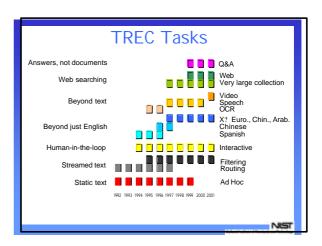


Donna Harman, Ellen Voorhees Retrieval Group Information Access Division National Institute of Standards and Technology

NB



### **TREC-9** Tracks

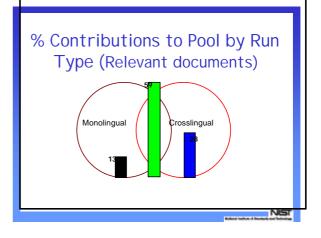
- · Cross-language (English to Chinese)
- Filtering
- Interactive
- Query
- Question Answering
- Spoken Document Retrieval
- Web

# Cross Language Track

- Task: ad hoc search for documents written in one language using topics in another language
  - 25 topics in English created by bilingual assessors; Chinese version also available
  - 126,937 documents; 188 MB in BI G5
  - Hong Kong newspapers donated by Wiser Ltd.
    - Hong Kong Commercial Data (Aug 98-Jul 99)
    - Hong Kong Daily News (Feb 99-July 99)
    - Takongnao (Oct 98-Mar 99)

#### **Relevance Judgments**

- Judged highest priority mono- and crosslingual run from each group
  - 39 cross (75%) / 13 mono (25%)
  - 51 auto / 1 manual (Thank you, Berkeley!)
- Added top 50 documents from each judged run to the pool
- Mean actual pool size = 598 (39% of max) within expected range



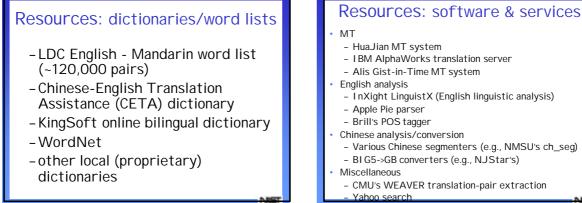
#### Participants

**BBN** Technologies Fudan University IBM T.J. Watson Research Center Johns Hopkins University Korea Advanced Institute of Science and Technology Microsoft Research, China MNI S-TextWise Labs National Taiwan University

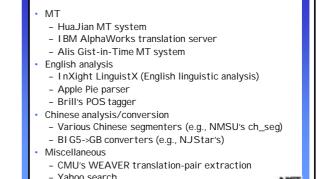
#### More participants

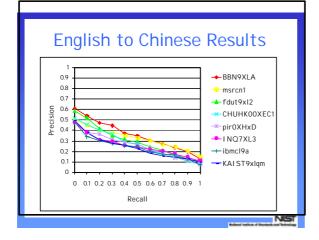
Queens College, CUNY **RMIT** University Telcordia Technologies, Inc. The Chinese University of Hong Kong Trans-EZ Inc. University of California at Berkeley University of Maryland University of Massachusetts

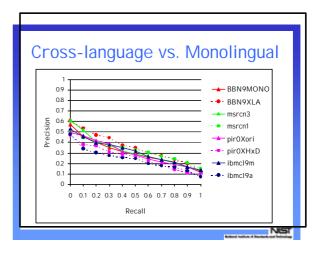
NE

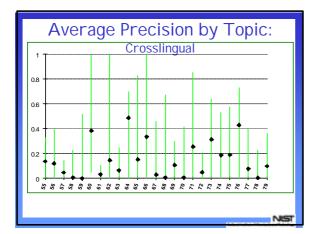


N









# What was learned from the Chinese CLIR track?

- Many approaches to English to Chinese topic translation, including use of various dictionaries, word lists, parallel text, and commercial MT systems
- Extensive set of Chinese retrieval experiments performed ranging from various n-gram methods to word based to complete language modeling
- Because of the tight focus of this track, cross-system comparison is possible

NIST

## TREC 2001

- Cross language
  - Chinese ∠ NTCIR workshop (NII, Japan)
    TREC task will be English, French<sup>†</sup> Arabic
- Filtering track using new Reuters corpus
- Interactive to investigate live web
- Expanded web and QA tracks
- New video track

