

# ParaMor & Morpho Challenge 2008

**Christian Monson**

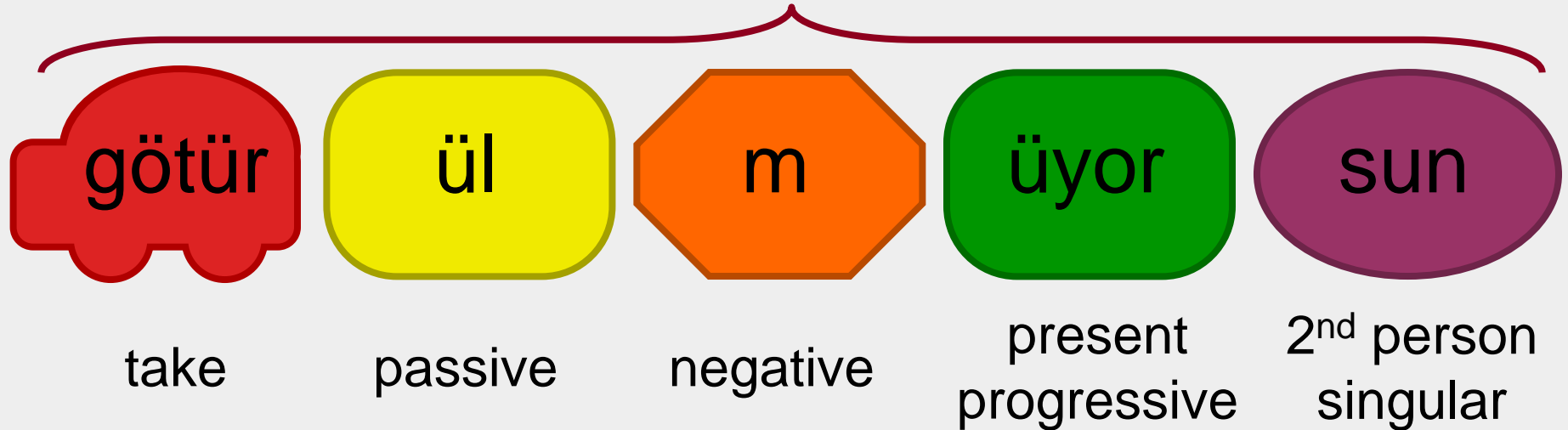


Jaime Carbonell, Alon Lavie, Lori Levin



# Turkish Morphology – Beads on a String

One Turkish Word



**You are not being taken**

# Computational Morphology Improves:

## Machine Translation

Turkish-English (Oflazer, 2007)

Czech-English (Goldwater and McClosky, 2005)

## Information Retrieval

English, German, Finnish (Kurimo et al., 2008)

## Speech Recognition

Finnish (Creutz, 2006)

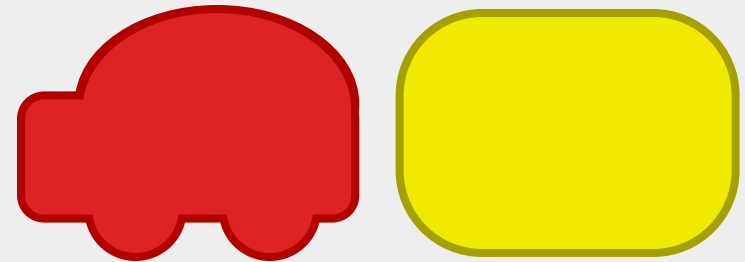
## Grapheme-to-Phoneme Conversion

German (Demberg, 2007)

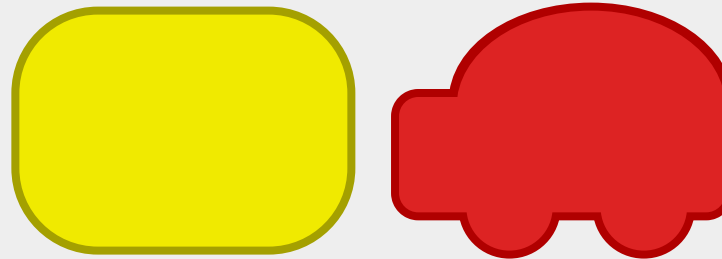
# Morphology is Complex – Operations



**Suffixation**



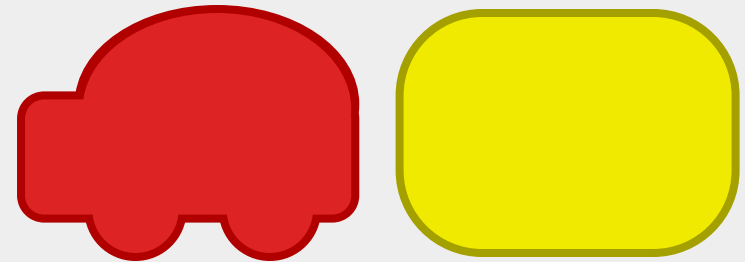
**Prefixation**



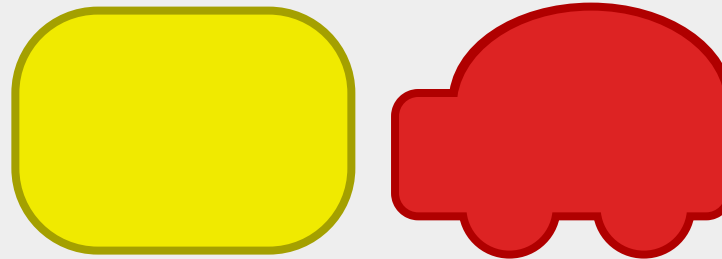
# Morphology is Complex – Operations



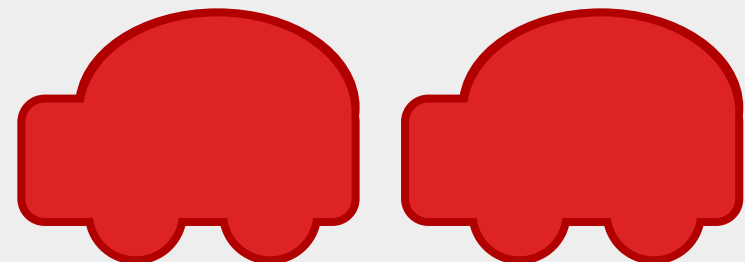
**Suffixation**



**Prefixation**



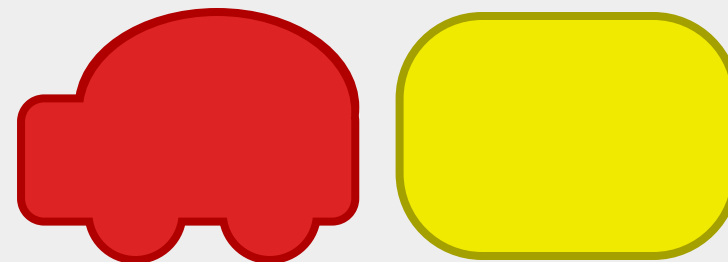
**Reduplication**



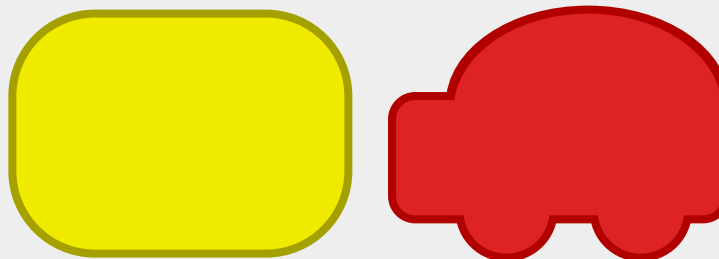
# Morphology is Complex – Operations



**Suffixation**



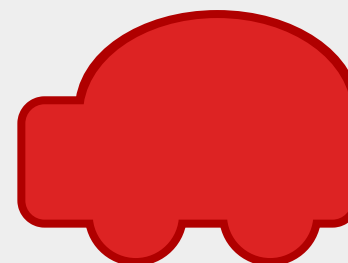
**Prefixation**



**Reduplication**



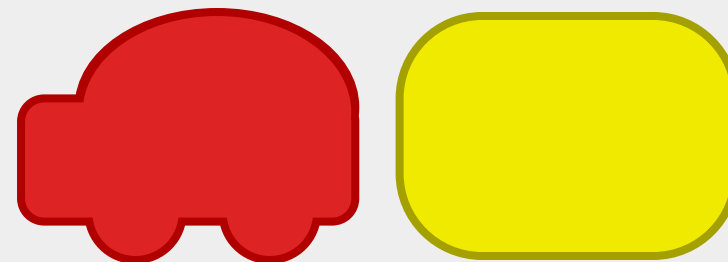
**Infixation**



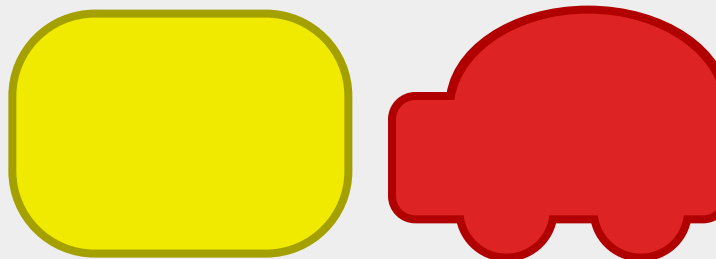
# Morphology is Complex – Operations



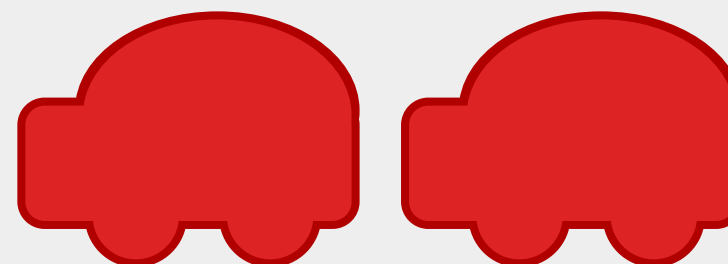
**Suffixation**



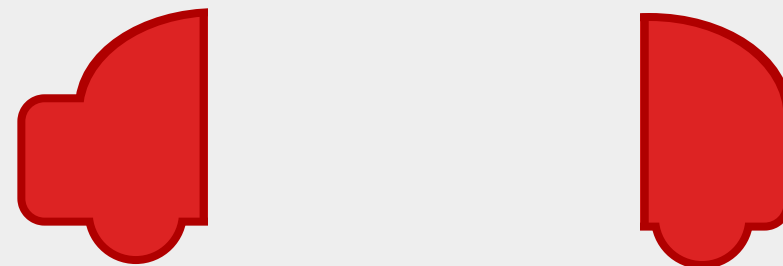
**Prefixation**



**Reduplication**



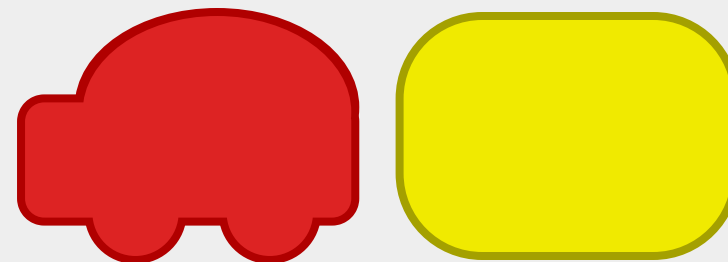
**Infixation**



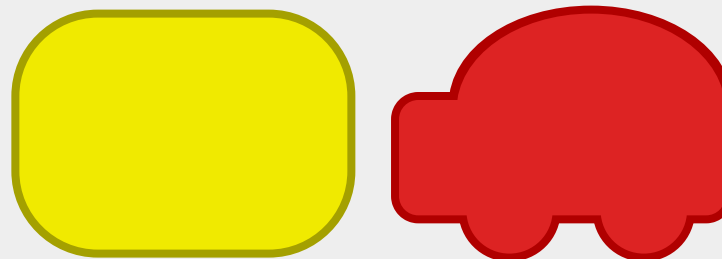
# Morphology is Complex – Operations



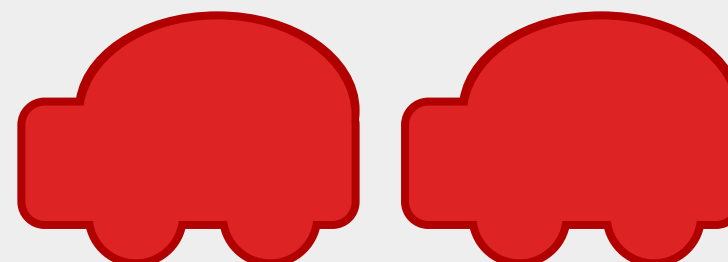
**Suffixation**



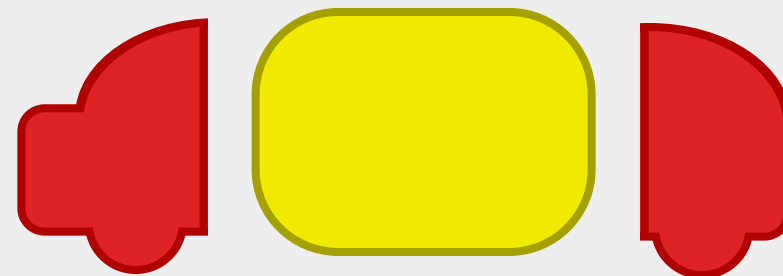
**Prefixation**



**Reduplication**

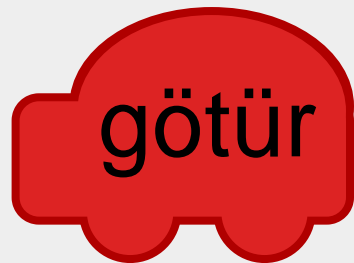


**Infixation**





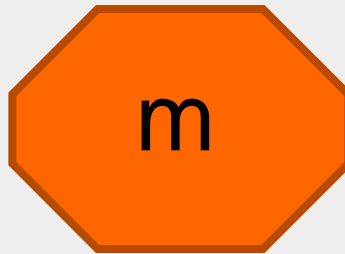
# Morphology is Complex – Morphophonology



take



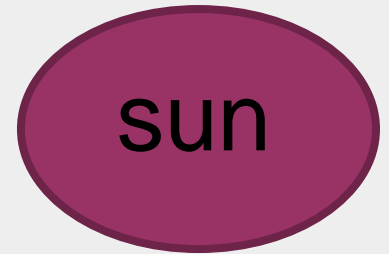
passive



negative



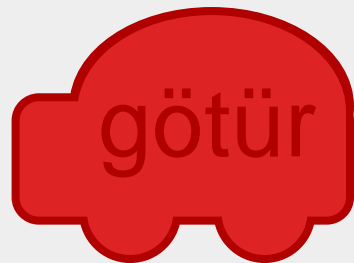
present  
progressive



2<sup>nd</sup> person  
singular

**You are not being taken**

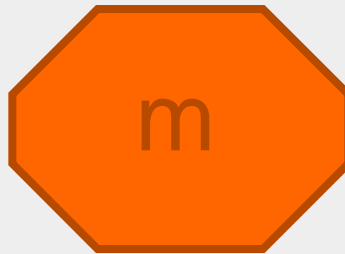
# Morphology is Complex – Morphophonology



take



passive



negative



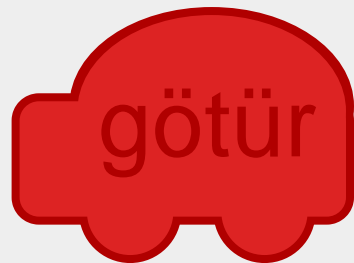
future



2<sup>nd</sup> person  
singular

You will not be taken

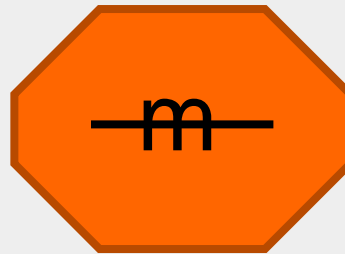
# Morphology is Complex – Morphophonology



take



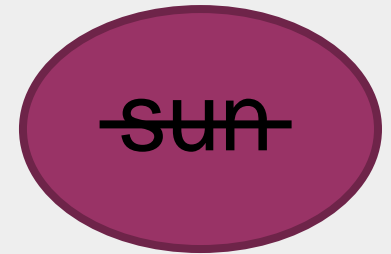
passive



negative



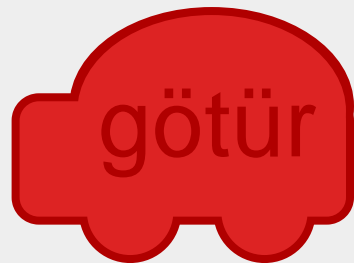
future



2<sup>nd</sup> person  
singular

**You will not be taken**

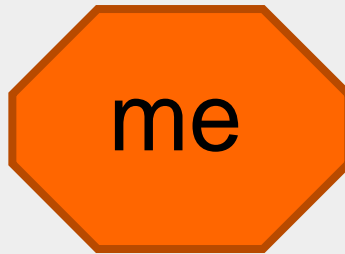
# Morphology is Complex – Morphophonology



take



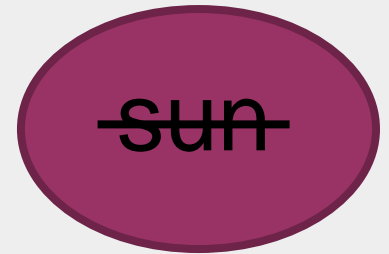
passive



negative



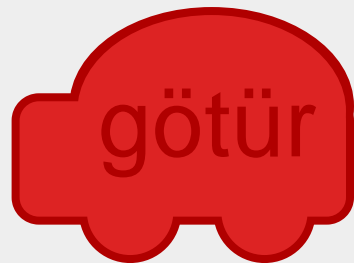
future



2<sup>nd</sup> person  
singular

**You will not be taken**

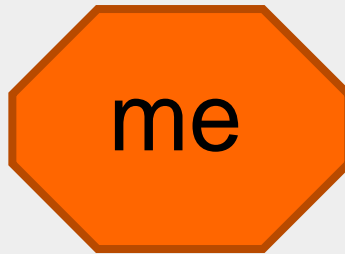
# Morphology is Complex – Morphophonology



take



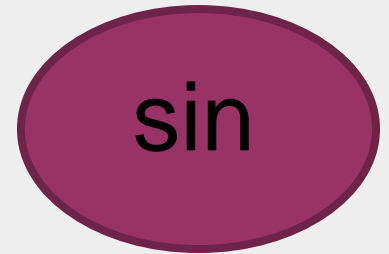
passive



negative



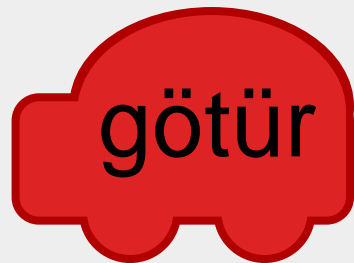
future



2<sup>nd</sup> person  
singular

**You will not be taken**

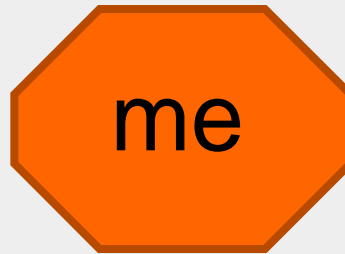
# Morphology is Complex – Morphophonology



take



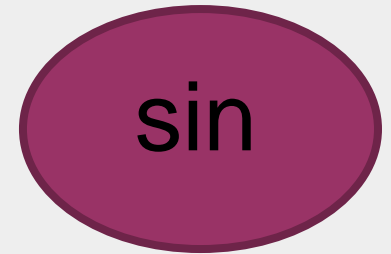
passive



negative



future



2<sup>nd</sup> person  
singular

**You will not be taken**

## Hungarian

mentek

men +tek

go +Present.2<sup>nd</sup>.Plural

‘yinz go’

## Hungarian

mentek

men +tek

go +Present.2<sup>nd</sup>.Plural

‘yinz go’

men +t

+ek

go +PastParticiple +Plural

‘those who have gone’



