

# Advanced Methods for Corpus Processing

Lluís Padró <padro@lsi.upc.edu>

Lluís Màrquez <lluism@lsi.upc.edu>

German Rigau <german.rigau@ehu.es>



**Centre de Tecnologies i Aplicacions  
del Llenguatge i la Parla**

# Foreword

- **UPC: Artificial Intelligence**
  - EMNLP: Empirical Methods for NLP
  - 2001/2002-2002/2003
    - Lluís Padró, Lluís Màrquez, German Rigau
  - 2003/2004-2004/2005
    - Lluís Padró, Lluís Màrquez, Neus Català, German Rigau
  - 2005/2006
    - L. Padró, L. Màrquez, X. Farreres, J. Daudé, G. Rigau
- **EHU: NLP**
  - AMCM: Advanced Methods for Corpus Management
  - 2004/2005-...-2011/2012
    - Lluís Padró, Lluís Màrquez, German Rigau

# Content

- Statistical methods for NLP (Lluís Padró)
  - 6/2, 7/2 : 16h - 19:30h
- Machine Learning for NLP (Lluís Màrquez)
  - 8/2, 9/2 : 16h - 19:30h
- Ontologies & large-scale Knowledge Bases for NLP (German Rigau)
  - 6/3, 8/3, 13/3, 15/3, 19/3, 20/3, 21/3: 16:00h - 18:00h
- Concluding remarks (German Rigau)
  - 22/3: 16:00h - 18:00h

# Content

- Statistical methods for NLP (Lluís Padró)
  - Mathematical Foundations
  - Linguistic Foundations
  - Corpora
  - Statistical Inference
  - Markov Models
  - Tagging
  - Maximum Entropy Models

# Content

- Machine Learning for NLP (Lluís Màrquez)
  - Introduction: Machine Learning and Machine Learning for NLP
  - Machine Learning: Classical Methods from AI
  - Margin-based Machine Learning Algorithms
  - Machine Learning for NLP
  - Applications

# Content

- Ontologies & Knowledge-based for NLP (German Rigau)
  - Words & Works
  - Ontologies:
    - Mikrokosmos
  - Large-scale Knowledge Bases:
    - WordNet & EuroWordNet
  - More large-scale resources
    - ConceptNet, Framenet, VerbNet, PropBank, ...
  - Building Wordnets
  - WordNet extensions:
    - SUMO ontology, eXtended WordNet, Meaning project
- Concluding remarks (German Rigau)
  - Reasoning