

# iCLEF at Sheffield

Zoë Bathie  
Mark Sanderson

presented by Daniela Petrelli

University of Sheffield  
United Kingdom

## The Question

query formulation  
and translation

English



French



Is automatic translation  
sufficient to allow  
the user spotting the  
relevance?

presentation  
and validation

## The Experiment

- 8 English speakers
- No French
- 2 broad + 2 narrow
- 2 systems



IRS



LA Times  
English



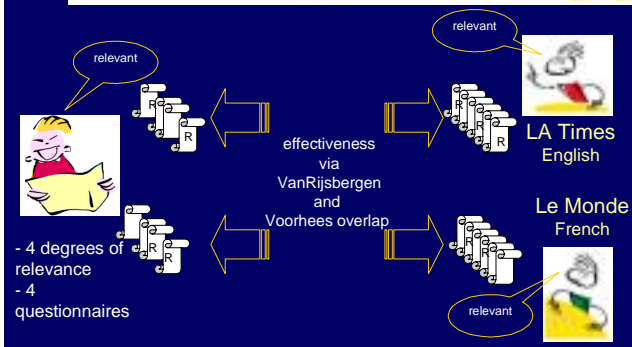
IRS



Le Monde  
French

French to English  
automatic translation

## The Experiment



## Warnings

- Different text collections means (slightly) different content
- Cultural differences and background may affect assessors and users more than in IR
- Different IR systems used for retrieving imply possibly different documents proposed

## Results: documents rated relevant

- better judgement on translated documents  
BUT  
the difference is not significant

	F a=0.2 Recall preferred	F a=0.8 Precision preferred
Le Monde	0.49	0.60
LATimes	0.40	0.46

Automatic translation quality is sufficient  
for users to make relevant judgments

## Results: documents rated relevant + somewhat relevant

- partial relevance increases accuracy on user's own language

	F a=0.2 Recall preferred	F a=0.8 Precision preferred
Le Monde	0.65	0.59
LATimes	0.58	0.52

Does automatic translation smooth language expressiveness?

## Results: user vs. assessors

- Users feel lack of support when dealing with translated documents

Degree of overlap between user's and assessors' judgement [0.39-0.47]

Voorhees's overlap [0.42-0.49]

Overlap is the intersection of the two sets divided by the union

## Looking forward: an HCI perspective (interface design issues)

query formulation  
and translation

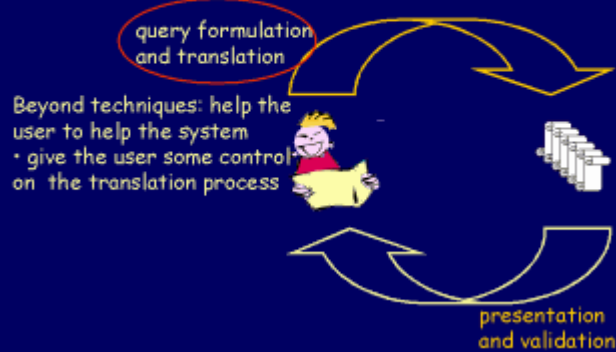
The human factor

- linguistic skills
- language power
- bad memory
- good sense maker

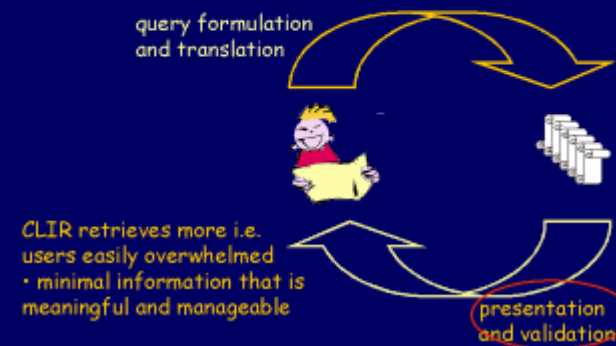


presentation  
and validation

## Looking forward: an HCI perspective (interface design issues)



## Looking forward: an HCI perspective (interface design issues)



## CLARITY CLIR and organization of text and audio documents

- IST European founded project aiming at developing a CLIR system in English, Spanish and Finnish
- Partners: Uni Sheffield, Uni Tampere, SICS, Uni Mass (USA), AlmaMedia, BBCmonitoring
- Site: <http://clarity.shef.ac.uk/>