

HUMAN LANGUAGE TECHNOLOGY EVALUATIONS

Rationales & Experiences

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.elda.org/> or <http://www.elra.info/>

Presentation Outline

- ❑ (Quick description) European language Resources Association For archives!
- ❑ Human Language Technologies
- ❑ **Human Language Technologies Evaluation (s)**
 - ❑ What, why, for whom, how
- ❑ **Examples of Evaluation campaigns**
- ❑ **ELDA Online (UIMA-based) evaluation platforms**
- ❑ **Conclusions**

European Language Resource Association

An Improved infrastructure for Data sharing & HLT evaluation



➤ **An Association of users of Language Resources, established in 1995**

⇒ **A Repository Center:**

⇒ **Technical & Logistic issues**

⇒ **Commercial issues (prices, fees, royalties)**

⇒ **Legal issues (Licensing, IPR)**

⇒ **Information Dissemination**

➤ **Infrastructure for the evaluation of Human Language Technologies**
providing resources, tools, methodologies, logistics,
Capitalization on evaluation packages

- **Membership Drive:**

- ⇒ **Non for profit organization ... association of users of LRs**

- ⇒ **Pay per Resource**

- **Some of the benefits of becoming a member:**

- ✓ Discount on LRs prices
 - ✓ Identification & Production of specific resources
 - ✓ Legal assistance
 - ✓ Evaluation of specific Human language Technologies
 - ✓ Market Analysis (ELRA reports)
 - ✓ LREC and other events favorable condition (registration fees, proceedings, ...)
 - ✓ Technical reports from validation, production committees
 - ✓ Journal of LR and Evaluation (favorable conditions)
 - ✓ Fidelity Program
 - ✓ Universal Catalogue

ELRA: An efficient infrastructure to serve the HLT Community

5 years of establishment



- **1994-1995 RELATOR** a European-wide consortium with the support of the European Commission, striving to establish a **European repository of linguistic resources**.
- **1995** Set up of the European Language Resources Association (**ELRA**) as an Archiving House, set up of a Distribution Agency (**ELDA**)
- **1996** Our first “List” of identified Language Resources followed by a first Catalogue
- **1997** Negotiation of Distribution rights, Start of Distribution activities, R&D projects on issues related to LRs (co-funded EU, French agencies, etc.)
- **1998** 1st LREC,
Introduction of the BLARK Concept (with ELSNET), Later on ELARK
Production & commissioning the production of LRs (Internal & network of production units)
Market Analysis , Technologies-State of the Art
Evaluation Activities (Else project Jan’98), Integration of ASR & MT (Systran & Dragon), TTS
- **1999** Strong focus on Production (**SpeechDat** family, LRs-P&P, SpeeCon, etc.),
Validation of LRs and LRs quality assessment, Set up of validation units, Bug reporting & Correction patches

ELRA: An efficient infrastructure to serve the HLT Community

5.... years of consolidation and new era



- **2000- The European HLT landscape – HLT benchmark (HLT players Directory - EUROMAP)**
Start of an active partnership with LDC: Coordination of legal & production issues
Launch of **EVALUATION Campaigns** (Amaryllis, Aurora, ...)
- **2001- Launch of CLEF : Cross-Lingual Evaluation Forum**
Identification of New types of Resources **Multimodal Resources** (Isle project)
HLT & LRs **roadmaps**, survey of existing national activities,
- **2003- Evaluation projects: Technolangue, CHIL, TC-STAR, Over 20 technology components**
LangTech2003 (Paris),
Boost of networking actions/Partnership Europe Mediterranean countries (**NEMLAR**)
Meta-Data for LRs cataloguing (**Intera**)
- **2004- Universal Catalogue**
- **2005 HLT Portal: Information, evaluation packages, evaluation services**
LRs Unification Project
- **2006-2007: more resources on evaluation and on-demand evaluation services**

- ✓ **Need for applications/technologies to "better" produce/access information**
 - Information workflow
 - Information retrieval, Question-answering, Multilingual Information access
 - Machine translation, Text/audio Summarization,

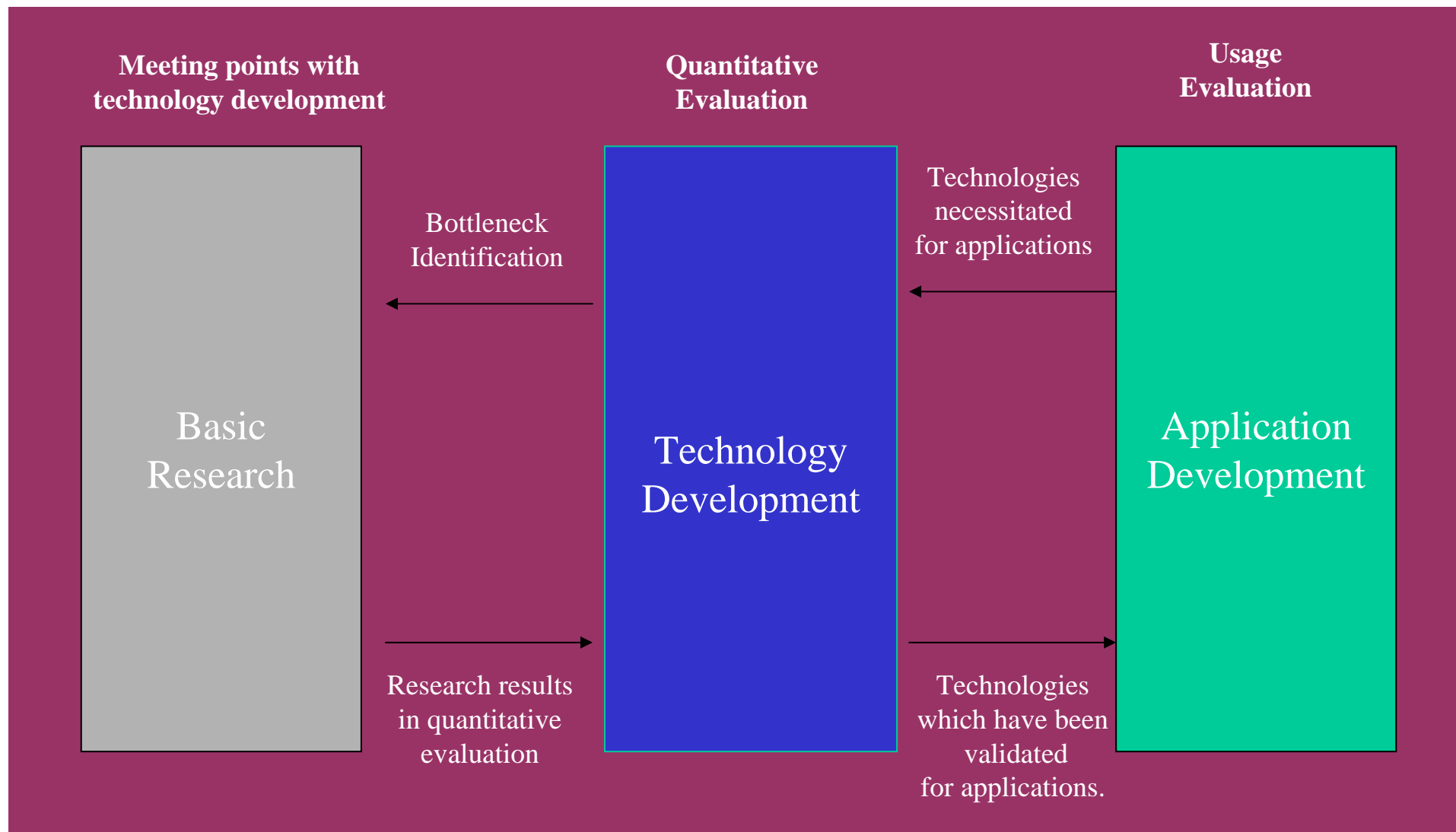
- ✓ **Need access in one's mother tongue**
 - Automatic transcription of audio data (and translation): Speech to Speech Translation

