

# Evaluation of LRAs & LRAs for evaluation

*Edouard Geoffrois, DGA  
COCOSDA/WRITE Workshop  
June 1<sup>st</sup>, 2008*



DÉLÉGATION GÉNÉRALE POUR L'ARMEMENT



# Outline

- Language resources, evaluation and infrastructure
  - Resources for evaluation vs. other resources
  - Evaluation of resources vs. other evaluations
- Organizational issues

# ▶ General definitions and properties

- Resources:
  - something that can be used to create another thing, to accomplish a goal
  - in limited quantity, possible restrictions on availability
- Infrastructure
  - underlying base or foundation, esp. for an organization or system
  - wide array, large network
  - long and indefinite life, sustainable
  - public activities to support private sector

# Types of evaluation

Evaluation through publications

Interpret results, disseminate knowledge

Reproduce and analyze results, measure progress, determine maturity

Measure and analyze user perception, refine definition of needs

Technology evaluation

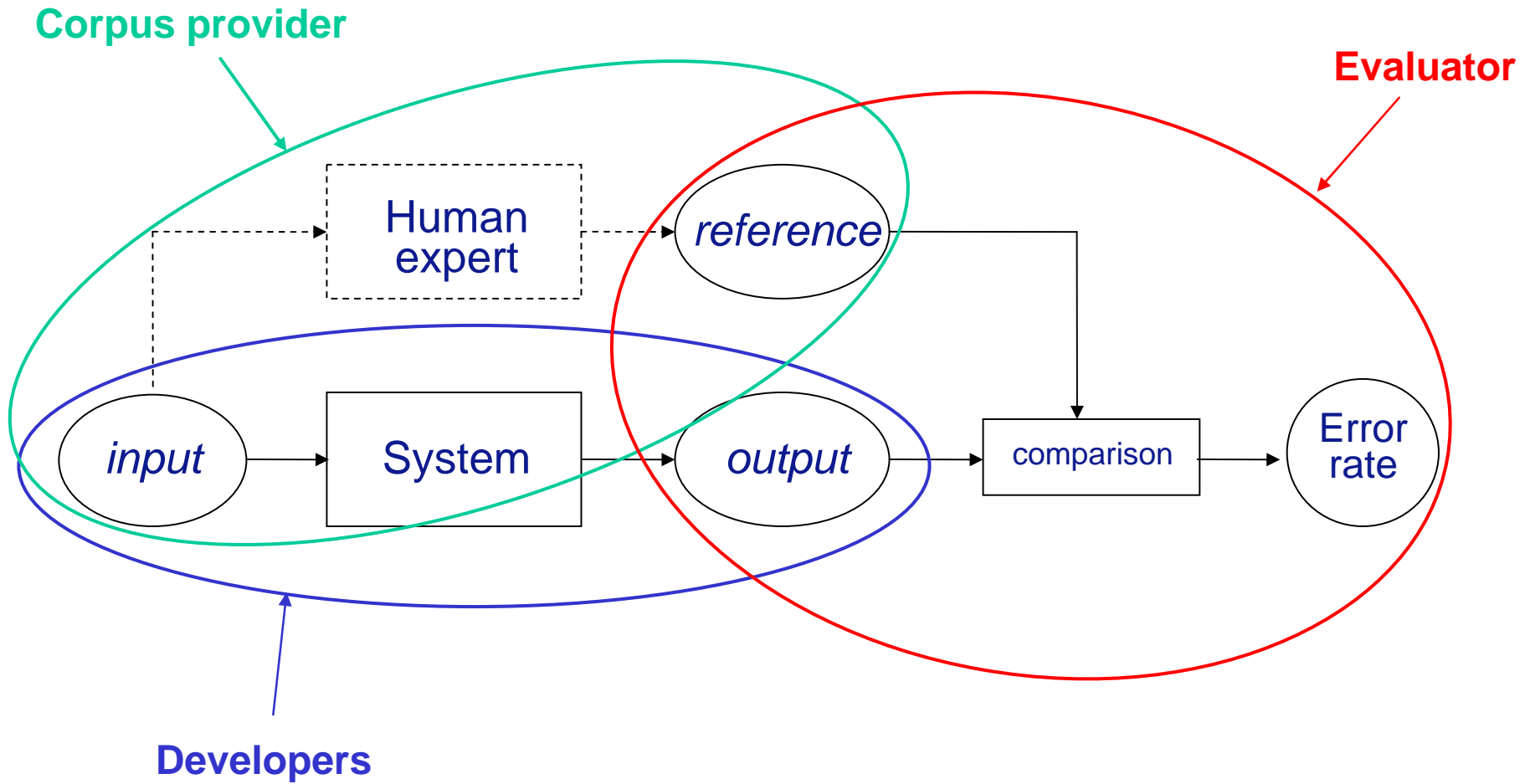
Usage evaluation

Theoretical

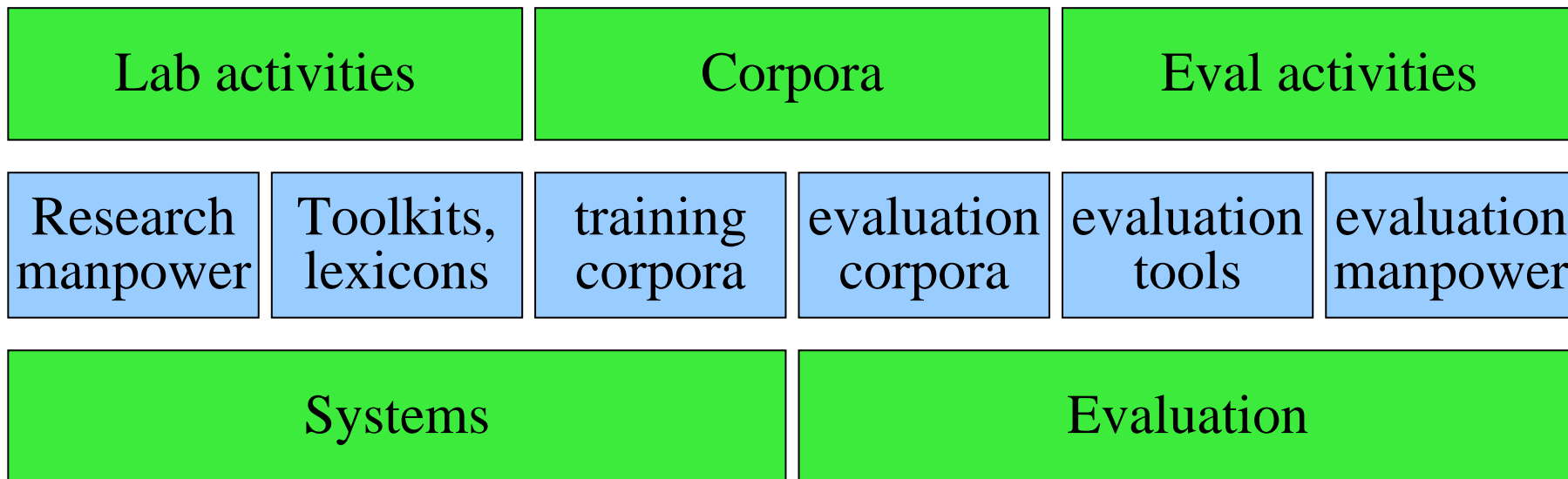
Objective (measuring instruments)      Experimental      Subjective (user panels)



# Roles in evaluation



# ► Ressources & evaluation





# Corpora for training vs. for evaluation

	<i>Corpora for training</i>	<i>Corpora for evaluation</i>
quality	can be traded-off with quantity	expected to be flawless
features	include single reference	include variants and any specific feature for evaluation
size	large enough for wide coverage	adjusted for test significance
reusability	cumulative	renewed for each evaluation
distribution	can be charged at "market price"	must be freely available for the sake of reproducibility

# Evaluation of LRs

- Evaluation through systems
  - more representative of actual goal
  - indirect
- Intermediate evaluation/validation
  - direct analysis of the data
  - partial validation



# ► Underlying issues

- Scientific
  - Reproducibility of experiments
  - Unbiased results
- Specific domain
  - Measurement depend on specific data
  - Data must include human added value
  - Data must be released for progress
  - Data must be unknown to avoid bias
- Organizational
  - availability, sustainability
  - sharing cost

# ▶ Consequences for evaluation

- third-party, synchrony, need to organize in the form of campaigns
- market failure, needs strong support by those who fund the research
- needs to be organized in close connection with program management

# ▶ Some best practice to generalize

- Integrated projects gathering the R&D activities on a few common goals and the related shared resources and evaluation activities
- Funding evaluation activities with a scheme similar to project management ?



# QUESTIONS ?