

An Evaluation Infrastructure for HLT in Europe



Khalid CHOUKRI
ELRA/ELDA

55 Rue Brillat-Savarin, F-75013 Paris, France
Tel. +33 1 43 13 33 33 -- Fax. +33 1 43 13 33 30

Email: choukri@elda.org

Web: <http://www.elra.info> & www.elda.org/

Evaluation in USA & Europe

- **USA:** DARPA (NIST/LDC)
- Top-down, finance, permanent infrastructure, govt. influence
- **Europe:** no equivalent infrastructure, fragmented, isolated projects, initiatives
- National: GRACE, Aupelf ARC; Verbmobil; Morpholympics; Senseval/Romanseval,
- EU: CLEF, SQALE, TSNLP, EAGLES, ETSI/Aurora, DiET, DISC, TEMAA, SPARKLE
- **ELSE: Recommendations for permanent European HLT evaluation infrastructure; comparative vs. competitive**

Why ELRA/ELDA

- In Agreement with ELRA's mission: Language Resources and Evaluations are closely connected: users of LRs are developers of HLT.
- A clear side effect of evaluation is often the production of high quality resources.

Why ELRA/ELDA

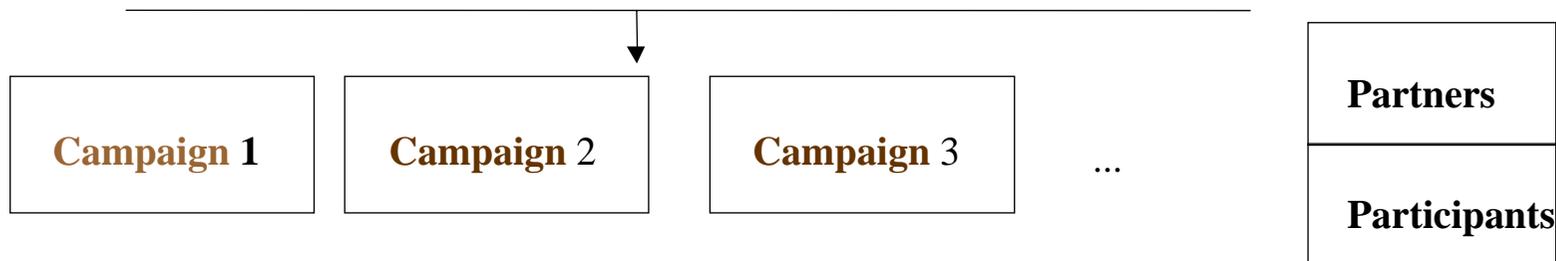
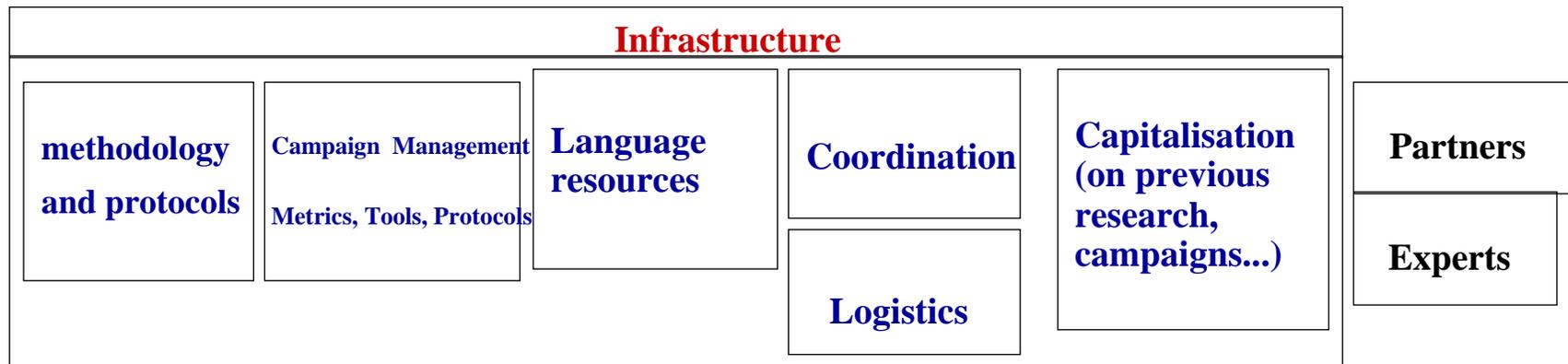
- Drawing on its experience in national and Europe-wide evaluation projects
- on its experience in the production, validation, packaging and distribution of language resources (+legal issues)
- Set up **network of evaluation** (R&D) centres providing evaluation resources, software, methodologies, protocols
- Carry out independent evaluation (ethical)

ELRA/ELDA (evaluation department) has the capacities to set up a European clearing house for HLT evaluation ... based on a network of active specialized partners ...

.... in the same way that ELDA has become a major clearing house for language resources (distribution team, validation units, production partners, ...).

ELRA Evaluation networks

Steering committee, scientific organising committee



ELRA's Contribution to Evaluation

-
- **2000-2001** Launch of **EVALUATION** Campaigns
(Amaryllis, ARCADE/ROMANSEVAL, Aurora, CLEF...)
 - **2003- Evaluation projects:**
Technolanguage, CHIL, TC-STAR, Over 20 technology components
 - **2005 HLT Portal:**
Information, evaluation packages, evaluation services

The BLARK Concept: Three steps initiative:

- Define the BLARK, i.e. for every language a specification of the minimum general text or spoken corpus, basic tools to manipulate it and skills required to be able to do any pre-competitive research for the language.
- Identify existing data collections, tools, for that language (considering multilingual and cross-lingual aspects).
- Initiate co-ordinated actions to fill in the gaps (funding schemes, etc.)

**Extend the BLARK Concept to HLT Evaluation (BLARK Eval),
Basic HLT Components Morphosyntactic analyzers
More sophisticated Components Speech to Speech Translation**

Conclusion

- Permanent Evaluation Infrastructure
- Experience in LR production, validation, distribution etc.

Networks of active specialized partners

- Inspiration: DARPA, ELSE, EC and national evaluation projects (Technolanguge, TC-STAR, CHIL, etc;)
- Reactive and proactive approach
- Evaluation packages (resources, scoring software, protocols, agreed methodologies, best practices, know-how etc.) => ELRA catalogue of evaluation packages
- HLT Evaluation Portal