# In Morphology Systems for New Languages

**Complexity → Time + Expertise**

# In Morphology Systems for New Languages

**Complexity → Time + Expertise**

**Kemal Oflazer**

**E**xpert on

**T**urkish

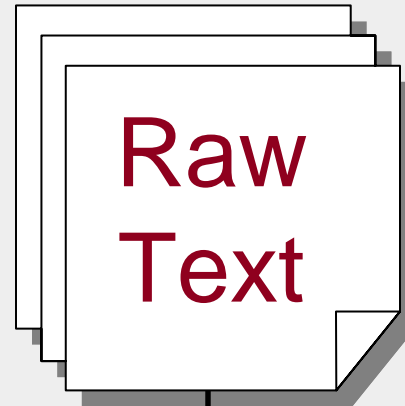
**C**omputational morphology

**T**ime

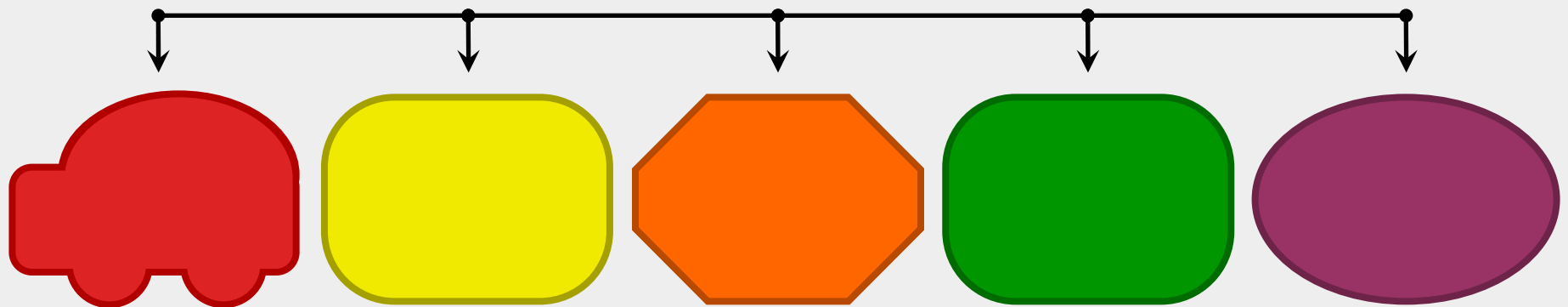
**3 - 4 M**onths to manually build a basic Turkish analyzer

**P**lus lexicon development and maintenance

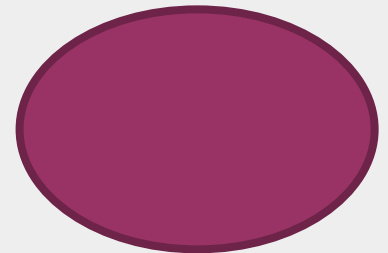
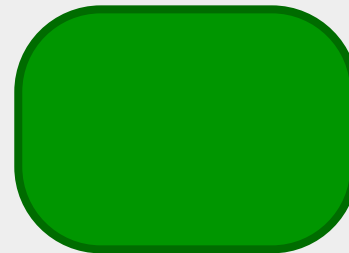
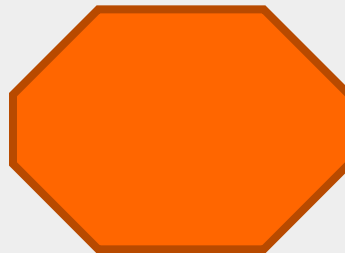
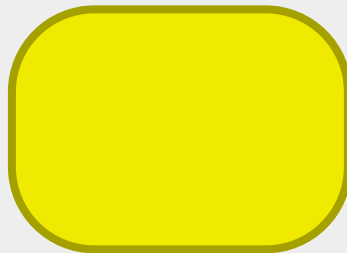
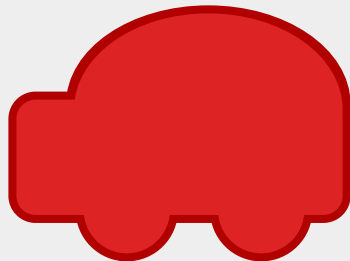
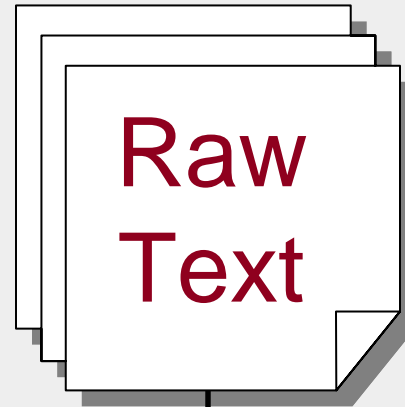
# The Solution



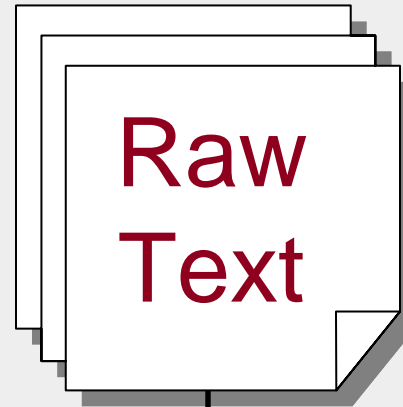
Unsupervised  
Morphology  
Induction



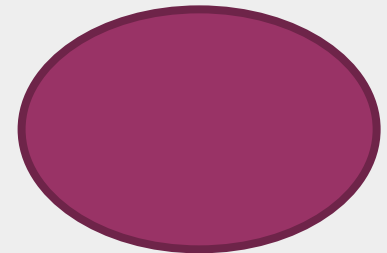
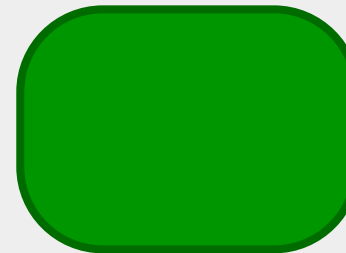
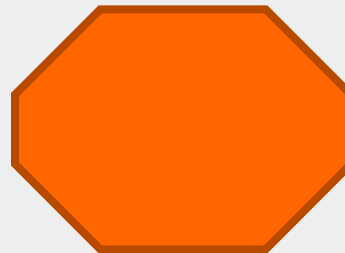
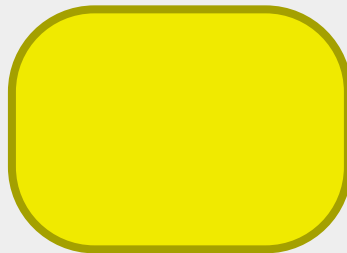
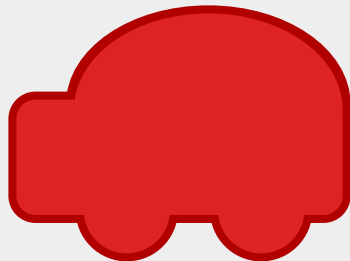
# The Solution



# The Solution



Language  
Structure



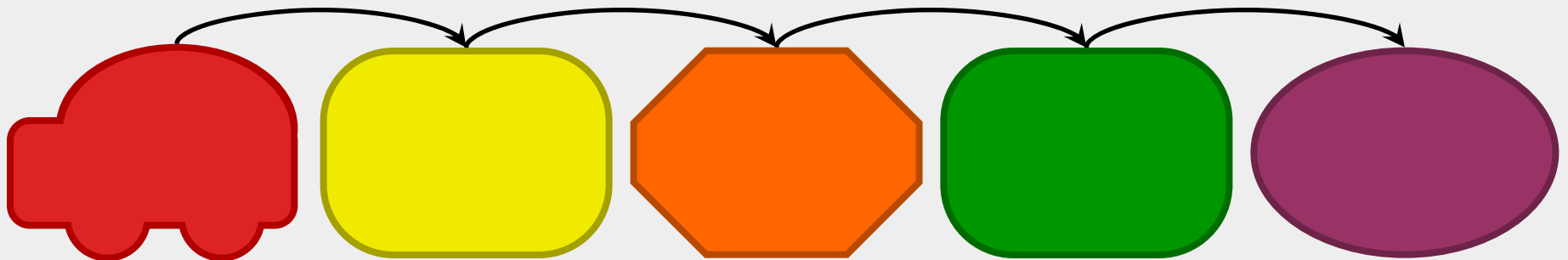
# Techniques for Unsupervised Morphology Induction



## Transition Likelihood

**H**arris (1955) – Finite State Automata

**B**ernhard (2007)



# Techniques for Unsupervised Morphology Induction



## Transition Likelihood

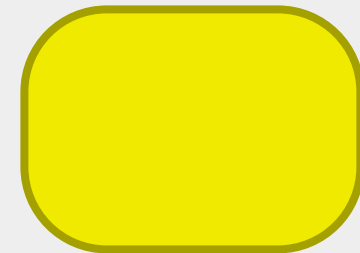
Harris (1955) – Finite State Automata

Bernhard (2007)

## Minimum Description Length

**G**oldsmith (2001, 2006)

**C**reutz's Morfessor (2006)



# Techniques for Unsupervised Morphology Induction



## Contextual Similarity

**W**icentowski (2002)

**S**chone (2002)





# Techniques for Unsupervised Morphology Induction



## Contextual Similarity

**W**icentowski (2002)

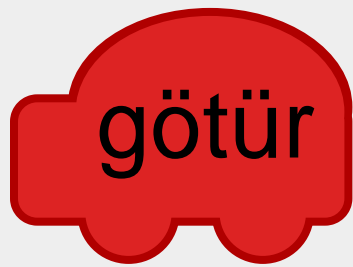
**S**chone (2002)

## The Paradigm

**S**nover (2002)

**P**araMor (2007)

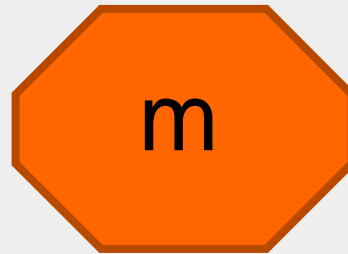
# What is a Paradigm?



take



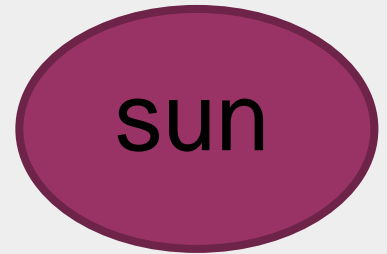
passive



negative

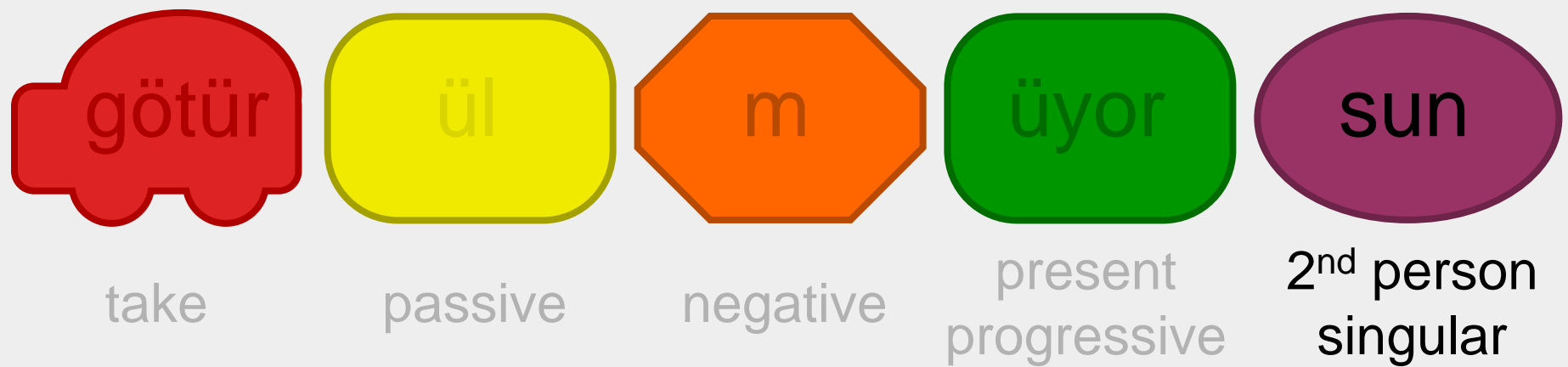


present  
progressive

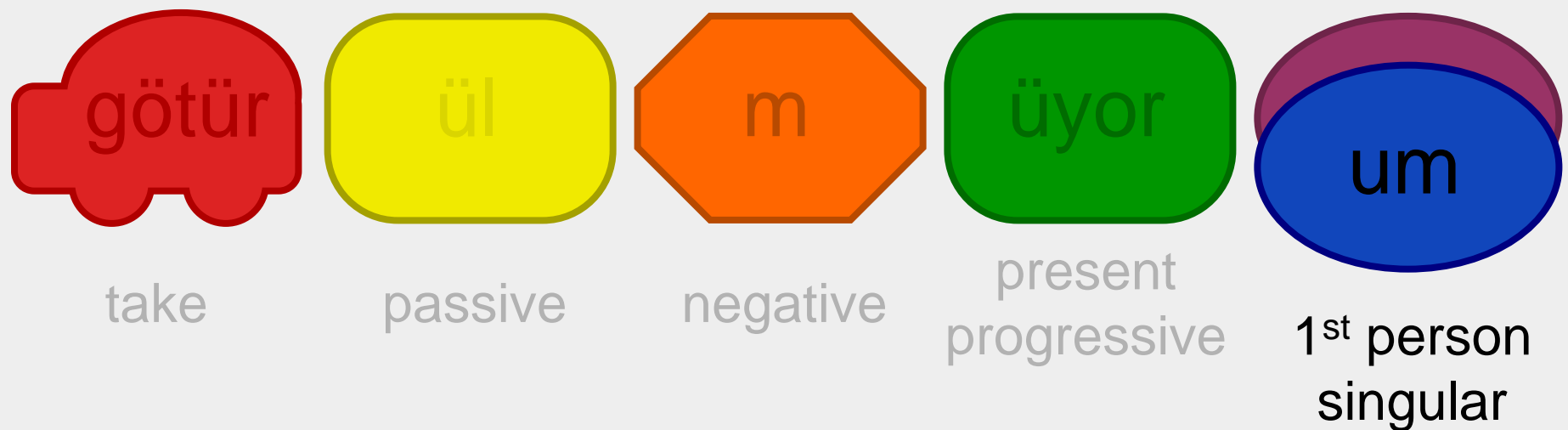


2<sup>nd</sup> person  
singular

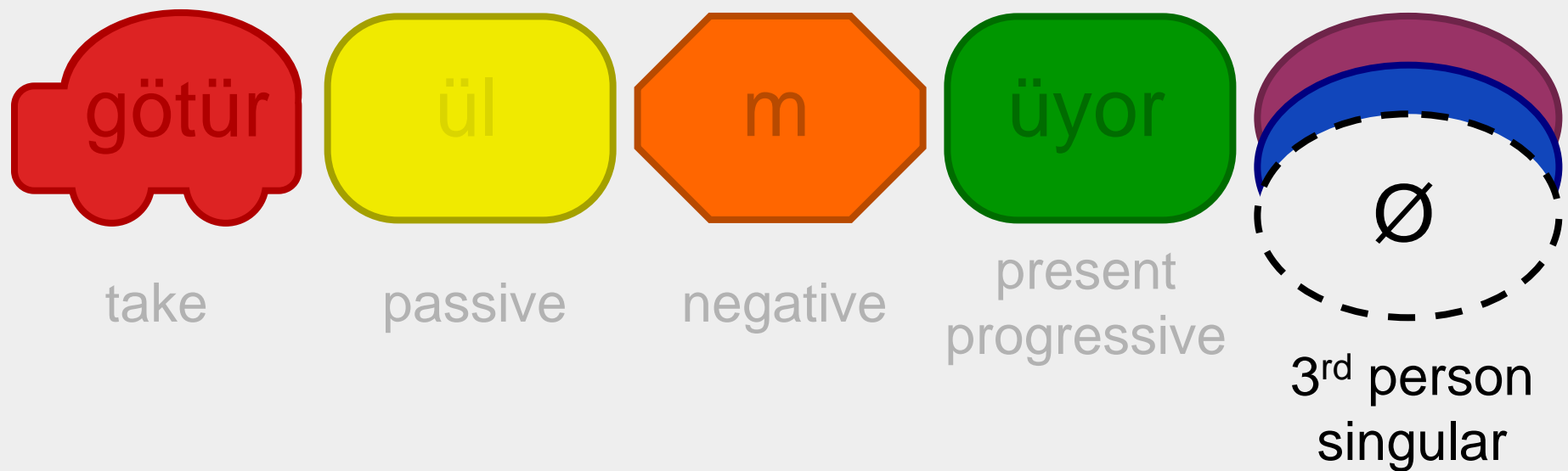
# Paradigms Structure Inflectional Morphology



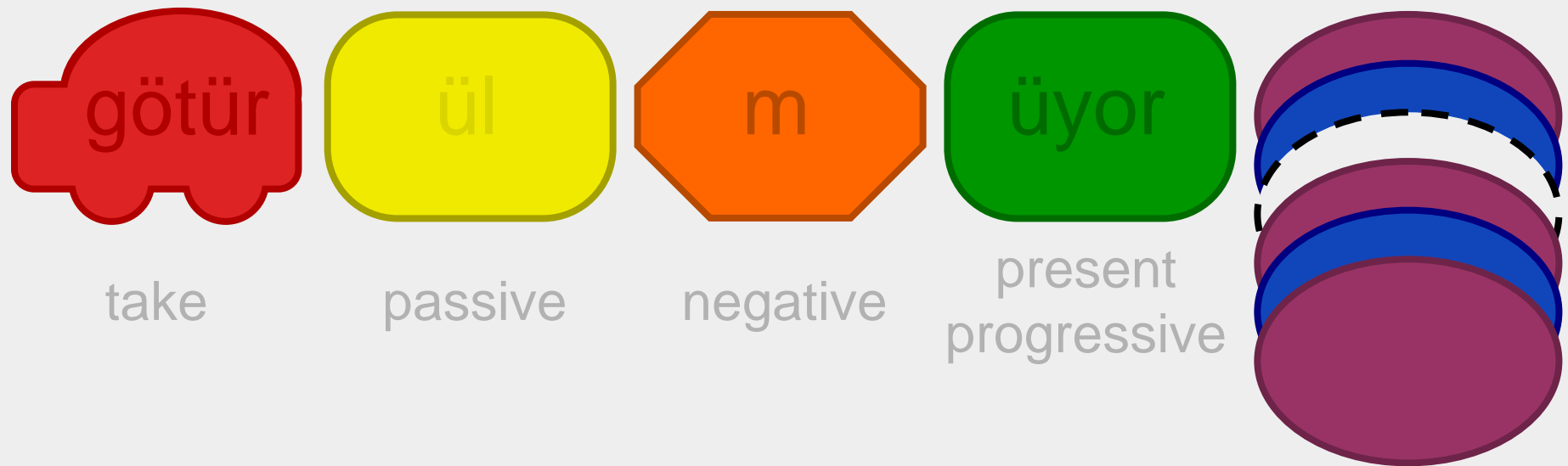
# Paradigms Structure Inflectional Morphology



