

# The CLEF 2001 Interactive Track

Track Coordinators:

Douglas W. Oard and Julio Gonzalo

# Participants

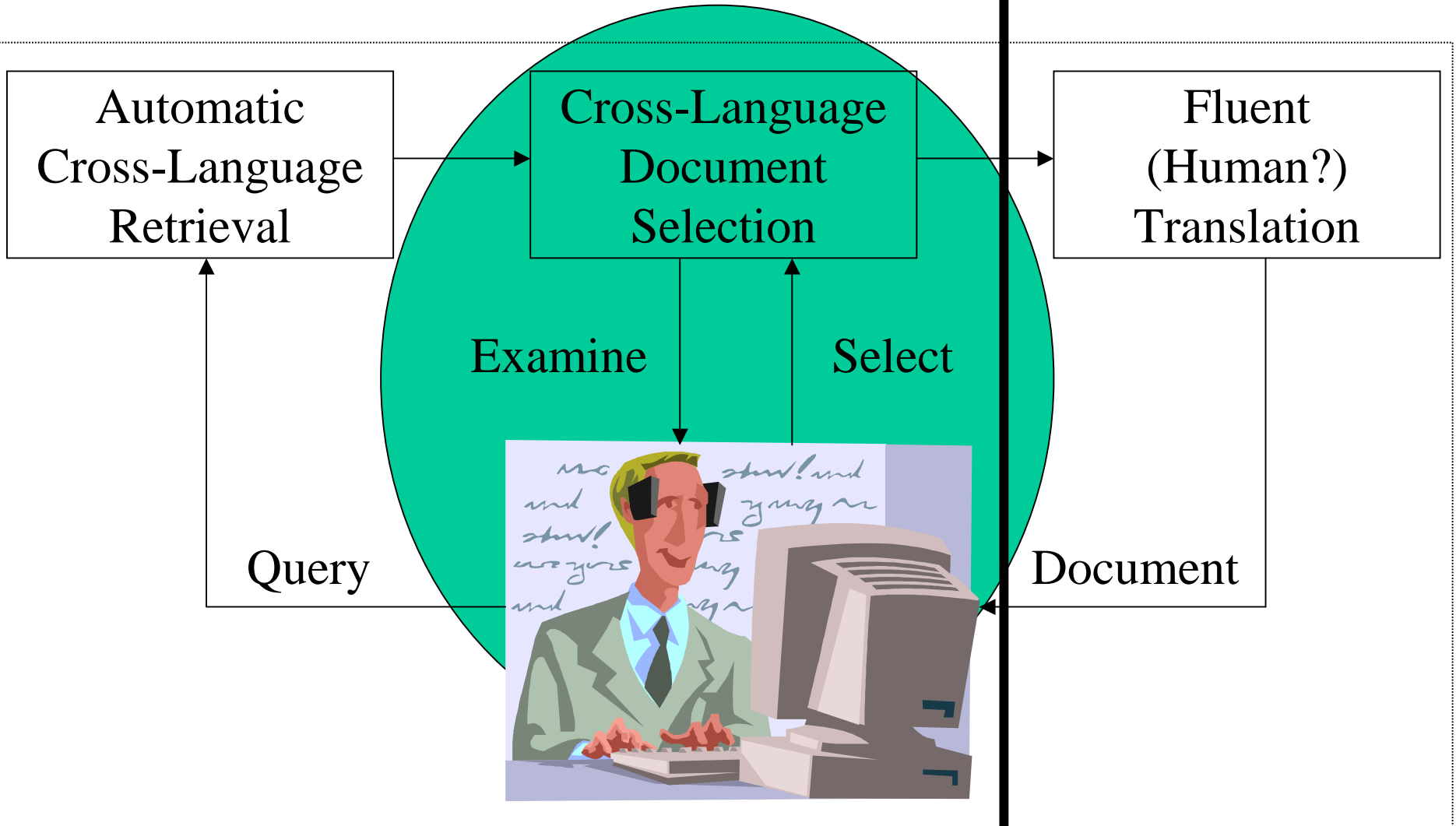
- Spain
  - Universidad Nacional de Educación a Distancia (UNED)
- United Kingdom
  - University of Sheffield
- United States of America
  - University of Maryland

# Outline

- Task
- Experiment design
- Results overview
- Next steps

# Cross-Language Information Access

# Information Use



# Cross-Language Browsing Task

- Goal: find documents in unfamiliar language
  - Automatic CLIR system nominates ranked candidates
  - Searcher selects relevant documents
- Design challenges
  - Translation strategy
  - User interface design

# iCLEF French Collection

- 5 topic descriptions in any CLEF 2000 language
  - 2 narrow, 2 broad, 1 practice
- 5 ranked lists of documents
  - 50 French documents per topic
  - English machine translations (Systran Professional 3.0)
- Complete relevance judgments
  - Reused from CLEF 2000 (unjudged documents removed)

# iCLEF English Collection

- Same 5 topics
- 5 ranked lists of documents
  - 50 English documents per topic
  - Spanish machine translations (Systran Professional 3.0)
- Complete relevance judgments
  - Reused from CLEF 2000 (unjudged documents removed)