- ✓ **Need for interactive communication services (Multilingual access & speech enabled applications, multimodal dialogue, ...)**

- ✓ **Security & privacy (e.g. biometrics, multimodal interfaces)**

- ✓

What to evaluate ... Levels of Evaluation



What to evaluate ... Levels of Evaluation

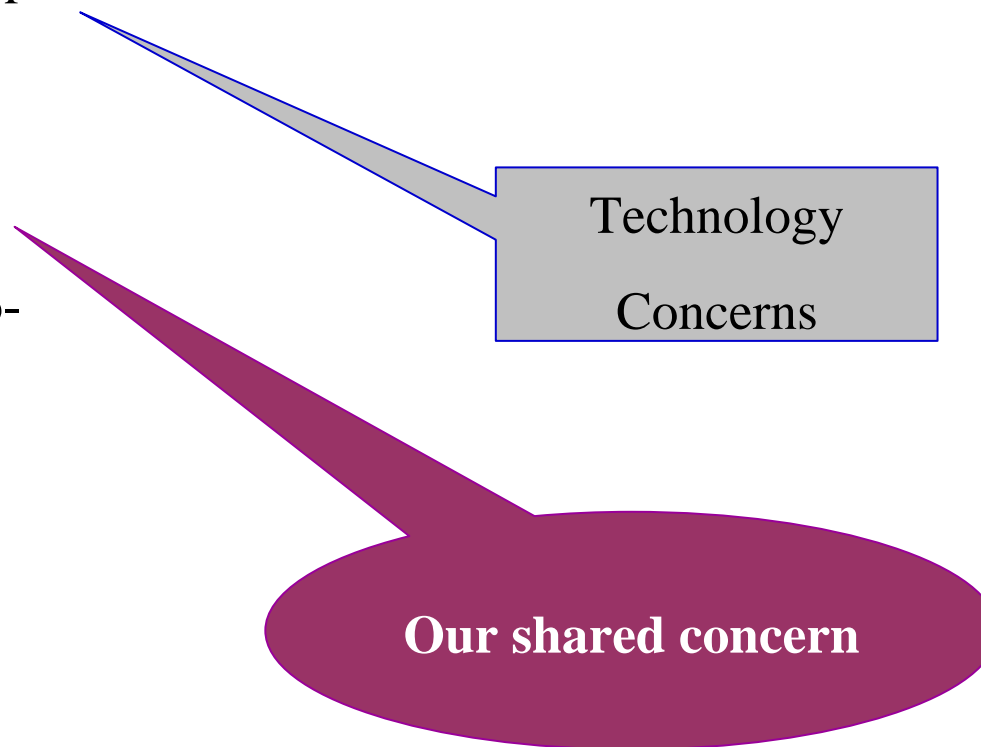
-Basic Research Evaluation
(validate research direction)

-Technology Evaluation
(assessment of solution for
well defined problem)

-Usage Evaluation (end-
users in the field)

-Impact Evaluation (socio-
economic consequences)

-Programme Evaluation
(funding agencies)



What Evaluation is NOT

© Courtesy G.Thurmair Malta workshop.

Latest German MT evaluation campaign

Übersetzungssoftware – Checkliste						
	gPromt	Hexaglot Volltext-Übersetzer 3-9	Langenscheidt T1 Englisch 3-9	Personal Translator	PowerTranslator 9	translate 8
Hersteller	Promt	Hexaglot	Langenscheidt	linguatec	LEC	Lingensio
Vertrieb					Avanquest	digital publishing
Web-Adresse	www.promt.de	www.hexaglot.de	www.langenscheidt.de	www.linguatec.de	www.avanquest.com/de	www.lingensio.de
Registrierung erforderlich	–	–	–	–	✓	–
Lizenzmodell						
Dokumentation gedruckt/auf CD	Handbuch (360 S.) ✓	Booklet (14. S.)/ PDF (205 S.)	–/PDF (147 S.)	Handbuch (117 S.) ✓	Booklet (81 S.) ✓	Handbuch (144 S.) –
Satzarchiv	✓	–	–	–	–	–
Add-ins für Word/Excel/PowerPoint/Outlook	✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓	–	✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓
weitere Add-ins	MS FrontPage, Adobe Reader	Internet Explorer	Internet Explorer, MS Access	Adobe Reader	Internet Explorer	–
Sprachausgabe/OCR-Software	–	–/–	–	✓/–	–/–	–/–
Rechtschreibkontrollen/vorschläge	✓ ✓	–	–	✓ ✓	–	–
Wörterbücher						
Nachschlagen System-WB/bekannte Formen/per Mausclick	✓ ✓ ✓	–/–/–	✓ ✓ ✓	✓ ✓ ✓ ¹	–/–	✓ ✓ ✓ ¹
Idiomatikwörterbuch	–	–	–	–	–	–
Benutzerwörterbuch/mehrere möglich	✓ ✓	✓/–	✓/–	✓	–	✓
Alternativinträge möglich	✓	–	–	✓	–	–
Import von Wortlisten	–	–	✓	–	–	–
Verarbeitbare Formate						
Lesen: TXT/RTF/DOC/HTML/PDF	✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓
Schreiben: TXT/RTF/DOC/HTML/PDF	✓ ✓ ✓ ✓ ✓	–/–/–/–	✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓
Relatives (RMP, TIF, JPG)	✓	–	✓	–	–	–
Import/Export von TMX-Daten	–	–	–	–	–	–
Bedienung						
kontextsensitive Hilfe	✓	–	–	–	–	–
Editor/zwei Texte horizontal/vertikal	✓ ✓ ✓	–/–/–	–/–/–	✓ ✓ ✓	–/–/–	–/–/–
Wörter anschauen per Markierung	✓	–	–	✓	–	–
Übersetzen in Word: als Einfügemarken/Markierung/Abatz	✓ ✓ ✓	–/–/–	–/–/–	✓ ✓ ✓	–/–/–	–/–/–
Übersetzen in Excel: Markierung/Blatt/Mappe	✓ ✓ ✓	–/–/–	–/–/–	✓ ✓ ✓	–/–/–	–/–/–
Übersetzen in PowerPoint: Markierung/Folie/Präsentation	✓ ✓ ✓	–/–/–	–/–/–	✓ ✓ ✓	–/–/–	–/–/–
Zeilebegrenzung einstellbar	–	–	–	–	–	–
Speichern Quelle/Ziel/Beides/Tabelleform	✓ ✓ ✓ ✓	✓ ✓ ✓ ✓	✓ ✓ ✓ ✓	✓ ✓ ✓ ✓	✓ ✓ ✓ ✓	✓ ✓ ✓ ✓
gemeinsam Projekt speichern/Backup-Funktion/Stapelverarbeitung	✓ ✓ ✓	–/–/–	–/–/–	✓ ✓ ✓	–/–/– ⁴	–/–/–
Übersetzung						
automatische Erkennung der Sprachrichtung	✓	–	–	✓	–	–
Sachgebiet: manuelle Wahl/autom. Erkennung	✓ ✓	–/–	–/–	✓ ✓	–/–	–/–
Eigennamen: Kennzeichnung möglich/autom. Erkennung	✓/–	✓/–	✓/–	✓	–	–
Synchronisierte Funktion	✓	–	–	✓	–	–
Wahlmöglichkeit bei Pronomina/Anrede	✓ ✓	–/–	–/–	✓ ✓	–/–	–/–
als Quelltext: Wahlmöglichkeit bei Komposita/alte oder neue BS	✓ ✓	–/–	–/–	✓ ✓	–/–	–/–
engl. Quelltext: Wahlmöglichkeit bei Imperativ/UK- oder US-Englisch	–	–	–	–	–	–
unterschiedliche Markierungen für Alternativen/unbek. Wörter/Eigennamen/Komposita im Zieltext	✓ ✓ ✓ ✓	–/–/–/–	✓ ✓ ✓ ✓	–/–/–/–	–/–/–/–	–/–/–/–
unbekannte Wörter: suchen/Liste drucken	✓ ✓	–/–	✓ ✓	✓/–	–/–	–/–
Zeilenumbruch als Satzende wahlweise	✓	–	–	–	–	–
Bewertung						
Dokumentation	○	⊖	○	⊕	⊖	⊕
Bedienung	○	⊖	⊖	⊕	○	⊕
Integration in andere Anwendungen	⊕⊕	⊕⊕	○	⊕	⊕	⊕
Übersetzung Alltagstexte	⊕⊕⊕	⊕	⊕	⊕⊕⊕	⊕	⊕⊕⊕
Übersetzung technische/wissenschaftliche Texte	⊕⊕	○	○	⊕	○	⊕
Wortschatz	⊕⊕⊕	⊕	⊕	⊕⊕⊕	○	⊕⊕⊕
Preis						
getestete Version	Professional 500 €	Office 60 €	Professional 300 €	Professional 250 €	Standard 50 €	translate pro 250 €
weitere	Office 100 € Experts 690 €	–	Standard 50 € Standard plus 100 €	Home 50 € Office 100 €	Professional 200 € (Engl., Frz., Italienisch)	translate 50 € translate plus 100 €

