

know-how

innovation

# *CLEF 2002*

## *Overview of Results*

*Martin Braschler*

*Eurospider Information Technology AG*

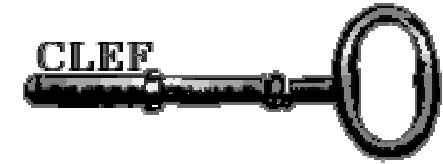
*8006 Zürich, Switzerland*

*[martin.braschler@eurospider.com](mailto:martin.braschler@eurospider.com)*

 *eurospider*  
relevance retrieval

solution

*Outline*

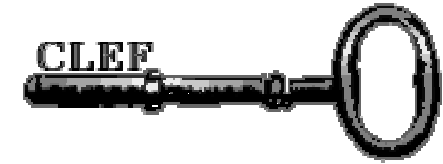


- **Participants**
- **Experiment Details**
- **Trends**
- **Effects**
- **Results**
- **Conclusions, Outlook**



# Participants

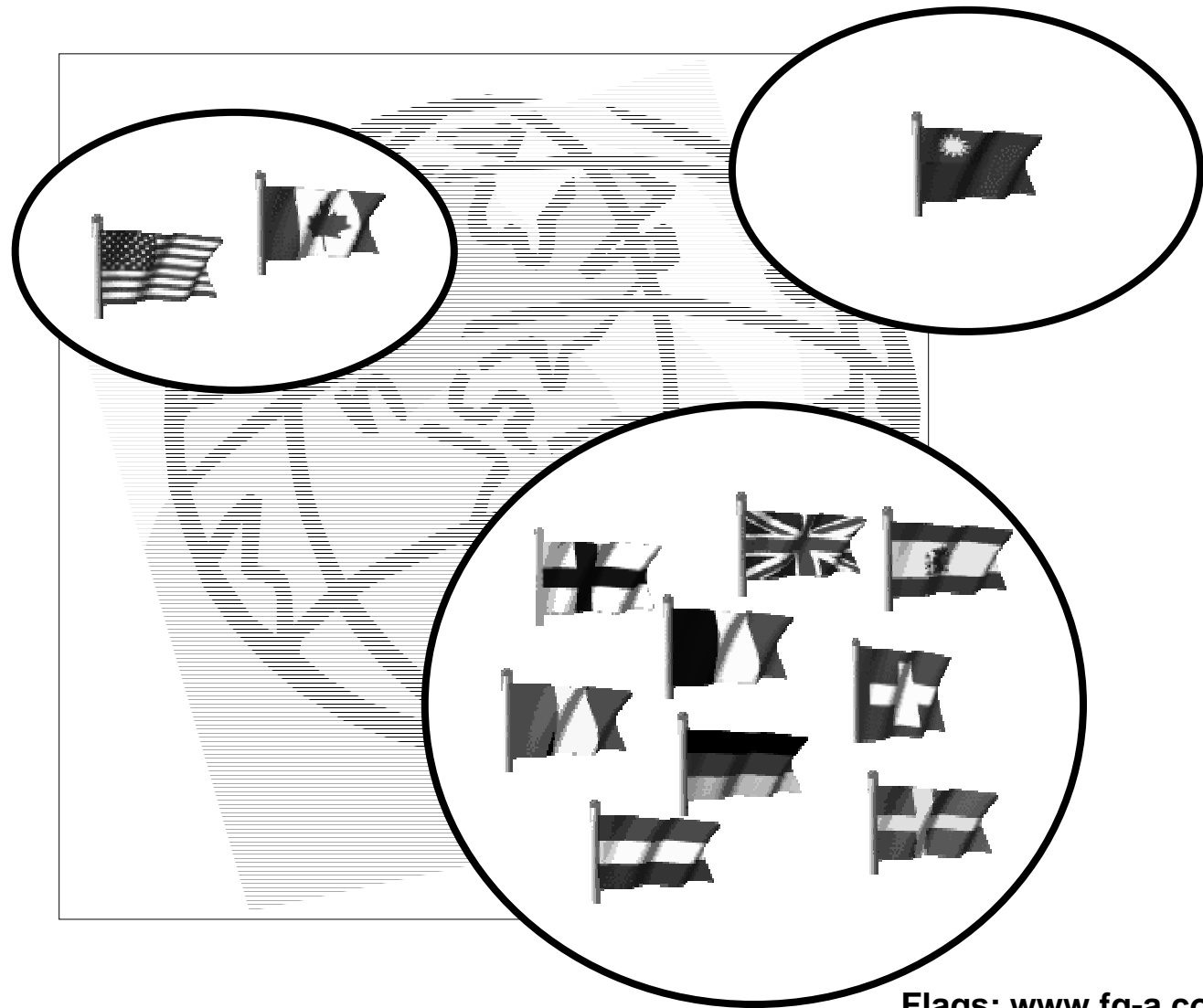
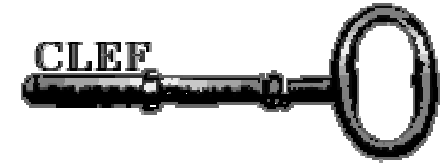
City University	SINAI/U Jaen *
Clairvoyance Corp.	Tagmatica
COLE Group/U La Coruna	Thomson Legal **
CWI/CNLP *	U Alicante *
Eurospider **	U Amsterdam *
Fondazione Ugo Bordon	U Dortmund *
Hummingbird *	U Exeter *
IMBIT	U Hildesheim
IMS U Padova	U Maryland **
IRIT **	U Montreal/RALI **
ITC-irst **	U Neuchâtel *
JHU-APL **	U Salamanca **
Lexware	U Sheffield **
Medialab *	U Tampere **
Middlesex U	U Twente/TNO **
National Taiwan U *	UC Berkeley (2 groups) **
OCE Tech. BV *	UNED *
SICS/Conexor *	Xerox *



37 participants, 12 different countries.  
(\*/\*\* = one or two previous participations)



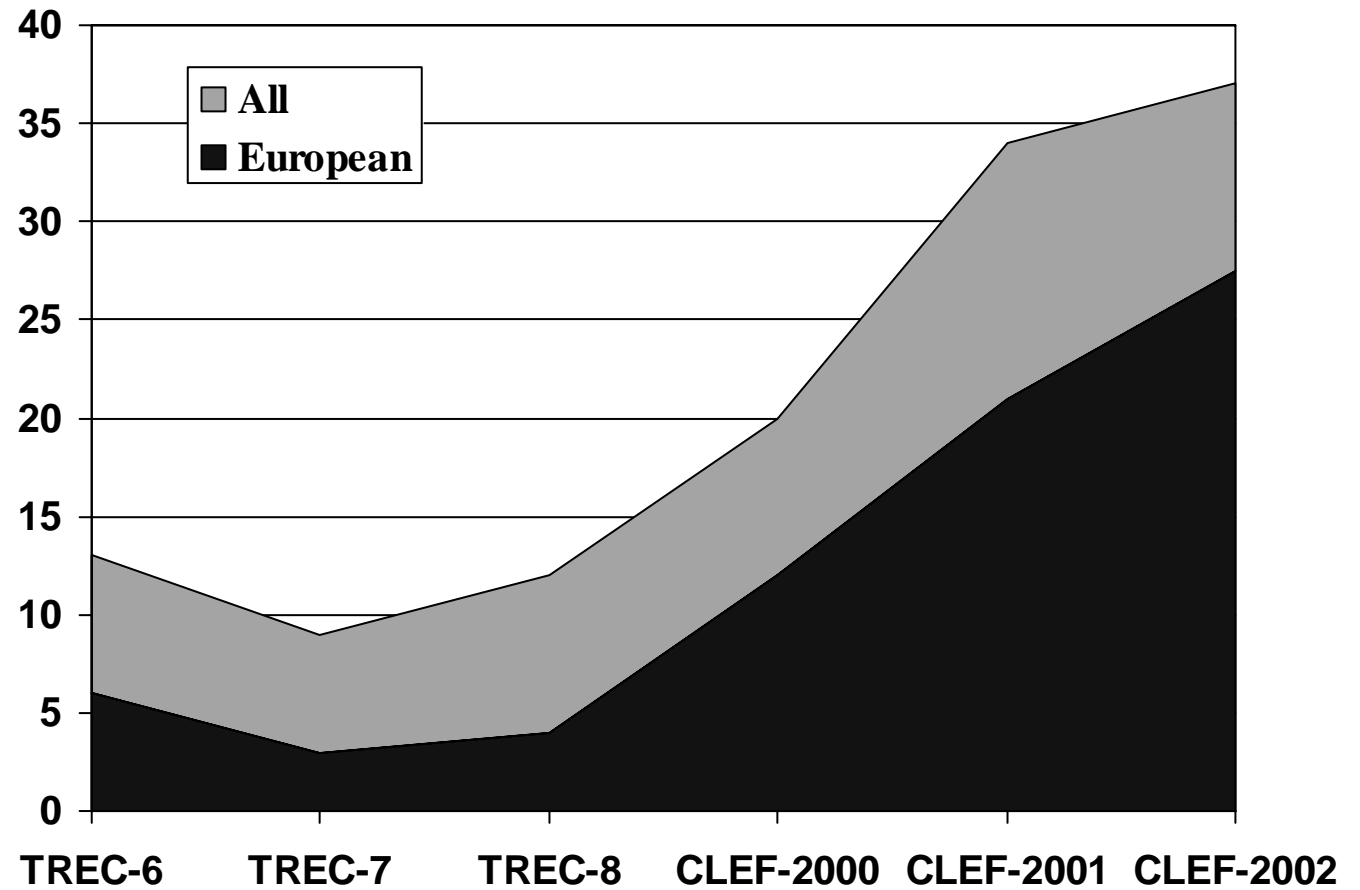
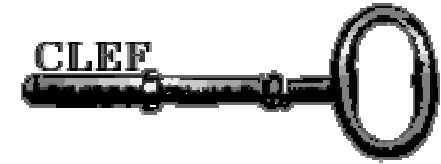
*CLEF's  
Global Reach*



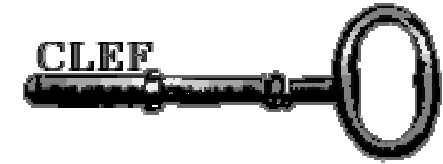
**Flags: [www.fg-a.com](http://www.fg-a.com)**

# CLEF Growth

(Number of Participants)



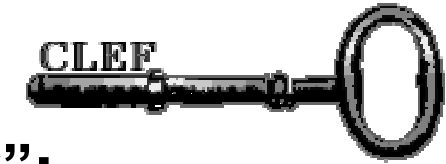
*The CLEF  
Multilingual  
Collection*



	# part.	# lg.	# docs.	Size in MB	# assess.	# topics	# ass. per topic
<b>CLEF 2002</b>	<b>34</b>	<b>8</b>	<b>1,138,650</b>	<b>3011</b>	<b>140,043</b>	<b>50 (30)</b>	<b>~2900</b>
<b>CLEF 2001</b>	<b>31</b>	<b>6</b>	<b>940,487</b>	<b>2522</b>	<b>97,398</b>	<b>50</b>	<b>1948</b>
<b>CLEF 2000</b>	<b>20</b>	<b>4</b>	<b>368,763</b>	<b>1158</b>	<b>43,566</b>	<b>40</b>	<b>1089</b>
<b>TREC8 CLIR</b>	<b>12</b>	<b>4</b>	<b>698,773</b>	<b>1620</b>	<b>23,156</b>	<b>28</b>	<b>827</b>
<b>TREC8 AdHoc</b>	<b>41</b>	<b>1</b>	<b>528,155</b>	<b>1904</b>	<b>86,830</b>	<b>50</b>	<b>1736</b>
<b>TREC7 AdHoc</b>	<b>42+4</b>	<b>1</b>	<b>528,155</b>	<b>1904</b>	<b>~80,000</b>	<b>50</b>	<b>~1600</b>



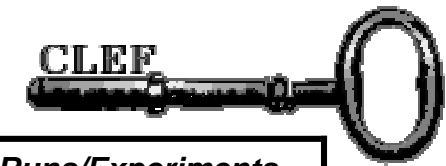
*Tasks in CLEF  
2002*



- **Multilingual as “main task”:**  
documents in 5 languages, topics in 11 languages
- **Bilingual tasks: any combination of 8 target languages, 11 topic languages**
  - English as target language: only newcomers or special cases
- **Monolingual tasks: 7 target languages**
- **Domain-specific: GIRT (German docs.) and Amaryllis (French docs.), bi- and monolingual, extra resources available**
- **Interactive track: special overview talk**



