

Overview of the Multilingual Question Answering Track

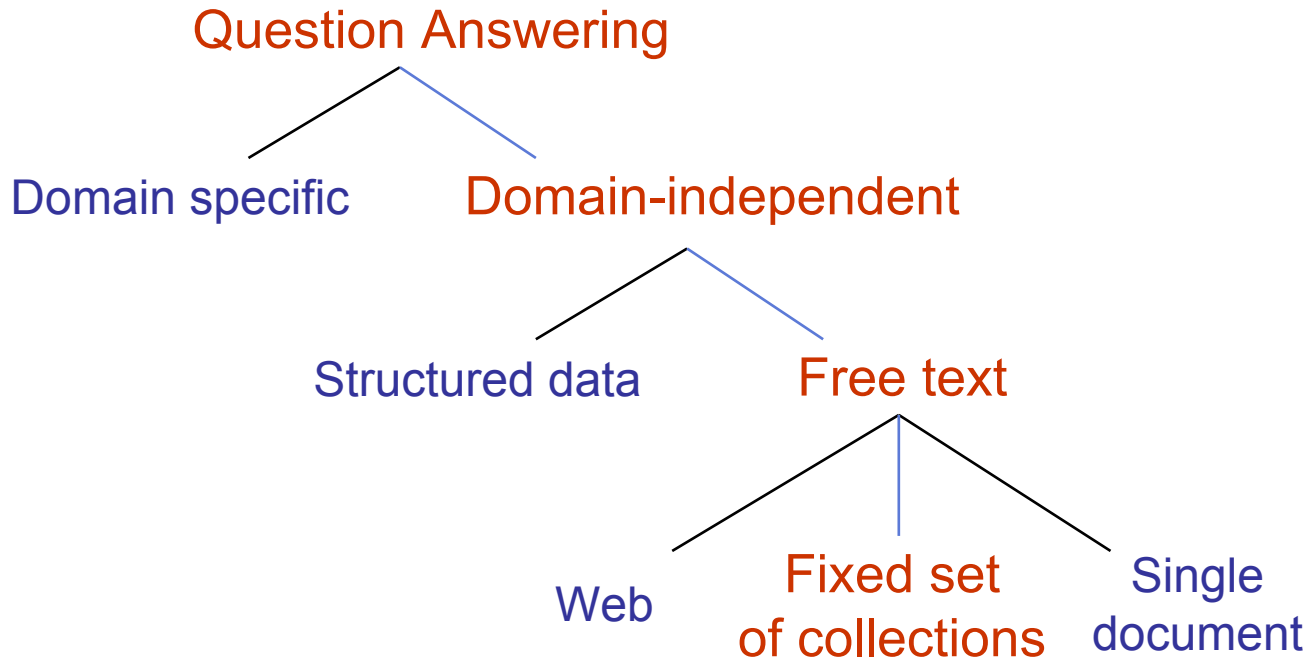
Alessandro Vallin, Danilo Giampiccolo, Bernardo Magnini

CELCT - ITC-irst
(Trento, Italy)

Outline

- Multilingual Question Answering
- Task
- Participants
- Results
- Approaches
- Conclusions and Future Directions

Open Domain Question Answering



QA at CLEF: Focus on **multilingual question answering**

- monolingual tasks in other languages than English
- cross-language QA: questions in a source language, answers in a target language