APPENDIX A - COMPARISON TABLE

BW=Browser base, DB=Deskbar, FB=Floating bar, SA=Standalone Application, TB=Toolbar

Criteria / Desktop Search Tools	Ask Jeeves	Blinkx	Copernic	diskMETA	dtSearch	Enfish	Google	HotBot	ISYS	MSN	Likasoft	Yahoo!
1. Usability	4.25	3.00	4.80	2.50	2.50	3.40	4.40	2.00	1.75	4.40	4.75	4.00
1.1. Application Types	SA	SA, TB	SA, DB	SA	SA	SA	BW, DB, FB	BW, TB, DB	SA	BW, TB, DB	SA	SA, TB
1.2. Features	■■■	■■■	■■■■■	■■■	■■■■	■■■■■	■■■	■■■	■■■	■■■■	■■■■■	■■■■■
1.3. Simplicity	■■■■■	■■■■■	■■■■■	■■■■■	■■■■	■■■	■■■■■	■■■	■■■	■■■■■	■■■■■	■■■■■
1.4. Navigations	■■■■■	■■■■	■■■■■	■■■	■■■	■■■■	■■■■■	■■■	■■■	■■■■■	■■■■■	■■■■■
1.5. Aesthetic	■■■■■	■■■■	■■■■■	■■■	■■■	■■■■■	■■■■■	■■■■	■■■	■■■■■	■■■■■	■■■■■
1.6. Others in Usability			■■■■■			■■■■■	■■■■■			■■■■■		
2. Versatility	2.57	3.75	4.14	2.43	3.25	3.38	3.50	3.00	3.75	3.63	3.38	3.88
2.1. Supported PC Environment	■■■	■■■■■	■■■	■■■■	■■■■	■■■■	■■■	■■■■	■■■■	■■■■	■■■■	■■■
2.2. Supported Files	■■■	■■■■■	■■■■■	■■■■■	■■■■■	■■■■	■■■■	■■■■	■■■■■	■■■■	■■■■■	■■■■■
2.3. Media Support	■■■	■■■■■	■■■■■	■■■■	■■■	■■■	■■■■	■■■	■■■■	■■■■■	■■■	■■■■■
2.4. Application Support	■■■	■■■■■	■■■■■	■■■■	■■■■■	■■■■■	■■■■	■■■	■■■■	■■■■	■■■■■	■■■■■
2.5. Multi-language Support	■■■	■■■■	■■■	■■■	■■■■	■■■■	■■■■■	■■■■	■■■■■	■■■■	■■■■■	■■■■
2.6. Web Integration	■■■	■■■■■	■■■■■	■■■	■■■■■	■■■■	■■■■■	■■■■■	■■■	■■■■	■■■	■■■■
2.7. E-mail Integration	■■■	■■■■■	■■■■■	■■■	■■■■■	■■■■■	■■■■	■■■■	■■■■■	■■■■■	■■■■■	■■■■
2.8. Others in Versatility		■■■■			■■■■	■■■■■	■■■■	■■■■	■■■■■	■■■■	■■■■■	■■■■
3. Accuracy	3.20	2.60	4.50	2.60	3.50	2.40	3.20	2.20	2.80	4.20	3.20	3.20
3.1. Word Accuracy	■■■	■■■■■	■■■	■■■■■	■■■■■	■■■	■■■■	■■■	■■■	■■■■■	■■■■	■■■■
3.2. Additional Word Support	■■■	■■■	■■■■■	■■■■■	■■■■■	■■■	■■■■	■■■	■■■■■	■■■■	■■■■	■■■■
3.3. Index Accuracy	■■■■	■■■	■■■■■	■■■	■■■	■■■	■■■■	■■■	■■■	■■■■■	■■■	■■■
3.4. Output Format	■■■	■■■	■■■■■	■■■	■■■■	■■■	■■■■	■■■■	■■■	■■■■	■■■■■	■■■■■
3.5. Filter & Sort	■■■	■■■■	■■■■■	■■■	■■■	■■■■	■■■	■■■	■■■	■■■■■	■■■■	■■■■
3.6. Others in Accuracy			■■■■■		■■■■■							
4. Efficiency	3.80	1.80	4.20	3.40	3.00	3.40	3.33	2.00	3.00	3.60	4.40	3.60
4.1. Download/Installed File Size	■■■■■	■■■	■■■■■	■■■■■	■■■■	■■■■	■■■■■	■■■■■	■■■	■■■■■	■■■■■	■■■■
4.2. Indexed File Size	■■■■■	■■■	■■■■■	■■■■■	■■■	■■■■	■■■■	■■■	■■■■	■■■■■	■■■■■	■■■■
4.3. Initial Index Time	■■■■	■■■	■■■■■	■■■■■	■■■	■■■■	■■■■	■■■	■■■■	■■■■	■■■■■	■■■■■
4.4. Index Controls	■■■■■	■■■	■■■■■	■■■	■■■■■	■■■■■	■■■	■■■	■■■■	■■■■■	■■■■■	■■■■■
4.5. Memory & CPU Usages	■■■■■	■■■	■■■■	■■■■■	■■■■	■■■■	■■■■	■■■	■■■■	■■■■■	■■■■■	■■■■■
4.6. Others in Efficiency					■■■■■		■■■■■		■■■■■			
5. Security	3.14	2.63	3.00	2.86	2.88	3.00	3.13	2.86	3.00	2.86	3.00	3.29
5.1. HTTPS Cache Indexing	■■■	■■■	■■■■	■■■	■■■■	■■■■	■■■■■	■■■■	■■■■	■■■■	■■■■	■■■■
5.2. Personal Folder Search	■■■■	■■■	■■■■	■■■■	■■■■	■■■■	■■■■	■■■■	■■■■	■■■■■	■■■■	■■■■
5.3. Possible Intrusion	■■■■	■■■	■■■■	■■■■	■■■■	■■■■	■■■	■■■	■■■■	■■■	■■■■	■■■■
5.4. Protection Features	■■■■■	■■■■	■■■■	■■■■	■■■■	■■■■■	■■■■	■■■■	■■■■	■■■■	■■■■	■■■■■
5.5. Privacy	■■■	■■■■	■■■	■■■■	■■■	■■■■	■■■■	■■■■	■■■■■	■■■■	■■■	■■■■■
5.6. Spyware & Adware	■■■■	■■■■	■■■■	■■■■	■■■■	■■■■	■■■■	■■■■	■■■■	■■■■	■■■■	■■■■
5.7. Product Update	■■■■■	■■■■■	■■■■■	■■■■	■■■■	■■■	■■■■	■■■■	■■■	■■■	■■■■■	■■■
5.8. Others in Security		■■■■■			■■■■■		■■■■■					
6. Enterprise Readiness	2.00	2.00	4.00	2.00	3.00	3.00	2.00	2.00	4.00	2.00	2.00	4.00
6.1. Enterprise Products	■■■	■■■	■■■■■	■■■	■■■■	■■■■	■■■	■■■	■■■■	■■■	■■■	■■■■■
Overall Scores	3.16	2.63	4.11	2.63	3.02	3.10	3.26	2.34	3.05	3.45	3.62	3.66
Costs	Free	Free	Free	\$97.50	\$199.00	\$199.95	Free	Free	\$570.00	Free	\$25.00 - 45.00	Free
Overall Ratings	6	10	1	10	9	7	5	12	8	4	3	2