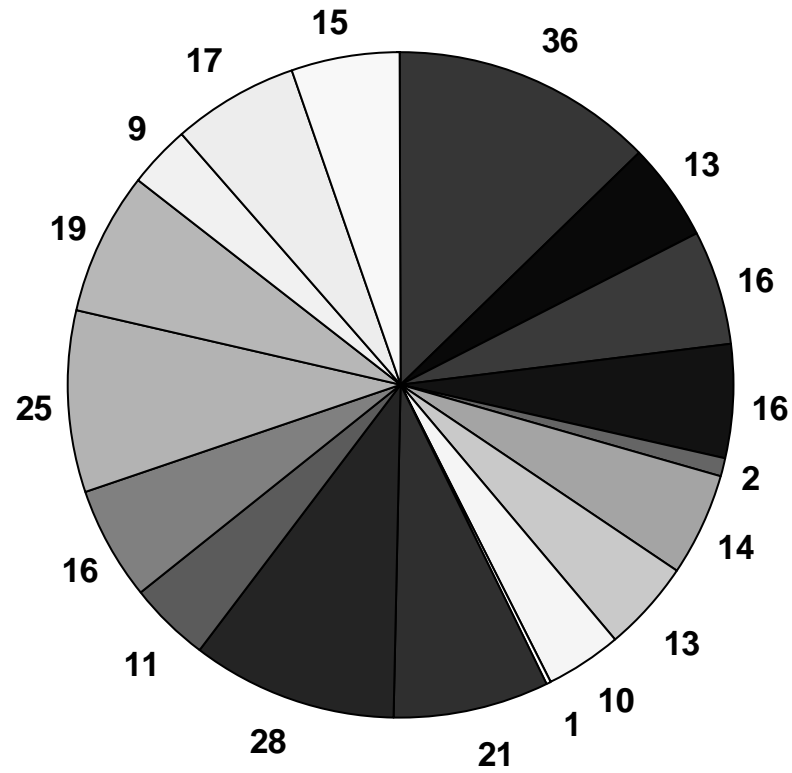
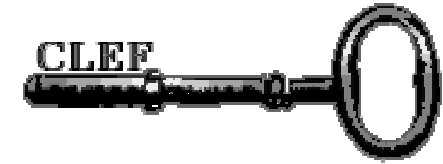
## Details of Experiments



<b>Track</b>	<b># Participants</b>	<b># Runs/Experiments</b>
<b>Multilingual</b>	<b>11</b>	<b>36</b>
<b>Bilingual to DE</b>	<b>6</b>	<b>13</b>
<b>Bilingual to EN</b>	<b>5</b>	<b>16</b>
<b>Bilingual to ES</b>	<b>7</b>	<b>16</b>
<b>Bilingual to FI</b>	<b>2</b>	<b>2</b>
<b>Bilingual to FR</b>	<b>7</b>	<b>14</b>
<b>Bilingual to IT</b>	<b>6</b>	<b>13</b>
<b>Bilingual to NL</b>	<b>7</b>	<b>10</b>
<b>Bilingual to SV</b>	<b>1</b>	<b>1</b>
<b>Monolingual DE</b>	<b>12</b>	<b>21</b>
<b>Monolingual ES</b>	<b>13</b>	<b>28</b>
<b>Monolingual FI</b>	<b>7</b>	<b>11</b>
<b>Monolingual FR</b>	<b>12</b>	<b>16</b>
<b>Monolingual IT</b>	<b>14</b>	<b>25</b>
<b>Monolingual NL</b>	<b>11</b>	<b>19</b>
<b>Monolingual SV</b>	<b>6</b>	<b>9</b>
<b>Domain-specific Amaryllis</b>	<b>3</b>	<b>15</b>
<b>Domain-specific GIRT</b>	<b>5</b>	<b>17</b>
<b>Interactive</b>	see interactive talk	see interactive talk



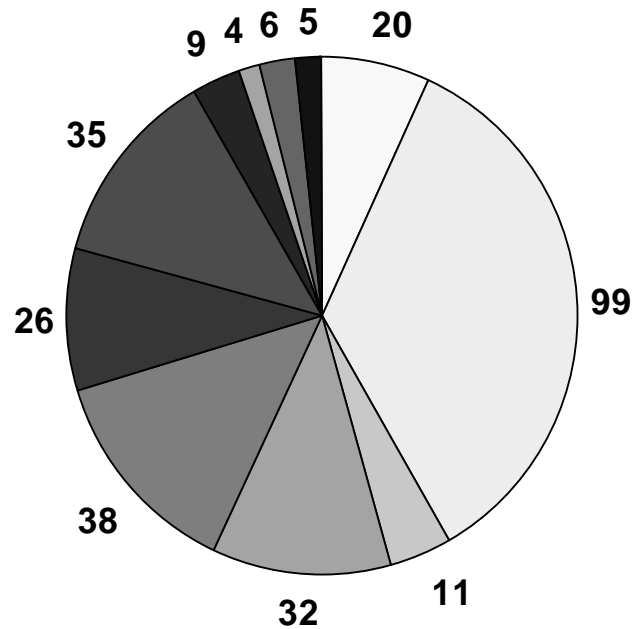
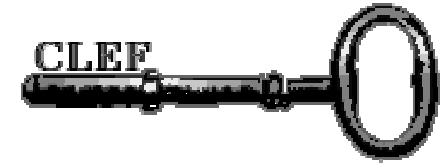
# Runs per Task



- Multi
- Bi-DE
- Bi-EN
- Bi-ES
- Bi-FI
- Bi-FR
- Bi-IT
- Bi-NL
- Bi-SV
- Mono-DE
- Mono-ES
- Mono-FI
- Mono-FR
- Mono-IT
- Mono-NL
- Mono-SV
- GIRT
- Amaryllis




# Runs per Topic Language

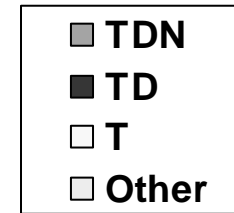
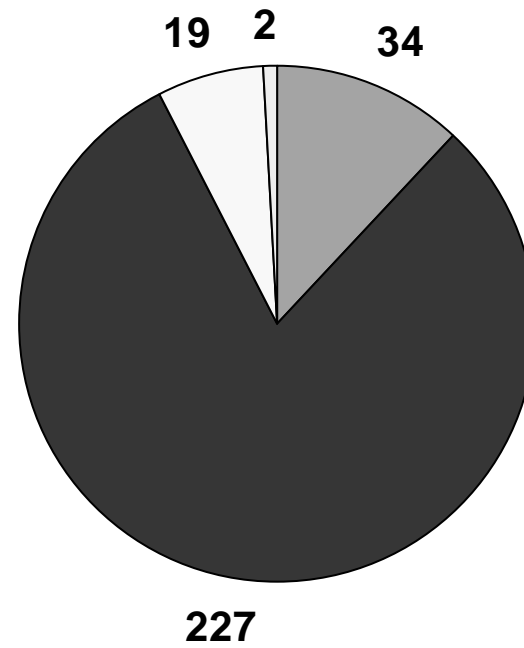
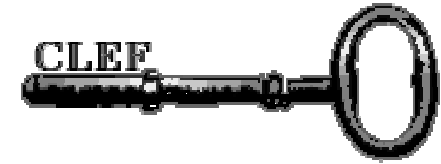


- Dutch
- English
- Finnish
- French
- German
- Italian
- Spanish
- Swedish
- Chinese
- Portuguese
- Russian



*Topic Fields*

 *eurospider*  
relevance retrieval



*Pooling*



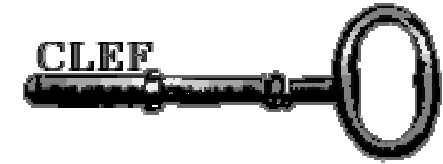
- **“Tool” to handle the size of relevance assessment work**
- **179 of 282 runs assessed**
- **Some tasks had all runs assessed: FI, SV, Amaryllis, GIRT**
- **Runs are pooled respecting nearly a dozen criteria:**
  - **participant’s preferences**
  - **“originality” (task, topic fields, languages, ..)**
  - **participant/task coverage**
  - **..**

*Preliminary  
Trends for  
CLEF-2002 (1)*



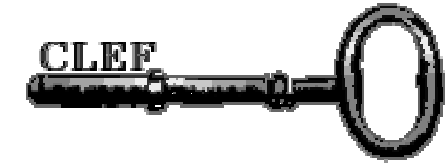
- **Less corpus-based features, still popular for combination (LSI, SimThes, stat. models)**
- **A few MT systems were very popular: Systran, LH Power Translator**
- **Lots of work on query expansion: blind feedback, concepts/synonyms, assoc. thesaurus, SimThes, etc. (Berekely, Océ, Fond. U. Bordoni, COLE, U Salamanca, ..)**
- **Weighting: tuning per language vs. same parameters for all languages**

*Preliminary  
Trends for  
CLEF-2001 (2)*



- **Diverse work on stemming: simple, elaborate (morphological analyzers), programmable, etc.**
- **Interesting work on decomposing (different results for different languages?)**
- **Merging: simple methods widely used, but some new ideas as well**
  - **unified index, reindex, prediction based on transl. quality, feedback-based**

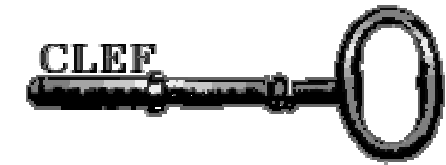
*CLEF-2002 vs.  
CLEF-2001*



- **Many participants were back**
- **More MT again, less corpus-based (approaches go into and out of favor)**
- **People try each other's ideas/methods:**
  - n-grams
  - combination approaches
  - similarity thesaurus
  - stemmers
- **Returning participants usually improve performance. (“Advantage for veteran groups”)**

*“Effect” of  
CLEF in 2001*

*(recycled slide)*

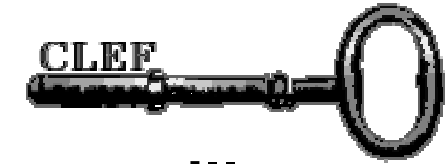


- **Many more European groups (21!)**
- **Dramatic increase of work in stemming/decompounding (for languages other than English)**
- **Work on mining the web for parallel texts**
- **Work on merging (breakthrough still missing?)**
- **Work on combination approaches**



