PASCAL
Pattern Analysis, Statistical Modelling and Computational Learning

Competition 2 – Information Retrieval

Mikko Kurimo, Sami Virpioja and Ville T. Turunen
Helsinki University of Technology, Finland

Overview

- Algorithms tested in an Information Retrieval task
  - English, German and Finnish
- Words in the documents and queries were replaced by the suggested segmentations
- If participant did not submit segmentations for the Competition 2 word list, shorter Competition 1 word list was used
  - Other words left unsegmented

Example

- Query: `Französische Atomtests`
- Doc 1: `Ein zweiter französischer Atomtest fand mit 15-20 kt Sprengkraft...`
- Doc 2: `Heim ist nicht automatisch ein gutes Heim...`

Example: Method A

- Query: `französisch +e atom test +s`
- Doc 1: `ein zwei +t +er` **`französisch`** `+er` **`atom test`** `fand mit 15-20 kt spreng kraft...`
- Doc 2: `heim ist nicht automat isch ein gut +es heim...`

Example: Method B

- Query: `fran zö sische a tom tes ts`
- Doc 1: `ein z weiter` **`fran zö`** `sischer` **`a tom`** `test fand mit 15-20 kt spr eng kraf t...`
- Doc 2: `heim ist nicht au` **`tom a`** `tisch ein gu` **`tes`** `heim...`

Setup

- LEMUR-toolkit: http://www.lemurproject.org/
- Okapi BM25 ranking
- Stoplist for the most common morphemes
  - a fixed threshold for corpus frequency

Evaluation

- Mean Average Precision (MAP)

Rank	Doc. ID	Relevant?	Recall	Precision	Total Rel.
					6
1	17		0	0	
2	22	YES	0.17	0.5	
3	4		0.17	0.33	
4	37		0.17	0.25	
5	34	YES	0.33	0.4	
6	8		0.33	0.33	
7	14		0.33	0.29	
8	43		0.33	0.25	
9	18		0.33	0.22	
10	31		0.33	0.2	

Evaluation

- Mean Average Precision (MAP)

Rank	Doc. ID	Relevant?	Recall	Precision	Total Rel.
					6
1	17		0	0	
2	22	YES	0.17	0.5	
3	4		0.17	0.33	
4	37		0.17	0.25	
5	34	YES	0.33	0.4	
6	8		0.33	0.33	
7	14		0.33	0.29	
8	43		0.33	0.25	
9	18		0.33	0.22	
10	31		0.33	0.2	