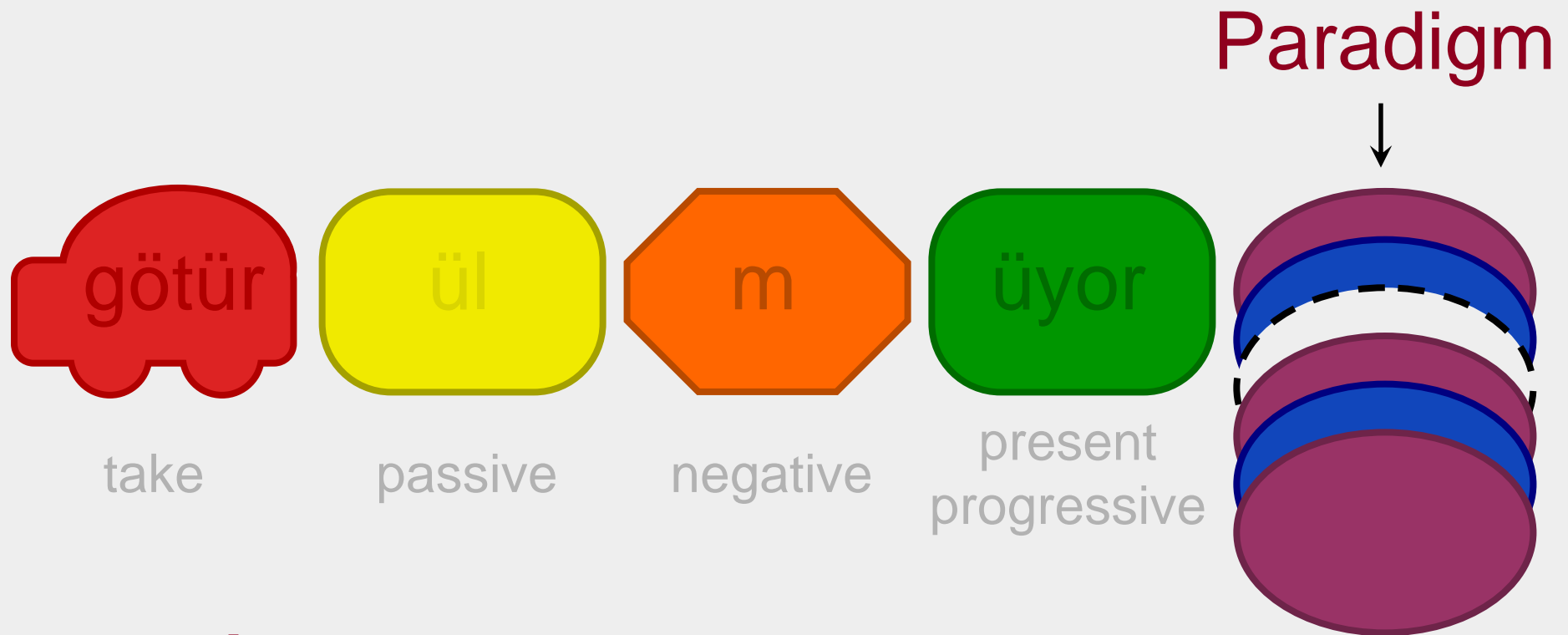
# Paradigms Structure Inflectional Morphology



# Paradigms Structure Inflectional Morphology



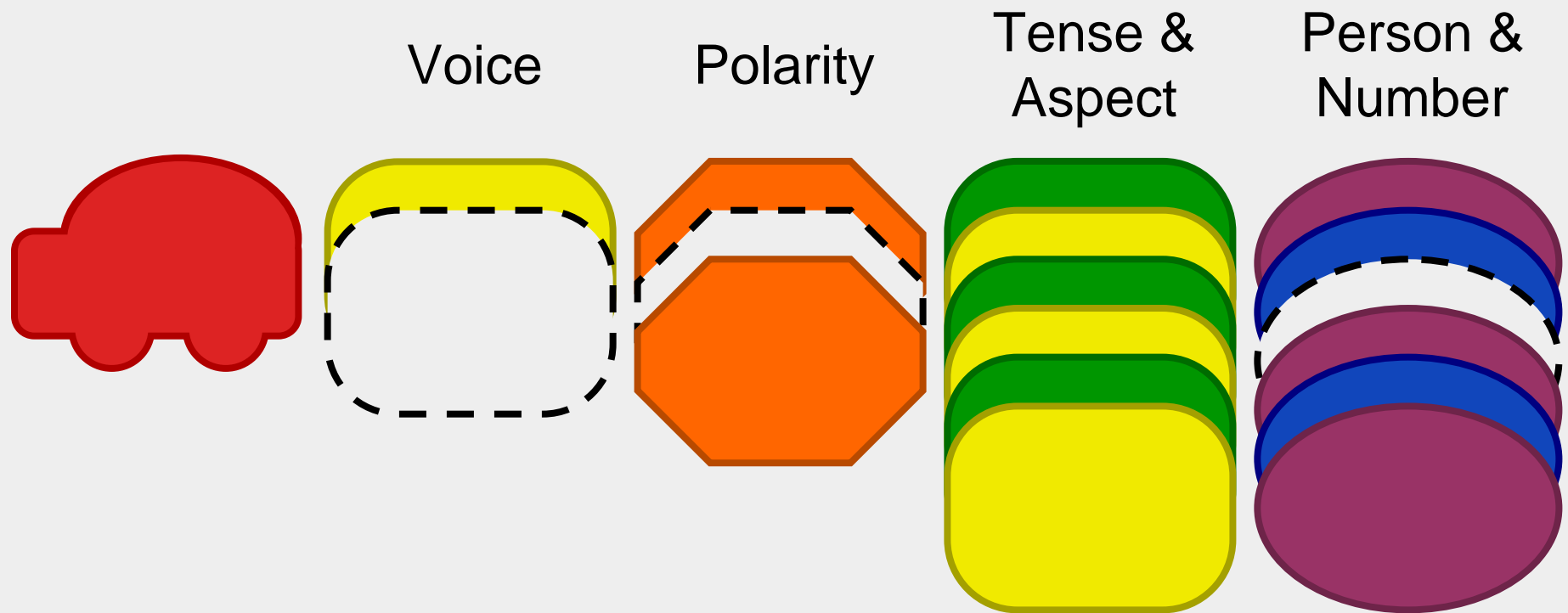
# Paradigms Structure Inflectional Morphology



## Paradigm

**M**utually substitutable morphological operations

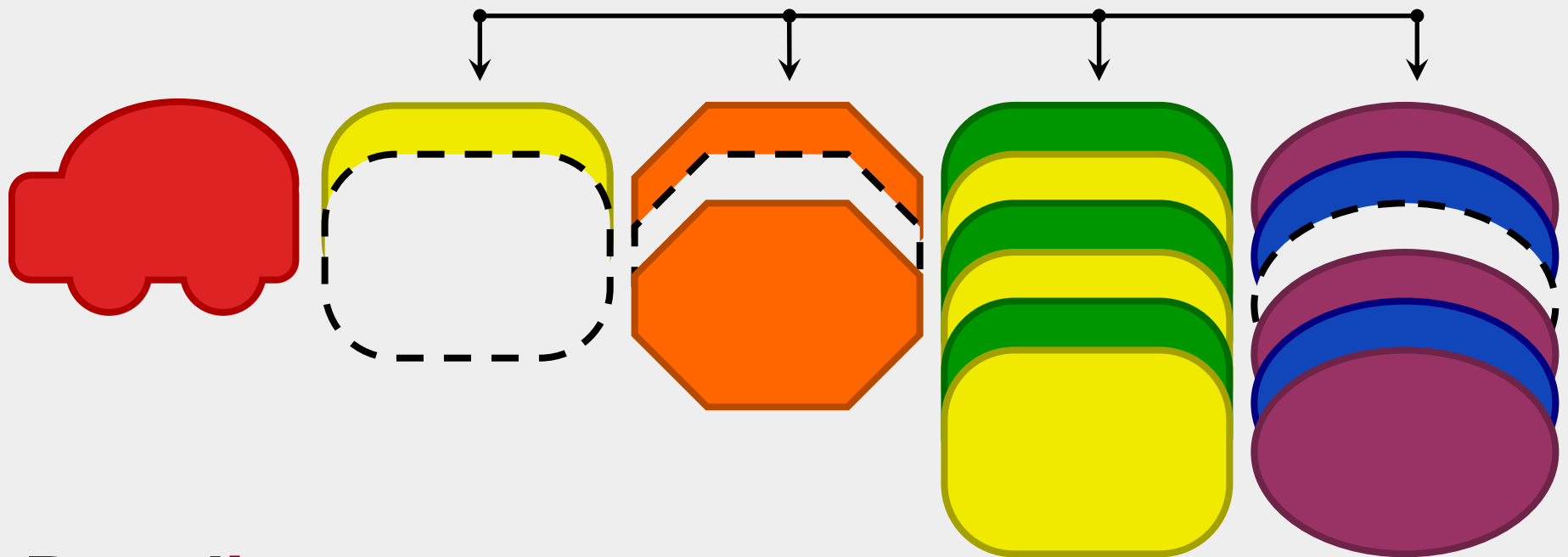
# Paradigms Structure Inflectional Morphology





# Paradigms Structure Inflectional Morphology

## Paradigms



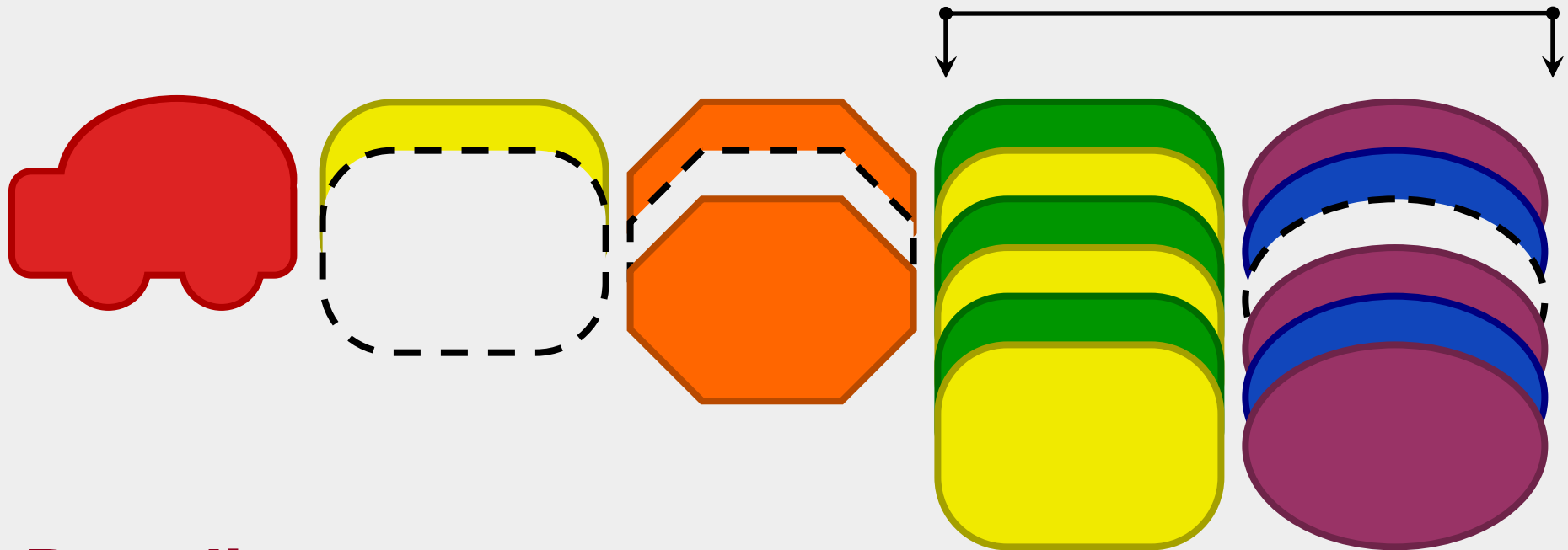
## Paradigm

**M**utually substitutable morphological operations

# The ParaMor Algorithm



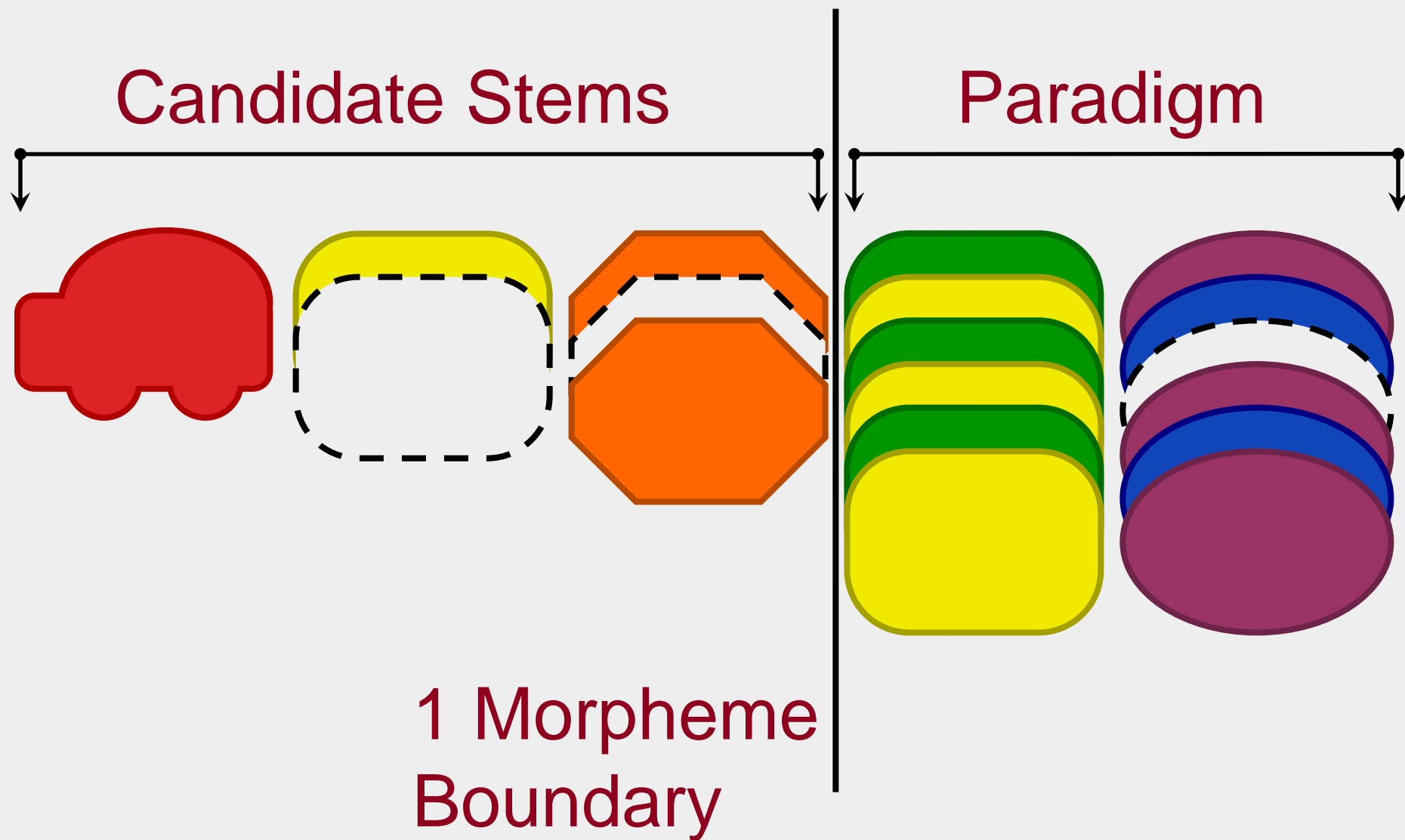
Paradigm



**Paradigm**

Mutually substitutable **strings**

# The ParaMor Algorithm



## Simplifying Assumptions

**S**uffixes only

70% of the World's Languages are  
Suffixing (Dryer, 2005)

**S**trict Concatenation

## Simplifying Assumptions

**S**uffixes only

70% of the World's Languages are  
Suffixing (Dryer, 2005)

**S**trict Concatenation

**Only a High-Level Overview**

# The ParaMor Algorithm



ParaMor  
Identify

## Identify Paradigms in 3 Steps

# The ParaMor Algorithm



ParaMor  
Identify  
Search

## Identify Paradigms in 3 Steps

1. **Search** for candidate paradigms

## Identify Paradigms in 3 Steps

1. **Search** for candidate paradigms
2. **Cluster** candidates modeling the same paradigm



## Identify Paradigms in 3 Steps

1. **Search** for candidate paradigms
2. **Cluster** candidates modeling the same paradigm
3. **Filter** least likely candidates

# The ParaMor Algorithm



**ParaMor**  
Identify  
Search  
Cluster  
Filter  
**Segment**

## Identify Paradigms in 3 Steps

1. **Search** for candidate paradigms
2. **Cluster** candidates modeling the same paradigm
3. **Filter** least likely candidates

## Segment Words

Using the discovered paradigms

# The ParaMor Algorithm



ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
Results

## Identify Paradigms in 3 Steps

1. **Search** for candidate paradigms
2. **Cluster** candidate paradigms  
same paradigm
3. **Filter**

**Today**

## Segment Words

Using the discovered paradigms

# The ParaMor Algorithm



ParaMor  
Identify  
**Search**  
Cluster  
Filter  
Segment  
Evaluation  
Results

## Identify Paradigms in 3 Steps

1. **Search** for candidate paradigms
2. **Cluster** candidates modeling the same paradigm
3. **Filter**

## Segment Words

Using the discovered paradigms

# Search for Candidate Paradigms



ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
Results

**Propose** a morpheme boundary at every character boundary in every word

**Consolidate** identical candidate suffixes into paradigm seeds

## Spanish Example

**Word List**  
50,000 Types

autorizaciones  
buscábamos  
costas  
importadoras  
vallas  
...

**s**  
10697

# Search for Candidate Paradigms



ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
Results

## Identify the most frequent mutually replaceable candidate suffix

**S**tems that occur with one suffix in a paradigm will likely occur with other suffixes in that paradigm

### Spanish Example

autorizaciones	
buscábamos	
costa <del>Ø</del>	costas
importadora <del>Ø</del>	importadoras
valla <del>Ø</del>	vallas
...	

**Ø** **s**  
5513

↑  
**s**  
10697

# Search for Candidate Paradigms



ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
Results

## A Parameter halts the introduction of suffixes

**W**hen the most frequent mutually replaceable candidate suffix severely decreases the stem count

autorizaciones  
buscabamos  
costar      costa~~Ø~~      costas  
importadora~~Ø~~ importadoras  
valla~~Ø~~      vallas  
...

