

# Language Technology Evaluation: which Funding Strategy?

*Edouard Geoffrois, DGA*

*FLaReNet launching event*

*Vienna, Feb 12<sup>th</sup>, 2009*





# Outline

- Rationale for evaluation campaigns
  - Evaluation in general is beneficial
  - Objective eval of LT requires synchronization
- Lack of LT evaluation infrastructures
  - Organization of evaluation for LT can be analysed as a market failure
  - Partial grants do not provide enough incentive
- Adapting the infrastructures to the needs
  - Grant 100% funding to significant efforts for evaluation of LT
  - Gather all R&D&I&E&C for a given technology in large integrated programs



# Some benefits of evaluation

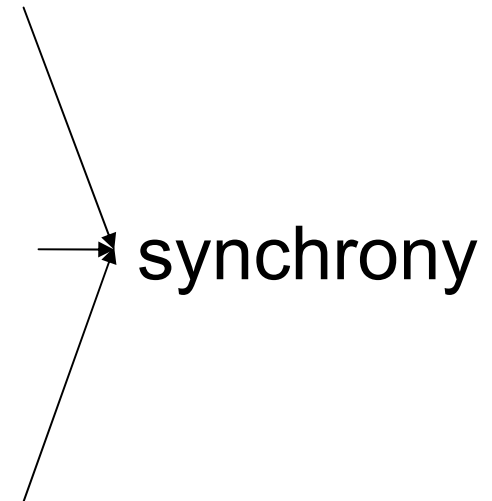
- Explicitly define problems
- Validate new ideas
- Compare approaches and systems
- Incite innovation
- Ease technology transfer
- Identify missing science
- Determine the maturity of a technology for a given application
- Organize the community
- Foster company competitiveness
- Monitor the efficient use of public funding

# ▶ Why a specific organization in the form of campaigns?

knowledge is highly variable → common test data

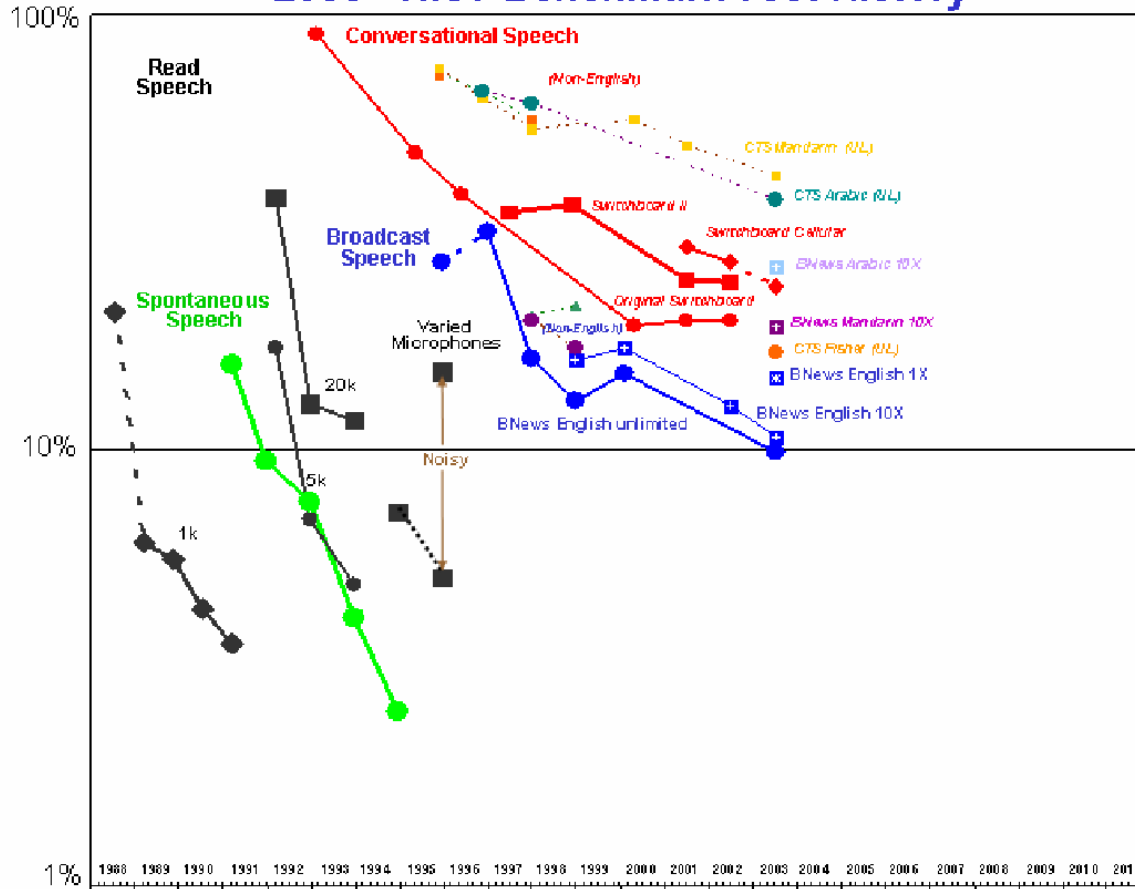
learning-based technologies → data unknown before test

scientific research → data published and discussed



# Illustration: two decades of speech recognition by DARPA/NIST

2003 - NIST Benchmark Test History



- Each curve shows the evolution of the best system performance over the years, with decreasing error rates
- When a given problem (one colored curve) is considered "solved", a more complex one is addressed

Source: <http://www.nist.gov/speech/history/index.html>



# ▶ LT evaluation as a public good

judge is human → high cost

software, low  
marginal cost → non-rival

foster open  
research → non-excludable

public  
good

# ▶ Adapted funding strategies

- Grant 100% funding to significant efforts for evaluation of LT
  - Relying on dedicated public bodies
  - Allowing for a special case (like management)
    - within main program contract
    - or next to it
- Gather all R&D&I&E&C for a given technology in large integrated program
  - Avoid splitting into per-application programs

# The different roles in technology development and their coordination