- AP = (0.5 + 0.4) / 6 = 0.15
- Mean over all queries

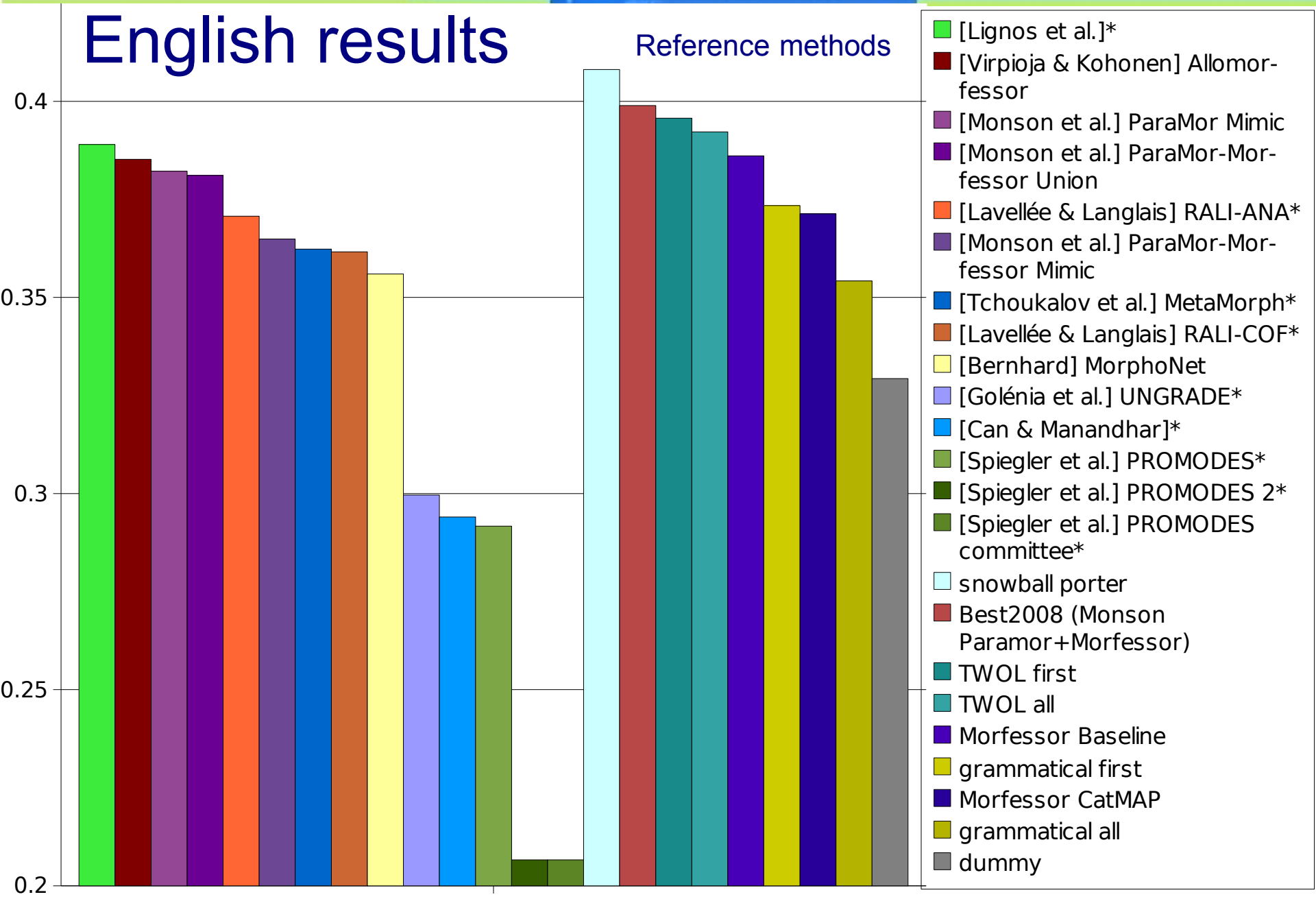
IR data sets (same as in 2007-2008)

- **Finnish (CLEF 2004)**
  - **55K documents from articles in Aamulehti 1994-95**
  - **50 test queries, 23K binary relevance assessments**
- **English (CLEF 2005)**
  - **107K documents from articles in Los Angeles Times 1994 and Glasgow Herald 1995**
  - **50 test queries, 20K binary relevance assessments**
- **German (CLEF 2003)**
  - **300K documents from short articles in Frankfurter Rundschau 1994, Der Spiegel 1994-95 and SDA German 1994-95**
  - **60 test queries, 23K binary relevance assessments**