Ø r s  
281

Ø s  
5513

s  
10697

# Search for Candidate Paradigms



ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
Results

Parameters set to produce

**H**igh-recall

**S**panish paradigms

**A**nd then **frozen**

costar      costa~~Ø~~      costas  
importadora~~Ø~~ importadoras  
valla~~Ø~~      vallas  
...  
autorizaciones  
buscabamos

Ø r s

281

Ø s

5513



**s**

10697



# Search for Candidate Paradigms



ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
Results

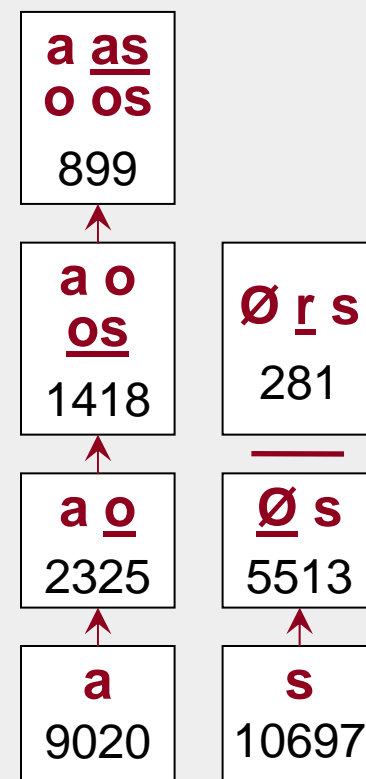
Move on to the next most  
frequent paradigm seed

	<div>Ø <u>r</u> s 281</div>
	<div>Ø <u>s</u> 5513</div>
<div>a 9020</div>	<div>s 10697</div>

# Search for Candidate Paradigms



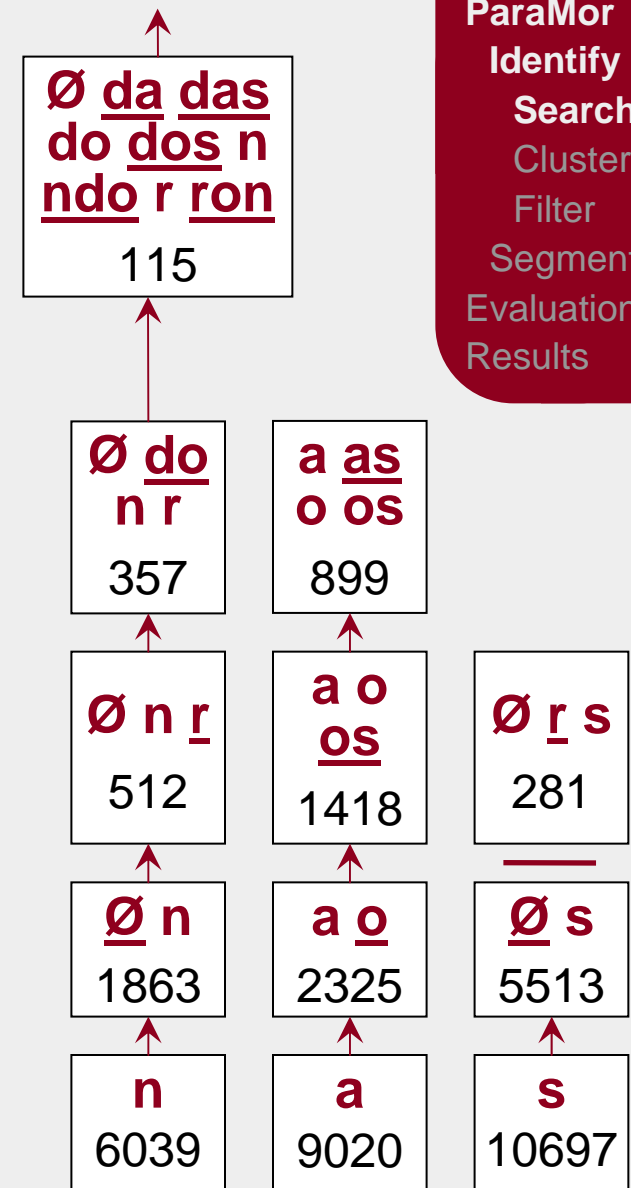
ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
Results



# Search for Candidate Paradigms



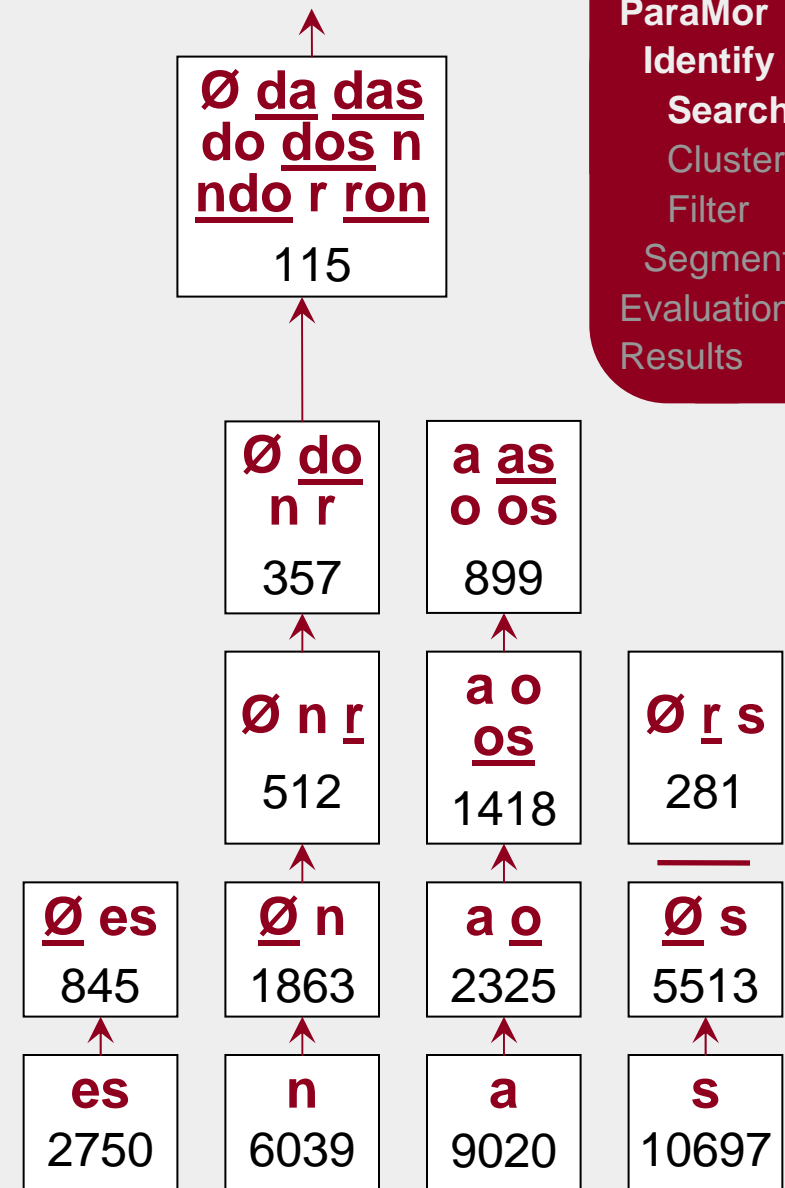
ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
Results



# Search for Candidate Paradigms



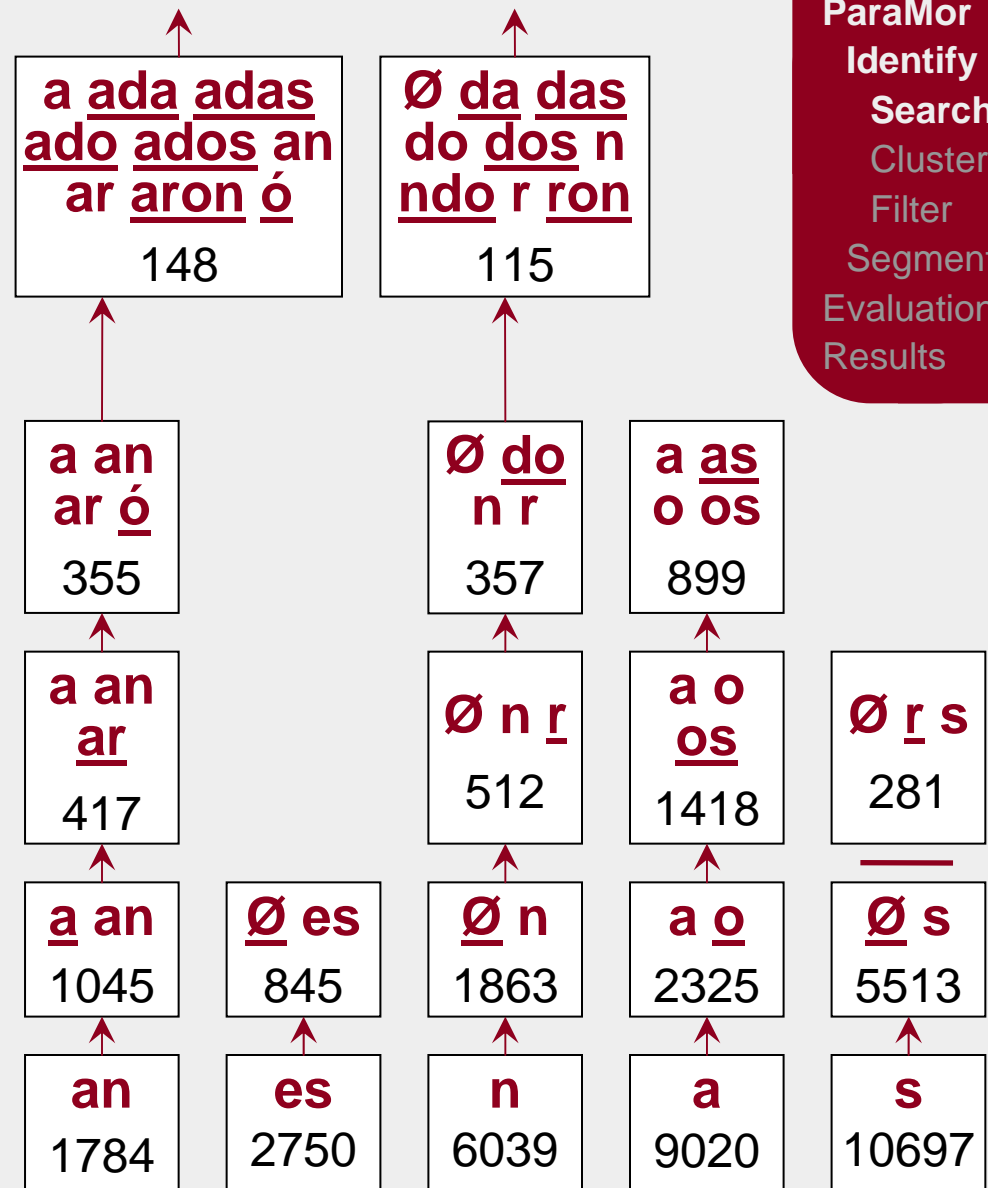
ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
Results



# Search for Candidate Paradigms



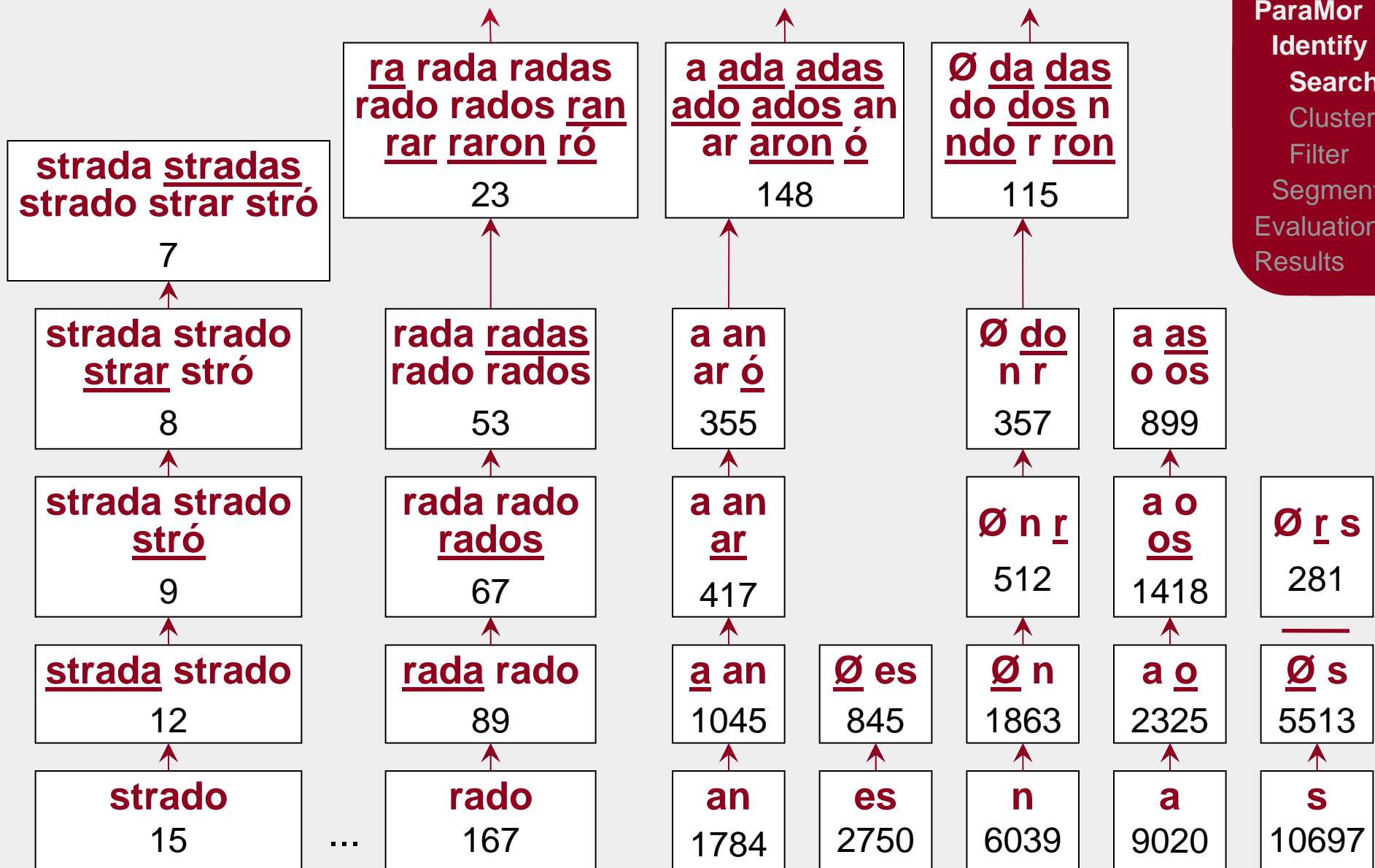
ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
Results



# Search for Candidate Paradigms



ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
Results



# Search for Candidate Paradigms



ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
Results

## Size of Search Space

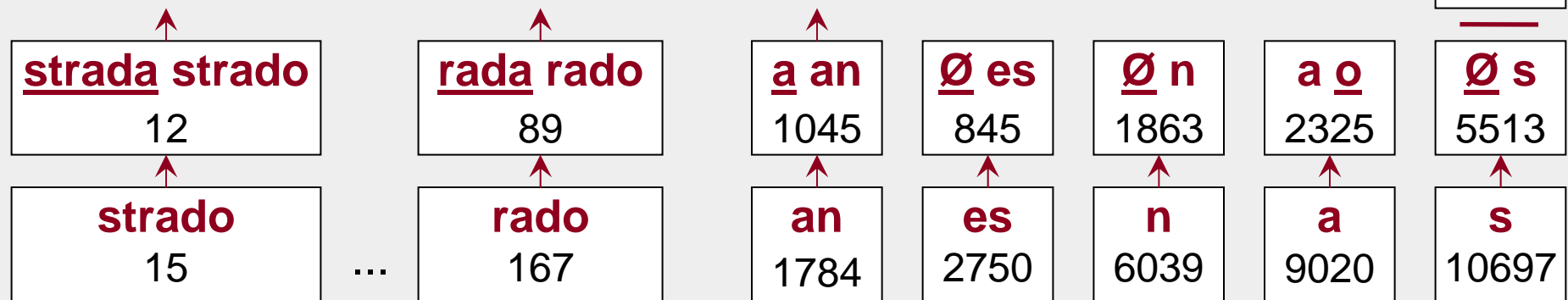
**H**uge:  $2^{| \text{candidate suffixes} |}$

**M**ost candidate suffixes have no common stems

**S**till Exponential

**G**reedily searched space:  $O(| \text{candidate suffixes} |)$

**T**his example is just 0.1% of the searched space



# Step 2: Clustering



ParaMor  
Identify  
Search  
**Cluster**  
Filter  
Segment  
Evaluation  
Results

## Identify Paradigms in 3 Steps

1. Search for candidate paradigms
2. **Cluster** candidates modeling the same paradigm
3. Filter

**Bottom-up  
Agglomerative  
Clustering**



## Segment Words

Using the discovered paradigms



# Step 3: Filtering



ParaMor  
Identify  
Search  
Cluster  
**Filter**  
Segment  
Evaluation  
Results

## Identify Paradigms in 3 Steps

1. **Search** for candidate paradigms
2. **Cluster** candidates modeling the same paradigm
3. **Filter** least likely candidates

## Segment Words

Using the discovered pa

**Adapted from**

Harris (1955) and  
Goldsmith (2006)

**Improved**

over 2007 Challenge

# A Few of the 42 Final Paradigms

**4 Suffixes** Ø menente mente s

**11 Suffixes** a amente as illa illas o or ora oras ores os

**41 Suffixes** a aba aban acion acciones acción ada adas ado ador  
adora adoras adores ados amos an ando ante antes ar  
ara aran aremos arla arlas arlo arlos arme aron arse ará  
arán aré aría arían ase e en ándose é ó

**29 Suffixes** e edor edora edoras edores en er erlo erlos erse erá  
erán ería erían ida idas ido idos iendo iera ieran ieron  
imiento imientos iéndose ió í ía ían

**20 Suffixes** ida idas ido idor idores idos imos ir iremos irle irlo irlos  
irse irá irán iré iría irían ía ían

**29 Suffixes** ce cedores cemos cen cer cerlo cerlos cerse cerá cerán  
cería cida cidas cido cidos ciendo ciera cieran cieron  
cimiento cimientos cimos ció cí cía cían zca zcan zco

**6 Suffixes** Ø es idad idades mente ísima

# A Few of the 42 Final Paradigms

4 Suffixes **Ø** niente mente **s**

11 Suffixes a amente as illa illas o or ora oras ores os

41 Suffixes a aba aban acion acciones acción ada adas ado ador  
adora addoras adores ados amos an ando ante antes ar  
ara aran aremos arla arlas arlo arlos arme aron arse ará  
arán aré aría arían ase e en ándose é ó

29 Suffixes

Number on  
Nouns

20 Suffixes ida idas ido idor idores idos imos ir iremos irle irlo irlos  
irse irá iran iré iría irían ía ían

29 Suffixes ce cedores cemos cen cer cerlo cerlos cerse cerá cerán  
cería cida cidas cido cidos ciendo ciera cieran cieron  
cimiento cimientos cimos ció cí cía cían zca zcan zco

6 Suffixes **Ø es** idad idades mente ísima

# A Few of the 42 Final Paradigms

4 Suffixes Ø menente mente s

11 Suffixes **a** amente **as** illa illas **o** or ora oras ores **os**

41 Suffixes a aba aban acion acciones acción ada adas ado ador  
adoraadorasadoresadosamosan ando ante antes ar  
ara aran aremos arla arlas arlo arlos arme aron arse ará  
arán aré aría arían ase e en ándose é ó

29 Suffixes

## Number & Gender on Adjectives

20 Suffixes ida idas ido idor idores idos imos ir iremos irle irlo irlos  
irse irá irán iré iría irían ía ían

29 Suffixes ce cedores cemos cen cer cerlo cerlos cerse cerá cerán  
cería cida cidas cido cidos ciendo ciera cieran cieron  
cimiento cimientos cimos ció cí cía cían zca zcan zco

6 Suffixes Ø es idad idades mente ísima

# A Few of the 42 Final Paradigms

4 Suffixes Ø menente mente s

11 Suffixes a amente as illa illas o or ora oras ores os

**41 Suffixes** a aba aban acion acciones acción ada adas ado ador  
adora adoras adores ados amos an ando ante antes **ar**  
ara aran aremos arla arlas arlo arlos arme aron arse ará  
arán aré aría arían ase e en ándose é ó

**29 Suffixes** e edor edora edoras edores en **er** erlo erlos erse erá  
erán ería erían ida idas ido idos tendo iera ieran ieron  
imiento imientos iéndose ió í ía ían

**20 Suffixes** ida idas ido idor idores idos imos **ir** iremos irle irlo irlos  
irse irá irán iré iría irían ía ían

29 Suffixes

## Verbal Suffixes

6 Suffixes Ø es idad idades mente ísima

# The ParaMor Algorithm



**ParaMor**  
Identify  
Search  
Cluster  
Filter  
**Segment**  
Evaluation  
Results

## Identify Paradigms in 3 Steps

1. Search for candidate paradigms
2. Cluster candidates modeling the same paradigm
3. Filter

**Segment Words** ← **Improved** over 2007 Challenge  
Using the discovered paradigms

# Segment Words Using the Paradigms



ParaMor

Identify

Search

Cluster

Filter

**Segment**

Evaluation

Results

4 Suffixes Ø menente mente s

11 Suffixes a amente as illa illas o or ora oras ores os

41 Suffixes a aba aban acion acciones acción ada adas ado ador  
adora adoras adores ados amos an ando ante antes ar  
ara aran aremos arla arlas arlo arlos arme aron arse ará  
arán aré aría arían ase e en ándose é ó