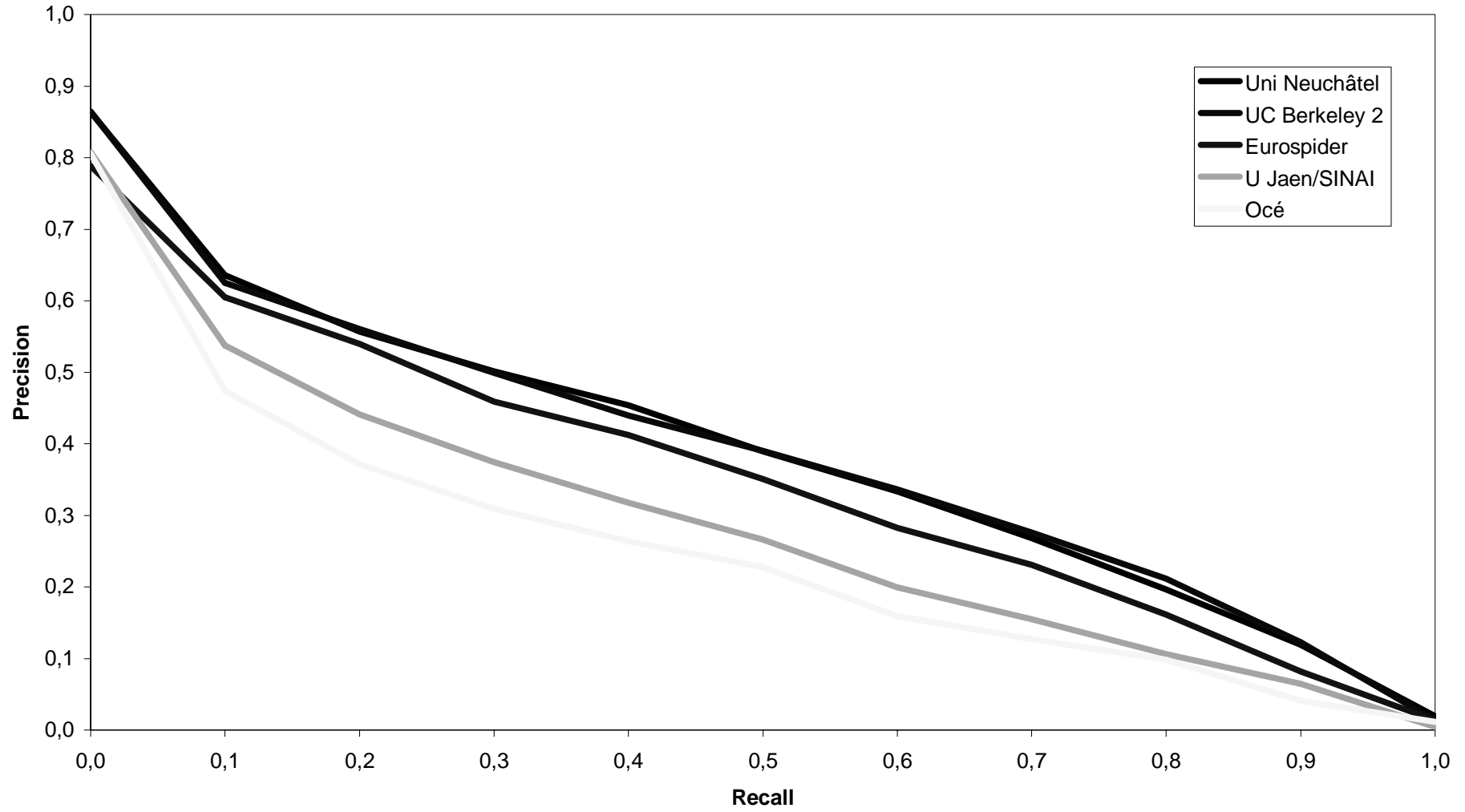


*“Effect” of  
CLEF in 2002*

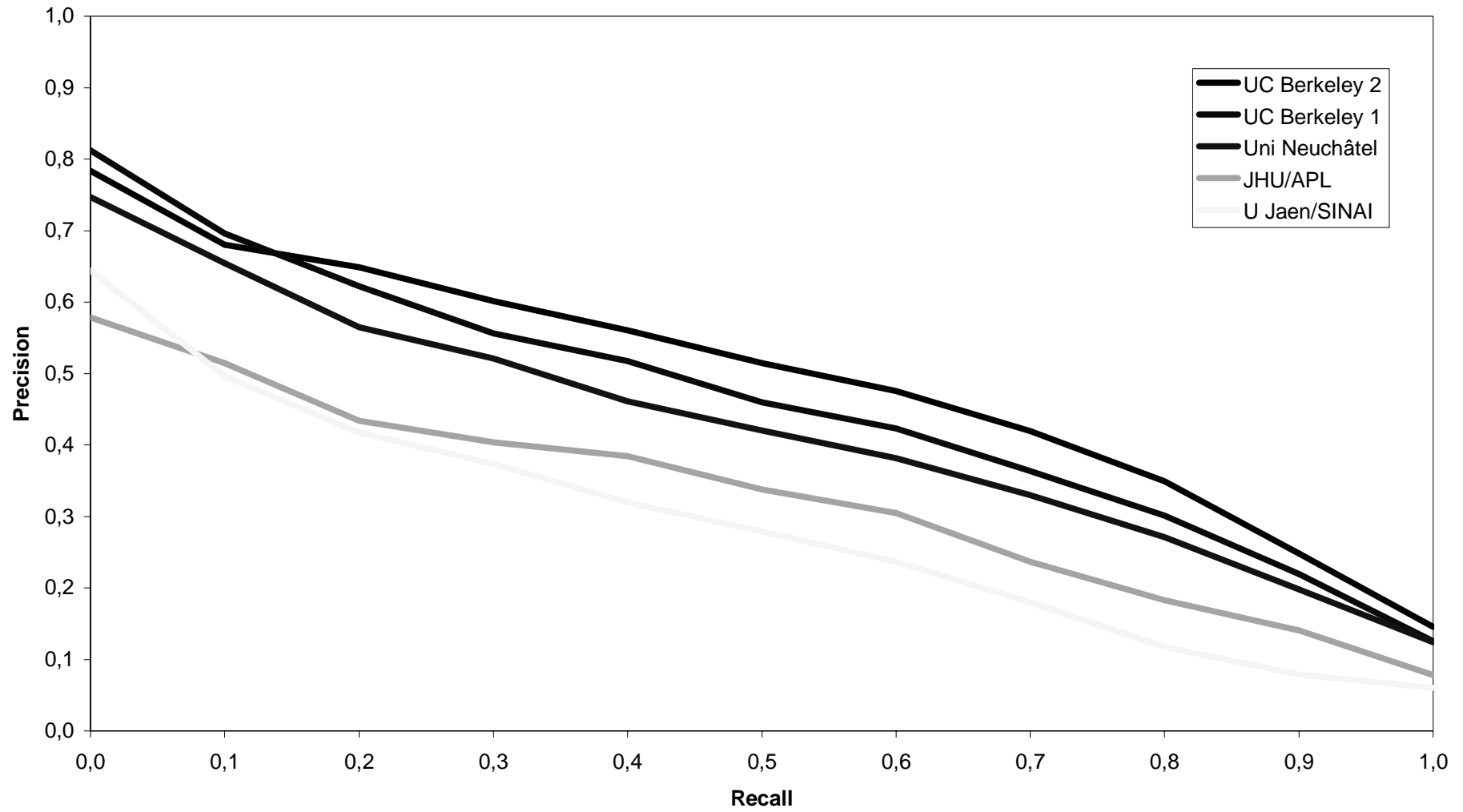


- **Number of European groups still growing (27,5!)**
- **Very sophisticated fine-tuning for individual languages**
  - **BUT: are we overtuning to characteristics of the CLEF collection?**
- **People show flexibility in adapting resources/ideas as they come along (architectures?)**
- **Participants move from monolingual → bilingual → multilingual**

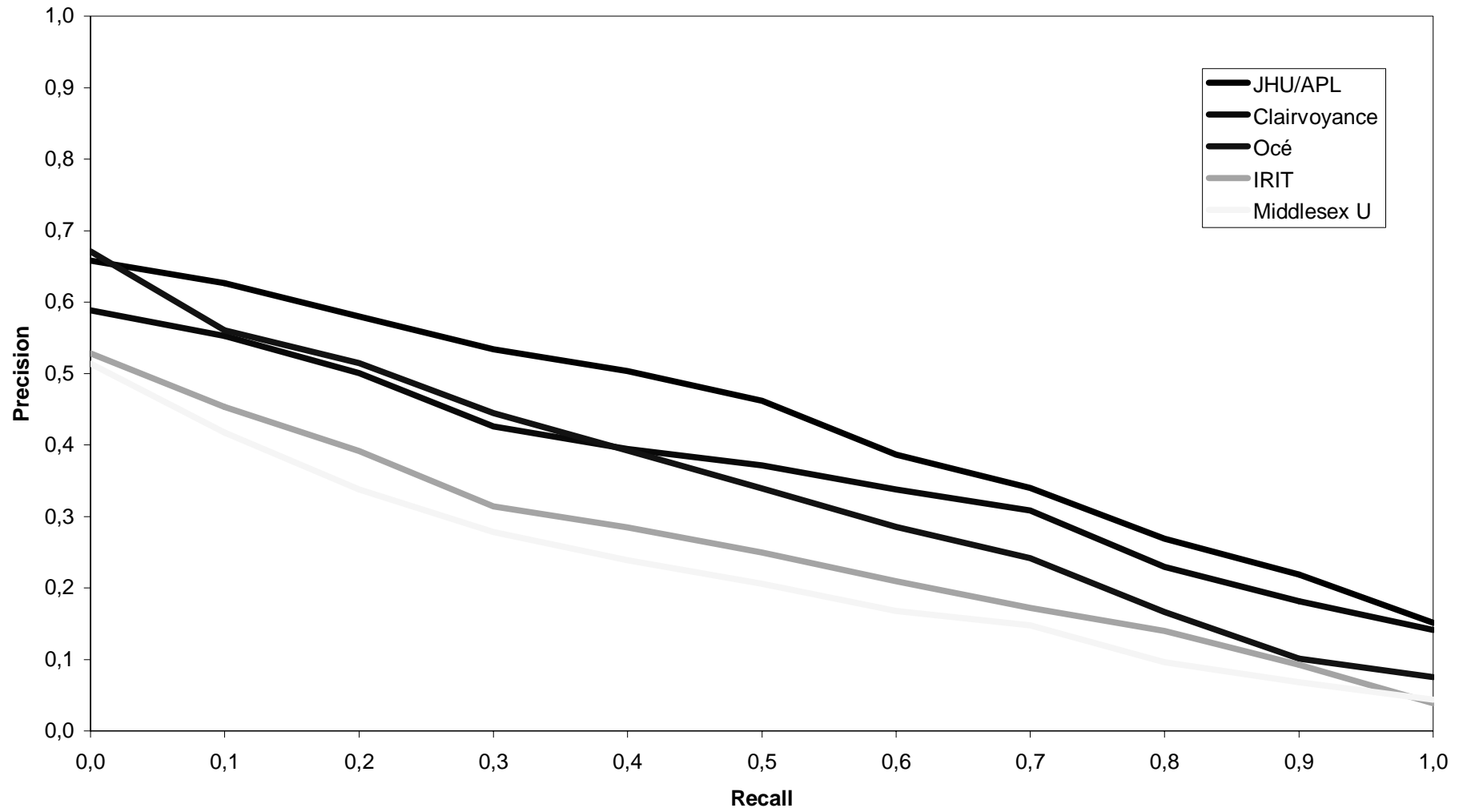
### CLEF 2002 Multilingual Track - Automatic



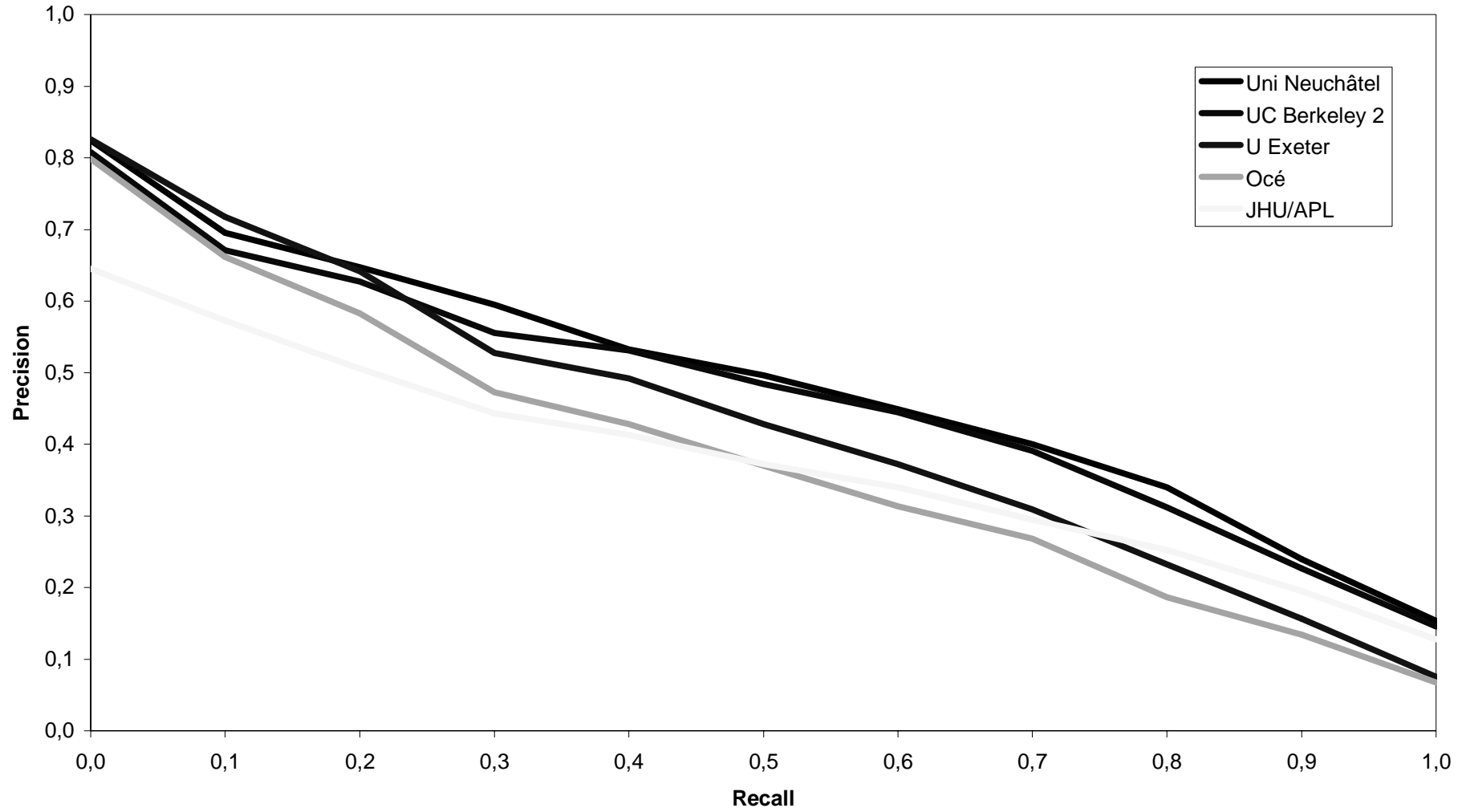
CLEF 2002 Bilingual to German - Automatic



