



UNIVERSITY OF PADUA
DEPARTMENT OF INFORMATION ENGINEERING

DEPARTMENT OF
INFORMATION
ENGINEERING
UNIVERSITY OF PADOVA



Ad-hoc Track Overview

Giorgio Maria Di Nunzio

dinunzio@dei.unipd.it

Nicola Ferro

ferro@dei.unipd.it

Information Management Systems Research Group

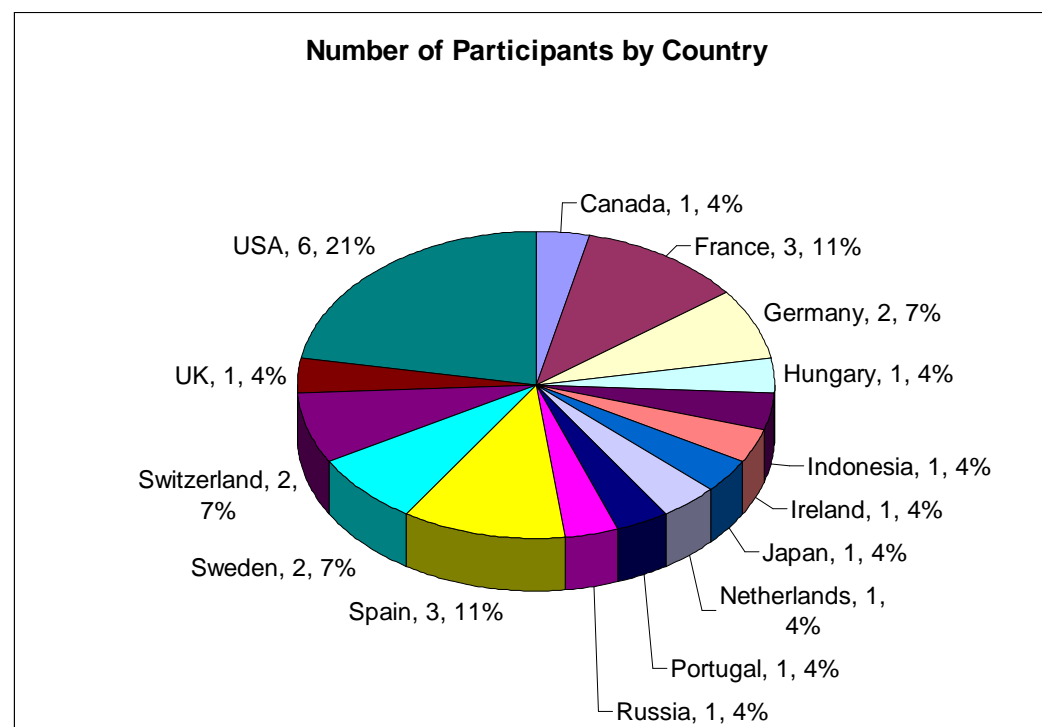


Outline

- ▶ Participants;
- ▶ Collection;
- ▶ Track Definition;
- ▶ Details of Experiments;
- ▶ Pooling;
- ▶ Results.

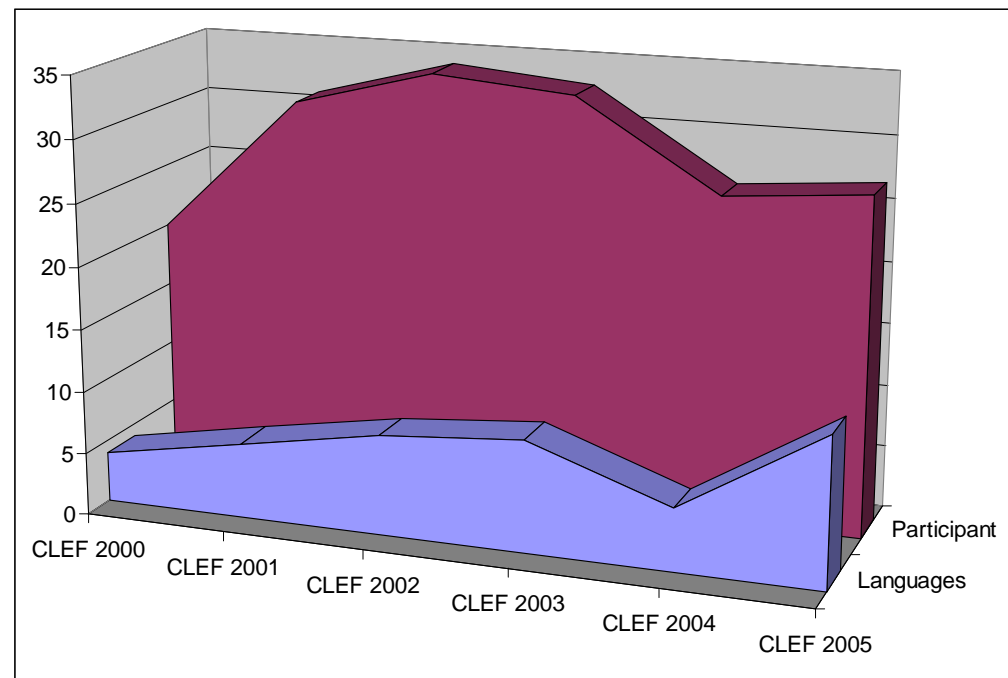
27 Participants - 15 Nations

Participant	Institution	Country
alicante	U. Alicante - Comp.Sci +	Spain
berkeley	U.Berkeley - IM&S	USA
berkeley-2	U.Berkeley - IM&S	USA
buffalo	U.Buffalo - SUNY - Informatics	USA
clips	CLIPS-Grenoble	France
cmu	CMU - Lang.Tech.	USA
cocri	ENS des Mines St. Etienne	France
dcu	Dublin City U. - Comp.Sci.	Ireland
depok	U.Indonesia - Comp.Sci	Indonesia
dsv-stockholm	U.Stockholm, NLP	Sweden
hagen	U.Hagen - Comp.Sci	Germany
hildesheim	U.Hildesheim - Inf.Sci	Germany
hummingbird	Hummingbird Core Tech.	Canada
ilps	U.Amsterdam - Informatics	The Netherlands
irit	Inst.Informatique,Toulouse	France
isi-unige	U.Geneva - Inf.Systems	Switzerland
jaen	U.Jaen - Intell.Systems	Spain
JHU/apl	Johns Hopkins U.	USA
miracle	Daedalus & Madrid Univs	Spain
msu-nivc	Moscow State U.-Computing	Russia
sics	Swedish Inst.Comp.Sci	Sweden
tlr	Thomson Legal Regulatory	USA
u.budapest	Budapest U. Tech. & Econom	Hungary
u.glasgow	U.Glasgow - IR	UK
u.surugadai	U.Surugadai - Cultural Inf.	Japan
unine	U.Neuchatel	Switzerland
xldb	U.Lisbon	Portugal



