



# The European Language Resources and Technologies Forum

## Session 3

### Evaluation and Validation

*Chair: Jan Odijk (Utrecht)*

*Rapporteur: Joseph Mariani (LIMSI & IMMI)*





# Session 3

- Validation
  - Usually restricted to LR (Data)
- Evaluation
  - Usually restricted to LT



# Validation



- Validation

- "The 'Standard Deviation' of LR Quality"

- Henk van den Heuvel (Radboud University Nijmegen, NL)

- "Towards More Effective LR Validation"

- └ Florian Schiel (Ludwig Maximilian Universität München - BAS, DE)





# Evaluation



## ■ Evaluation

- "Evaluation of Technology for Multilingual Information Access: the Next Step"
  - Carol Peters (ISTI-CNR, IT)
- "Can Evaluation Be Application-Independent?"
  - Bente Maegaard (Københavns Universitet - CST, DK)
- "Language Technology Evaluation: which Funding Strategy?"
  - Edouard Geoffrois (DGA, FR)
- "Toward an Integrated Evaluation Framework"
  - Bernardo Magnini (FBK, IT)





# Evaluation



- Evaluation (cont.)
  - "Evaluation: a Paradigm that Produces High Quality Language Resources"
    - Patrick Paroubek (LIMSI-CNRS, FR)
  - "A Proposal to Launch a Support Centre for 'Remote' Evaluation and Development of Language Technologies"
    - Harald Hoega (SVOX Deutschland GmbH, DE)
  - "Evaluation of HLT-Tools for Less Spoken Languages"
    - Cristina Vertan (Universität Hamburg, DE) Evaluation



# STOP HERE



# Validation (LR)

- 3.1. Which LR validation/evaluation methods are there already? Are such methods lacking for specific types of resources?
- 3.2. Can formal validation and content validation be approached in the same manner, or are there fundamental differences between them?
- 3.3. How can we measure the quality of LRs? What do we mean by quality?



# Validation (LR)



- 3.4. Have the LRs produced in the last decades indeed been used for the objectives they were intended for? And were they successful? What can we learn from this for future resources?
- 3.5. Can useful LRs be created without a specific objective in mind?
- 3.6. Are resources created with such an objective in mind not too limited in scope given the amount of effort and money invested in them?
- 3.7. Are LRs with pretty wide objectives such as BNC and the Dutch Spoken Corpus useful and for which objectives are they actually being used?



# Evaluation (LT)

- 3.8. Are the models as used in the various evaluation campaigns the right model? Are adaptations needed? Are there good experiences we should promote, or bad experiences that should be shared so that they can be avoided by others?
- 3.9. Are there urgent needs for evaluation data or tools?
- 3.10. Are there new trends or desires in methodologies for carrying out evaluation, and if so do they require new types of evaluation resources (data, tools, metrics)?



# Evaluation (LT)



- 3.11. Should the quality of the process of creating technology play a role in an overall evaluation, and if so, what recommendations can be made in this domain?
- 3.12. What should be the business model attached to LT evaluation? Should it be fully/partly supported by public funds? Should it be handled by public organizations? Should it be a prerequisite for any participation in public programs? How to port evaluation campaigns to other languages?