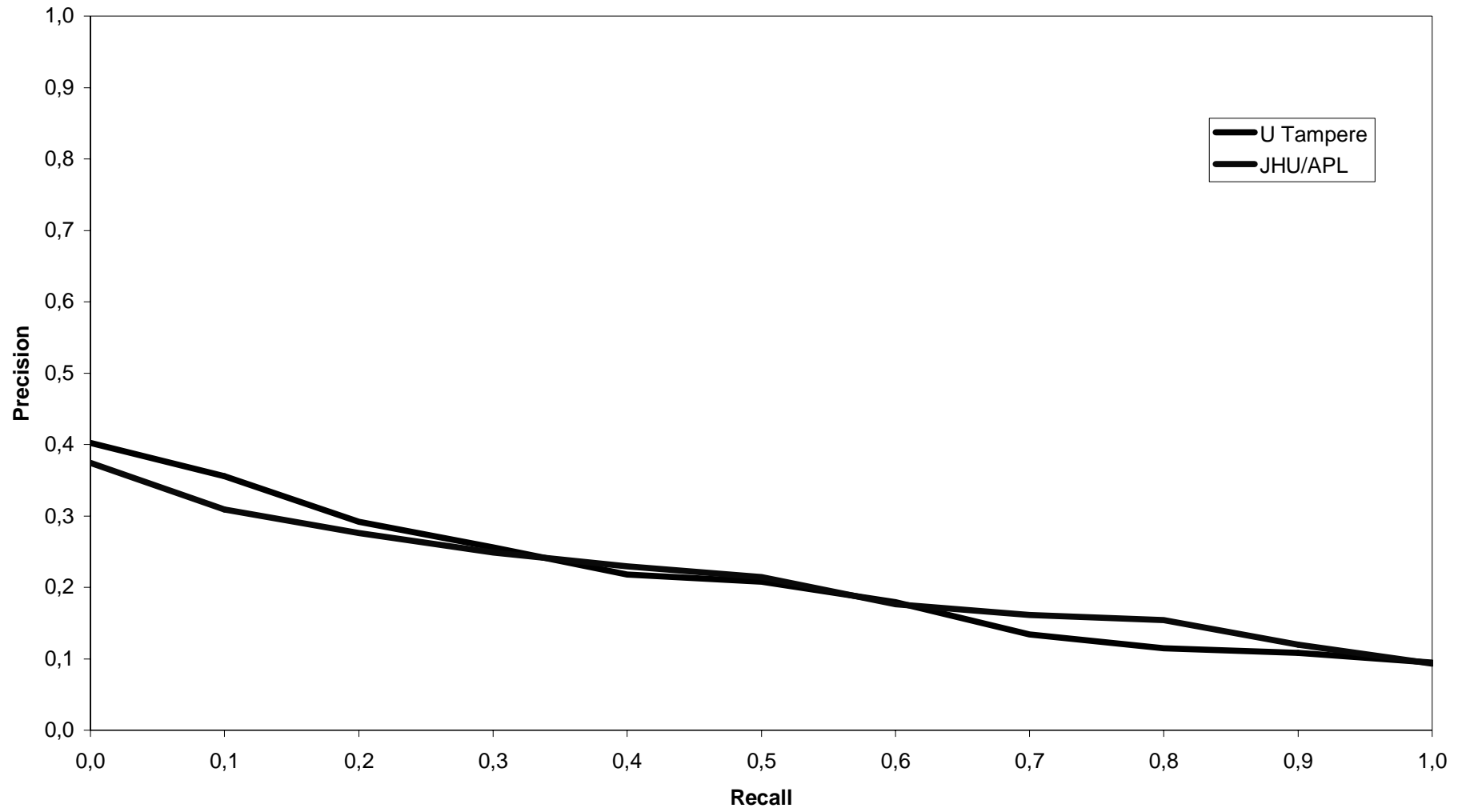
CLEF 2002 Bilingual to English - Automatic



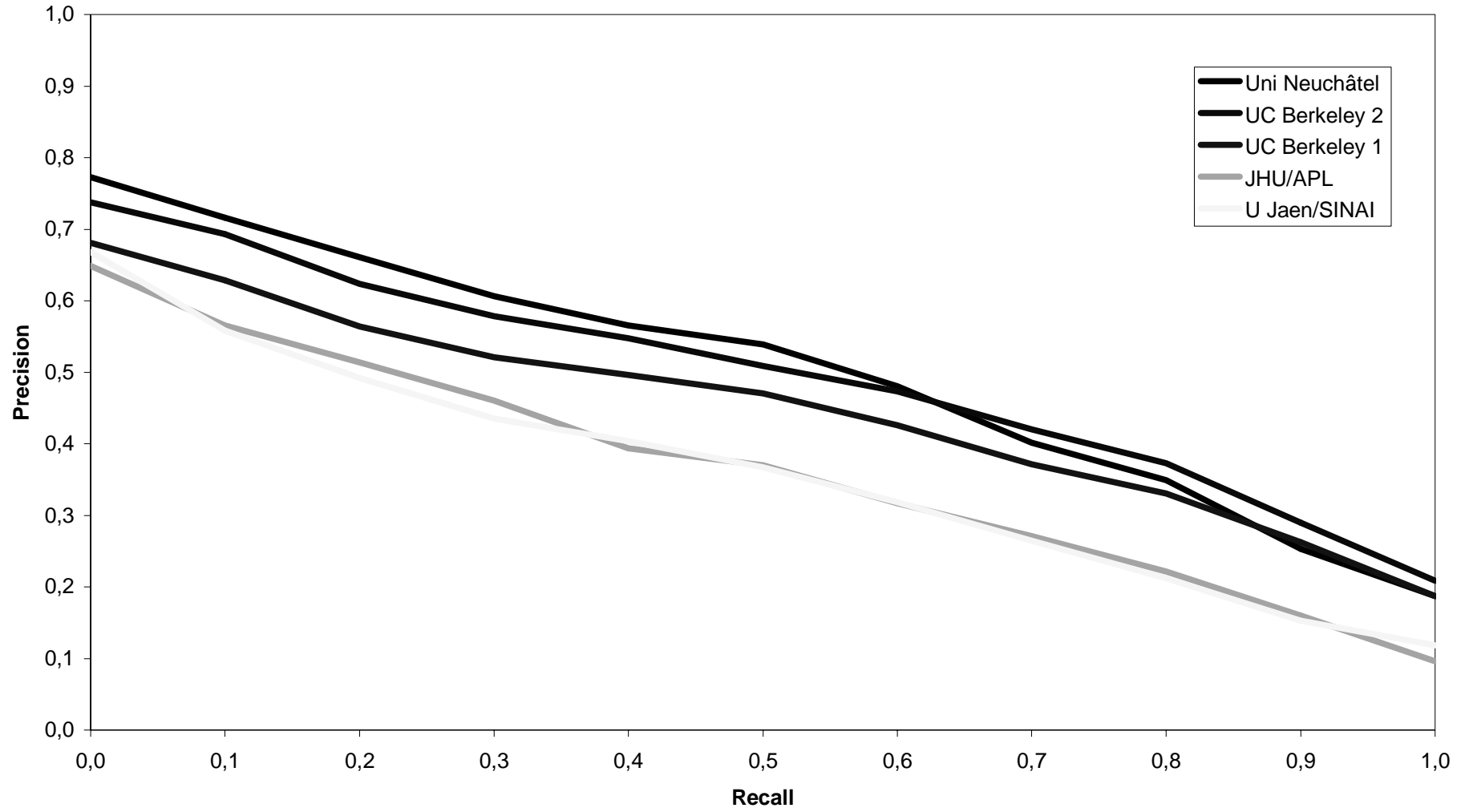
CLEF 2002 Bilingual to Spanish - Automatic



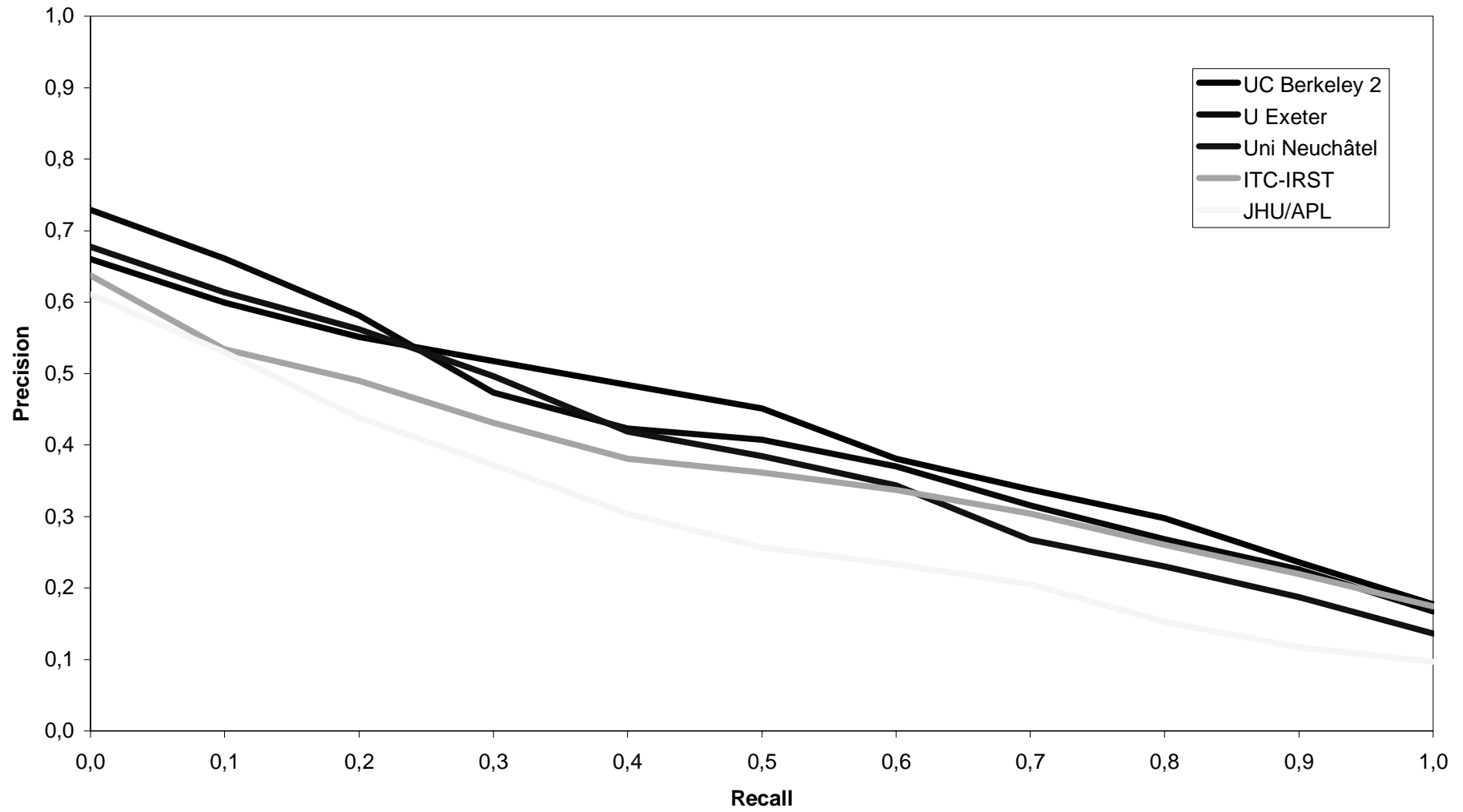
CLEF 2002 Bilingual to Finnish - Automatic