The CLEF 2005 Multilingual Collections

Edition	# Part.	# Lang.	# Docs	Size in MB	# assess.	# topics	#assess/topic
CLEF 2005	27	5 (12)	1,073,607 (2,313,442)	772 (1,789)	105,642	50 (60)	2,113
CLEF 2004	26	5	273,863	779	114,346	50	2,287
CLEF 2003	33	9	1,611,178	4,124	188,475	60 (37)	~ 3,100
CLEF 2002	34	8	1,138,650	3,011	140,043	50 (30)	~ 2,900
CLEF 2001	31	6	940,487	2,522	97,398	50	1,948
CLEF 2000	20	4	368,763	1,158	43,566	40	1,089



The CLEF 2005 Ad-hoc Tasks

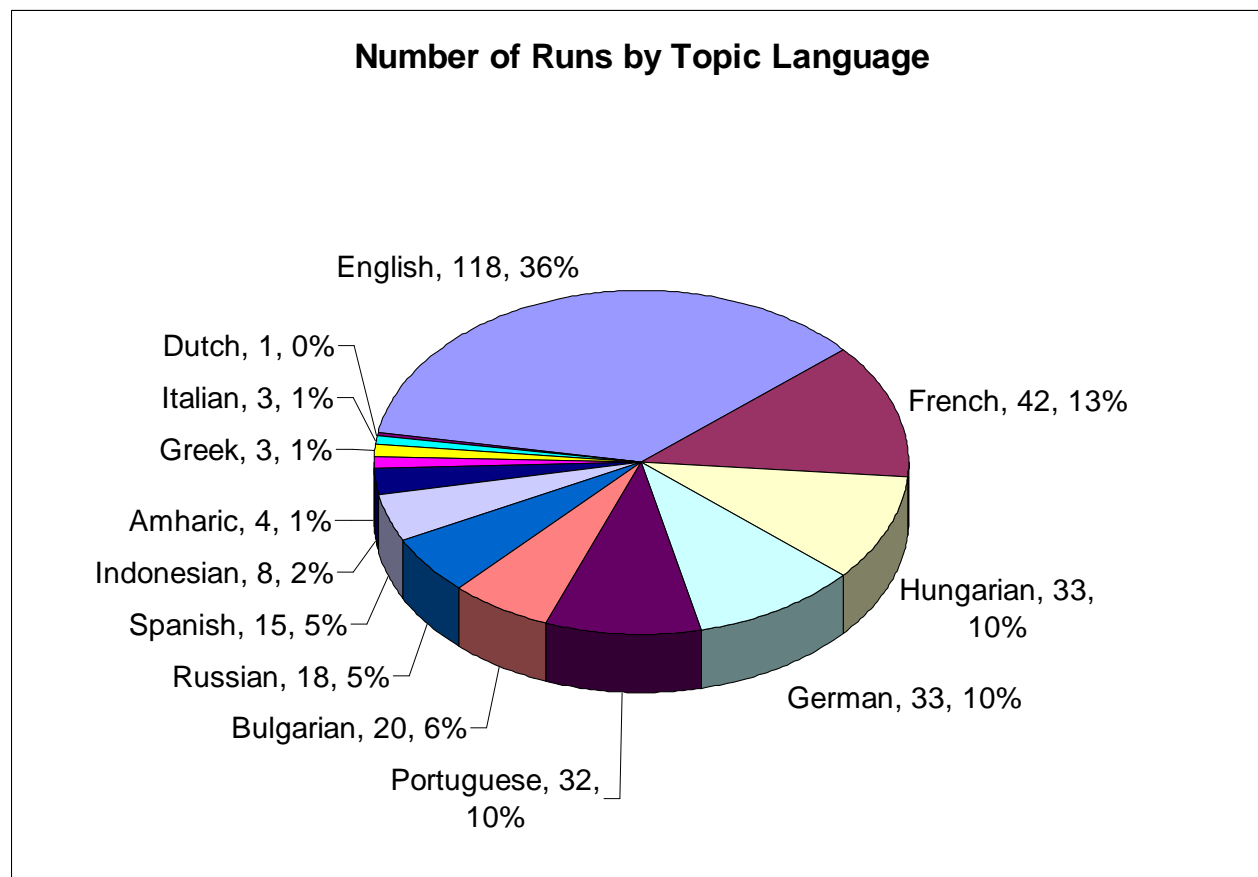
- ▶ Multilingual:
 - ▷ 2-years-on
 - ▷ Merging
- ▶ Bilingual:
 - ▷ X → Bulgarian
 - ▷ X → French
 - ▷ X → Hungarian
 - ▷ X → Portuguese
 - ▷ X → English (unusual source languages)
- ▶ Monolingual:
 - ▷ Bulgarian
 - ▷ French
 - ▷ Hungarian
 - ▷ Portuguese
- ▶ Domain Specific:
 - ▷ Multilingual (X → English ; German ; Russian)
 - ▷ Bilingual (X → English ; X → German ; X → Russian)
 - ▷ Monolingual (English ; German ; Russian)

Details of Experiments

Track	# Participants	# Runs	Avg. # Run per Part.
AH-2-years-on	4	21	5.25
AH-Merging	3	20	6.67
AH-Bilingual-X2BG	4	12	3.00
AH-Bilingual-X2FR	9	31	3.44
AH-Bilingual-X2HU	3	7	2.33
AH-Bilingual-X2PT	8	28	3.50
AH-Bilingual-X2EN	4	13	3.25
AH-Monolingual-BG	7	20	2.86
AH-Monolingual-FR	12	38	3.17
AH-Monolingual-HU	10	32	3.20
AH-Monolingual-PT	9	32	3.56
DS-Multi	1	3	3.00
DS-Bilingual-X2DE	5	15	3.00
DS-Bilingual-X2EN	4	13	3.25
DS-Bilingual-X2RU	3	5	1.67
DS-Monolingual-DE	6	17	2.83
DS-Monolingual-EN	6	15	2.50
DS-Monolingual-RU	5	8	1.60
Total		330	3.20