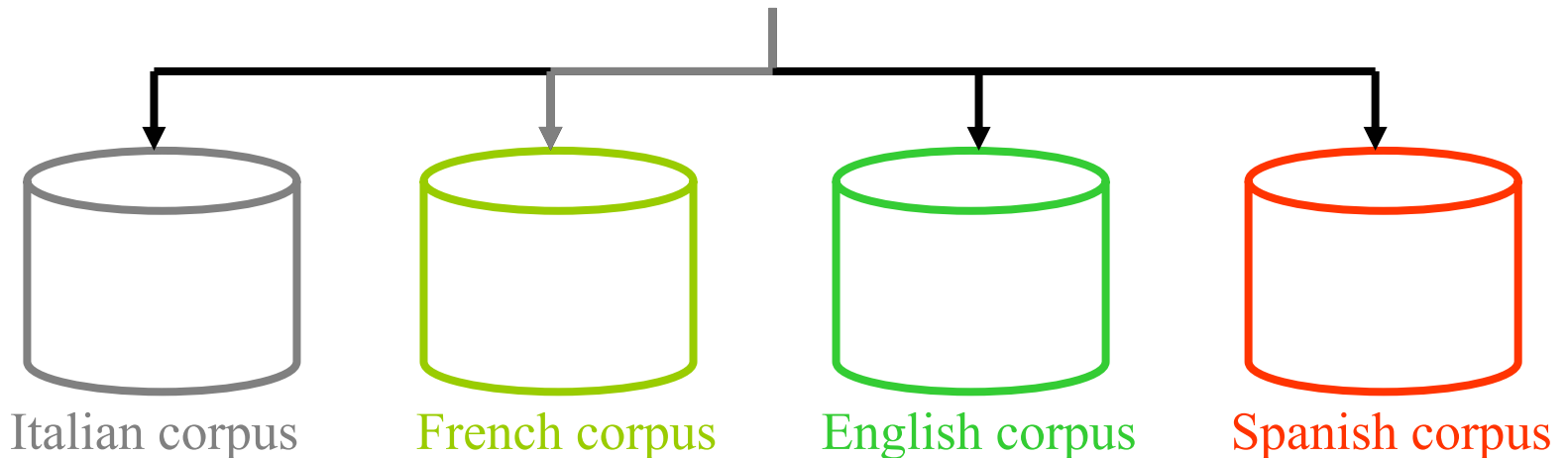
Multilingual Question Answering

- Answers may be found in languages different from the language of the question.
- Interest in QA systems for languages other than English.
- Force the QA community to design real multilingual systems.
- Check/improve the portability of the technologies implemented in current English QA systems.

Cross-Language QA

Quanto è alto il Mont Ventoux?

(How tall is Mont Ventoux?)



“Le Mont Ventoux, impérial avec ses *1909 mètres* et sa tour blanche telle un étendard, règne de toutes ...”

↓
1909 metri

QA at CLEF

- TREC QA style
 - Prevalence of Factoid questions
 - Exact answers + document id
- Use the CLEF corpora (news, 1994 -1995)
- Return the answer in the language of the text collection in which it has been found (i.e. no translation of the answer)
- QA-CLEF started as a pilot in 2003

QA-CLEF-05: Organization

Nine groups coordinated the QA track:

- **CELCT / ITC-irst** (A. Vallin, D. Giampiccolo): Italian
- **DFKI** (G. Erbach, B. Sacalenau): German
- **ELDA/ELRA** (C. Ayache): French
- **Linguatca** (D. Santos): Portuguese
- **UNED** (A. Penas): Spanish
- **U. Amsterdam** (M. De Rijke): Dutch
- **U. Limerick** (R. Sutcliff): English
- **Bulgarian Academy of Sciences** (P. Osenova): Bulgarian
- **U. Helsinki** (I. Aunimo): Finnish
- University of Indonesia: Indonesian as source

QA-CLEF-05: Task

Questions: 200 open domain questions; three kinds:

- **Factoid** (ca. 50%): *Which Arab country produces the most oil?*
 - **Definition** (ca. 25%): *Who is Josef Paul Kleihues?*
 - **Temporally restricted** (ca. 15%): by period, event and date
 - **NIL** (ca. 10%):
- **Answers:** exact answer in the target language
 - **Document collections:** open domain news corpora (on average 230MB)
 - **Evaluation:**
 - Each answer: Right, Wrong, ineXact, Unsupported
 - Two runs for participant

QA-CLEF-05: Temporally Restricted Questions

- **By period:** ex. *Which liberal politician was Italy's Health Minister from 1989 to 1993?*
- **By event:** ex. *Who was the Norwegian Prime Minister when the referendum on Norway's possible accession to the EU was held?*
- **By date:** ex. *Which country did Jimmy Carter visit in 1994 at Bill Clinton's request?*