Source: UW E-Business Consortium

HLT Evaluations For whom

-Basic Research Evaluation
(validate research direction)

-Technology Evaluation
(assessment of solution for well defined problem)

-Usage Evaluation (end-users in the field)

-Impact Evaluation (socio-economic consequences)

-Programme Evaluation
(funding agencies)

- *MT developers* want to improve the “quality” of MT output
- *MT users* (humans or software e.g. CLIR) want to improve productivity using the most suitable MT system (e.g. multilinguality)
-

Requirements for an evaluation campaign

- Select a common "task" Precompetitive ... Cooperation + Competition
- Attract enough Participants
- Organize the campaign (protocol/metrics/data)
- Follow-up workshop, interpret results and share info

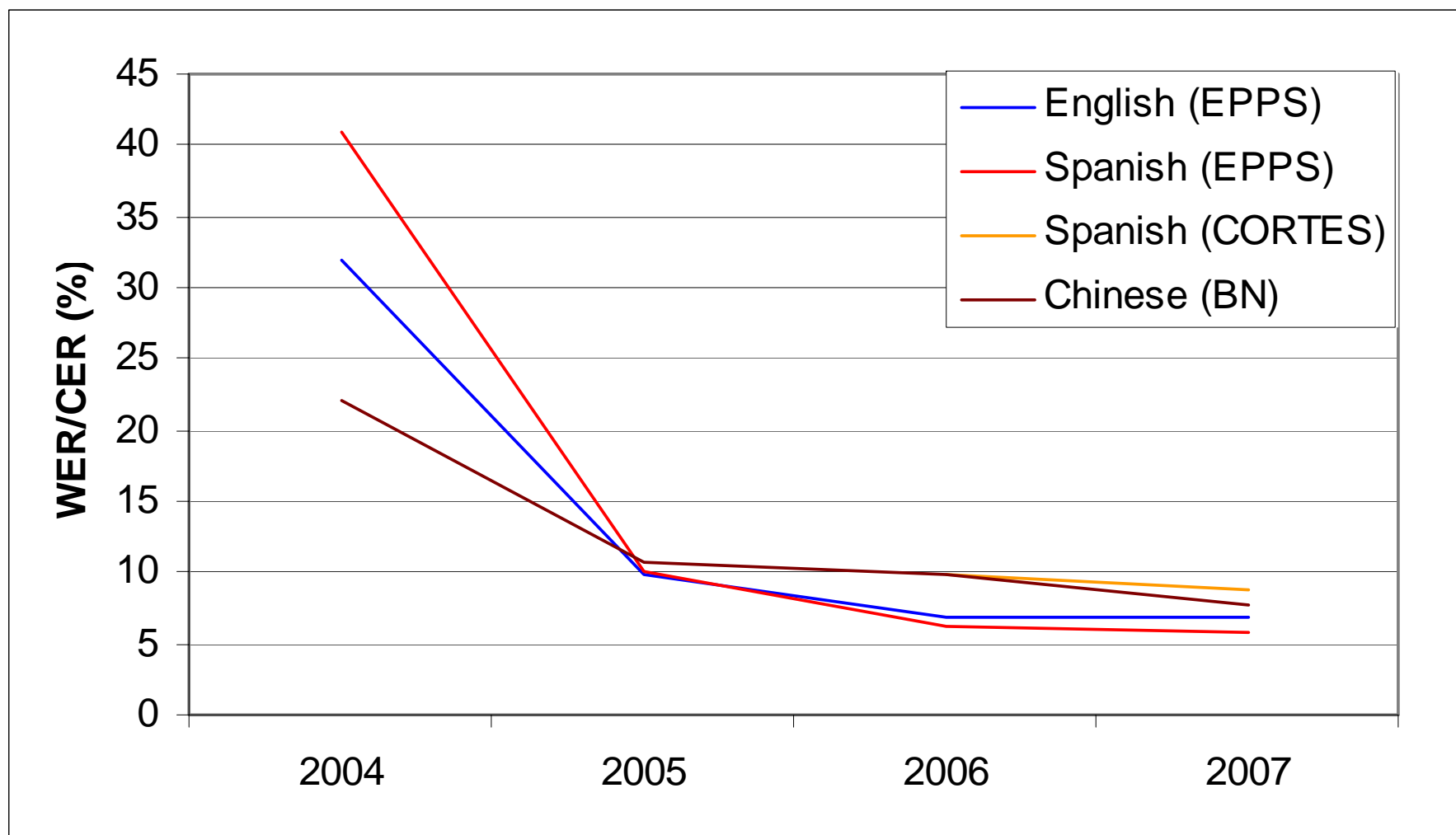
- Referential Language Resources (Data) (truth)
- Metric(s): Automatic, Human judgments ... scoring software
- scale/range of performance to compare with (Baseline)

