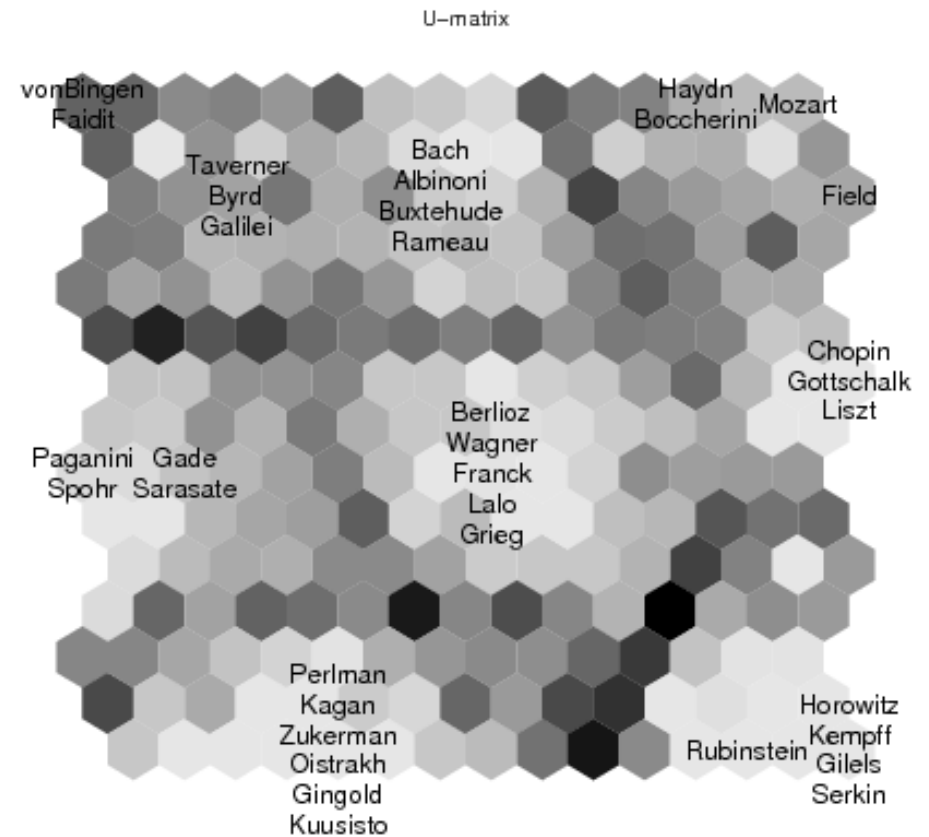
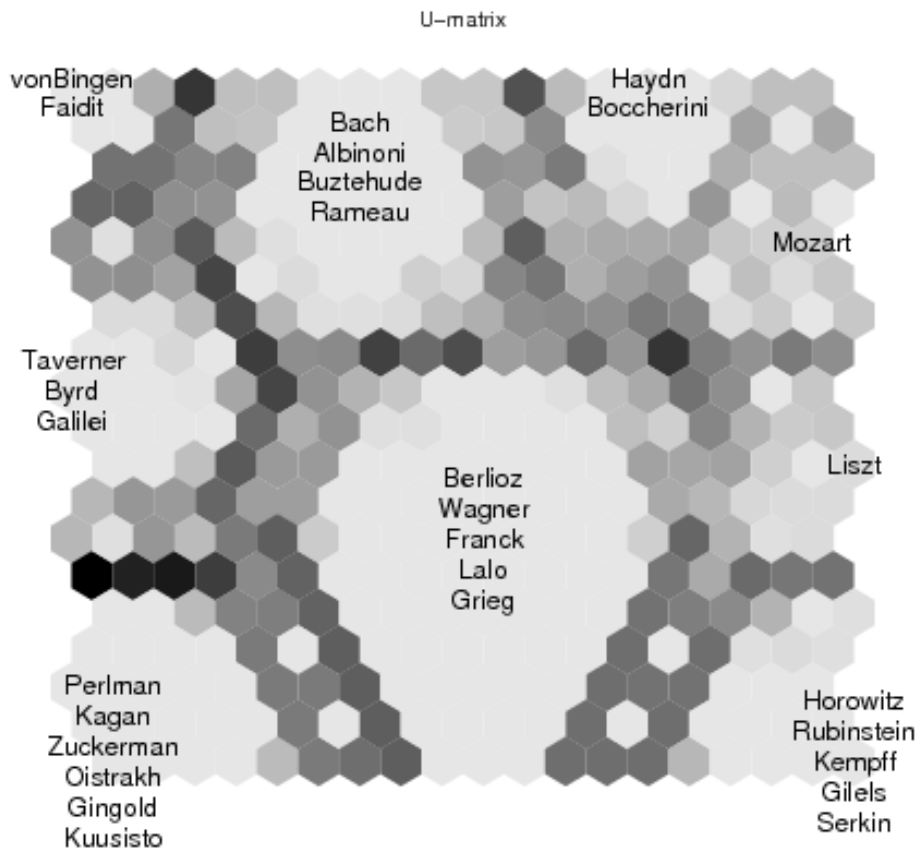
Composer/musician example