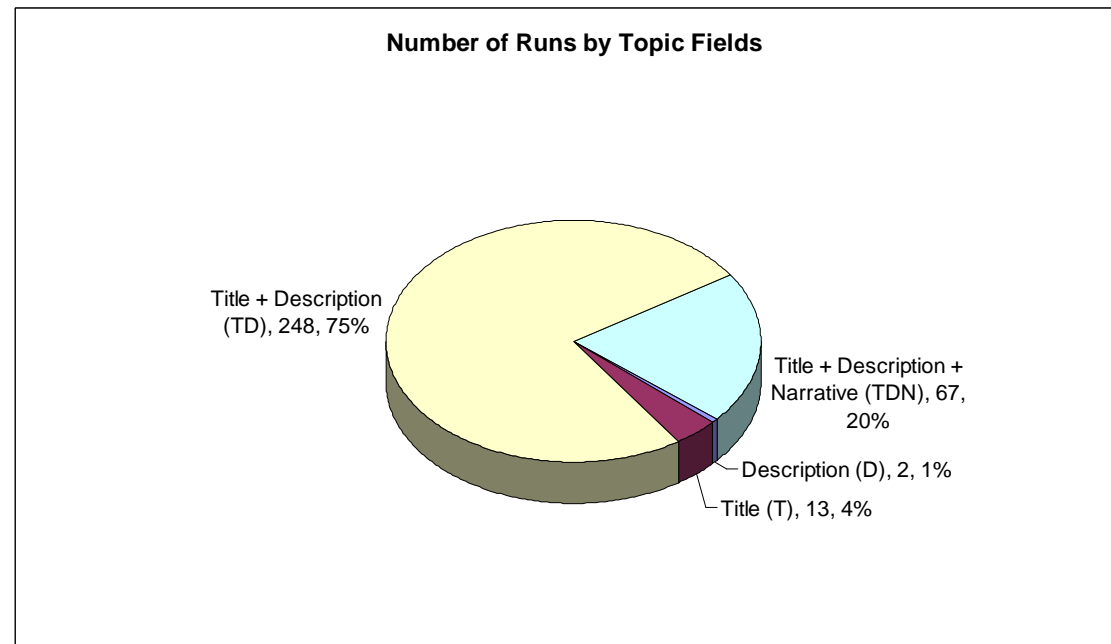
Runs by Topic Language

Topic Language	# Runs
English	118
French	42
Hungarian	33
German	33
Portuguese	32
Bulgarian	20
Russian	18
Spanish	15
Indonesian	8
Amharic	4
Greek	3
Italian	3
Dutch	1
Total	330



Runs by Topic Fields

Topic Field	# Runs
Title + Description (TD)	248
Title + Description + Narrative (TDN)	67
Title (T)	13
Description (D)	2
Total	330



CLEF 2005 Pooling

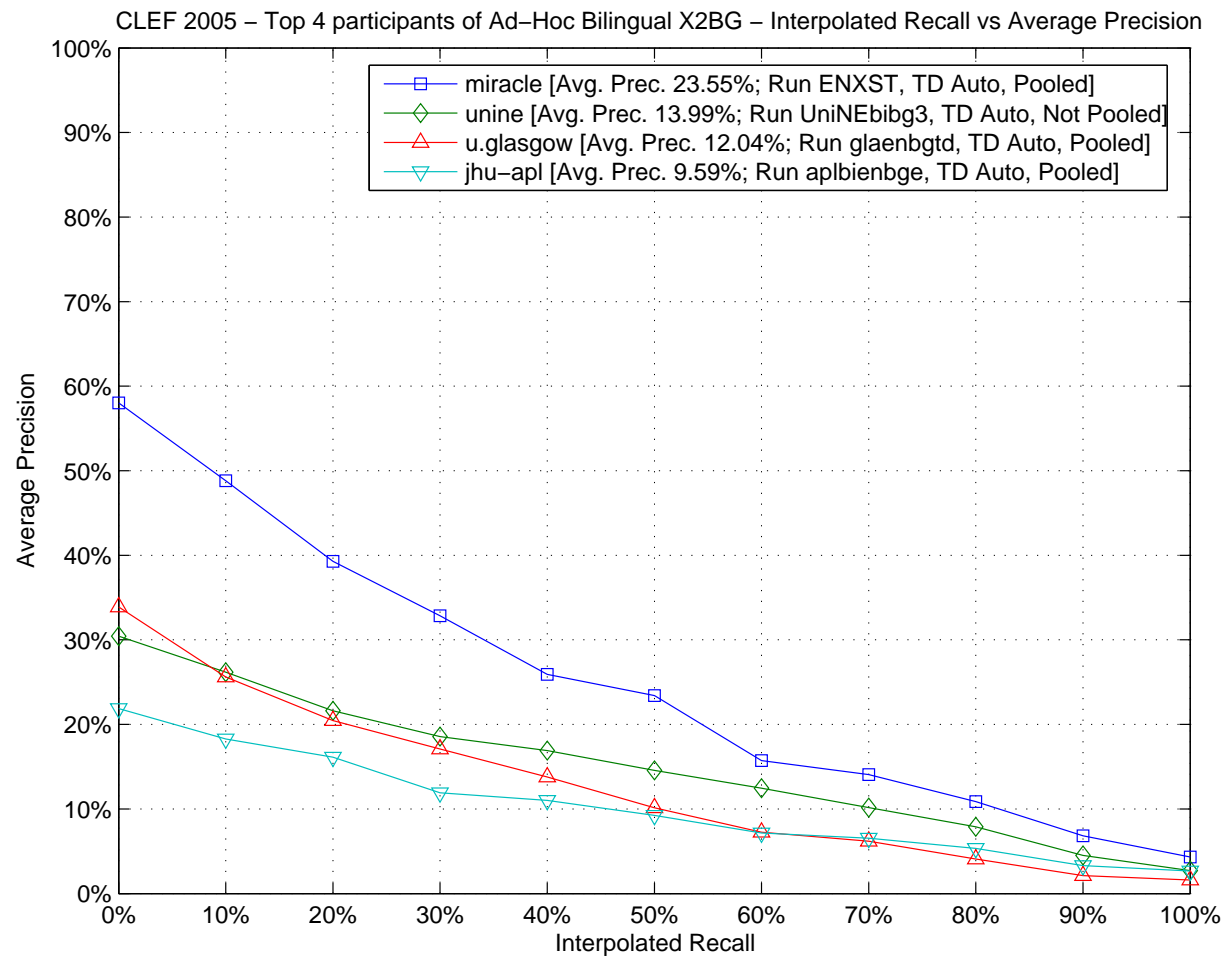
Pool	# Submitted Runs	# Pooled	Ratio
AH Bulgarian	32 (Mono 20 + Bi 12)	24 (Mono 15 + Bi 9)	75.00%
AH French	69 (Mono 38 + Bi 31)	21 (Mono 12 + Bi 9)	30.44%
AH Hungarian	39 (Mono 32 + Bi 7)	26 (Mono 23 + Bi 3)	66.67%
AH Portuguese	60 (Mono 32 + Bi 28)	20 (Mono 10 + Bi 10)	33.34%
AH English*	13 (Extended 31)	9 (Extended 15)	48.39%
DS German	32 (Mono 17 + Bi 15)	32 (Mono 17 + Bi 15)	100.00%
DS English	28 (Mono 14 + Bi 14)	28 (Mono 14 + Bi 14)	100.00%
DS Russian	13 (Mono 8 + Bi 5)	13 (Mono 8 + Bi 5)	100.00%
Total	286 (Mono 161 + Bi 125)	173 (Mono 99 + Bi 74)	60.49%