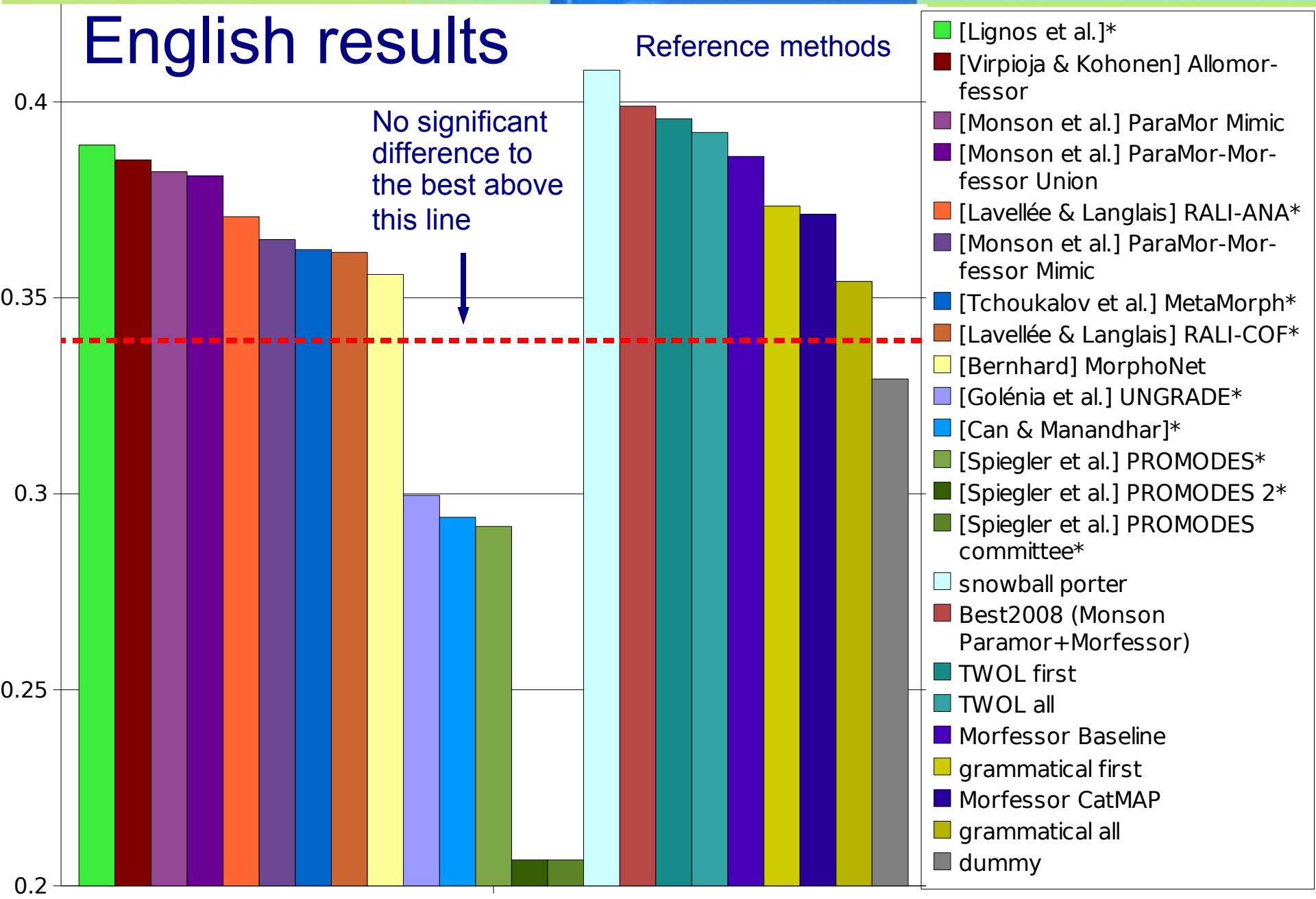
Reference methods

- **Morfessor Baseline:** our public code since 2002
- **Morfessor Categories-MAP:** improved, public 2006
- **dummy**: no segmentation, all words unsplit
- **grammatical**: full gold standard segmentation (reference of Competition 1)
  - all: all alternative segmentations included
  - first: only the first alternative chosen
- **TWOL**: word normalization by a commercial rule-based morphological analyzer (all & first)
- **Snowball**: Language specific stemming