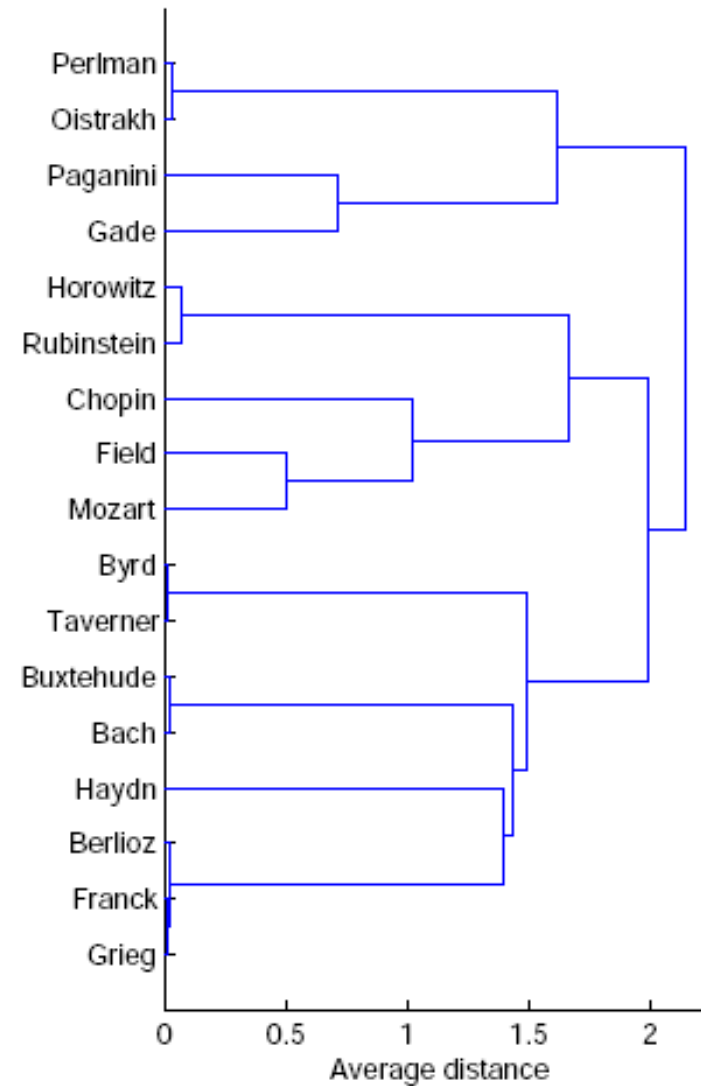
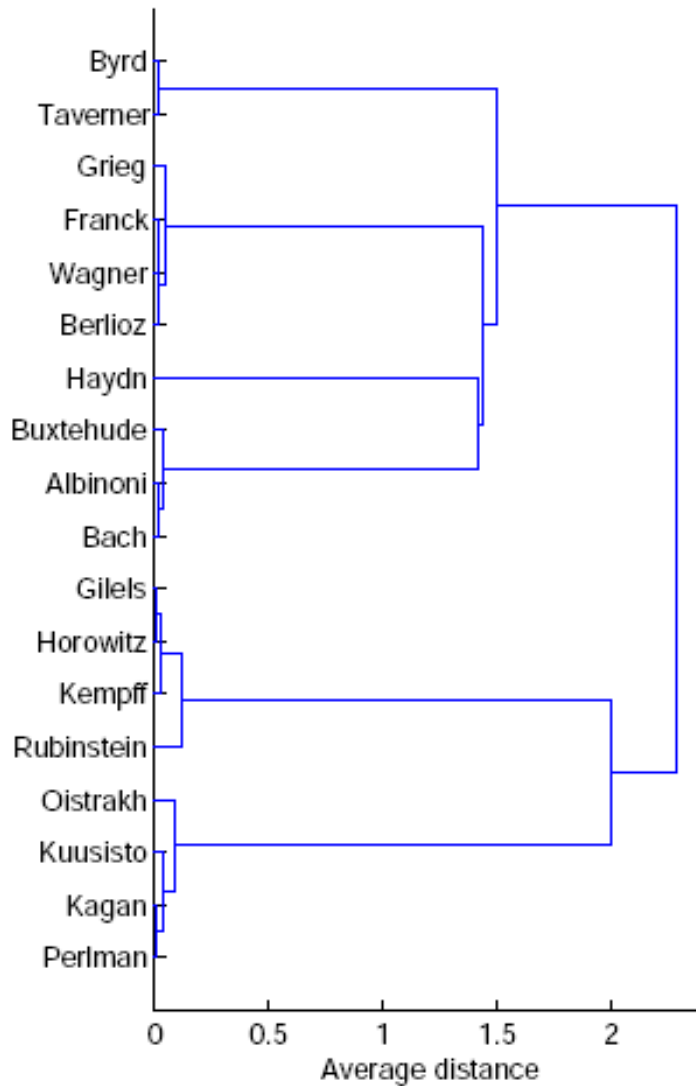
Composer/musician example



Composer/musician example



Dendrogram comparison





Thank you!