- Logistics' Management

- reliability assessment: independent body

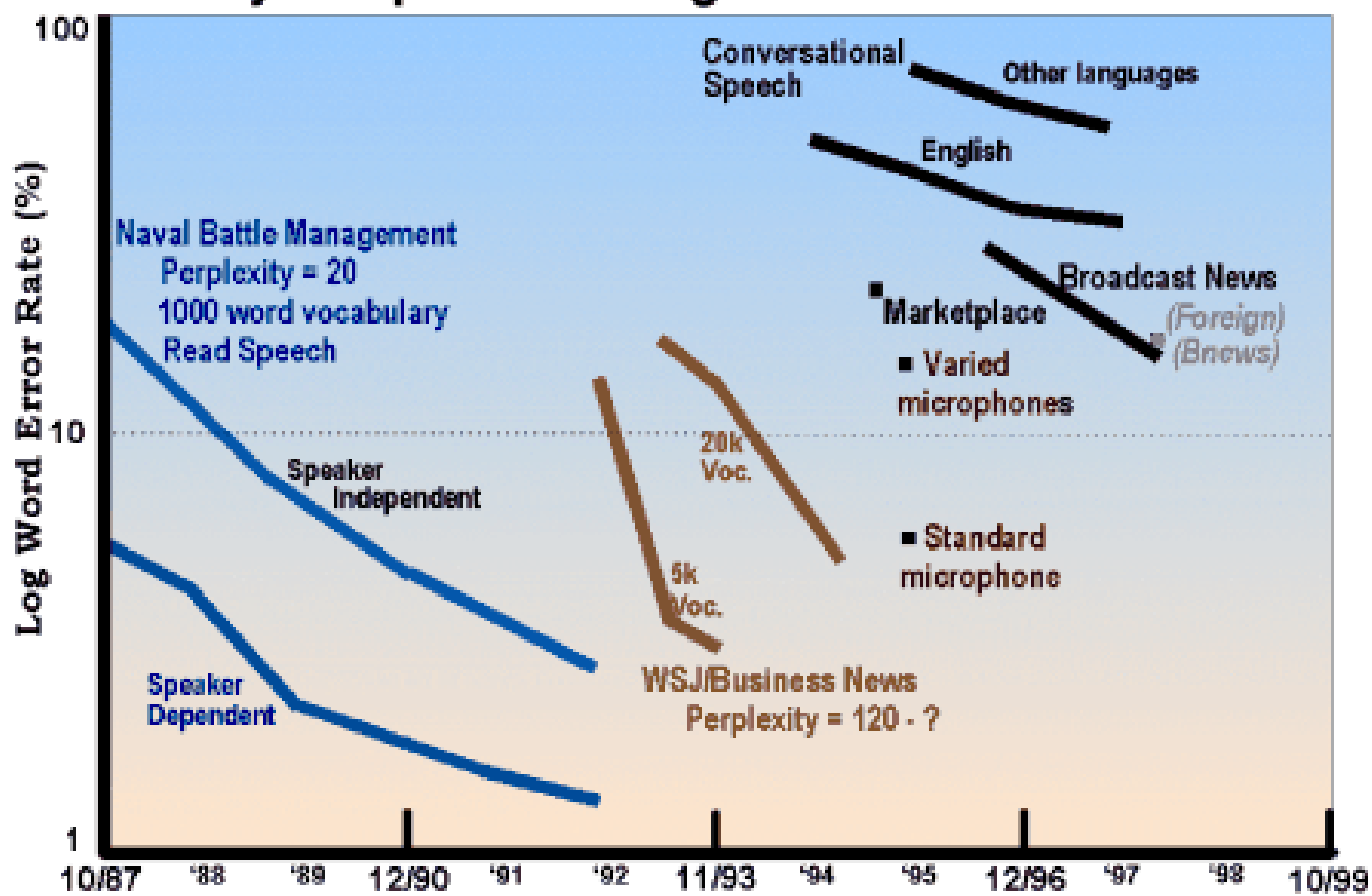
- **Participants: technology providers**

Evaluation & Progress measures



Progress & Evaluation (Courtesy Charles Wayne)

History of Speech Recognition Benchmark Tests



For whom ... essential for technology development

- Share of Information and knowledge between **participants**: (how to get the best results, access to data, scoring tools)
- Information obtained by **industrialists**: state of the art, technology choice, market strategy, new products.
- Information obtained by **funding agencies**: technology performance, progress/investment, priorities

Some types of evaluations

- Comparative evaluation
 - the same or similar control tasks and related data with metrics that are agreed upon
- Competitive vs Cooperative
- Black box evaluation ... Glass box
- Objective evaluation ... Subjective (Human-based)
- Corpus based (test suites)
- Quantitative measures ... Qualitative

- Used successfully in the USA by DARPA and NIST (since 1984)
- Similar efforts in Europe on a smaller scale, mainly projects (EU funded or national programs)
 - Select a common "task"
 - Attract enough Participants
 - Organize the campaign (protocol/metrics/data)
 - Follow-up workshop, interpret results and share info

➤ Speech & Audio/sound

✓ ASR: TC-STAR, AMI, CHIL, ESTER

✓ TTS: TC-STAR, EVASY

✓ Speaker identification (CHIL)

✓ Speech 2 Speech Translation

✓ Speech Understanding (Media)

✓ Acoustic Person tracking

✓ Speech activity detection,

✓

➤ Written NLP & Content

- ✓ IR, CLIR , QA, (Amaryllis, EQUER, CLEF)
- ✓ Text analysers (Grace, EASY)
- ✓ MT (CESTA, TC-STAR)
- ✓ Corpus alignment & processing (Arcade, Arcade-2, Romanseval/Senseval, ...)
- ✓ Term & Terminology extraction
- ✓ *Summarisation*

➤ Multimodal --- Video – Vision technologies

- ✓ Face Detection
- ✓ Visual Person Tracking
- ✓ Visual Speaker Identification
- ✓ Head Pose Estimation
- ✓ Hand Tracking

Some of the technologies being evaluated within CHIL ...<http://chil.server.de/>

A) Vision technologies

- A.1) Face Detection
- A.2) Visual Person Tracking
- A.3) Visual Speaker Identification
- A.4) Head Pose Estimation
- A.5) Hand Tracking

B) Sound and Speech technologies

- B.1) Close-Talking Automatic Speech Recognition
- B.2) Far-Field Automatic Speech Recognition
- B.3) Acoustic Person Tracking
- B.4) Acoustic Speaker Identification
- B.5) Speech Activity Detection
- B.6) Acoustic Scene Analysis