QA-CLEF-05: Evaluation

- Main evaluation measure was **accuracy** (fraction of Right responses).
- Whenever possible, a **Confidence-Weighted Score** was calculated:

$$\text{CWS} = \frac{1}{Q} \sum_{i=1}^Q \frac{\text{number of correct responses in first } i \text{ ranks}}{i}$$

- Beside CWS, two new measures were introduced, namely K1 and r value, to take into account both accuracy and confidence

QA-CLEF-05: MultiNine

```
<q answer_type="ORGANIZATION" category="D" cnt="0001"
restriction="NONE">
```

```
<language original="TRUE" val="BG">
```

```
<question group="BTB">Кой е Джулио Андреоти?</question>
```

```
<answer docid="SEGA20020315-052" n="1">бившия италиански премиер и пожизнен сенатор</answer>
```

```
</language>
```

```
<language original="FALSE" val="DE">
```

```
<question group="DFKI">Wer ist Giulio Andreotti?</question>
```

```
</language>
```

```
<language original="FALSE" val="EN">
```

```
<question group="BTB">Who is Giulio Andreotti?</question>
```

```
<answer docid="GH950303-000082" n="1">former Italian Prime
Minister</answer>
```

```
</language>
```

```
<language original="FALSE" val="ES">
```

```
<question group="UNED">¿Quién es Giulio Andreotti?</question>
```

```
<answer docid="EFE19951018-12131" n="1">ex primer ministro italiano</answer>
```

```
<answer docid="EFE19950921-13683" n="2">senador vitalicio</answer>
```

```
</language>
```

```
<language original="FALSE" val="FI">
```

```
<question group="HELS">Kuka on Giulio Andreotti?</question>
```

```
<answer docid="aamu19950927.sgml" n="1">Italian entinen pääministeri</answer>
```

```
</language>
```

```
<language original="FALSE" val="FR">
```

```
<question group="ELDA">Qui est Giulio Andreotti ?</question>
```

```
</language>
```

```
<language original="FALSE" val="IN">
```

```
<question group="UIN">Siapakah Giulio Andreotti itu?</question>
```

```
</language>
```

```
<language original="FALSE" val="IT">
```

```
<question group="CELCT">Chi è Giulio Andreotti?</question>
```

```
</language>
```

```
<language original="FALSE" val="NL">
```

```
<question group="UAM">Wie is Giulio Andreotti?</question>
```

```
<answer docid="AD19950128-0066" n="1">Italiaanse oud-premier</answer>
```

```
</q>
```

QA-CLEF-05: Potential Tasks

8 Monolingual and 73 bilingual tasks

S T	BG	DE	EN	ES	FI	FR	IN	IT	NL	PT
BG										
DE										
EN										
ES										
FI										
FR										
IT										
NL										
PT										

QA-CLEF-05: Participants

	America	Europe	Asia	Australia	TOTAL	submitted runs
TREC-1999	13	3	3	1	20	46
TREC-2000	14	7	6	-	27	75
TREC-2001	19	8	8	-	35	67
TREC-2002	16	10	6	-	32	67
TREC-2003	13	8	4	-	25	54
TREC-2004	?	?	?	?	28	63
CLEF 2003	3	5	-	-	8	17
CLEF 2004	1	17	-	-	18	48
CLEF 2005	1	22	1	-	24	67