CLEF 2002 Bilingual to French - Automatic

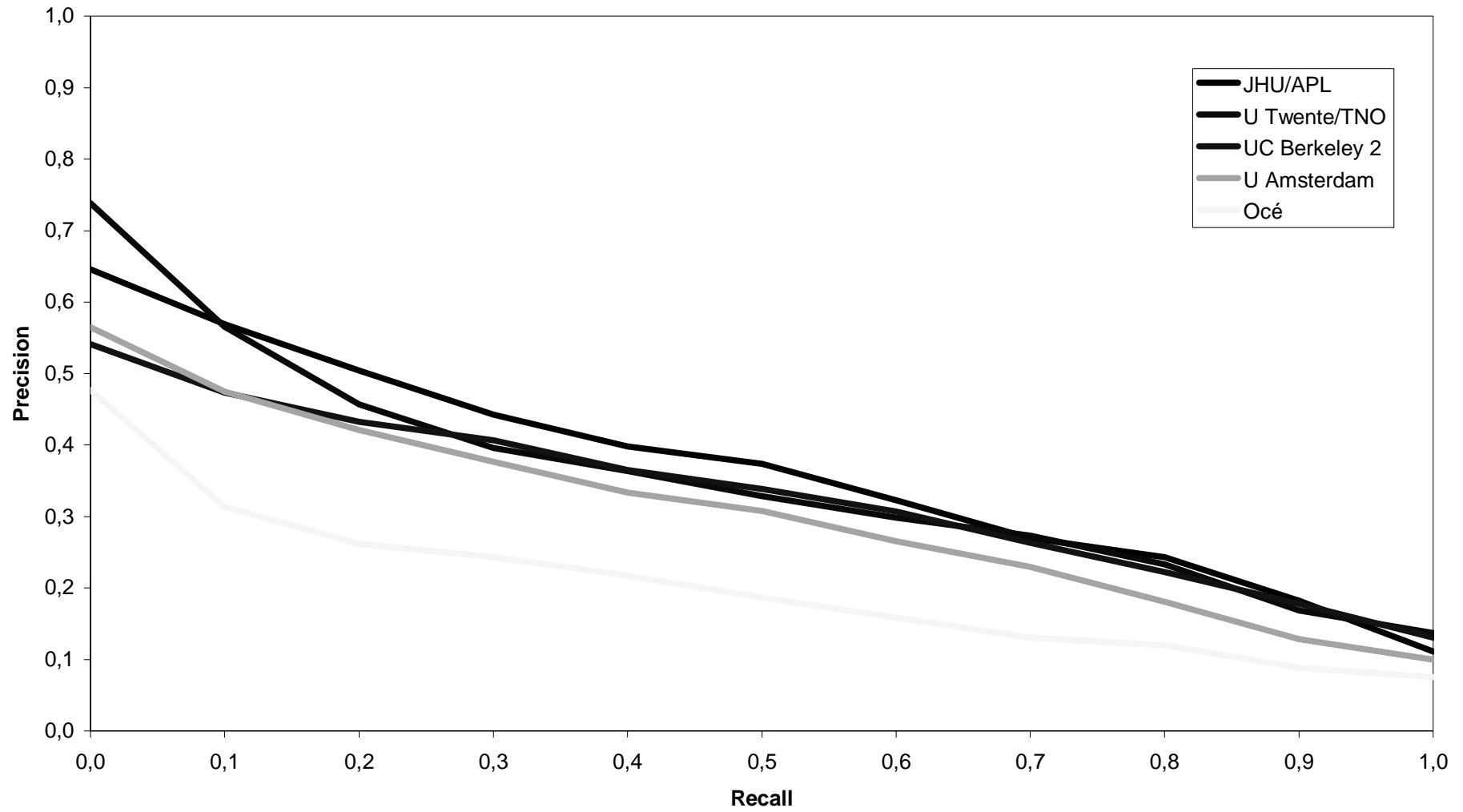


CLEF 2002 Bilingual to Italian - Automatic

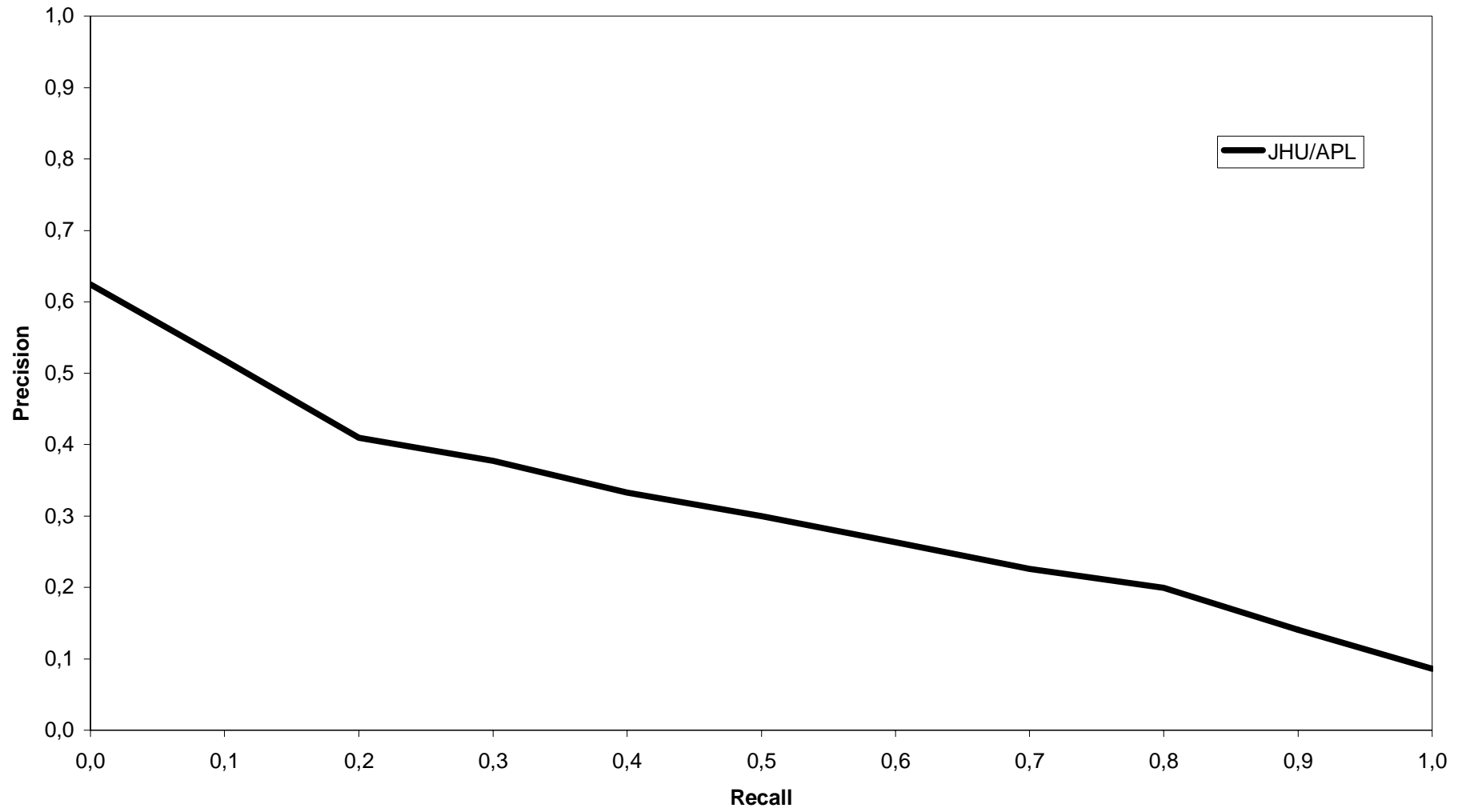




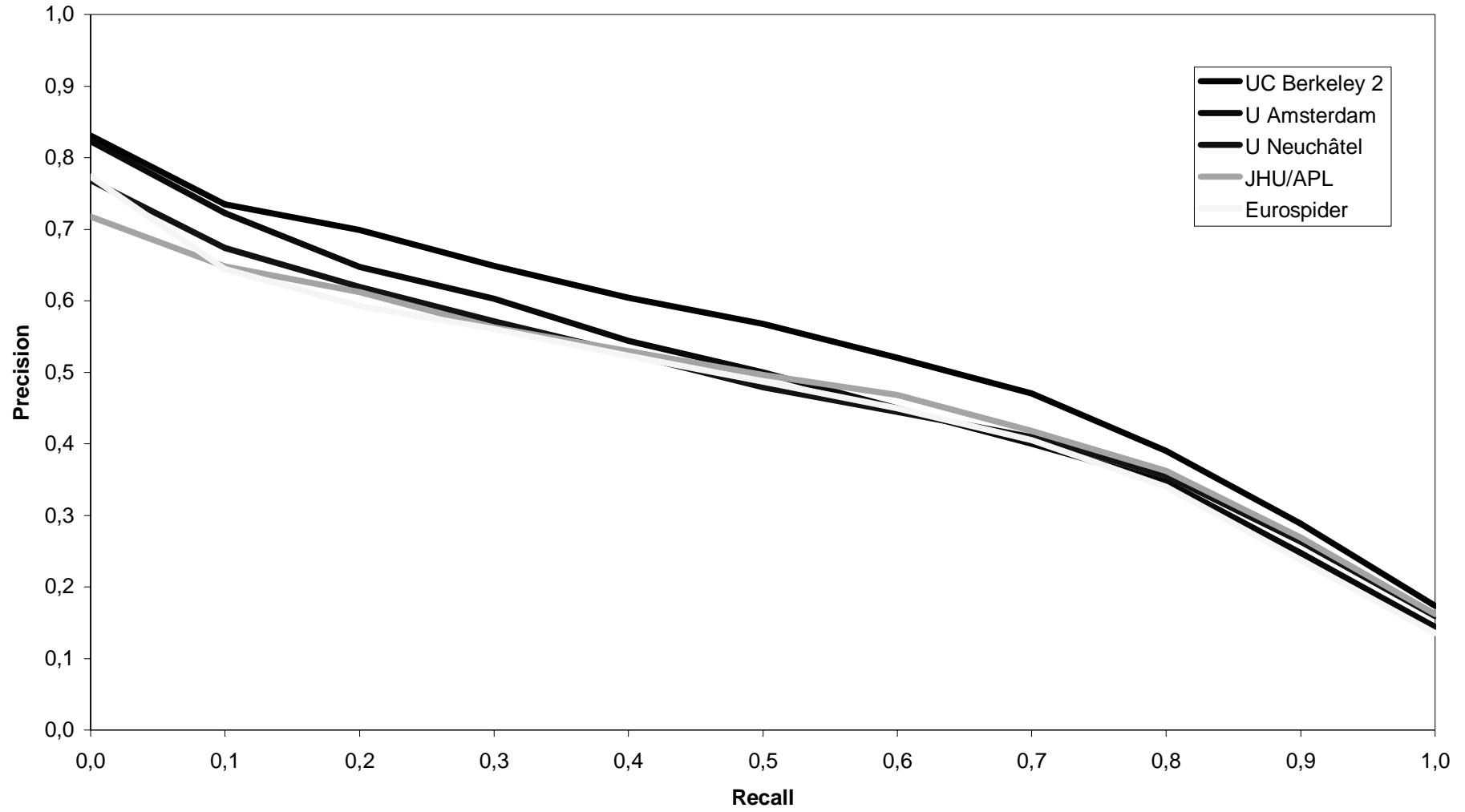
CLEF 2002 Bilingual to Dutch - Automatic