more at the CHIL/CLEAR workshops

C) Contents Processing technologies

- C.1) Automatic Summarisation ... Question Answering

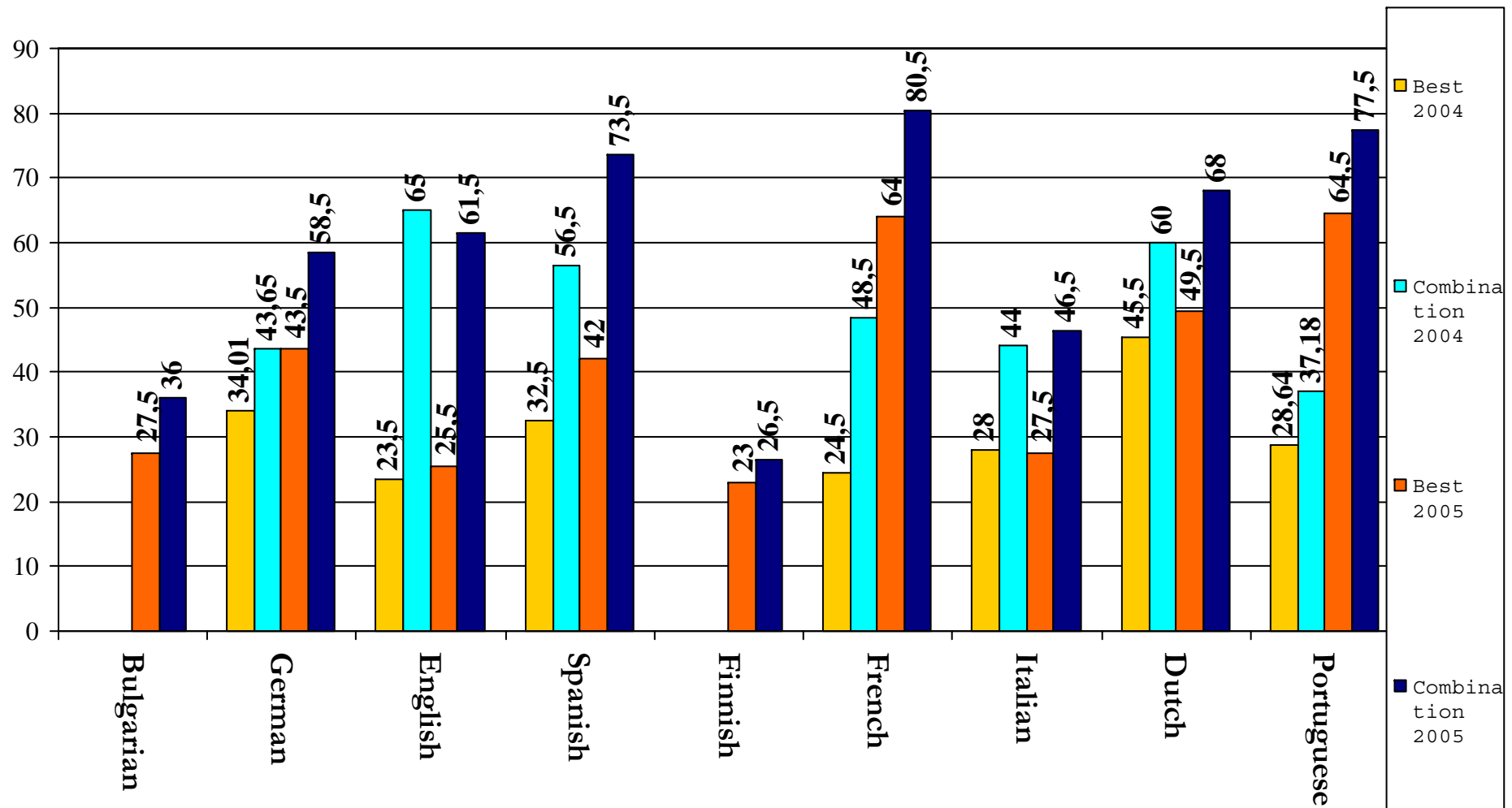
⇒ CLEF (Cross-Language Evaluation Forum)

Promoting research and development in Cross-Language Information Retrieval (CLIR)

(i) providing an infrastructure for the testing and evaluation of information retrieval systems - European languages - monolingual and cross-language contexts

(ii) creating test packages of reusable data which can be employed by system developers for benchmarking purposes.

QA-CLEF: State of the art & Improvements



TC-STAR evaluations

..... 3 Consecutive annual evaluations

1.SLT in the following directions

- i. Chinese-to-English (Broadcast News)
- ii. Spanish-to-English (European Parliament plenary speeches)
- iii. English-to-Spanish (European Parliament plenary speeches)

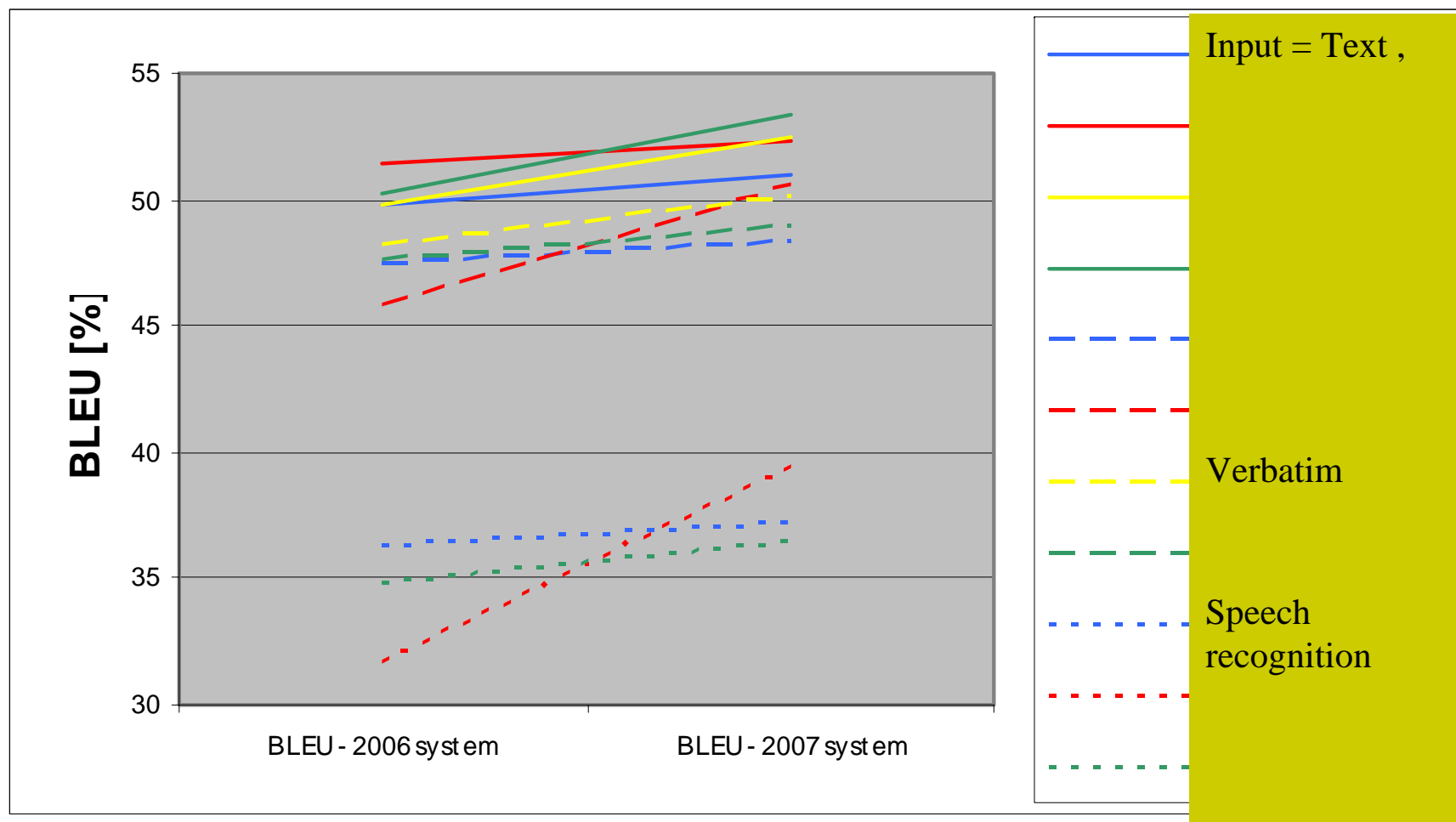
2.ASR in the following languages

- i. English (European Parliament plenary speeches)
- ii. Spanish (European Parliament plenary speeches)
- iii. Mandarin Chinese (Broadcast News)