QA-CLEF-05: Activated Tasks

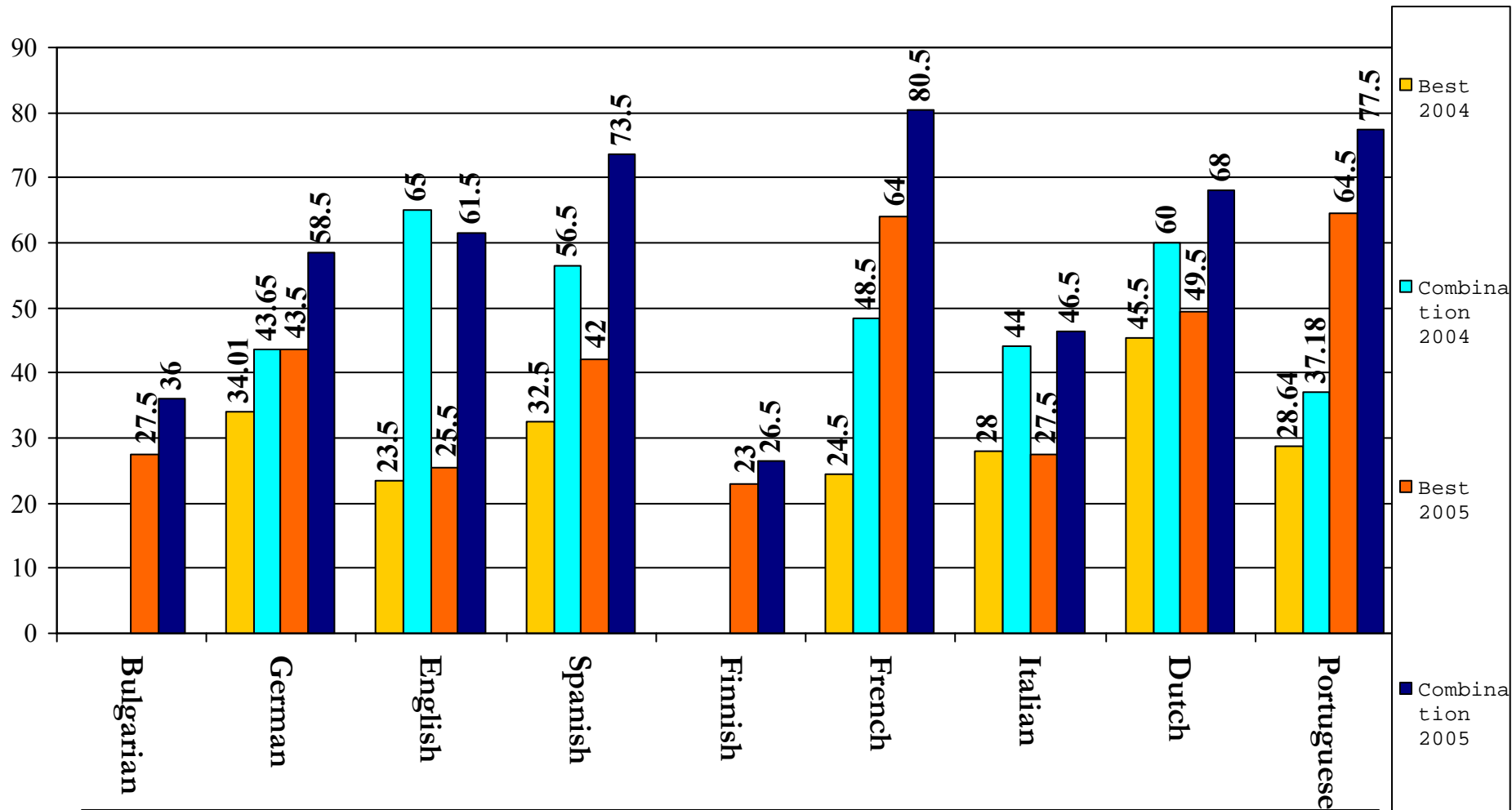
8 Monolingual and 15 bilingual tasks; (6 + 13 in 2004)

Tasks / participant = 1.8; (1.6 in 2004)