# iCLEF Topics

	<u>Relevant Fraction</u>	
	English	French
• Broad		
– 11: New constitution for South Africa	36/50	27/50
– 13: Conference on birth control	16/50	11/50
• Narrow		
– 17: Bush fire near Sydney	6/50	2/50
– 29: Nobel Prize for Economics	2/50	3/50



# Experiment Design

Participant

Task Order


1	<b>Topic11, Topic17</b>	<b>Topic13, Topic29</b>
2	<b>Topic11, Topic17</b>	<b>Topic13, Topic29</b>
3	<b>Topic17, Topic11</b>	<b>Topic29, Topic13</b>
4	<b>Topic17, Topic11</b>	<b>Topic29, Topic13</b>

Topic Key

Narrow: **11, 13**

Broad: **17, 29**

System Key

System A: 

System B: 

# An Experiment Session

- Task and system familiarization (30 minutes)
  - Gain experience with both systems
- 4 searches (20 minutes x 4):
  - Read topic description (in a language you know)
  - Examine translations (into that same language)
  - Select one of 5 relevance judgments (two clusters)
    - Relevant
    - Somewhat relevant, Not relevant, Unsure, Not judged
  - Instructed to seek high precision
- 8 questionnaires
  - Initial (1), each topic (4), each system (2), Final (1)

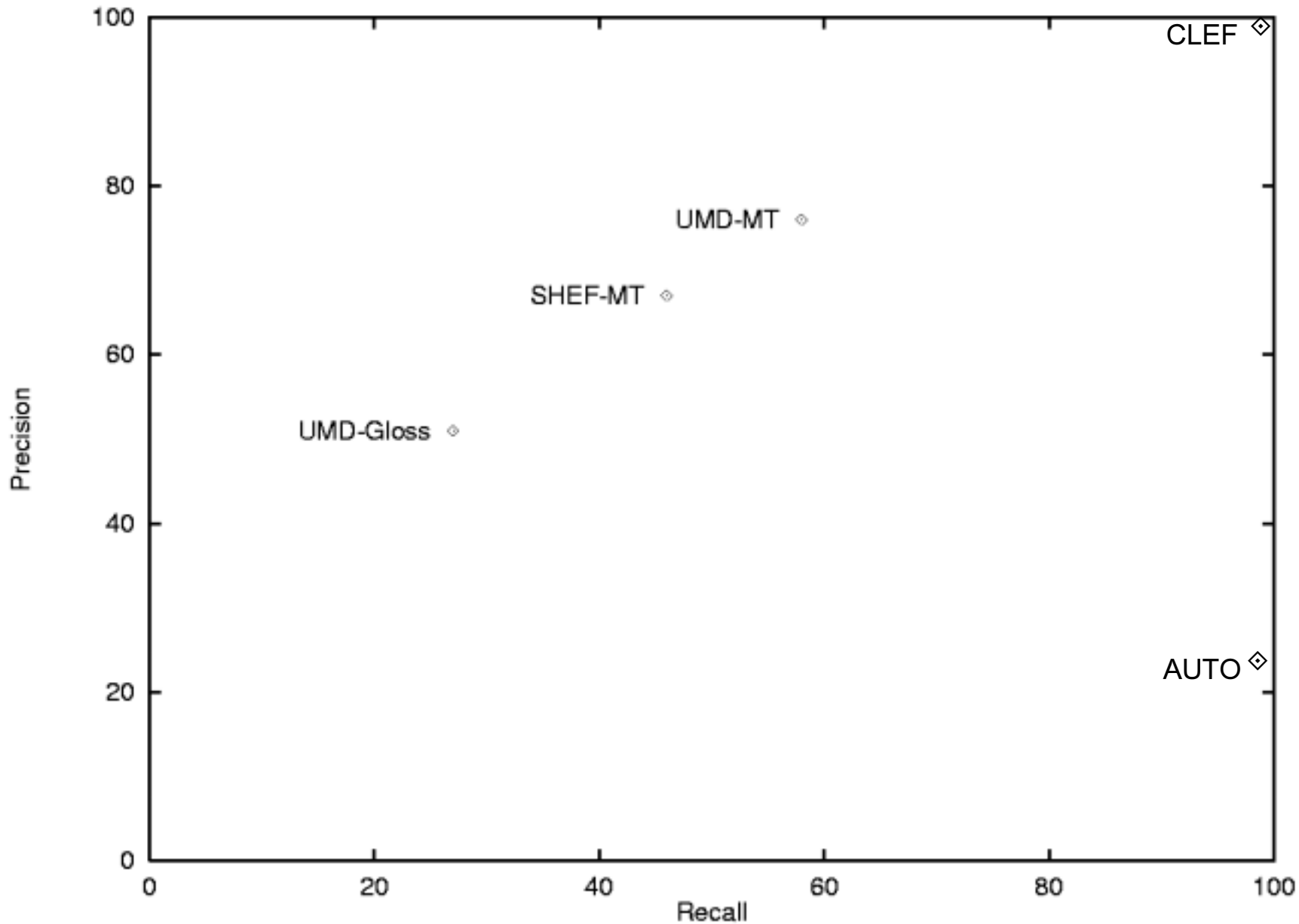
# Measure of Effectiveness

- Unbalanced F-Measure:
  - $P$  = precision
  - $R$  = recall
  - $\alpha = 0.8$ 
    - Favors precision over recall