29 Suffixes

**administradas**

20 Suffixes

**‘Feminine gender nouns  
under administration’**

29 Suffixes

6 Suffixes Ø es idad idades mente ísima

# Segment Words Using the Paradigms



**ParaMor**

Identify  
Search  
Cluster  
Filter

**Segment**  
Evaluation  
Results

4 Suffixes Ø menente mente s

11 Suffixes a amente as illa illas o or ora oras ores os

41 Suffixes a aba aban acion acciones acción ada adas ado ador  
adoraadorasadoresadosamosan ando ante antes ar  
ara aran aremos arla arlas arlo arlos arme aron arse ará  
arán aré aría arían ase e en ándose é ó

29 Suffixes

administr + ad + a + s

20 Suffixes

Past Participle

29 Suffixes

Feminine

Plural

6 Suffixes Ø es idad idades mente ísima



# Segment Words Using the Paradigms



ParaMor

Identify

Search

Cluster

Filter

**Segment**

Evaluation

Results

4 Suffixes Ø menente mente s

11 Suffixes a amente as illa illas o or ora oras ores os

41 Suffixes a aba aban acion acciones acción ada **adas** ado ador  
adora adoras adores ados amos an ando ante antes ar  
ara aran aremos arla arlas arlo arlos arme aron arse ará  
arán aré aría arían ase e en ándose é ó

29 Suffixes

administr**adas**

20 Suffixes

29 Suffixes

6 Suffixes Ø es idad idades mente ísima

# Segment Words Using the Paradigms



ParaMor

Identify

Search

Cluster

Filter

**Segment**

Evaluation

Results

4 Suffixes Ø menente mente s

11 Suffixes a amente as illa illas o or ora oras ores os

41 Suffixes a aba aban acion acciones acción **ada adas** ado ador  
adoraadorasadoresadosamosanandoanteantesar  
araaranaremosarlaarlasarloarlosarmearonarseará  
aránaréaríaaríanasee en ándose é ó

29 Suffixes

administr**adas** → administr**ada**

20 Suffixes

29 Suffixes

Also in corpus

6 Suffixes Ø es idad idades mente ísima

# Segment Words Using the Paradigms

## ParaMor

Identify  
Search  
Cluster  
Filter

**Segment**  
Evaluation  
Results

4 Suffixes Ø menente mente s

11 Suffixes a amente as illa illas o or ora oras ores os

41 Suffixes a aba aban acion acciones acción **ada adas** ado ador  
adoraadorasadoresadosamosan ando ante antes ar  
ara aran aremos arla arlas arlo arlos arme aron arse ará  
arán aré aría arían ase e en ándose é ó

29 Suffixes

administr**adas** → administr**ada**

20 Suffixes

Morpheme  
Boundary

29 Suffixes

6 Suffixes Ø es idad idades mente ísima

# Segment Words Using the Paradigms

## ParaMor

Identify  
Search  
Cluster  
Filter

**Segment**  
Evaluation  
Results

4 Suffixes Ø menente mente s

11 Suffixes **a** amente **as** illa illas o or ora oras ores os

41 Suffixes a aba aban acion acciones acción ada adas ado ador  
adoraadorasadoresadosamosan ando ante antes ar  
ara aran aremos aria arlas arlo arlos arme aron arse ará  
arán aré aría arían ase e en ándose é ó

29 Suffixes

administrad**as** → administrad**a**

20 Suffixes

Morpheme  
Boundary

29 Suffixes

6 Suffixes Ø es idad idades mente ísima

# Segment Words Using the Paradigms

## ParaMor

Identify  
Search  
Cluster  
Filter

**Segment**  
Evaluation  
Results

4 Suffixes Ø menente mente **s**

11 Suffixes a amente as illa illas o or ora oras ores os

41 Suffixes a aba aban acion acciones acción ada adas ado ador  
adoraadorasadoresadosamos an ando ante antes ar  
ara aran aremos arla arlas arlo arlos arme aron arse ará  
arán aré aría arían ase e enándose é ó

29 Suffixes

administrada**s** → administradaØ

20 Suffixes

Morpheme  
Boundary

29 Suffixes

6 Suffixes Ø es idad idades mente ísima

# Segment Words Using the Paradigms



ParaMor

Identify

Search

Cluster

Filter

**Segment**

Evaluation  
Results

4 Suffixes Ø menente mente s

11 Suffixes a amente as illa illas o or ora oras ores os

41 Suffixes a aba aban acion acciones acción ada adas ado ador  
adora adoras adores ados amos an ando ante antes ar  
ara aran aremos arla arlas arlo arlos arme aron arse ará  
arán aré aría arían ase e en ándose é ó

29 Suffixes

**administr + ad + a + s**

20 Suffixes

**R**ecovers multiple morpheme boundaries  
from candidate paradigms which each  
propose single morpheme boundaries

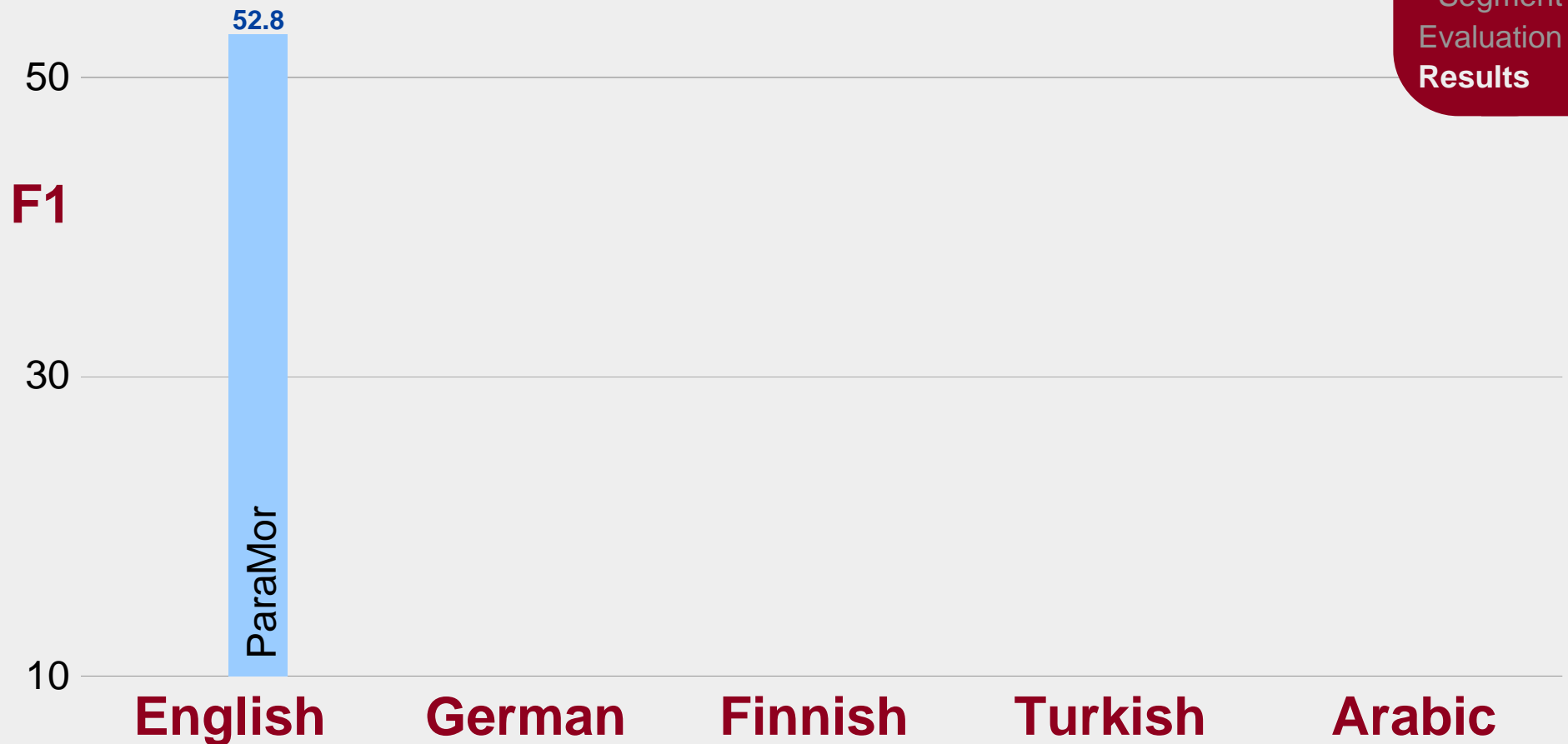
29 Suffixes

6 Suffixes Ø es idad idades mente ísima

# Linguistic Evaluation



ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**

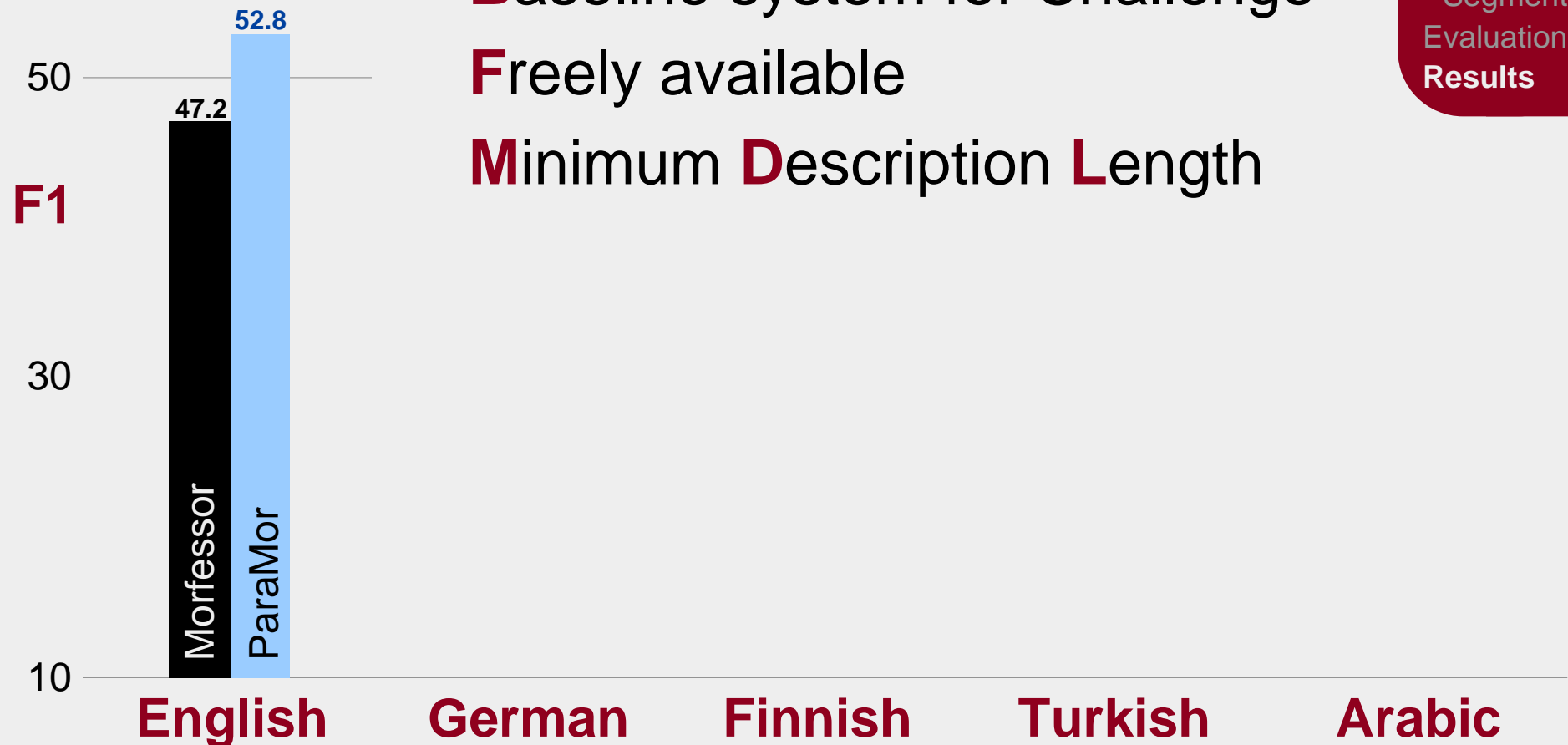


## Morfessor

**B**aseline system for Challenge

**F**reely available

**M**inimum **D**escription **L**ength



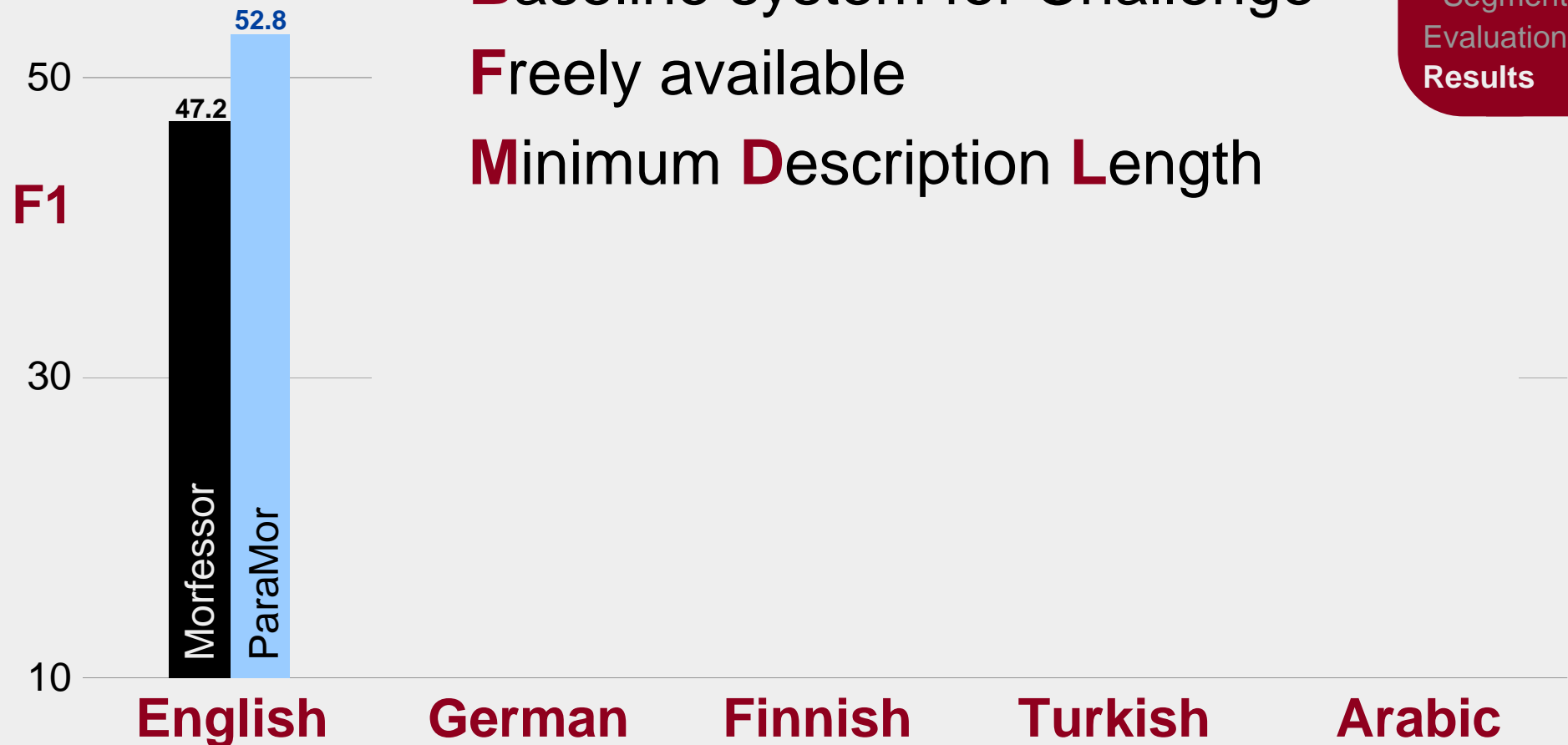


## Morfessor

**B**aseline system for Challenge

**F**reely available

**M**inimum **D**escription **L**ength



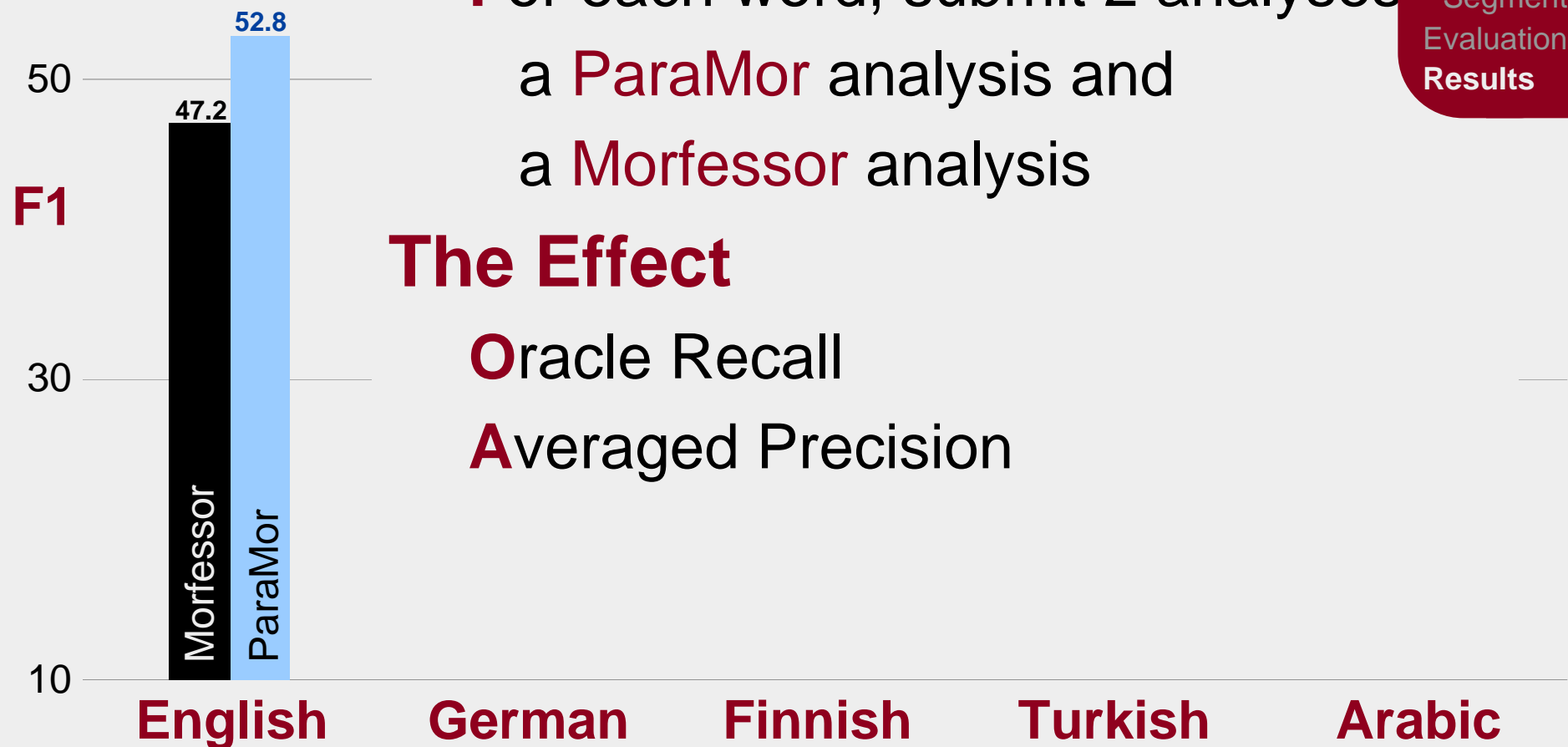
## Join ParaMor and Morfessor

For each word, submit 2 analyses  
a **ParaMor** analysis and  
a **Morfessor** analysis