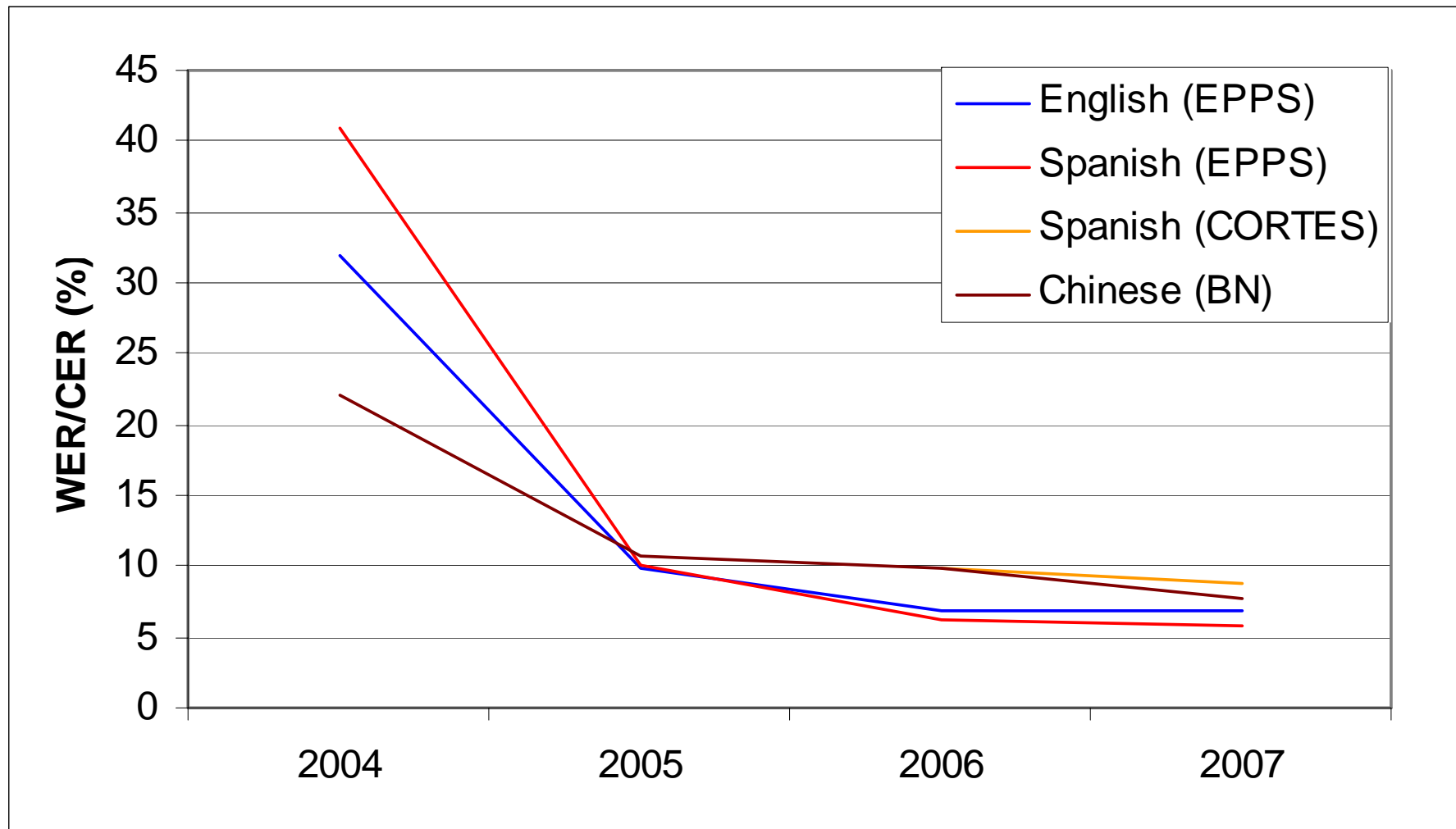
3.TTS in Chinese, English, and Spanish under the following conditions:

- i. Complete system
- ii. Voice conversion intralingual and crosslingual, expressive speech:
- iii. Component evaluation

Improvement of SLT Performances (En → Es)

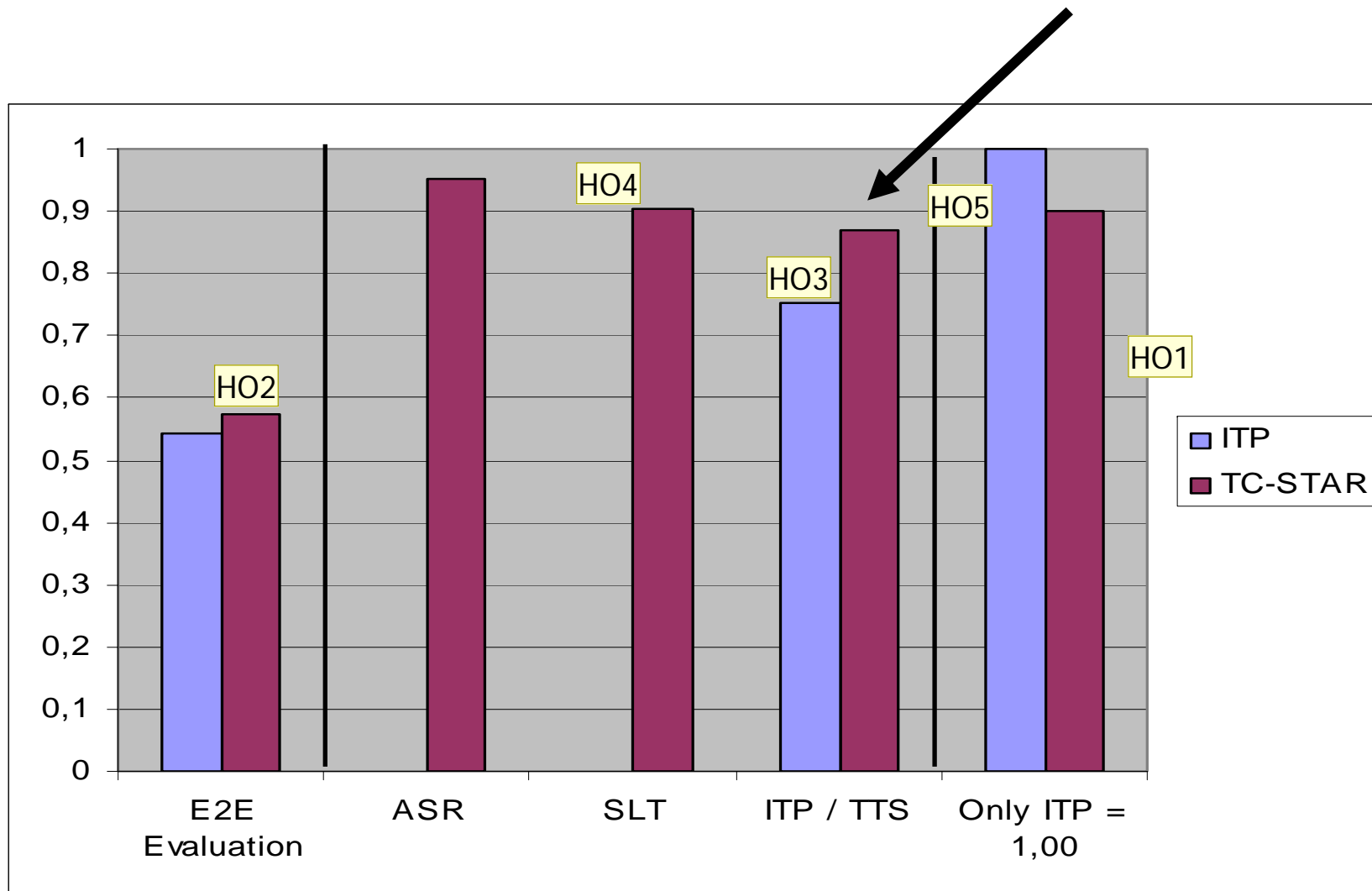


Improvement of ASR Performances (En→Es)