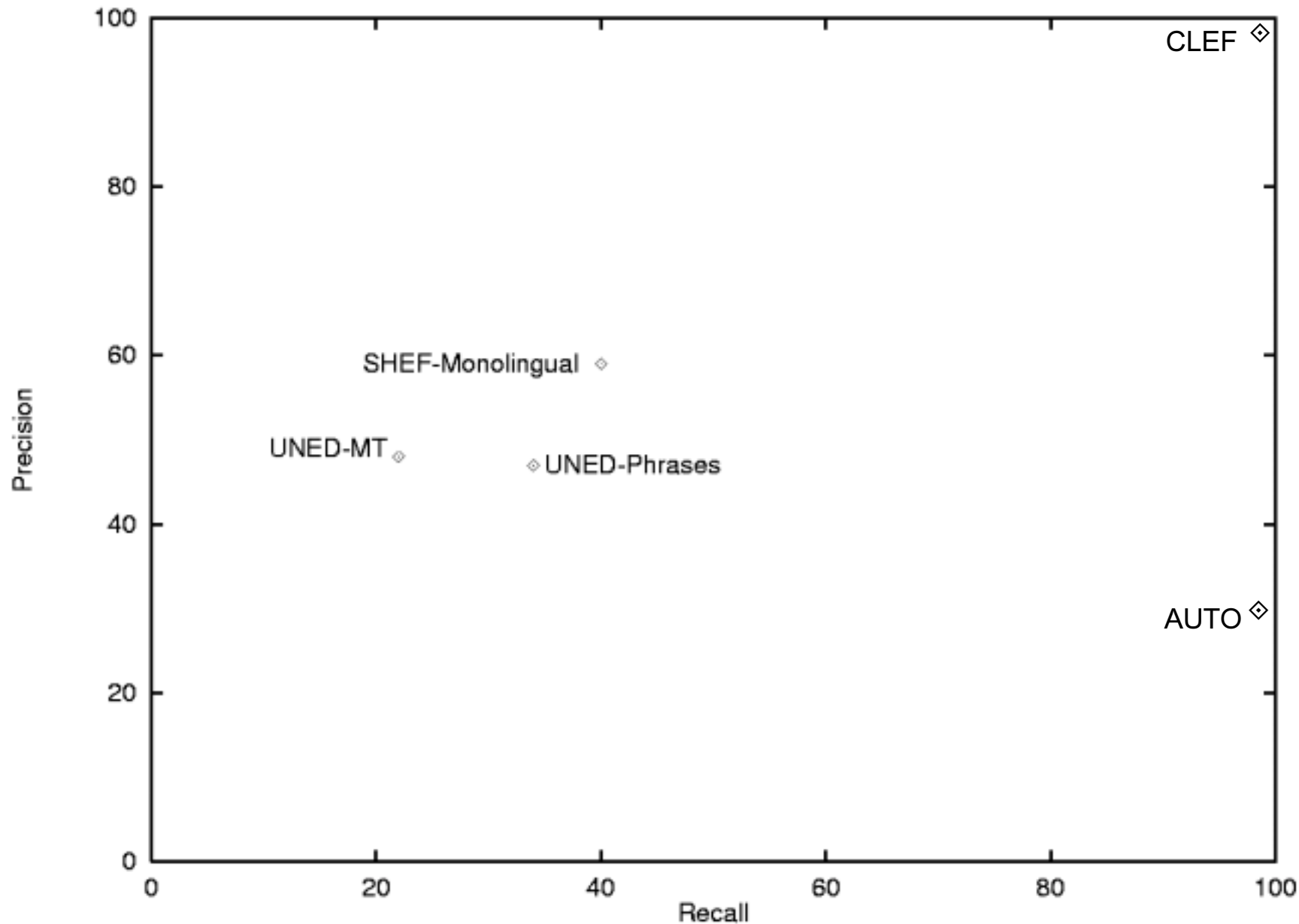
$$F_{\alpha} = \frac{1}{\frac{\alpha}{P} + \frac{1-\alpha}{R}}$$

- This models an application in which:
  - Fluent translation is expensive
  - Missing some relevant documents would be okay

# French Results Overview



# English Results Overview



# Some Observations

- Small agreement with CLEF assessments!
  - Time pressure, precision bias, strict judgments
- Systran was fairly consistent across sites
  - Only when the language pair was the same
- Monolingual > Systran > Gloss
  - In both recall and precision
- UNED's phrase translations improve recall
  - With no adverse affect on precision

# Looking to the Future

- There is a community of interest
  - Coming from both IR and machine translation
- The experiment design can be improved
  - Separately record assessment and confidence
  - Measures tuned for narrow topics
  - Agreed framework for statistical significance tests
- Some interesting new directions
  - Other topic types (shopping, tourism, ...)
  - Compare monolingual/CL using bilingual searchers