- ▶ *Only Bilingual Task. We asked some participants to send monolingual English runs to enrich the English pool.
 - ▷ many thanks to ilps, msu-nivc, alicante, buffalo, jaen.

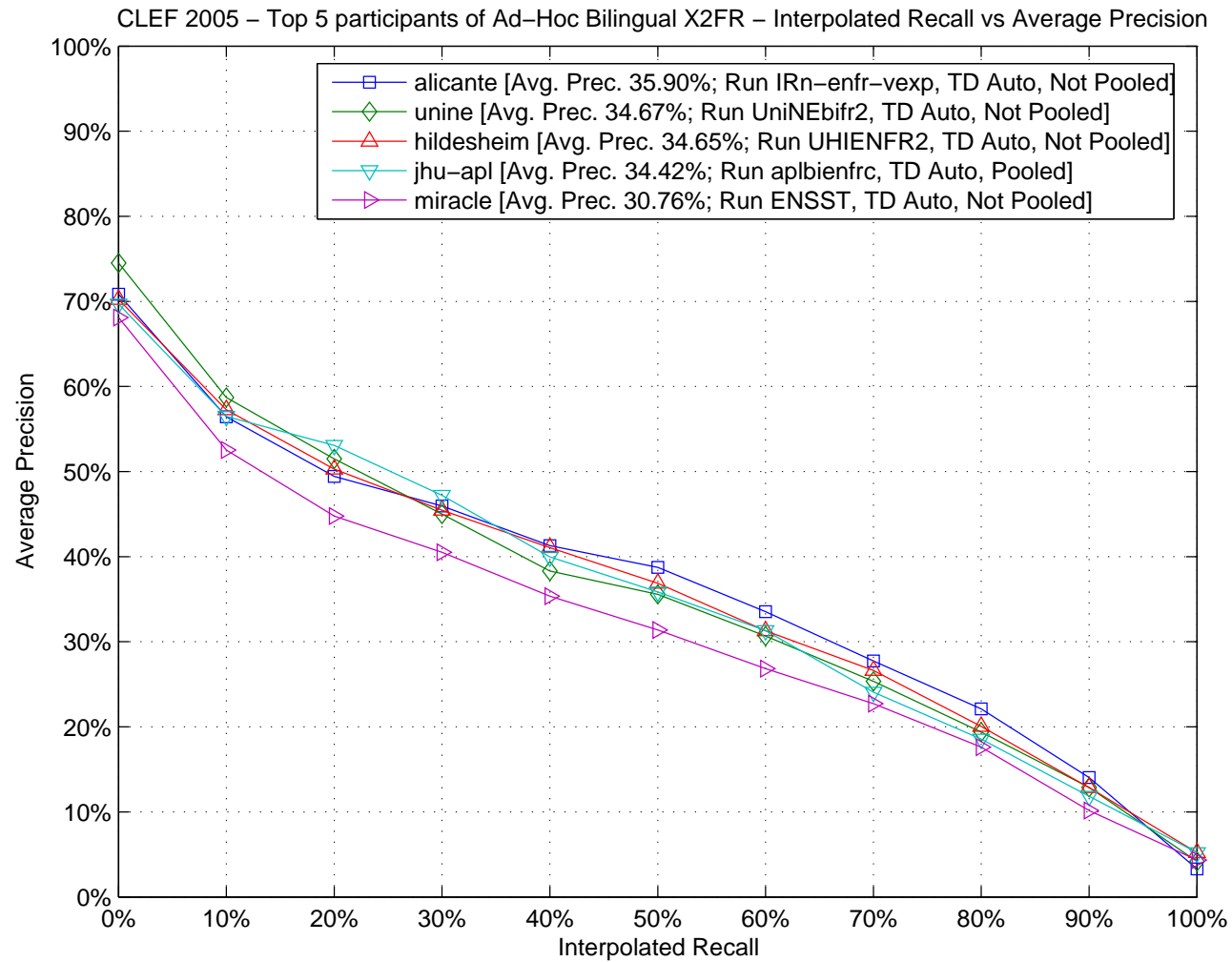
Bilingual Task

		Participant Rank					
Track		1st	2nd	3rd	4th	5th	Diff.
Bilingual Bulgarian	Part.	miracle	unine	u.glasgow	jhu/apl		1st vs 4th
	Avg. Prec.	23.55%	13.99%	12.04%	9.59%		145.57%
Bilingual French	Part.	alicante	unine	hildesheim	jhu/apl	miracle	1st vs 5th
	Avg. Prec.	35.90%	34.67%	34.65%	34.42%	30.76%	16.71%
Bilingual Hungarian	Part.	miracle	unine	jhu/apl			1st vs 3rd
	Avg. Prec.	30.16%	28.82%	24.58%			22.70%
Bilingual Portuguese	Part.	unine	jhu/apl	miracle	alicante	tlr	1st vs 5th
	Avg. Prec.	34.04%	31.85%	31.06%	29.18%	23.58%	44.36%
Bilingual English	Part.	jhu/apl	u.glasgow	depok			1st vs 3rd
	Avg. Prec.	33.13%	29.35%	12.85%			157.82%