## The Effect

**O**racle Recall

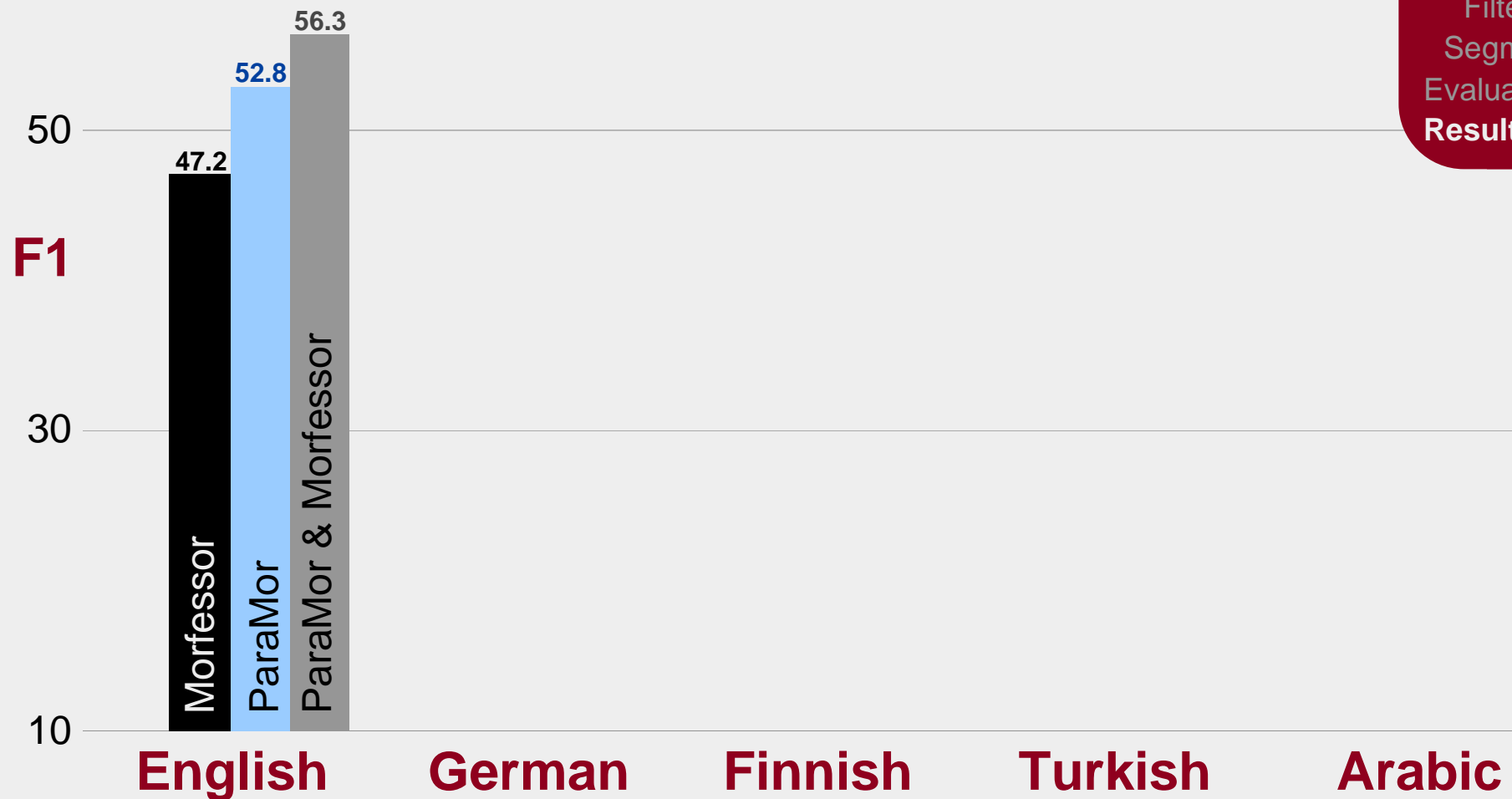
**A**veraged Precision



# Linguistic Evaluation



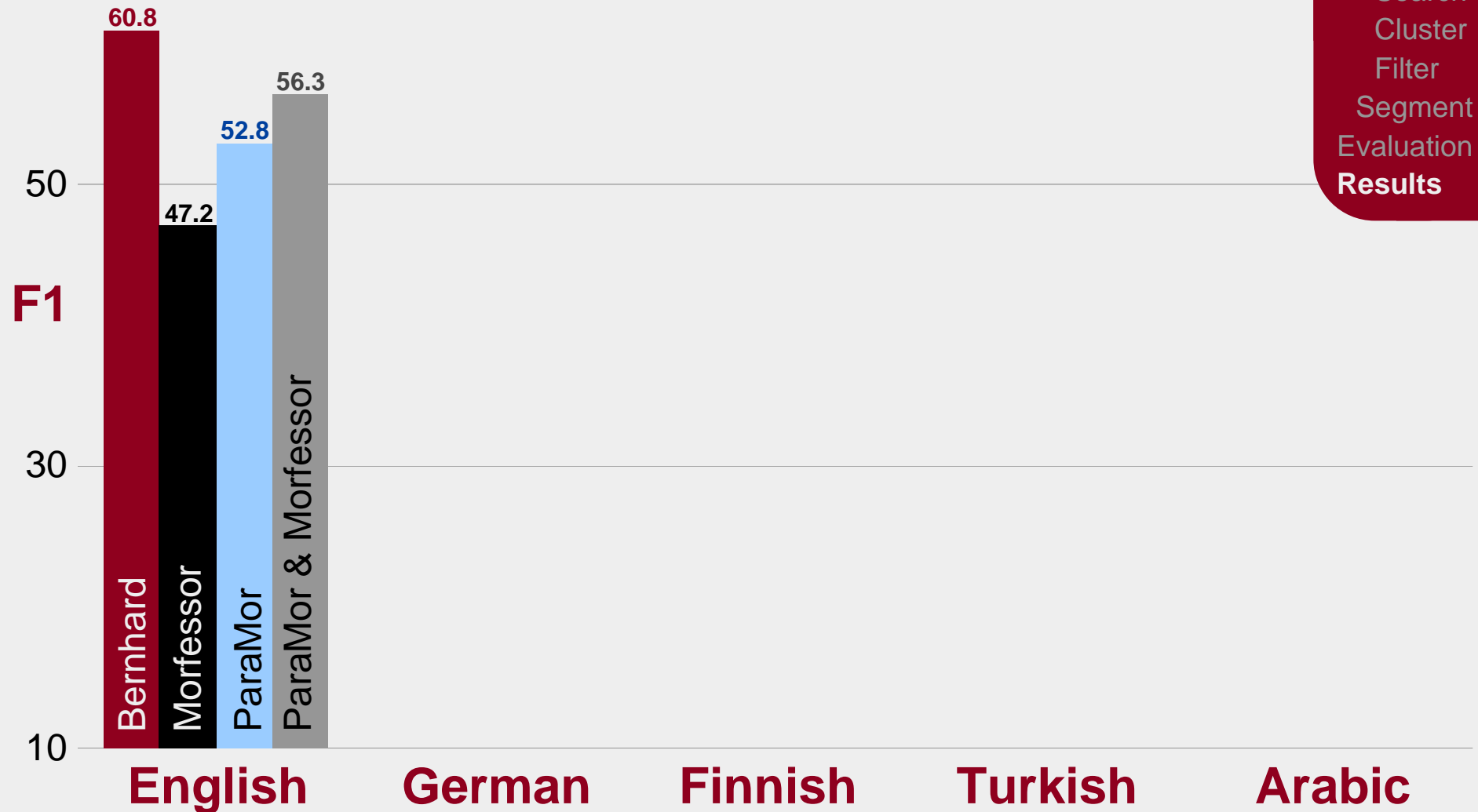
ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**



# Linguistic Evaluation



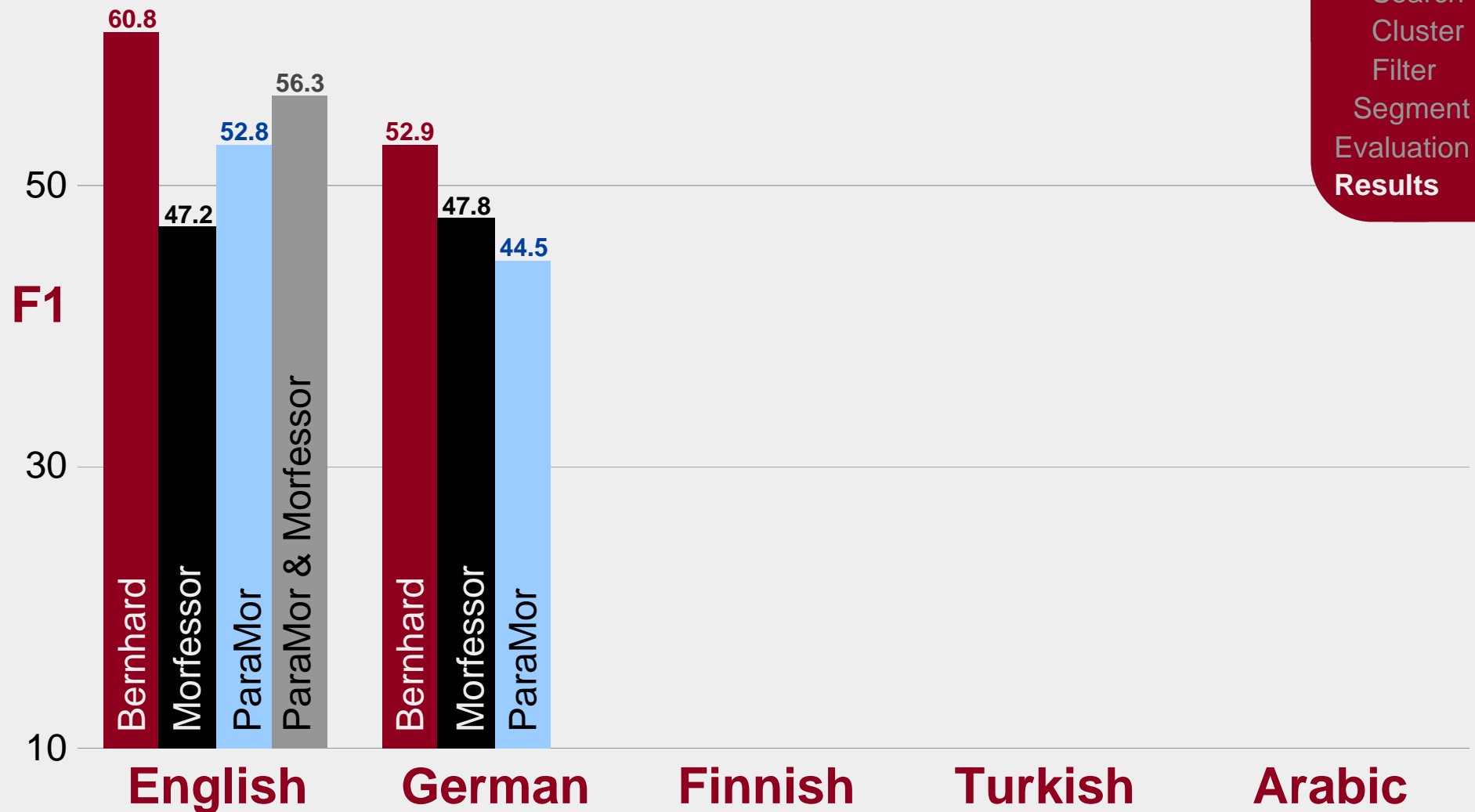
ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**



# Linguistic Evaluation



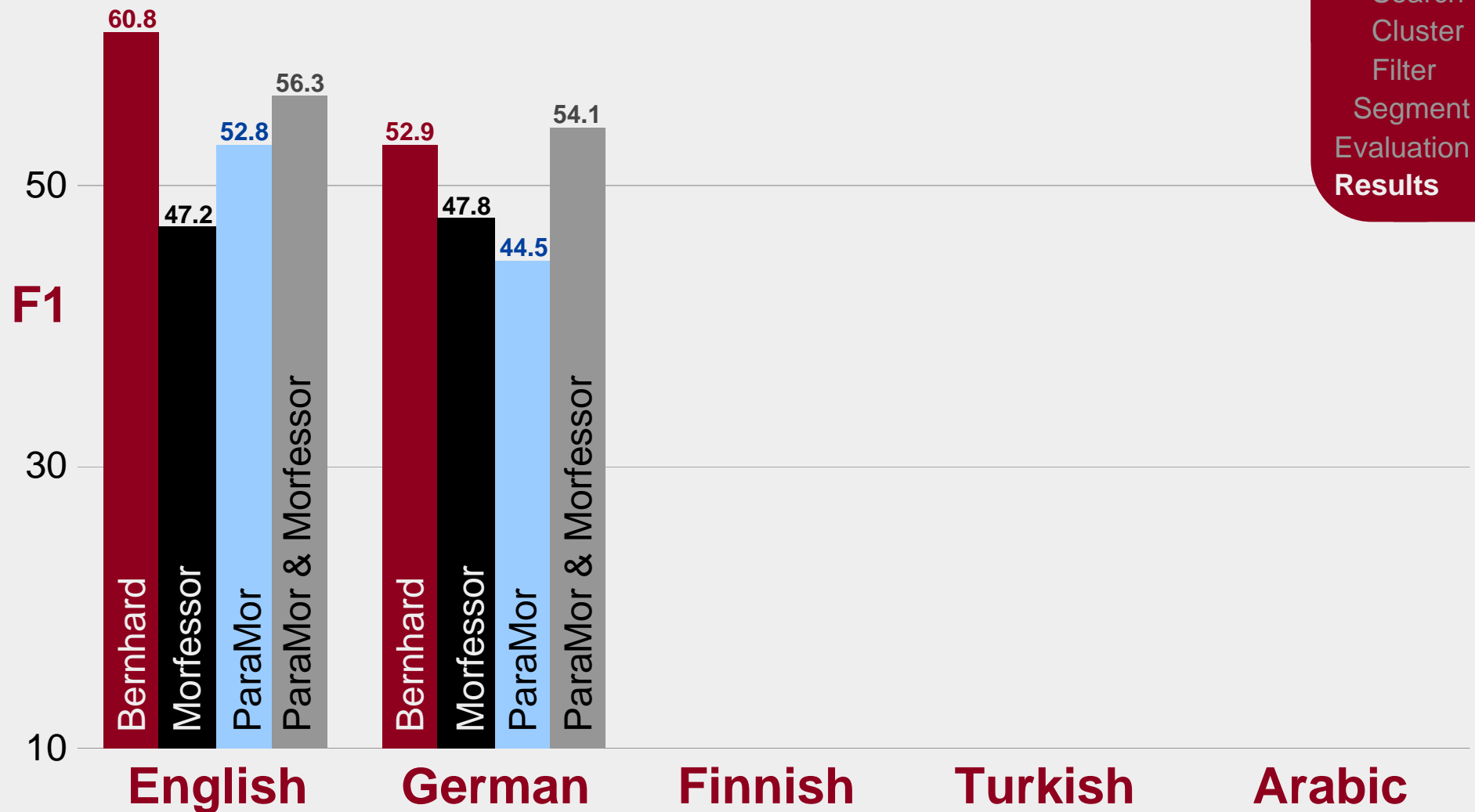
ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**



# Linguistic Evaluation



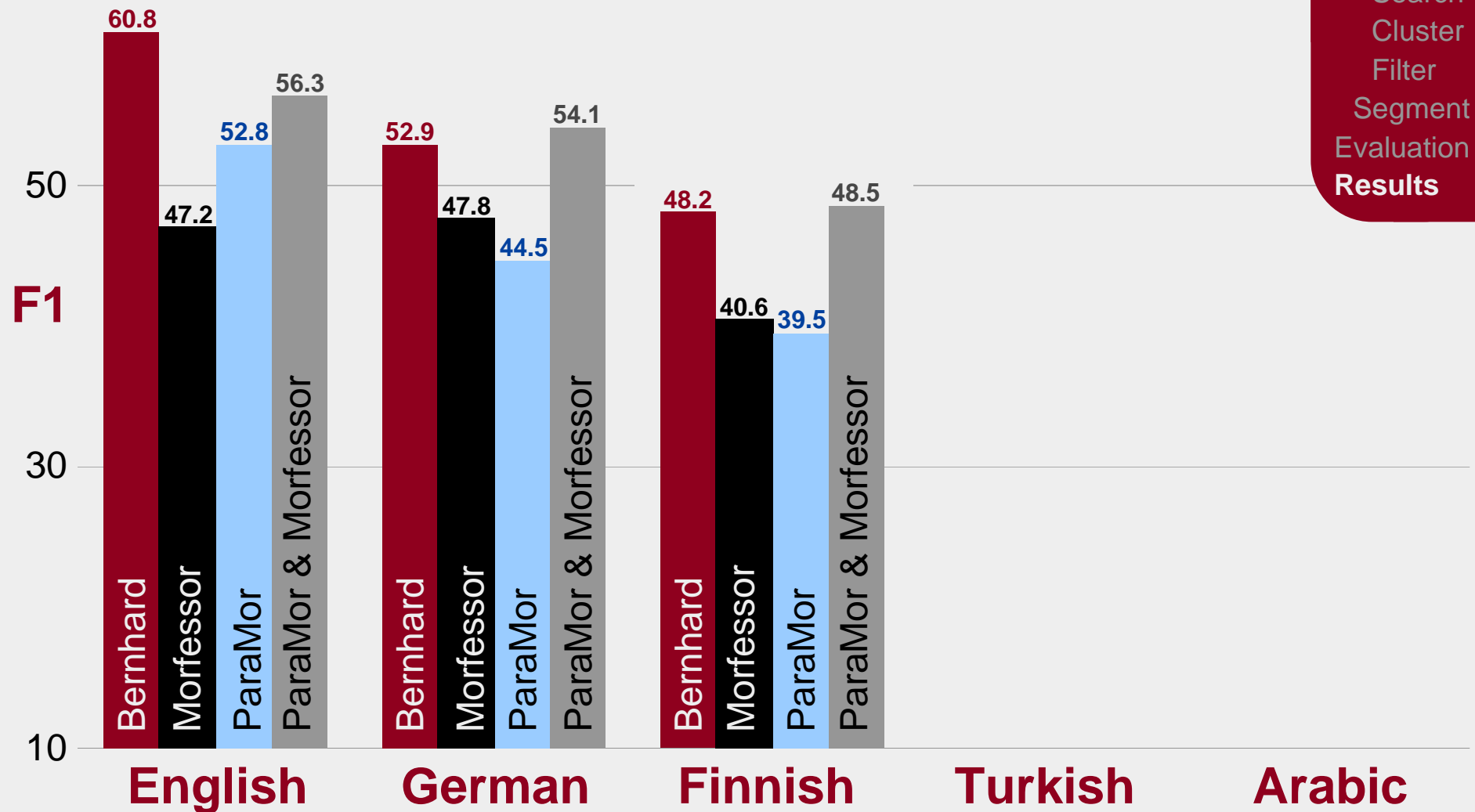
ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**