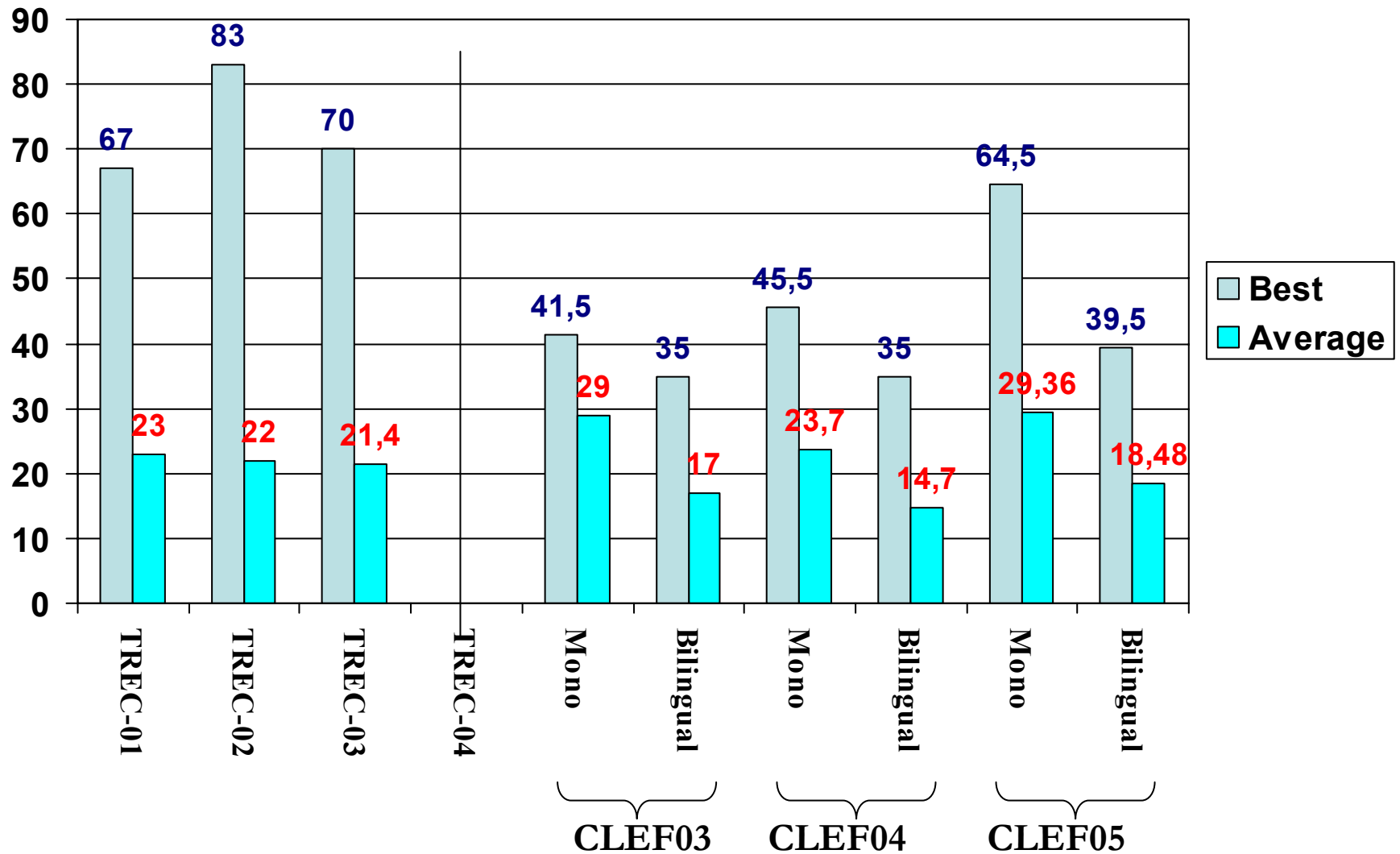
Comparability (tasks > 2 part. / tasks) = 0.2 (0.2 in 2004)

S T	BG	DE	EN	ES	FI	FR	IN	IT	NL	PT
BG	2		1							
DE		2	1							
EN	1	1		1	1	2	1	1		
ES			2	7				1		
FI					1					
FR			1			7		1		1
IT								3		
NL									2	
PT			1							3

QA-CLEF-05: Results



QA-CLEF-05: Results



QA-CLEF-05: Approaches (1)

- **Linguistic processors** and resources are used by most of the systems.
 - POS-tagging, Named Entities Recognition, WordNet, Gazetteers, partial parsing (chunking).
 - Deep parsing is adopted by many systems
 - Semantics (logical representation) is used by few systems
- **Answer patterns:**
 - superficial patterns (regular expressions)
 - deep (dependency trees): pre-processing the document collection, matching dependency trees, off-line answer patten retrieval.