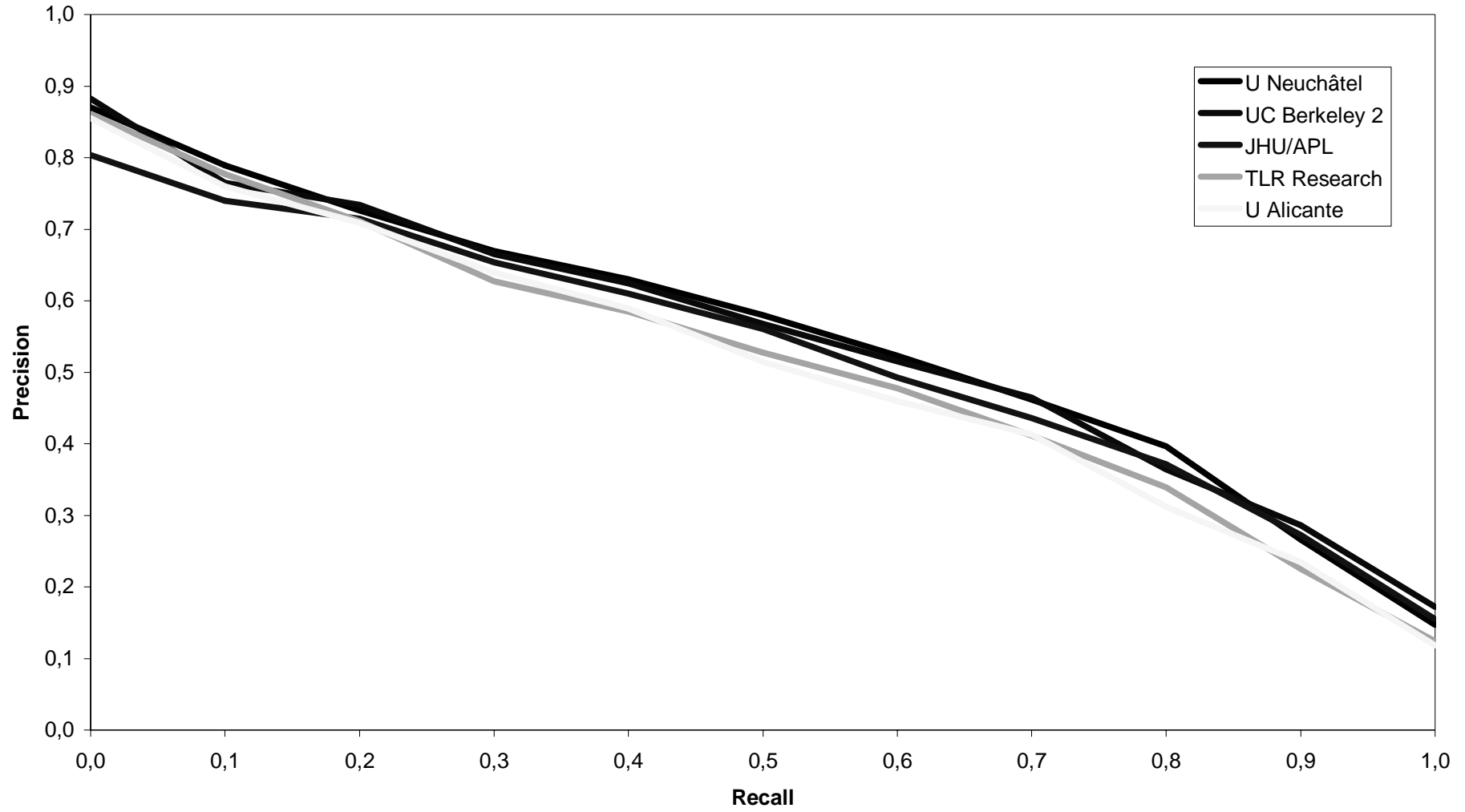
CLEF 2002 Bilingual to Swedish - Automatic



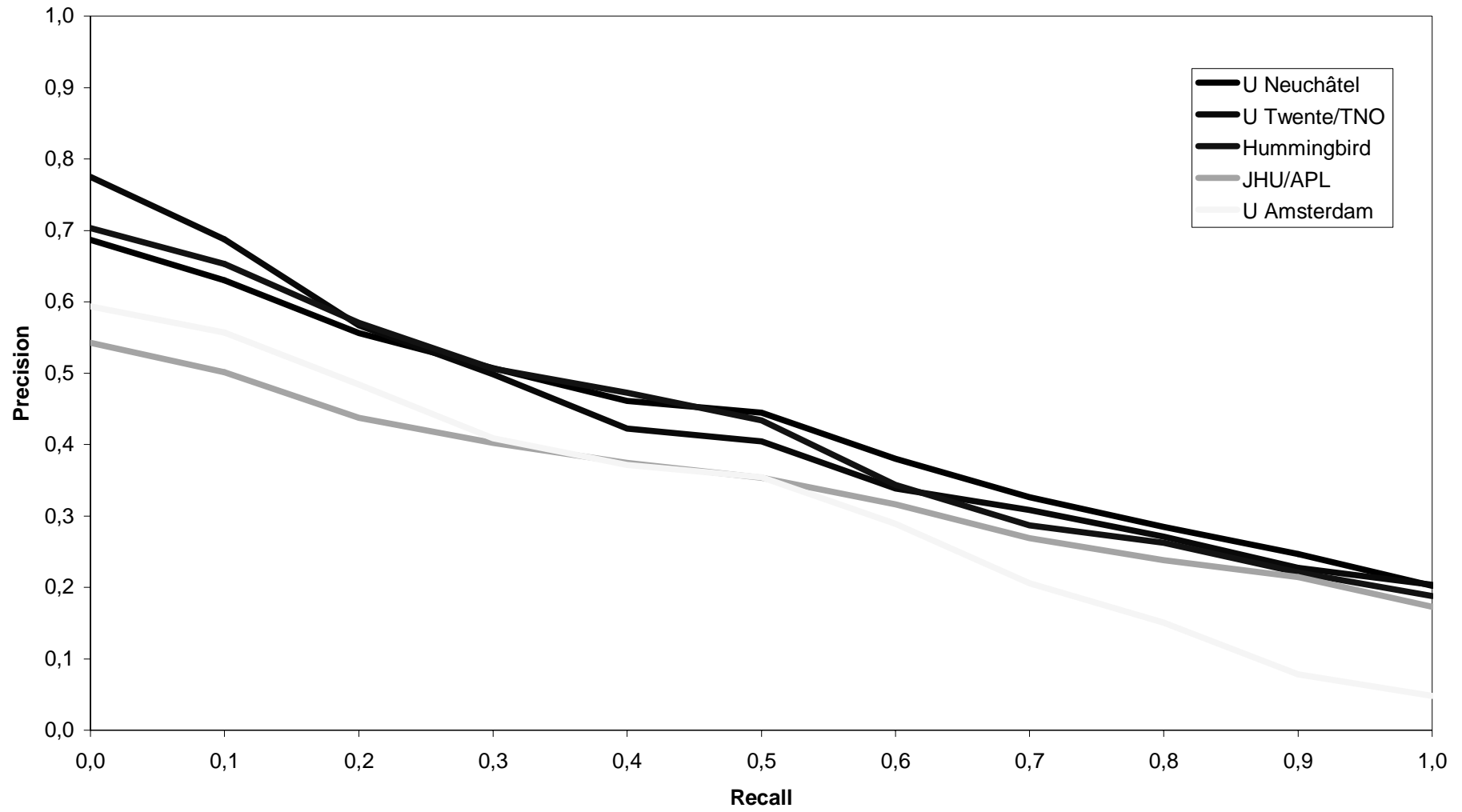
CLEF 2002 Monolingual German - Automatic



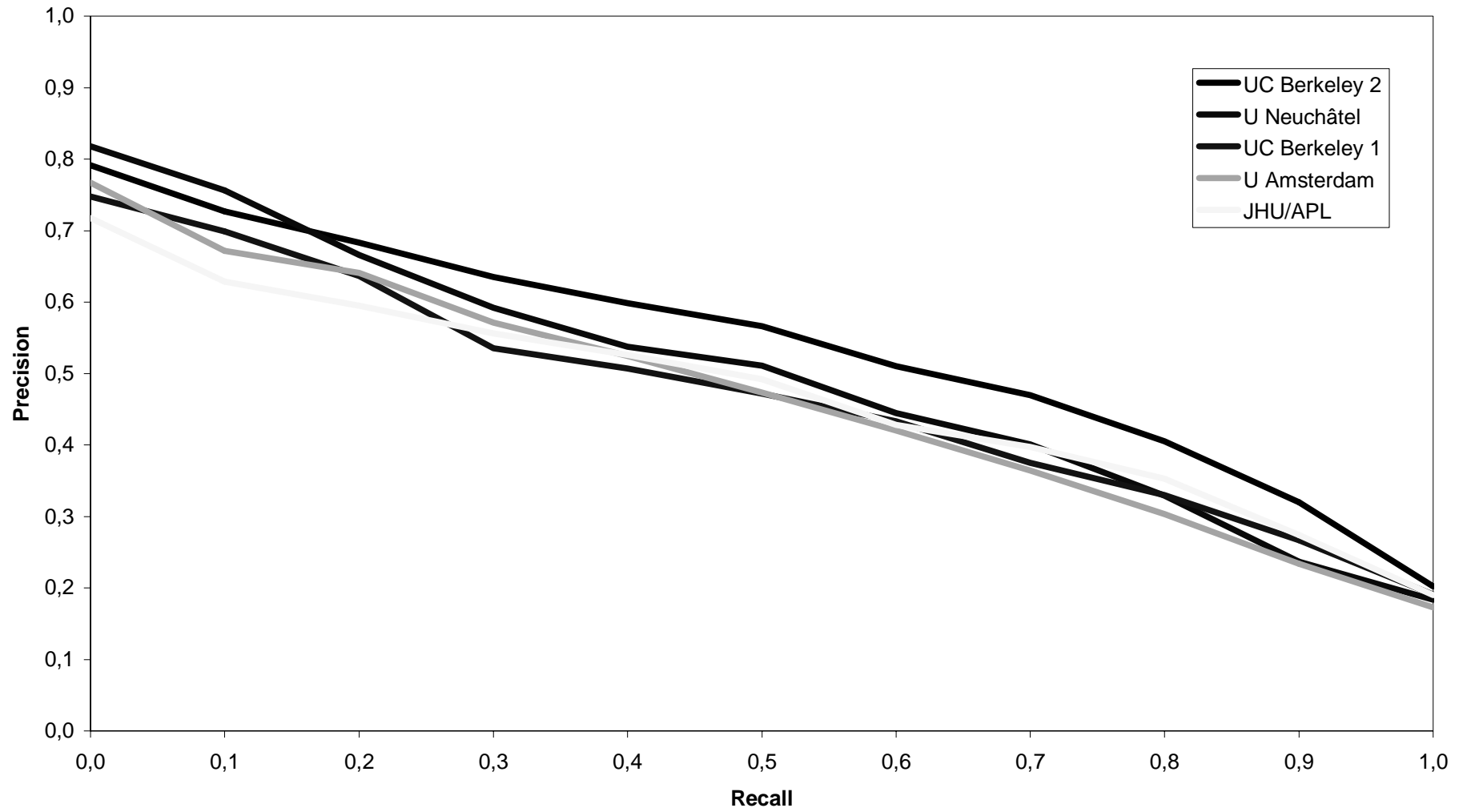
CLEF 2002 Monolingual Spanish - Automatic



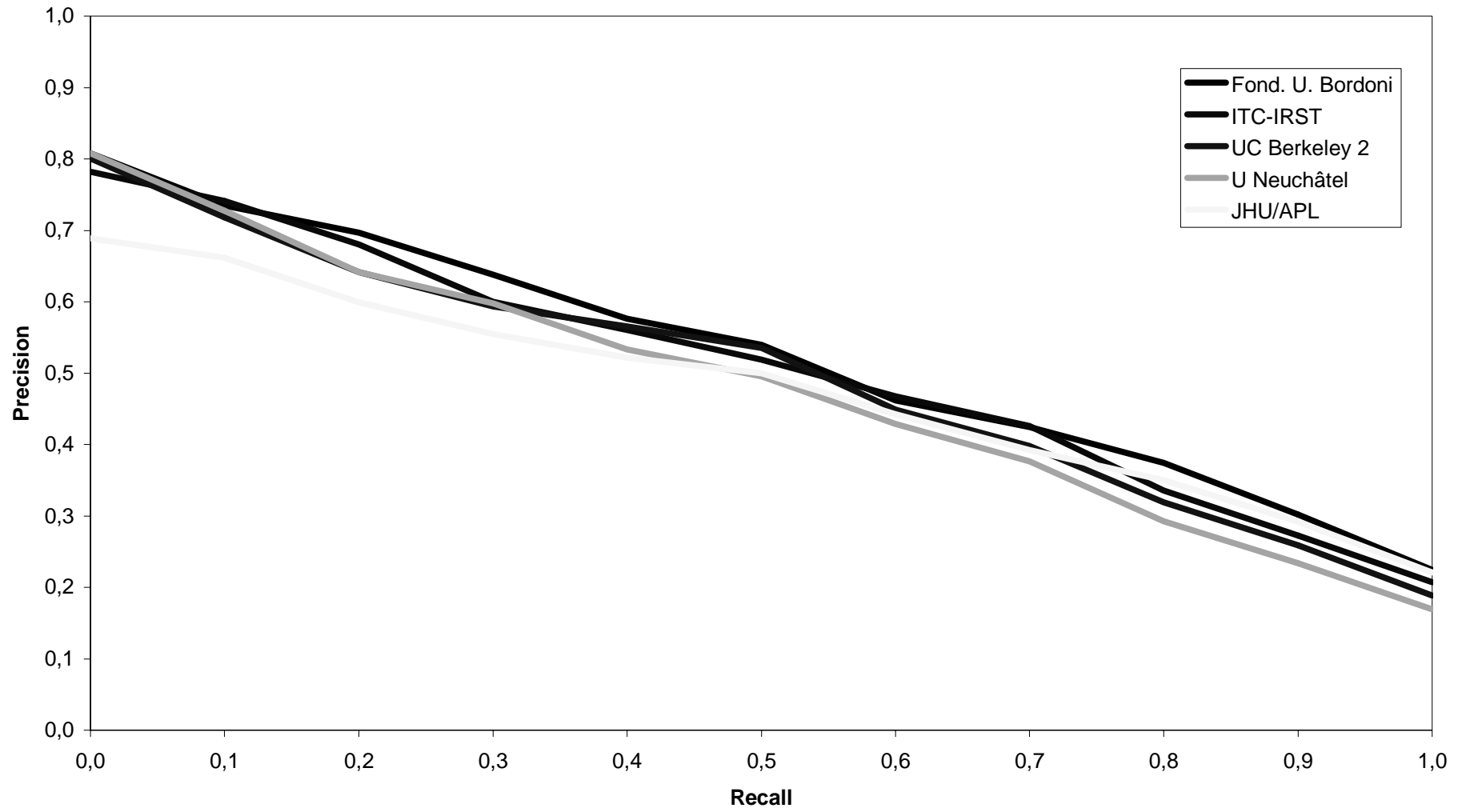
### CLEF 2002 Monolingual Finnish - Automatic