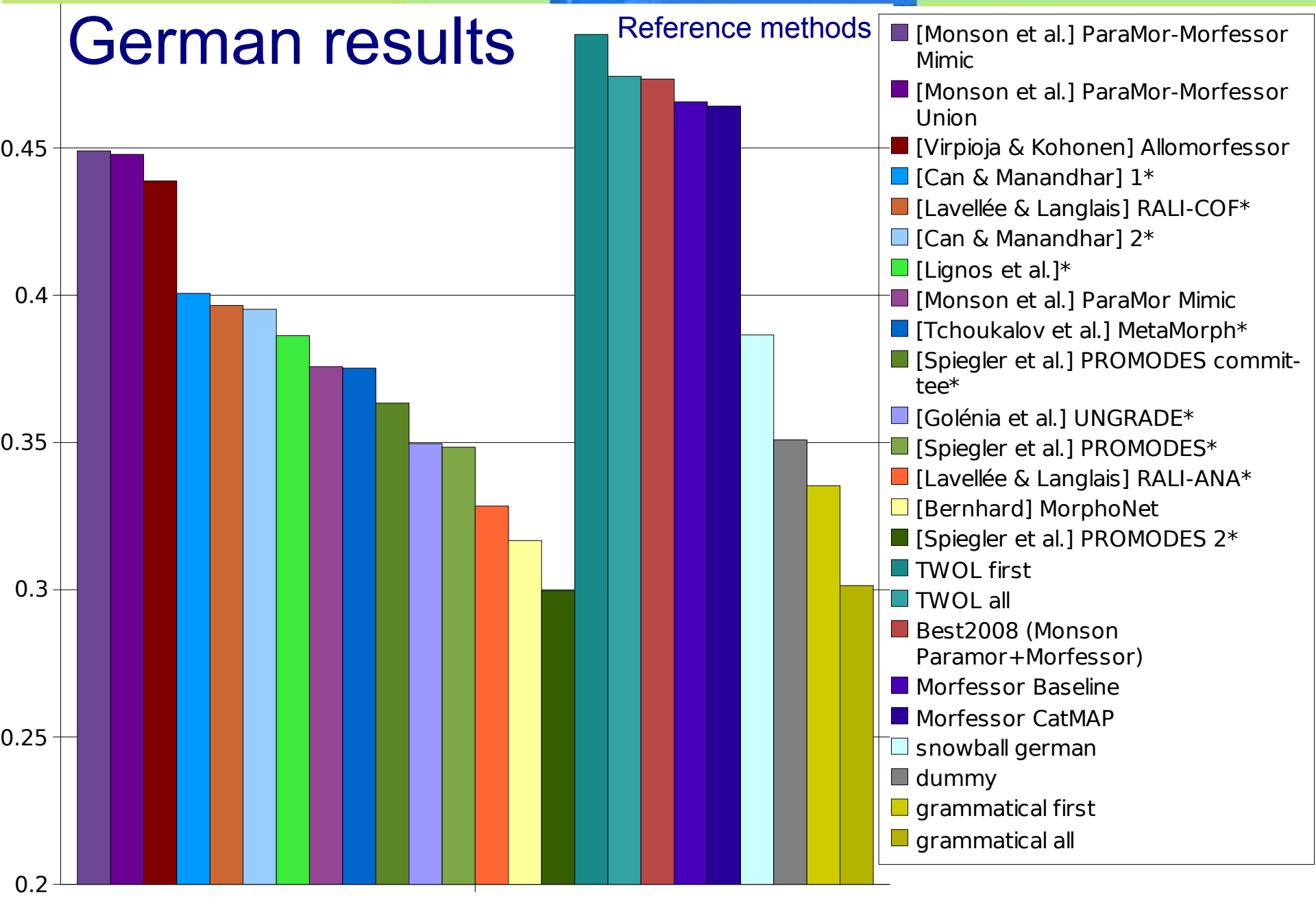
English results



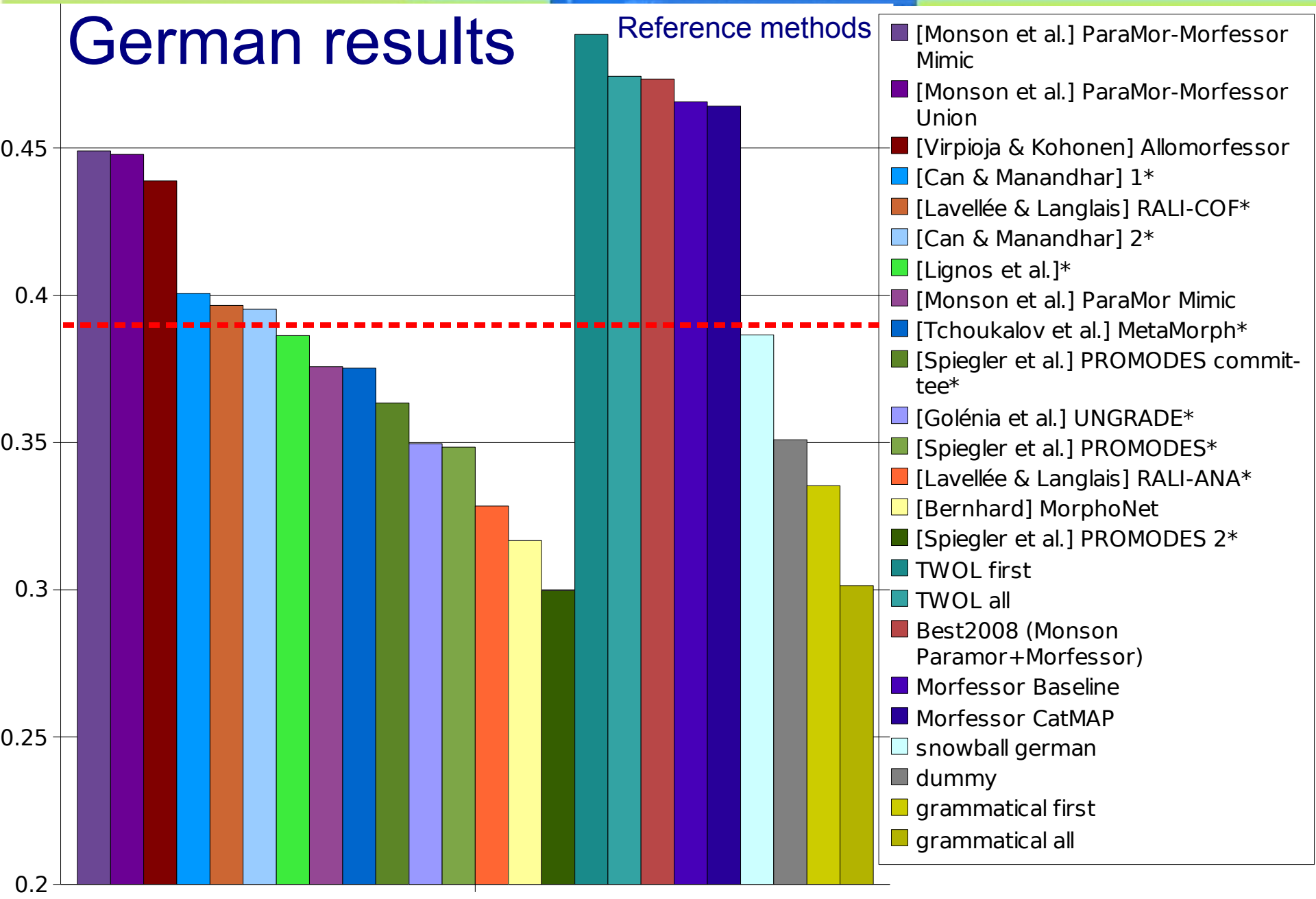
English results



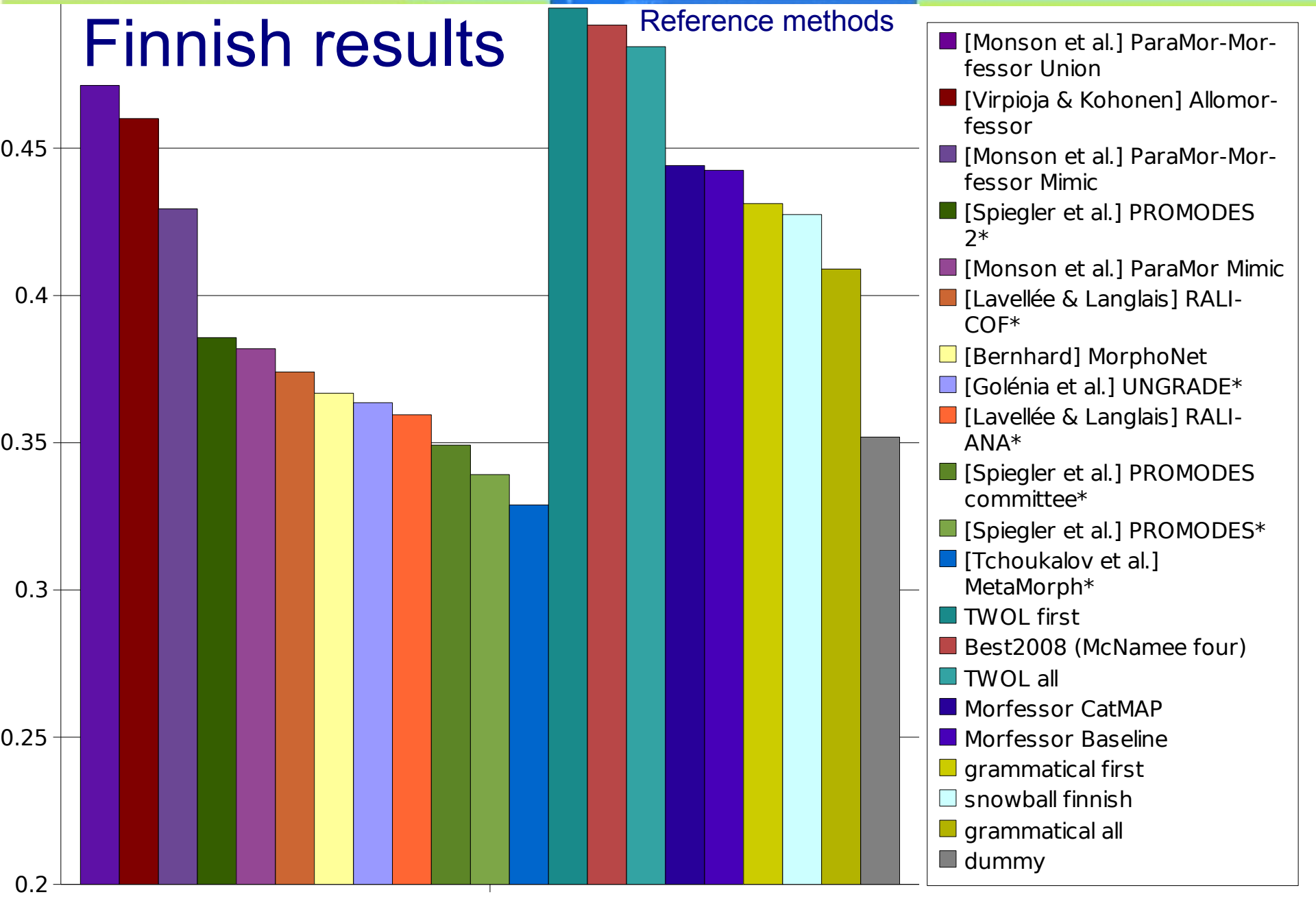
German results



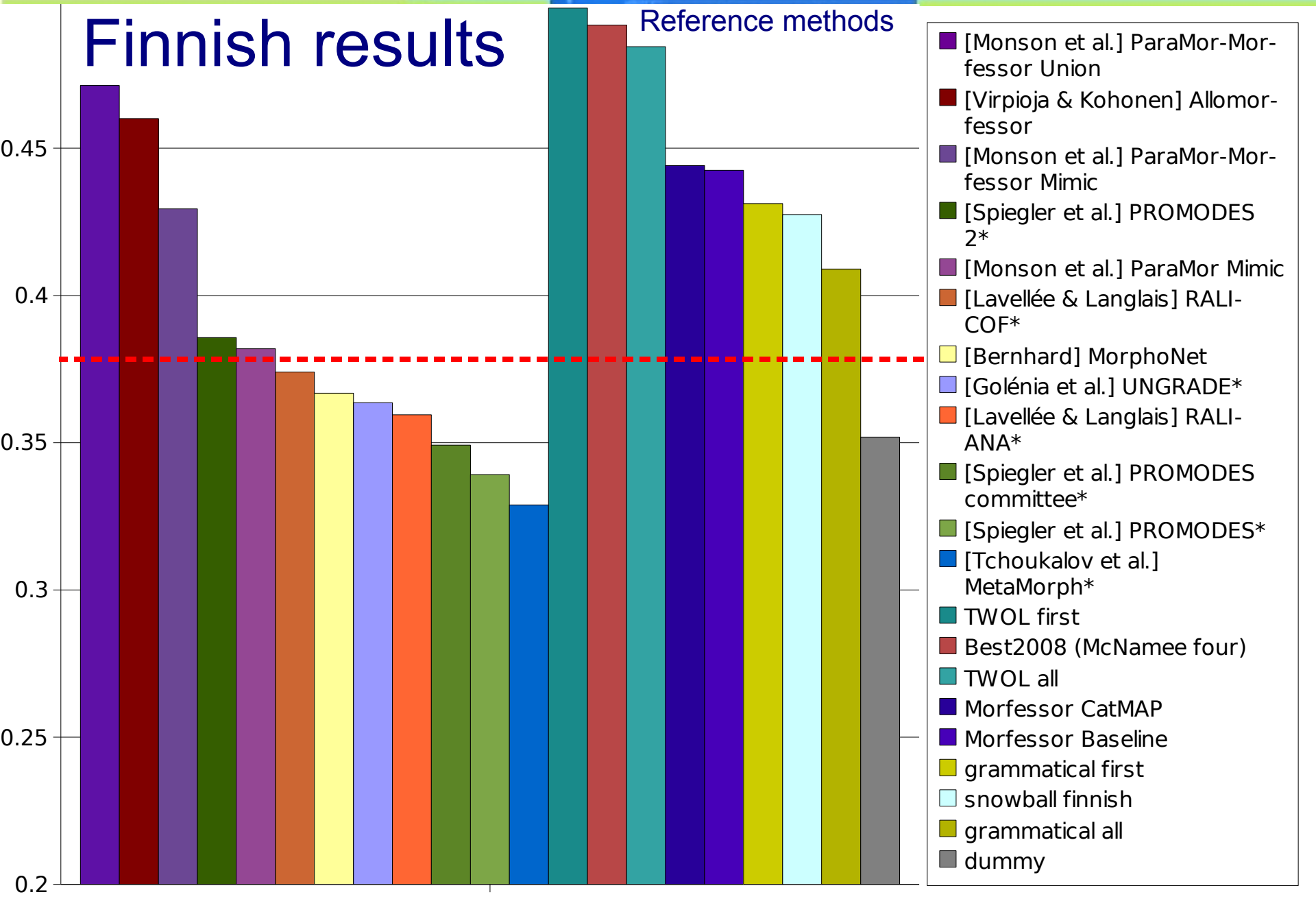
German results



Finnish results



Finnish results



Discussion

- Results not improved from last year
- Hard to achieve statistically significant differences
- No clear winner
- Strong in all languages:
  - “ParaMor-Morfessor Union” & ”Mimic”
  - ”Allomorfessor”
- Full word list not submitted by all participants
  - Comparison bit more difficult

Conclusions

- IR evaluations for 3 languages (out of 5)
- Good results in all languages by several algorithms

  => Unsupervised morphological analysis is a viable approach for IR
- Full report and papers in the CLEF proceedings
- Details, presentations, links, info at: http://www.cis.hut.fi/morphochallenge2009/