# Linguistic Evaluation



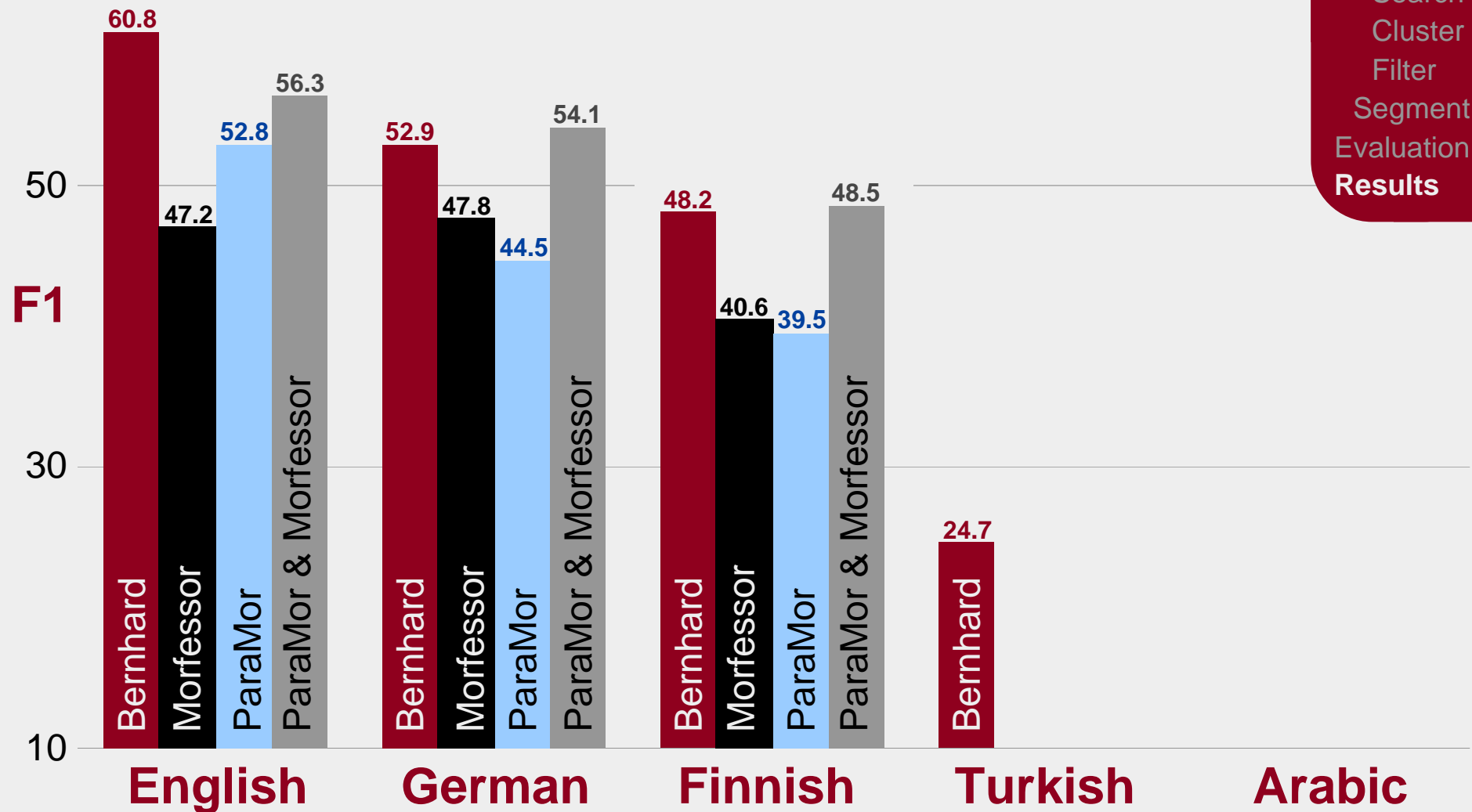
ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**



# Linguistic Evaluation



ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**

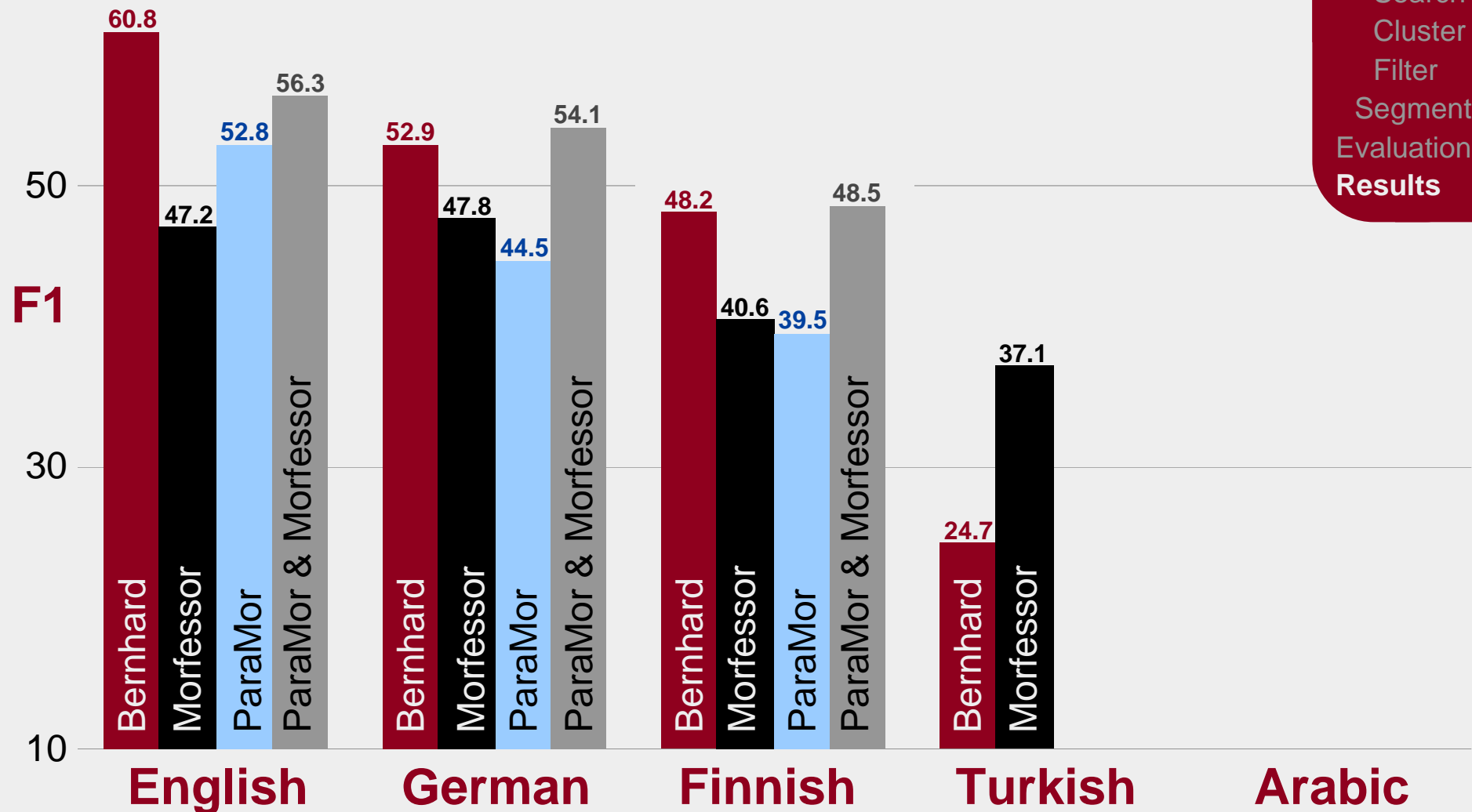




# Linguistic Evaluation



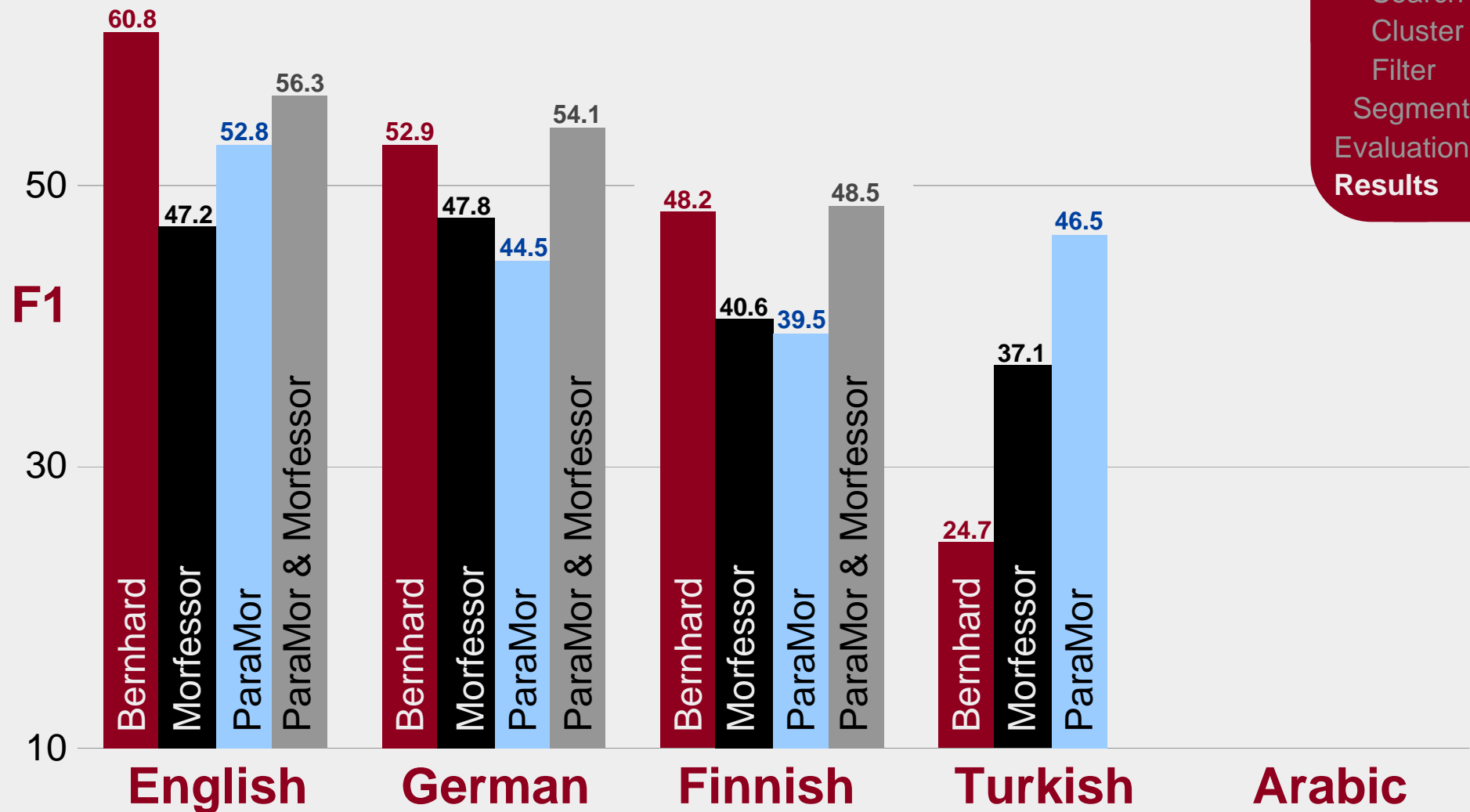
ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**



# Linguistic Evaluation



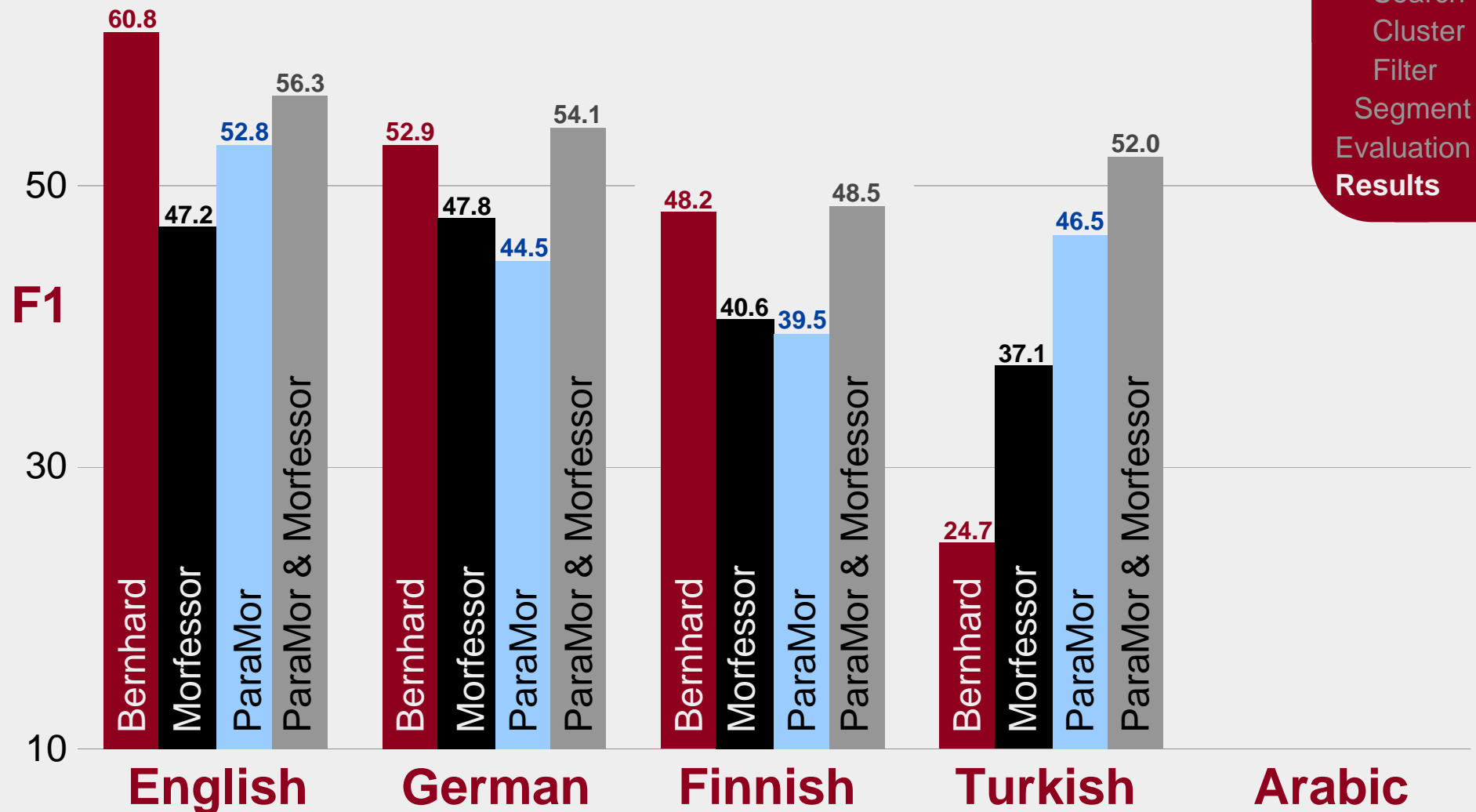
ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**



# Linguistic Evaluation



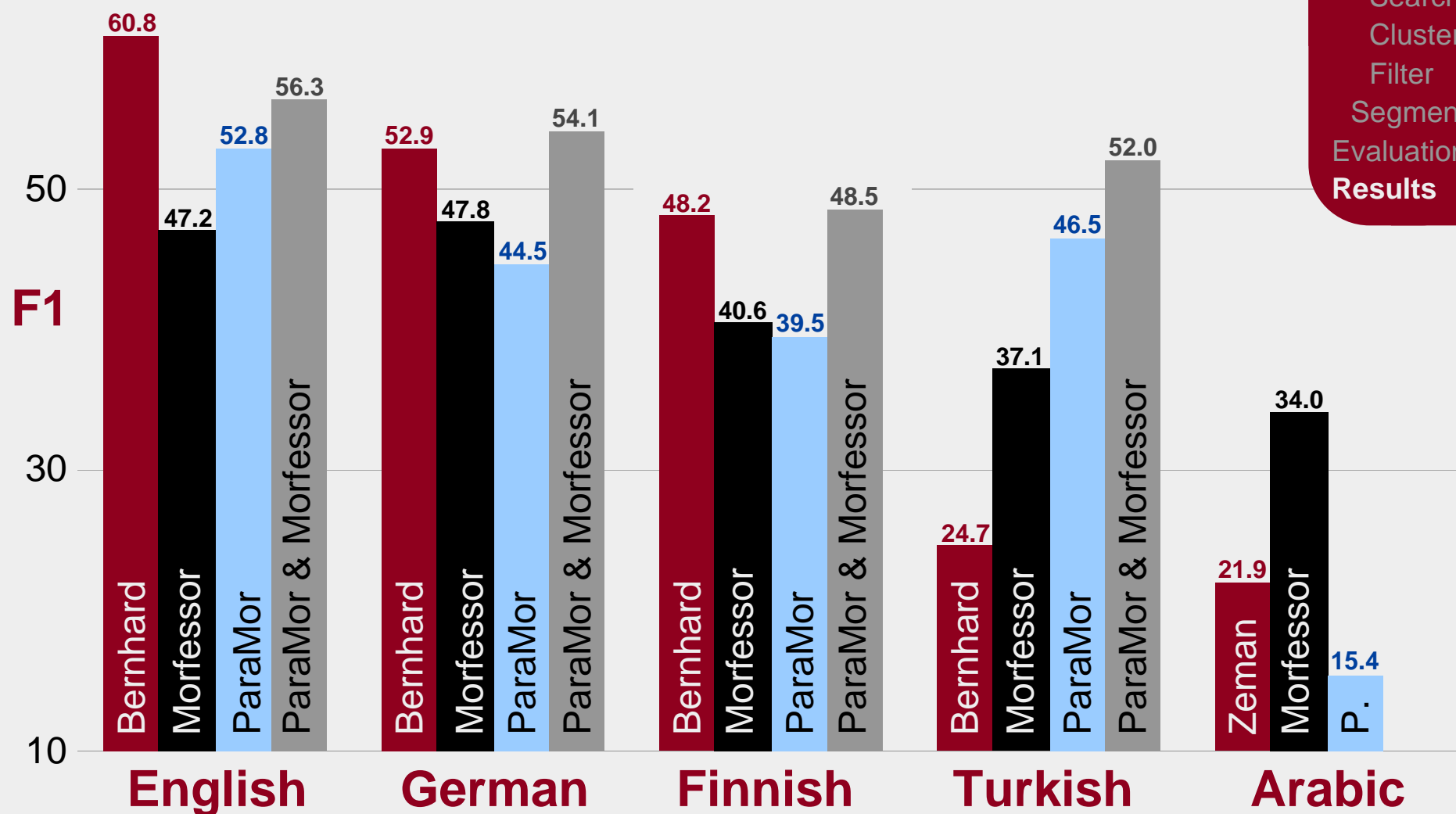
ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**



# Linguistic Evaluation



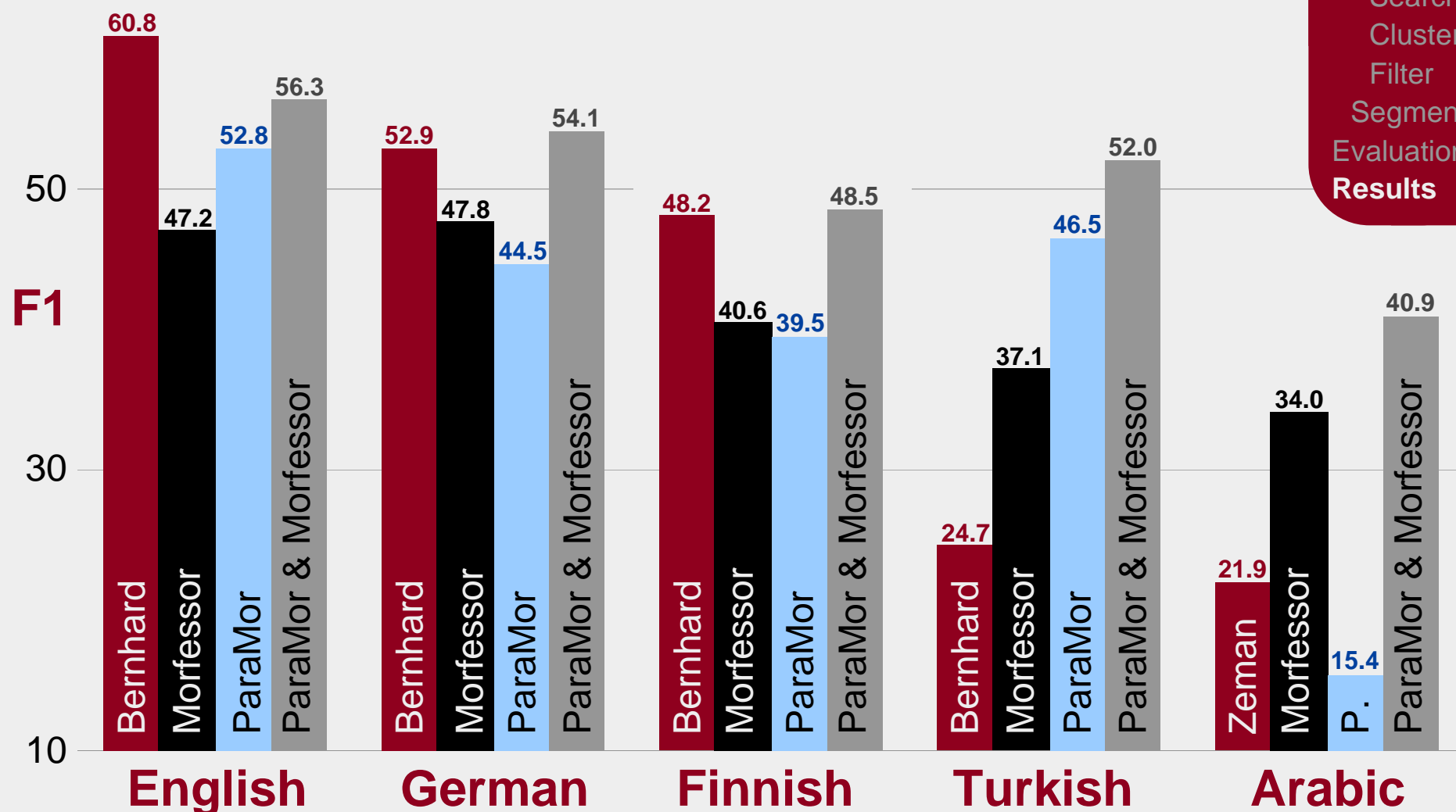
ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**