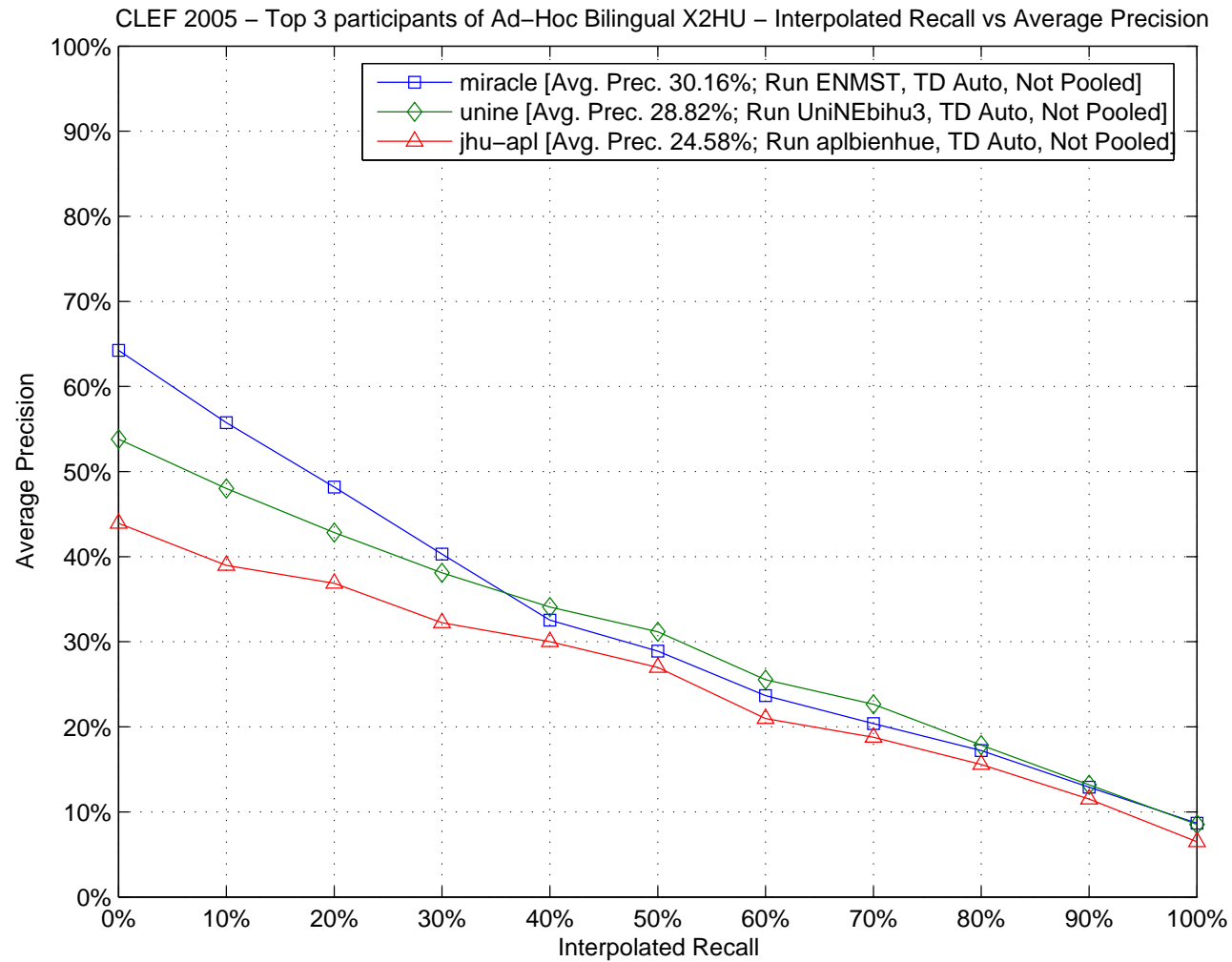
Bilingual X2BG



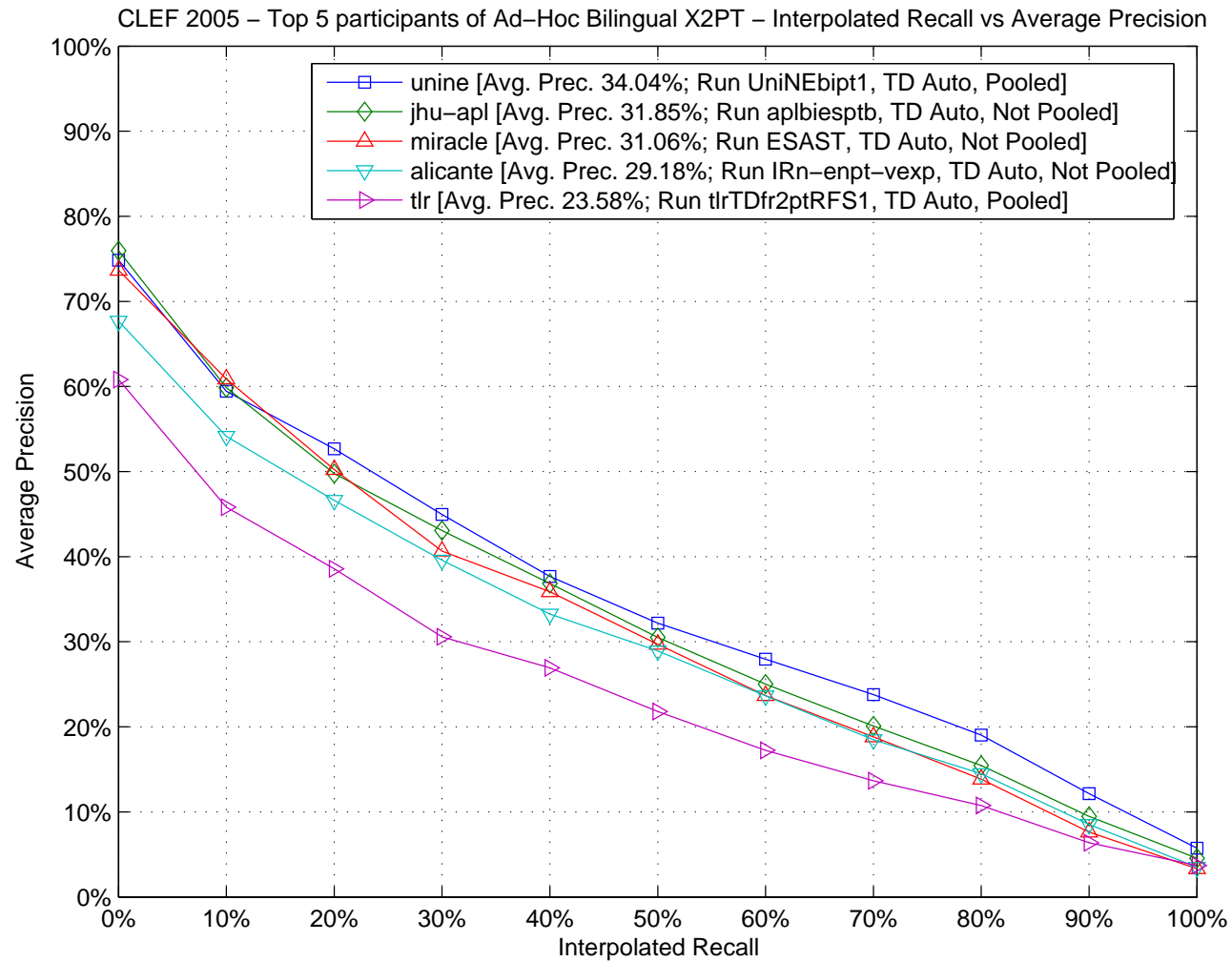
Bilingual X2FR