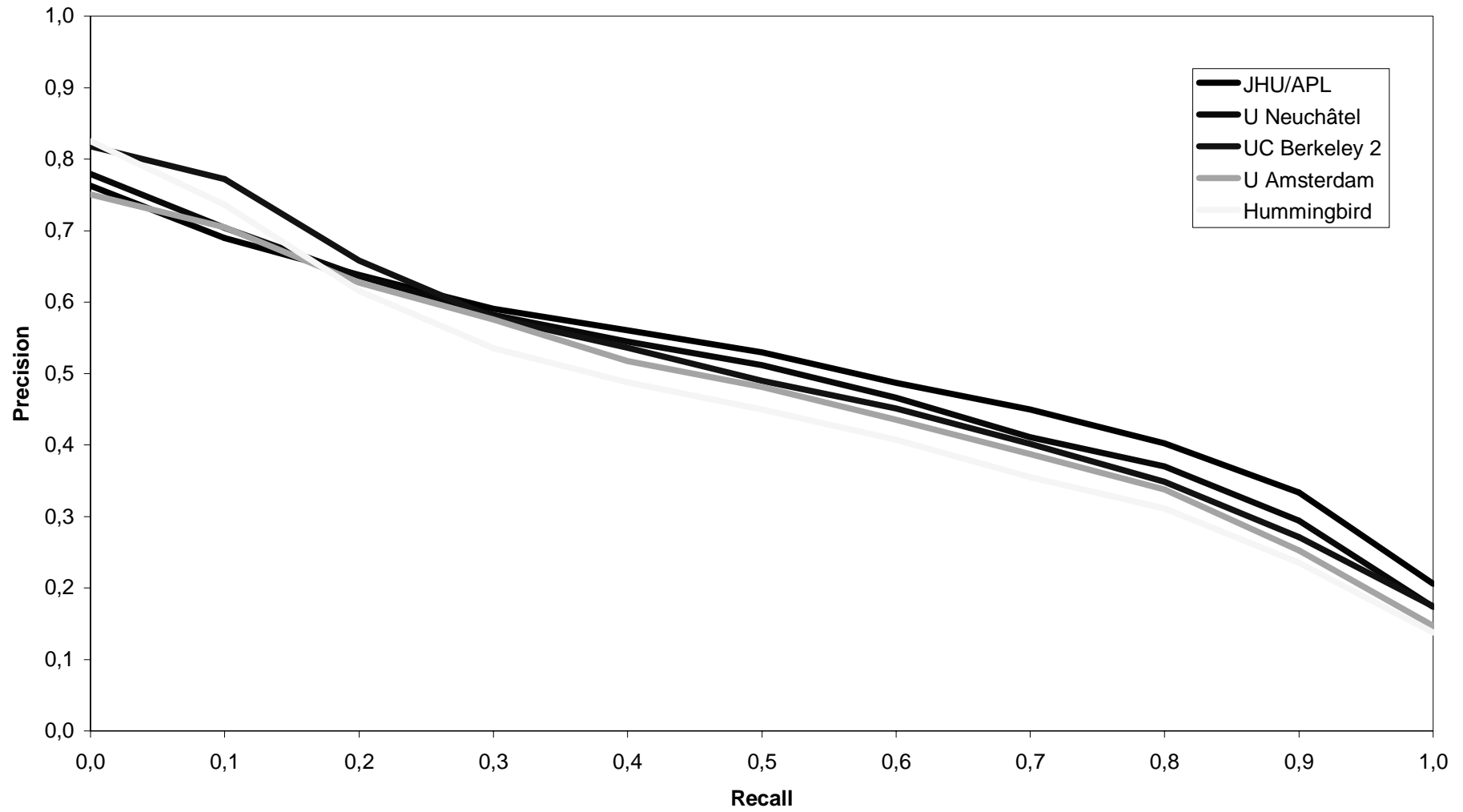
CLEF 2002 Monolingual French - Automatic