QA-CLEF-05: Approaches (2)

- Few system use some form of “**semantic**” indexing based on syntactic information or named entities
 - Few systems consult the **Web** at run-time
 - to find answers in specialized portals
 - to validate a candidate answer
- **Architecture** of QA modules: e.g. XML based
- **Cross-languages**
 - commercial translators, word by word translation
 - keyword translation

QA-CLEF-05: Conclusions

- Increasing interest in multilingual QA
 - More participants (+ 33%)
 - Two new languages as target (Bulgarian and Finnish) and one as source (Indonesian)
 - More activated tasks (44, they were 29 in 2004, + 51%)
 - Monolingual is the prevalent interest (61%)
 - Comparability among tasks is 20% (as in 2004)
 - Interesting results: 6 systems above 40% accuracy
- Future. Under discussion:
 - Add new languages (Romanian is a candidate)
 - Introducing pilots

Participants

Acronym	Name	Country
Alicante	U. Alicante - Comp.Sci +	Spain
INAOE	Ist.Nac.Astrofisica, Optica, Electronica	Mexico
LINGUATECA-N	Linguateca-Sintef	Norway
TALP	U.Politecnica Catalunya	Spain
MIRACLE	Daedalus & Madrid Univs	Spain
Jaen	U.Jaen - Intell.Systems	Spain
DEPOK	U.Indonesia - Comp.Sci	Indonesia
DIRO	U.Montreal	Canada
Hildesheim	U.Hildesheim - Inf.Sci	Germany
DFKI	Center Art. Intelligence	Germany
Helsinki-2	U.Helsinki - Comp.Sci	Finland
U.Glasgow	U.Glasgow - IR	UK
XLDB	U.Lisbon	Portugal
Hagen	U.Hagen - Comp.Sci	Germany
ILPS	U.Amsterdam - Informatics	Netherlands
ITC-irst	IRST-Trento	Italy
NICTA	U.Melbourne - Comp.Sci	Australia
Metacarta	Metacarta Inc.	USA
UNED	UNED-LSI	Spain
Priberam	Priberam Informatica	Portugal
SYNAPSE	SYNAPSE Development	France

Acronym	Name	Country
IRIT	Inst.Informatique,Toulouse	France
U.Evora	U.Evora - Informatics	Portugal
BUAP	U.Autonomous Puebla	Mexico
U.Sheffield	U.Sheffield - Inf.Studies	UK
U.Valencia	U.Valencia - AI	Spain
LINA	Lab Informatique-NLP	France
U.Groningen	U.Groningen - Letters	Netherlands
SICS	Swedish Inst.Comp.Sci	Sweden
Budapest U.	Budapest U. Tech.&Econom	Hungary
LIC2M	CEA	France
CLAWS	MS group-U.Amsterdam	Netherlands
Hummingbird	Hummingbird Core Tech.	Canada
CINDI	U.Concordia - Comp.Sci	Canada
REINA	U.Salamanca - Translation	Spain
Wolverhampton	U.Wolverhampton - CL	UK
BulTreeBank	Bulgarian Acad.Sciences	Bulgaria
LIR	LIMSI-LIR	France
U.Limerick	U.Limerick	Ireland
IMS Stuttgart	U. Stuttgart	Germany

Evaluation Exercise – CL Approaches

INPUT (source language)

question translation
into target language

U. Amsterdam
U. Edinburgh
U. Neuchatel

Question
Analysis /
keyword
extraction

translation of
retrieved data

Bulg. Ac. of Sciences
ITC-Irst
U. Limerick
U. Helsinki
DFKI
LIMSI-CNRS

OUTPUT (target language)

Candidate
Document
Selection

Candidate
Document
Analysis

Answer
Extraction

Document
Collection

Document
Collection
Preprocessing

Preprocessed
Documents