Bilingual X2HU

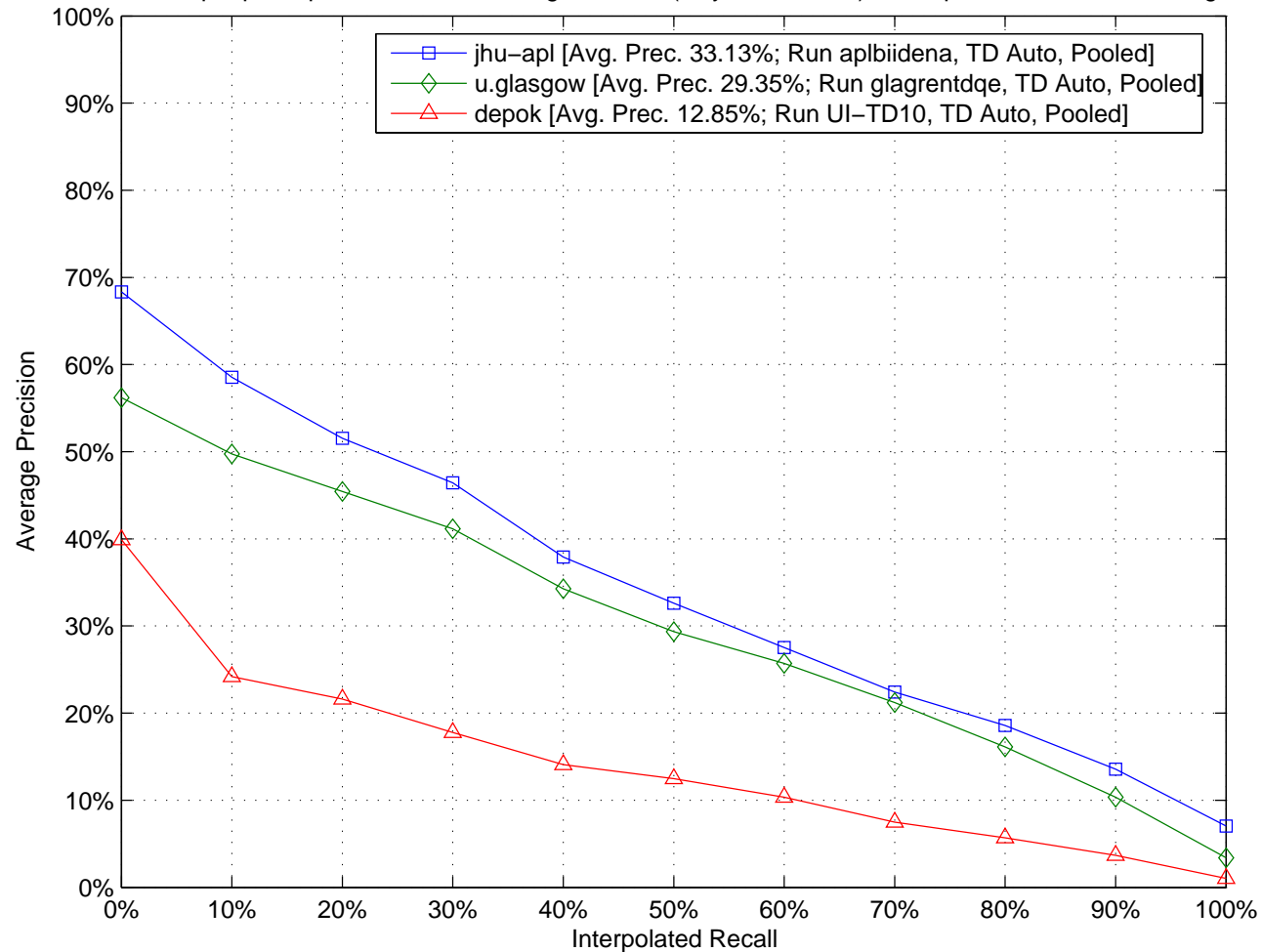


Bilingual X2PT



Bilingual X2EN

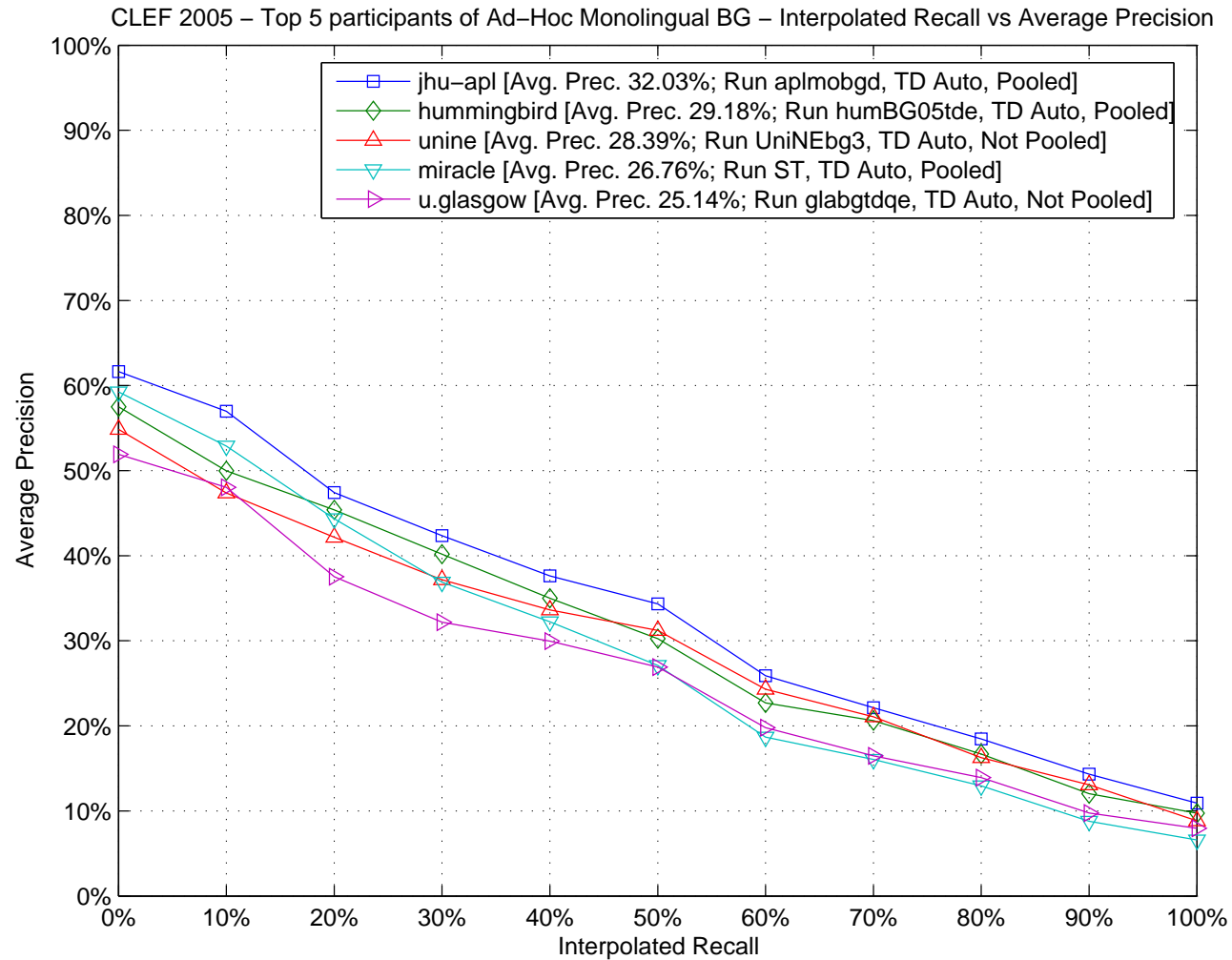
CLEF 2005 – Top 3 participants of Ad-Hoc Bilingual X2EN (only newcomers) – Interpolated Recall vs Average Pre