CLEF 2002 Monolingual Italian - Automatic

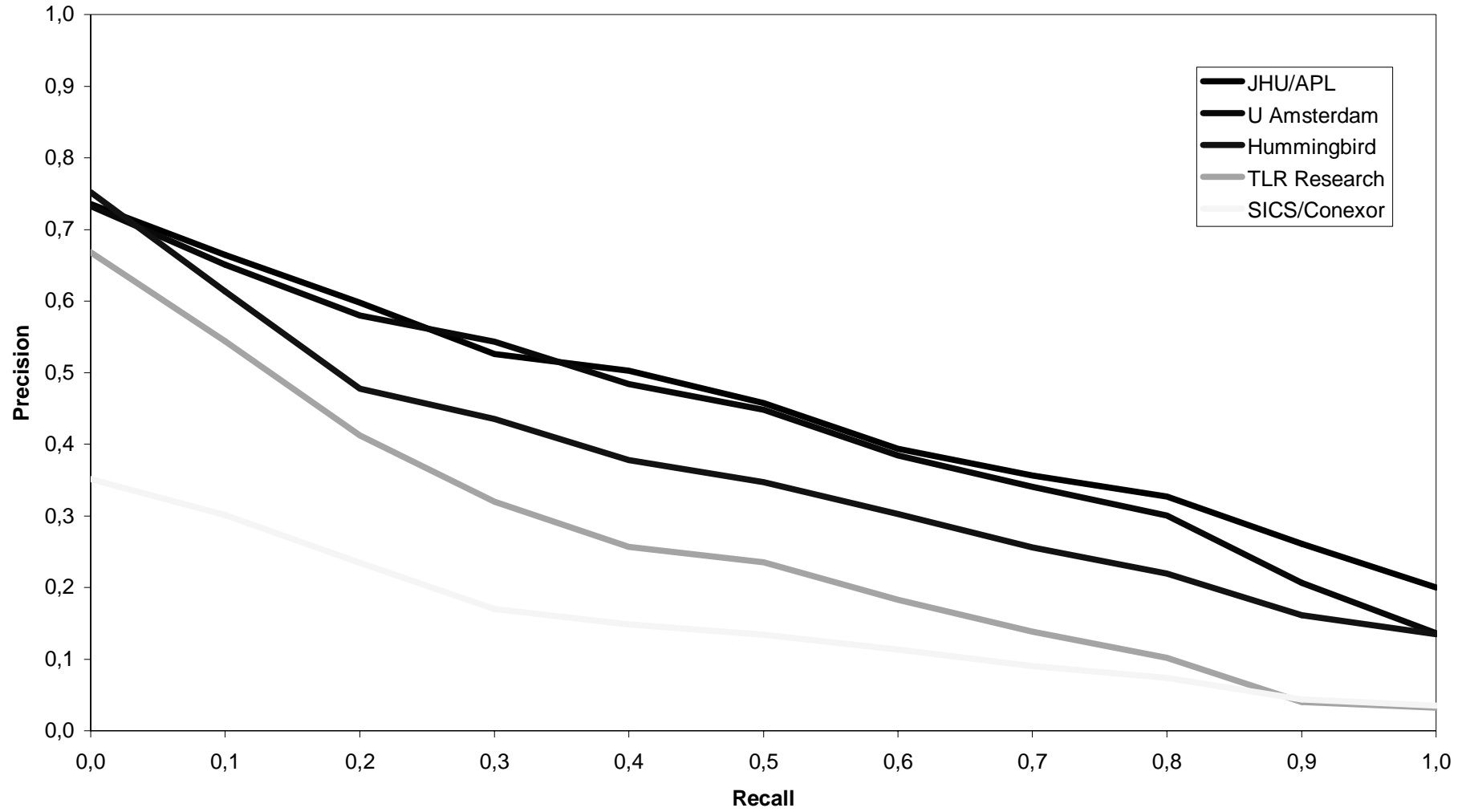


CLEF 2002 Monolingual Dutch - Automatic

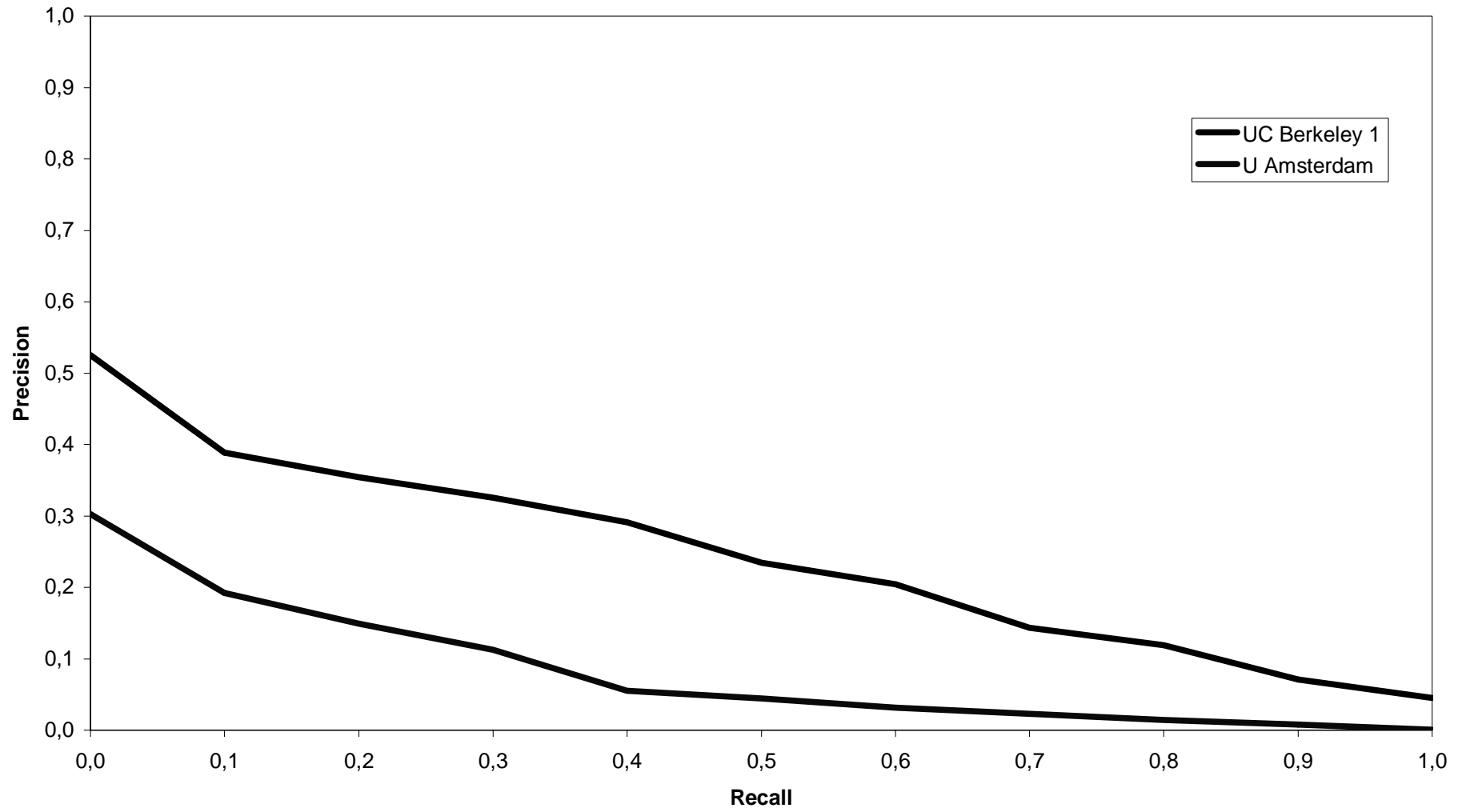




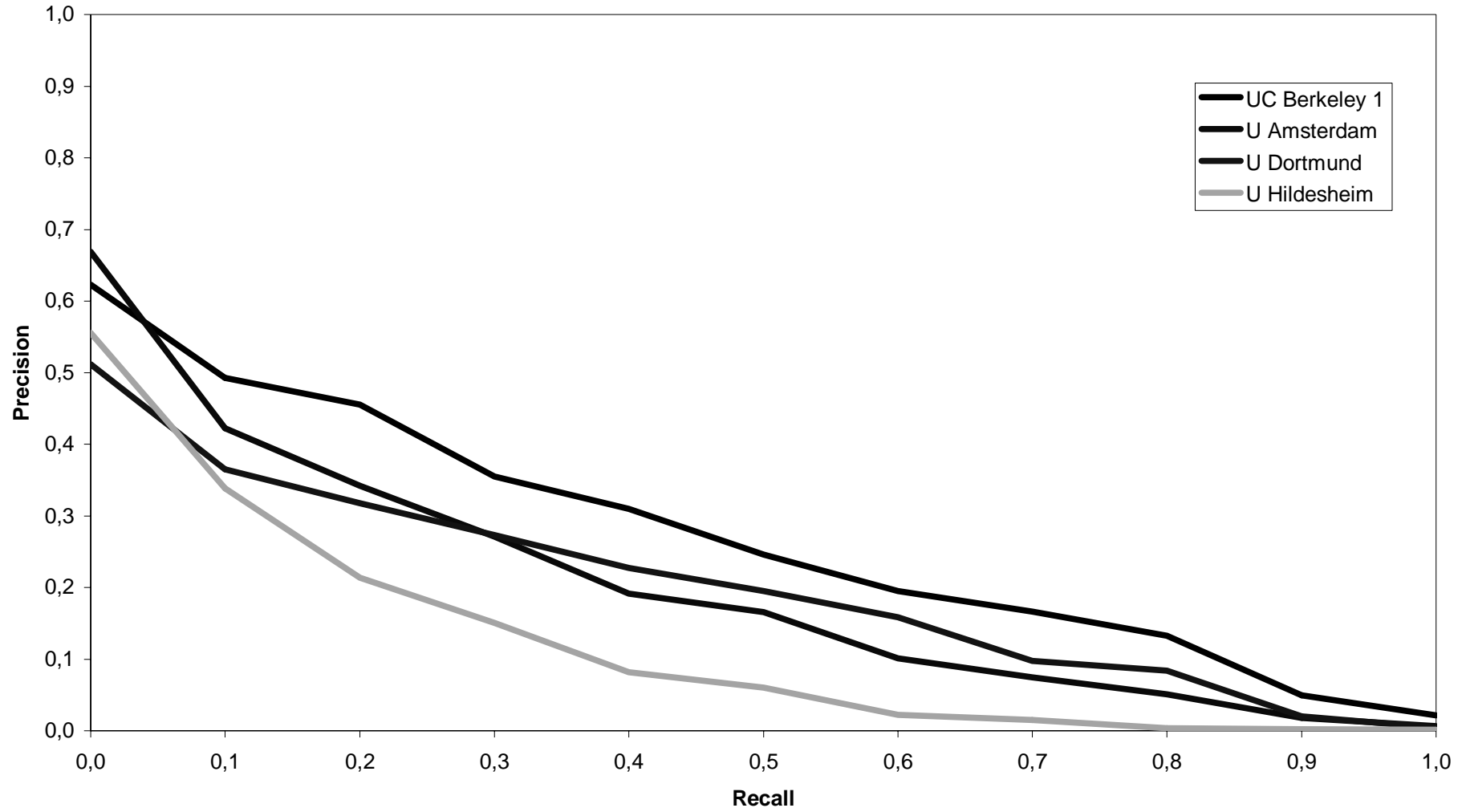
CLEF 2002 Monolingual Swedish - Automatic



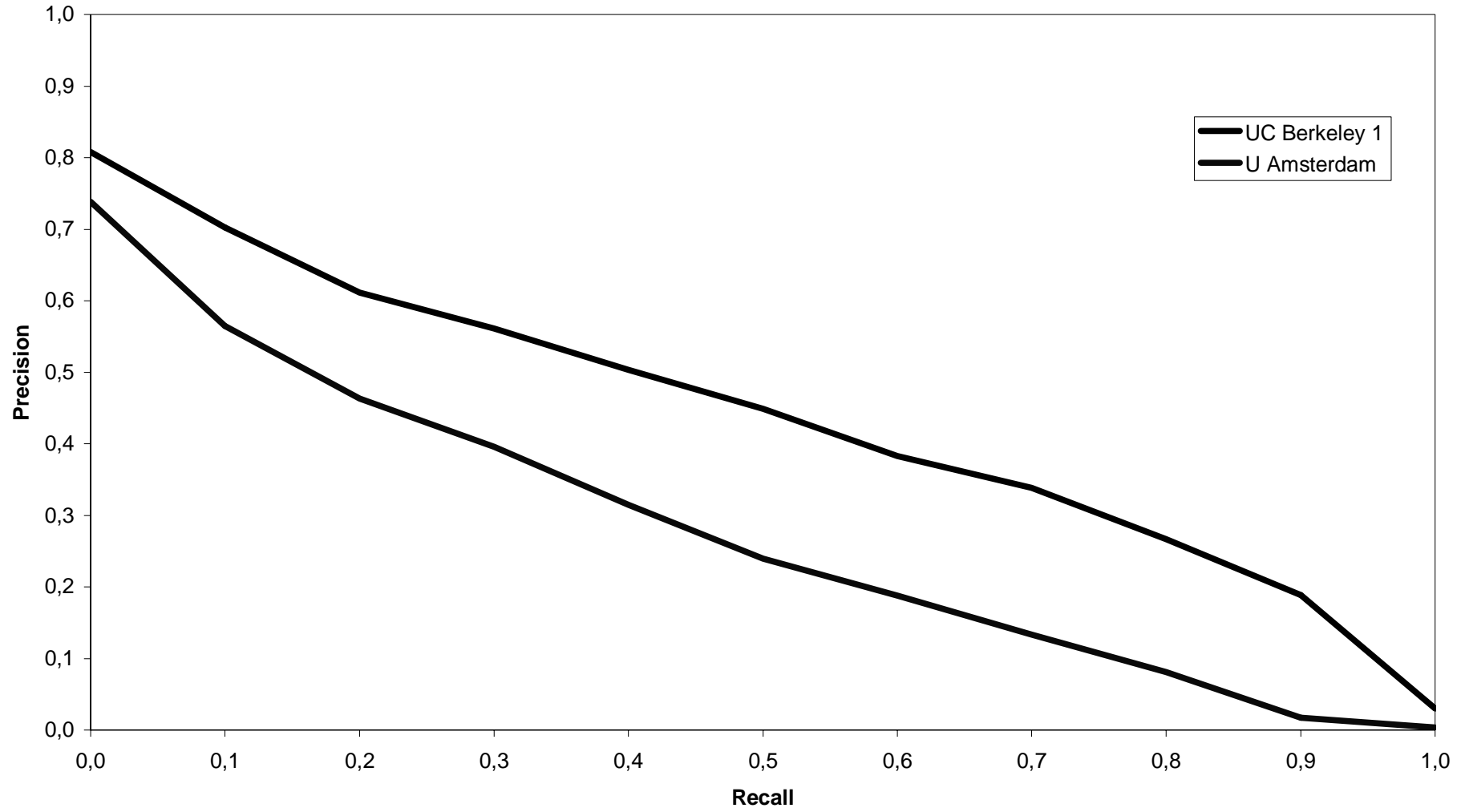
### CLEF 2002 GIRT Bilingual - Automatic



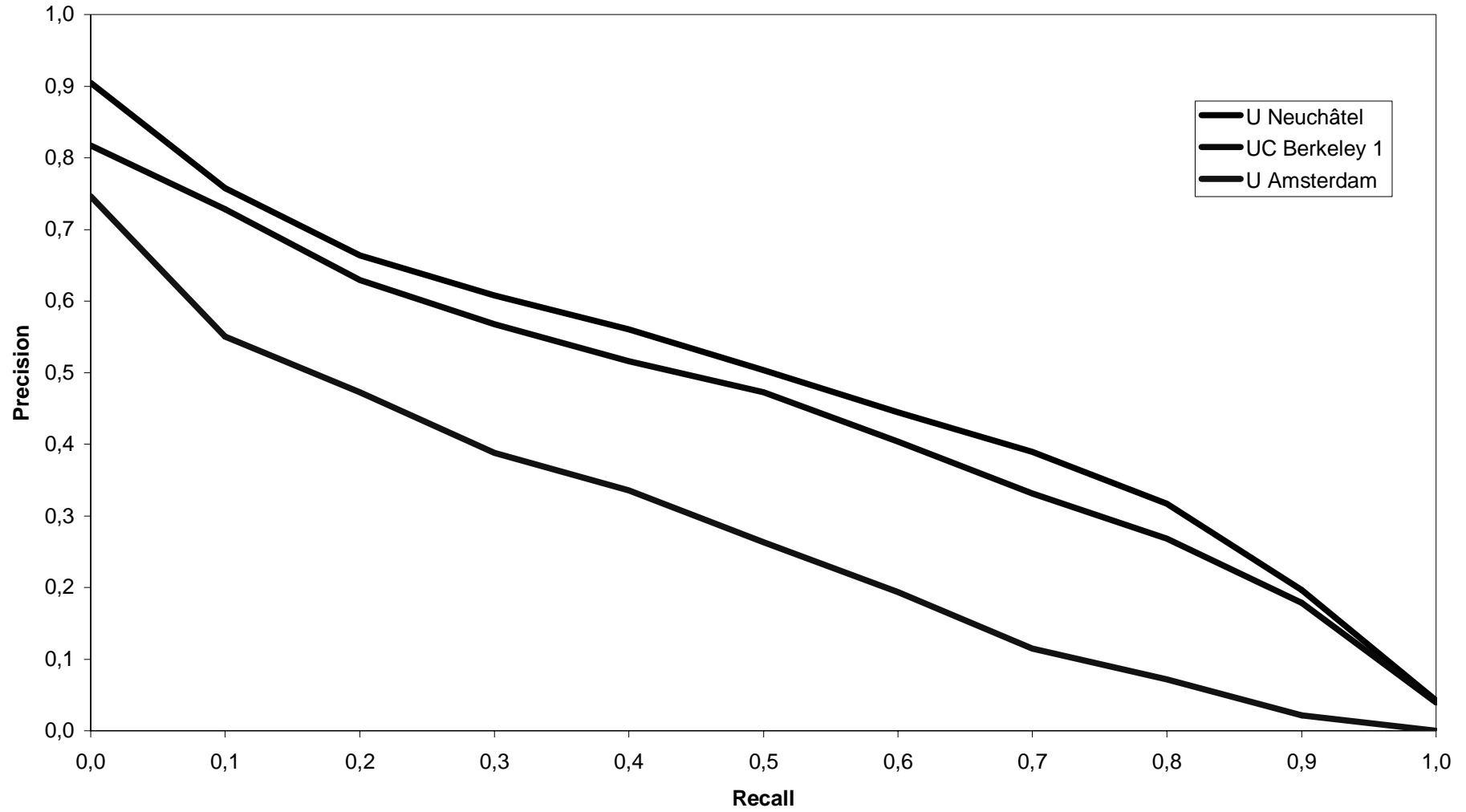
CLEF 2002 GIRT Monolingual - Automatic



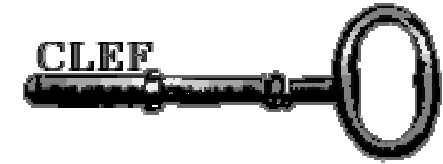
CLEF 2002 Amaryllis Bilingual - Automatic



CLEF 2002 Amaryllis Monolingual - Automatic



## Conclusions and Outlook



- **Returning groups improve performance and move to new tasks**
- **If you want top performance in the monolingual, you have to optimize the individual languages**
- **Groups show flexibility – they adapt lots of ideas, change approaches from year to year, combine new things**
- **Two of the main issues are merging and feedback**
- **Challenge will be how to handle the exploding number of “subtasks”**
- **How can we keep the tasks interesting for “veterans”, while still allowing new groups to catch up?**