- The end-to-end evaluation is carried out for 1 translation direction: English-to-Spanish
- Evaluation of ASR (Rover) + SLT (Rover) +TTS (UPC) system
- Same segments as for SLT human evaluation
- Evaluation tasks:
 - **Adequacy**: comprehension test → *Very surprising results!*
 - **Fluency**: judgement test with several questions related to fluency and also usability of the system

End to End results from 2006 tests



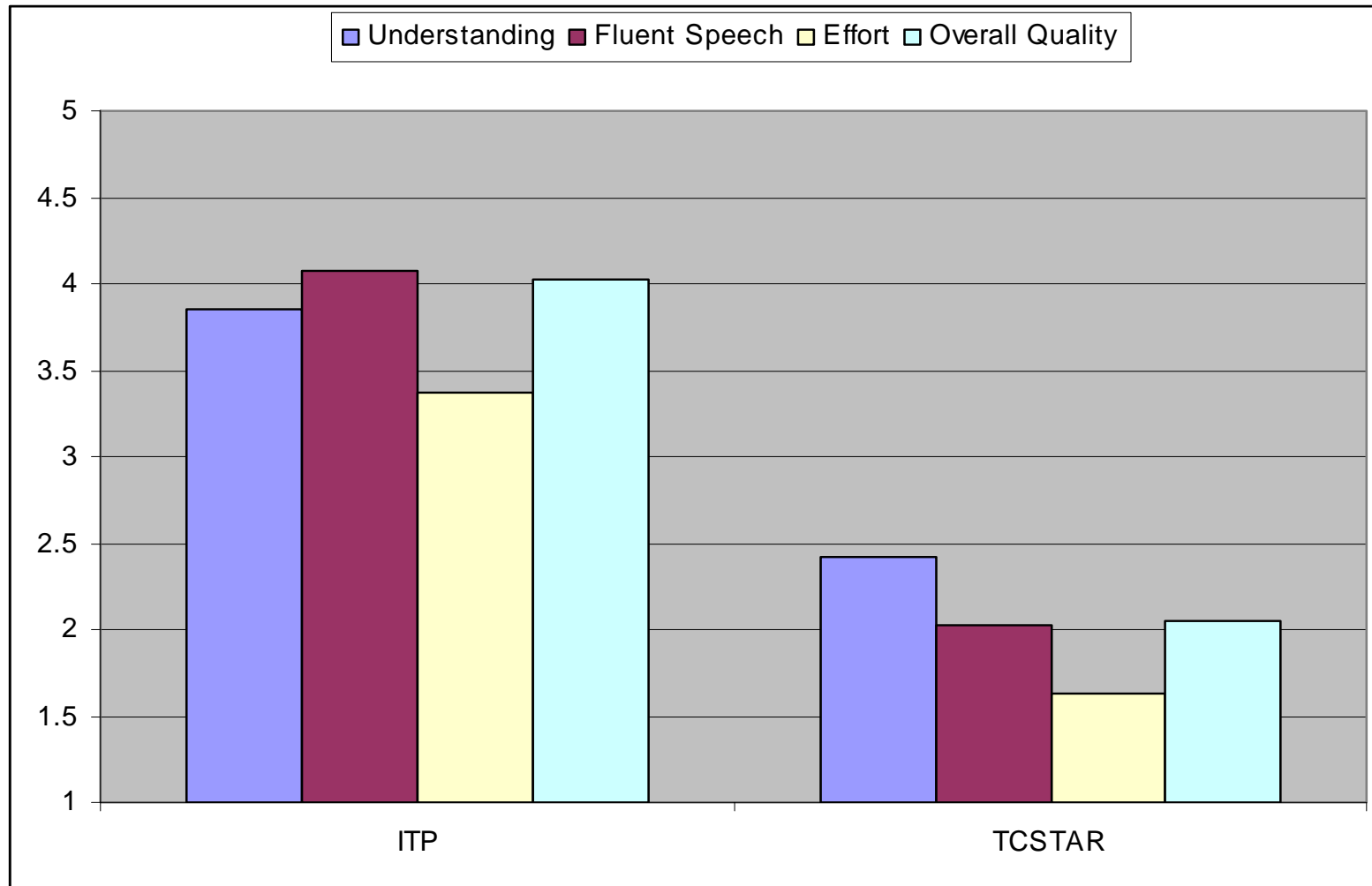
Diapositiva 29

- HO1** -2 evaluated systems: ITP for the interpreter version and TC-STAR for the automatic speech-to-speech translation system
Hamon Olivier; 16/06/2006
- HO2** - E2E Evaluation: the evaluation was done by the same assessors who did the subjective evaluation.
Hamon Olivier; 16/06/2006
- HO3** - ITP / TTS: as it was not foreseen that results would be better for TC-STAR than for ITP, the audio files had been validated to check whether they contained the answers to the questions. The first conclusions that can be drawn from this are: it was difficult for the assessors to find the answers (questions too hard?) and as the interpreter selects and reformulates the information, missing some details, then the question becomes too specific and not appropriate.
Hamon Olivier; 16/06/2006
- HO4** - TTS, SLT, ASR: in order to determine where the information was lost for the TC-STAR system, files from each component (recognized files for ASR, translated files for SLT, synthethized files for TTS) have been checked. The overall loss is 15% of the information, 5% being lost at each step.
Hamon Olivier; 16/06/2006
- HO5** - Only ITP: in the end, we used the questions whose answers were included in the interpreter files. So the TC-STAR system lost 10% of the information regarding the ITP evaluation (instead of 15%).
Hamon Olivier; 16/06/2006

Fluency questionnaire

- [Understanding] Do you think that you have understood the message?
1: Not at all ,5: Yes, absolutely
- [Fluent Speech] Is the speech in good Spanish?
1: No, it is very bad 5: Yes, it is perfect
- [Effort] Rate the listening effort
1: Very high 5: Low, as natural speech
- [Overall Quality] Rate the overall quality of this audio sample
1: Very badm unusable 5: It is very useful

End to End results 2007 results (subjective test: 1...5)



Online Evaluation platforms

- Download test suites, Upload results, receive your scores, support for Interpretation of results
- "UIMA" Based evaluation platform
- Fully automatic evaluation
- So far ASR, MT, Parsing,
- Validation on TC-STAR & Passage projects

Some concluding remarks on Technology evaluation

- It saves developers time and money
- It help assess progress accurately
- Identify promising technologies (market)
- It produces reusable evaluation packages
- Benchmarking ... state of the art
- Share knowledge ... dedicated workshops
- Feedback ... Funding agencies
- It helps to identify areas where more R&D is needed

<http://www.hlt-evaluation.org/>

- [Overview](#)
- [HLT Evaluations](#)
- [Activities by technology](#)
- [Activities by geographical region](#)
- [Players](#)
- [Evaluation resources](#)
- [Evaluation Services](#)

Let us list some well known campaigns

LREC 2008 (6th Edition)



LREC'2008
28 May- 1st June 2008
Marrakech, Morocco

Thank you very much for you attention

Slides will be made available