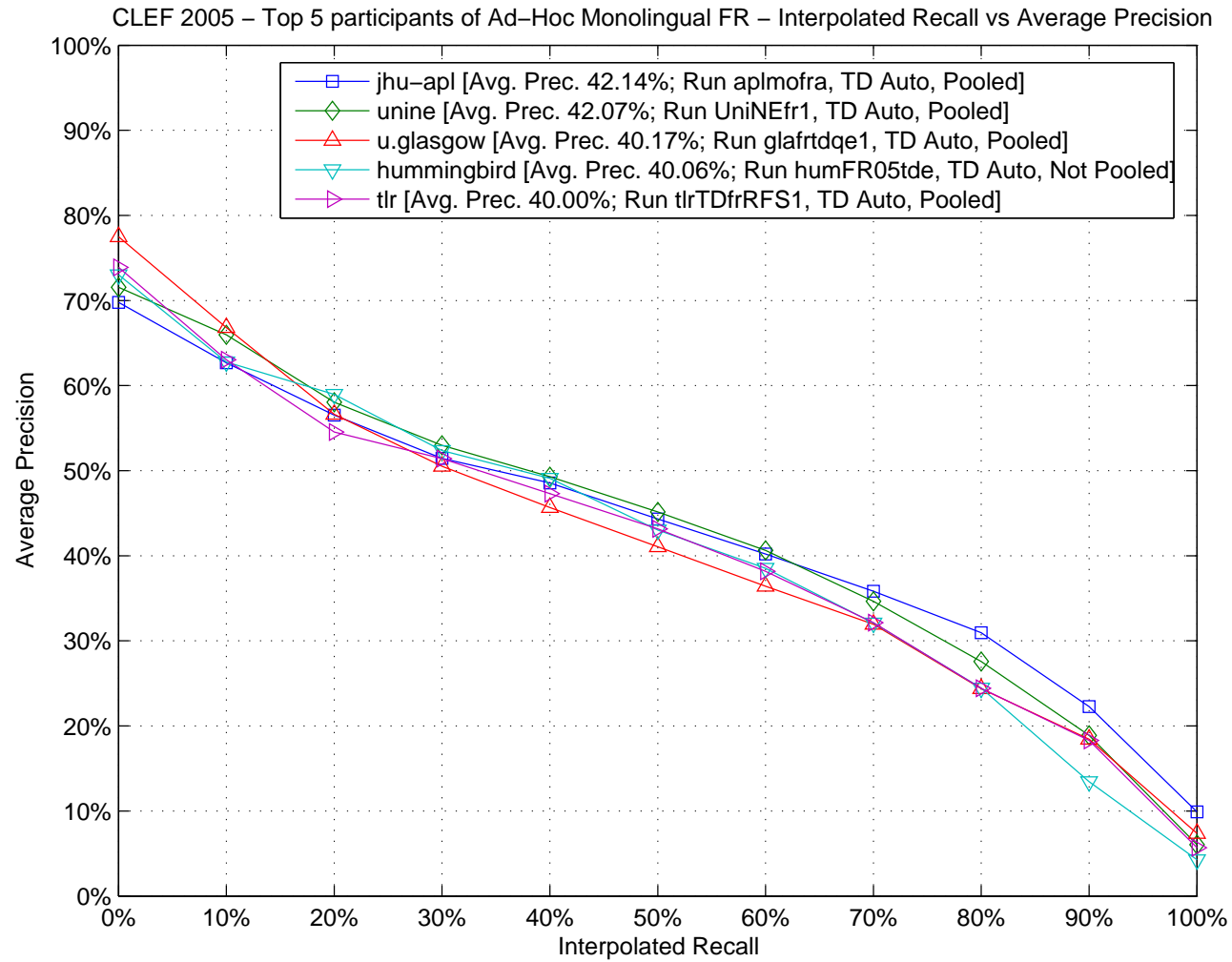
Monolingual Tasks

Track		Participant Rank					Diff.
		1st	2nd	3rd	4th	5th	
Monolingual Bulgarian	Part.	jhu/apl	hummingbird	unine	miracle	u.glasgow	1st vs 5th
	Avg. Prec.	32.03%	29.18%	28.39%	26.76%	25.14%	27.41%
Monolingual French	Part.	jhu/apl	unine	u.glasgow	hummingbird	tlr	1st vs 5th
	Avg. Prec.	42.14%	42.07%	40.17%	40.06%	40.00%	5.35%
Monolingual Hungarian	Part.	jhu/apl	unine	miracle	hummingbird	hildesheim	1st vs 5th
	Avg. Prec.	41.12%	38.89%	35.20%	33.09%	32.64%	25.98%
Monolingual Portuguese	Part.	unine	hummingbird	tlr	jhu-apl	alicante	1st vs 5th
	Avg. Prec.	38.75%	38.64%	37.42%	36.54%	36.03%	7.55%

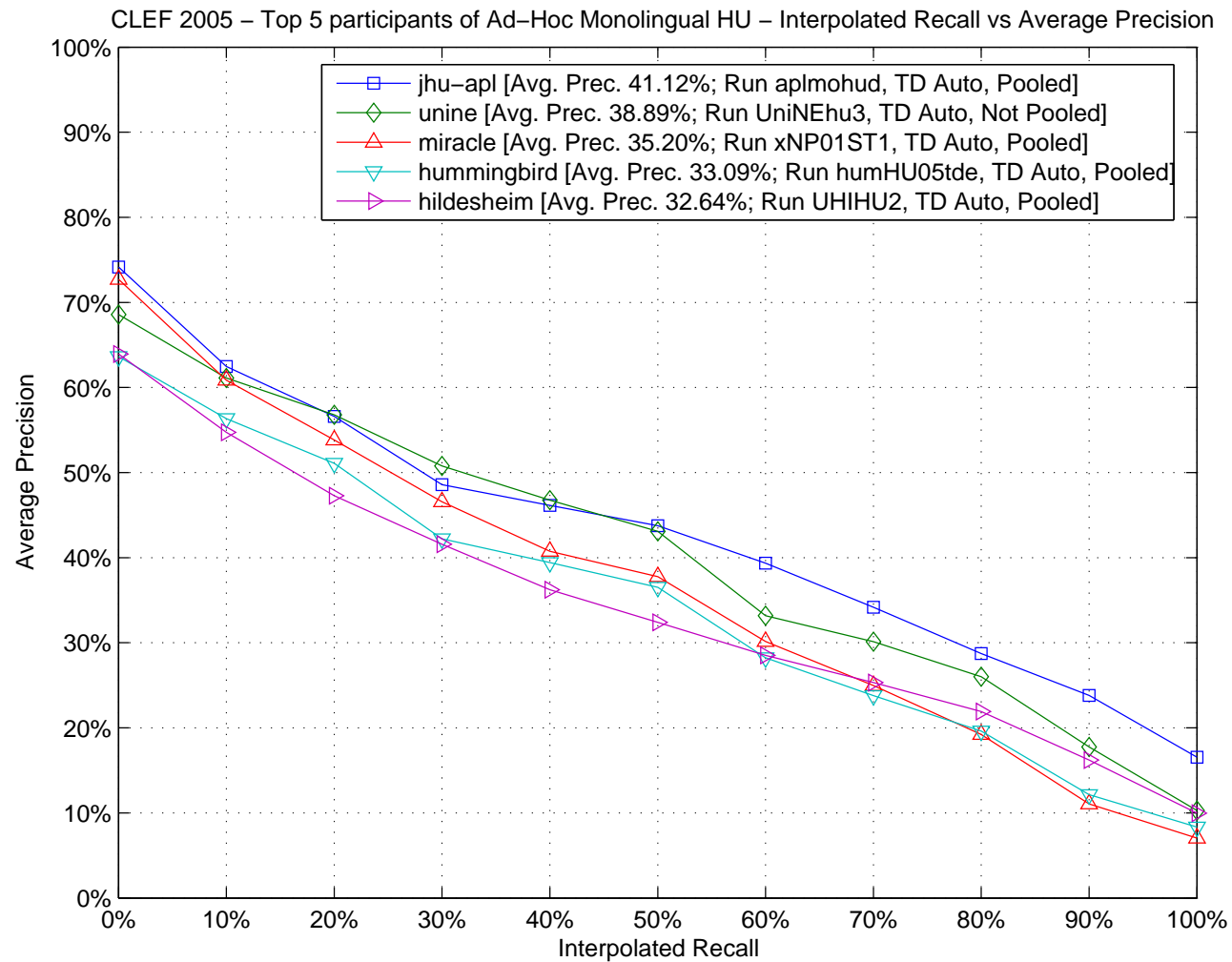
Monolingual BG



Monolingual FR



Monolingual HU



Monolingual PT