# Linguistic Evaluation



ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**

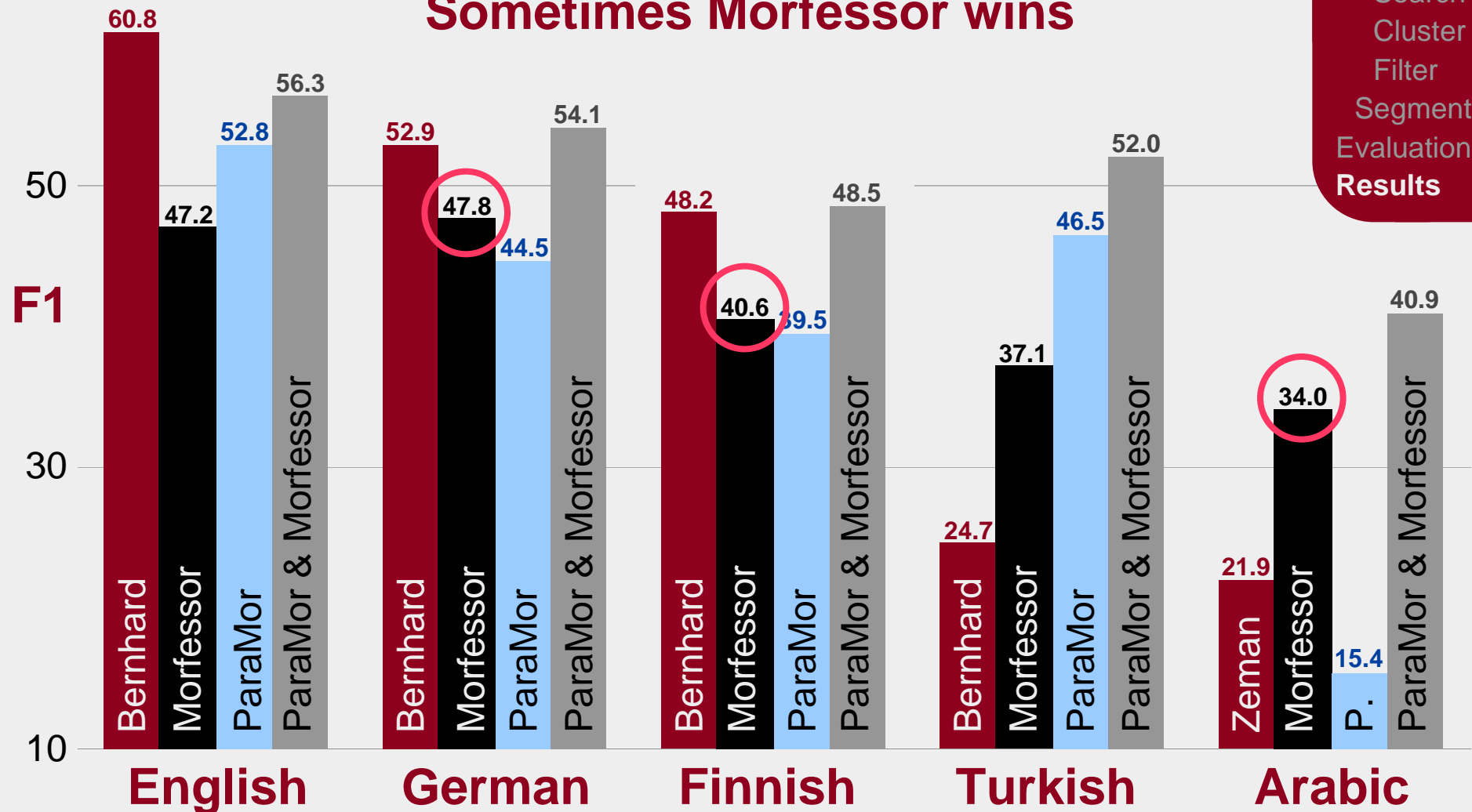


# Linguistic Evaluation



ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**

## Sometimes Morfessor wins

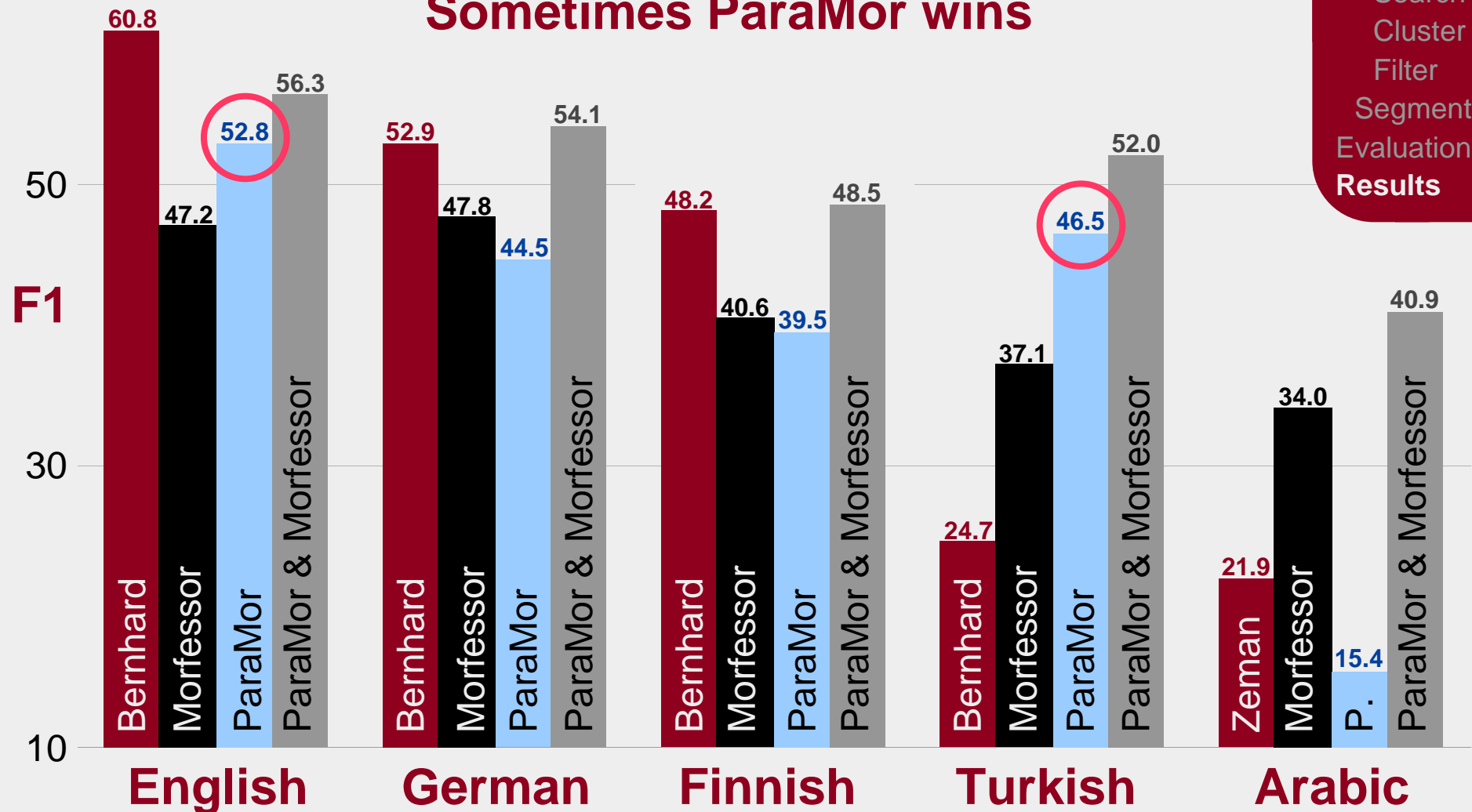


# Linguistic Evaluation



ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**

Sometimes ParaMor wins

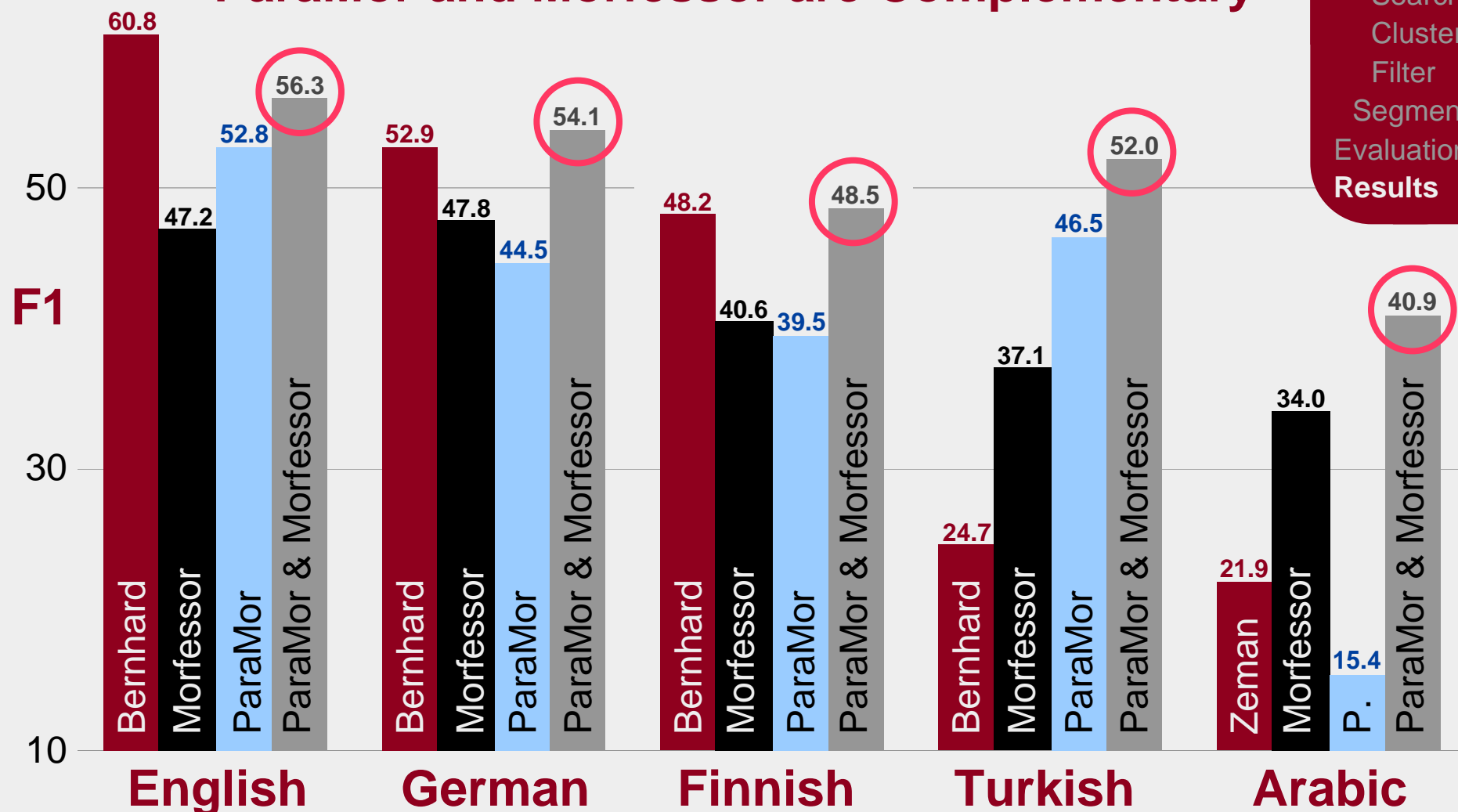


# Linguistic Evaluation



ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**

## ParaMor and Morfessor are Complementary

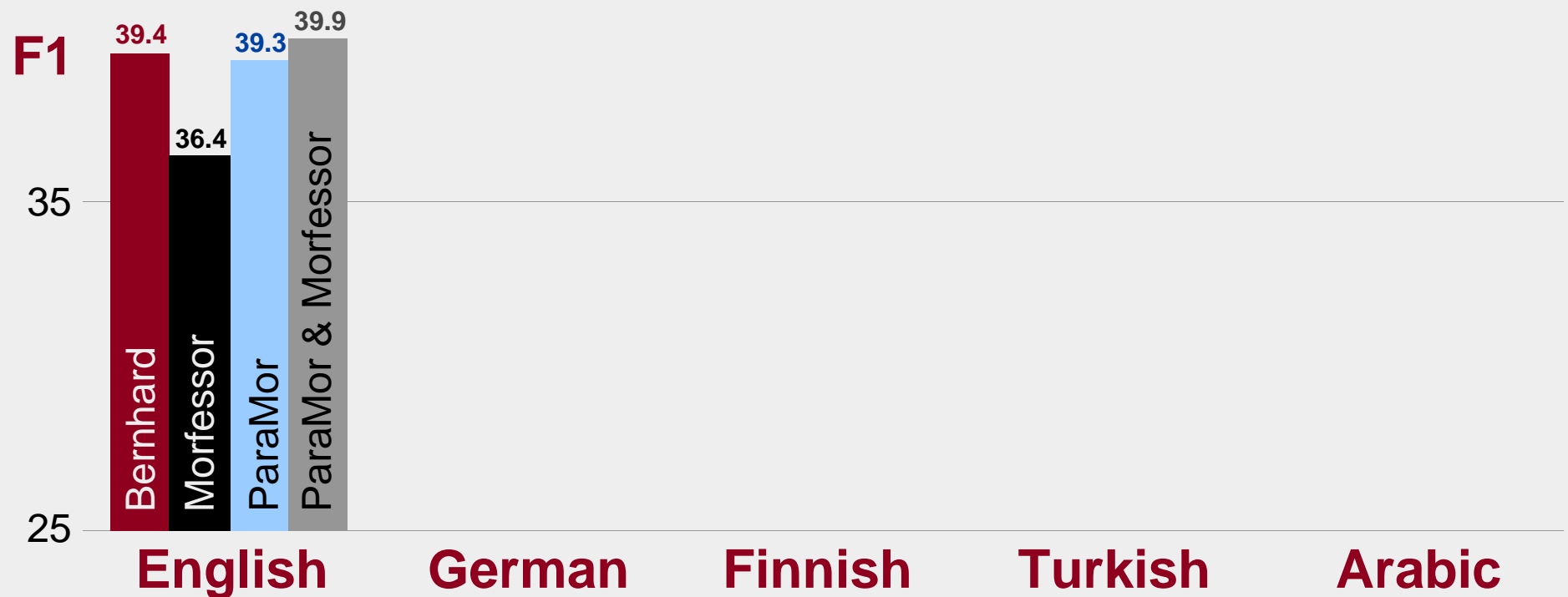




# IR Evaluation



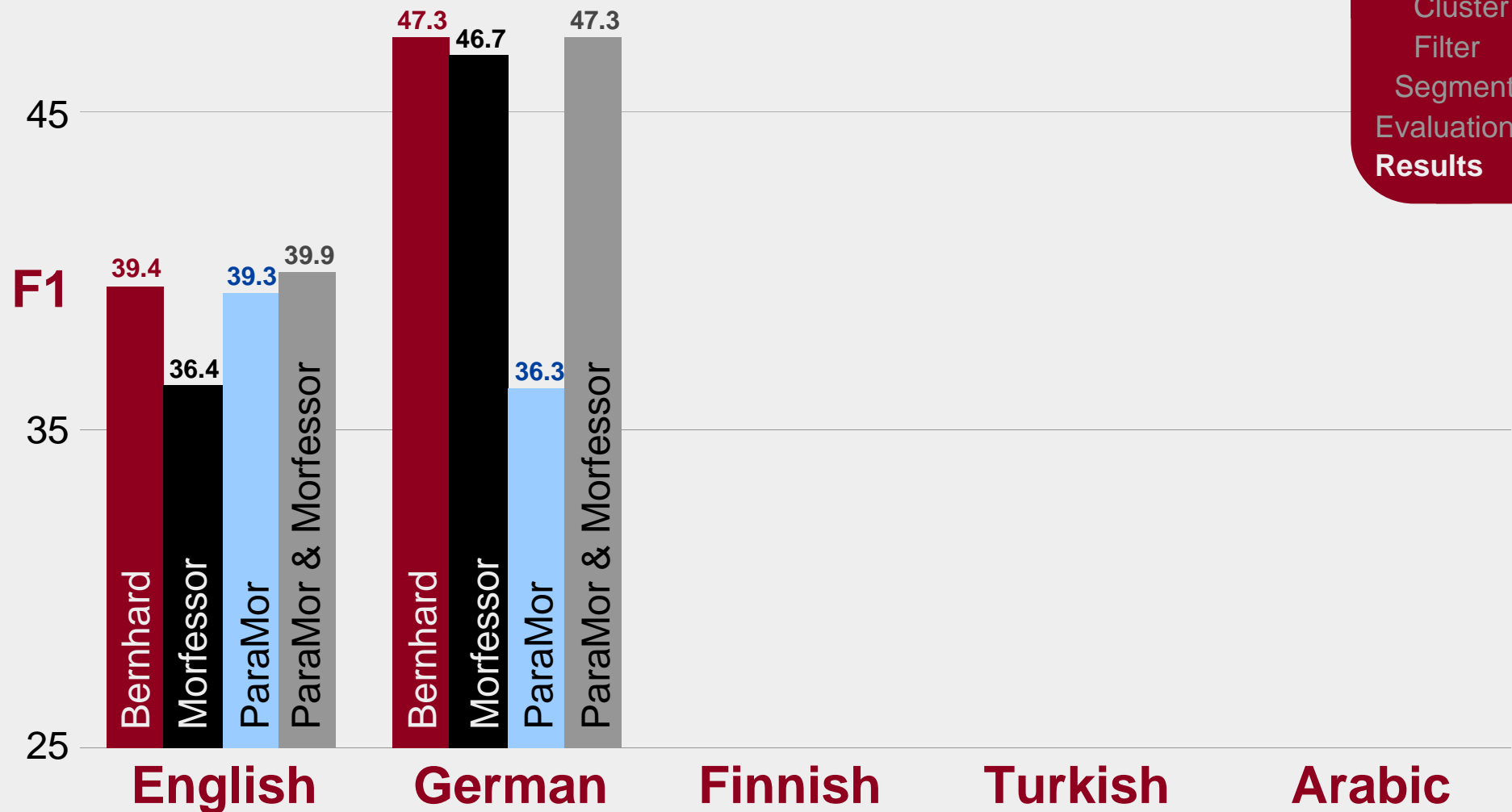
ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**



# IR Evaluation



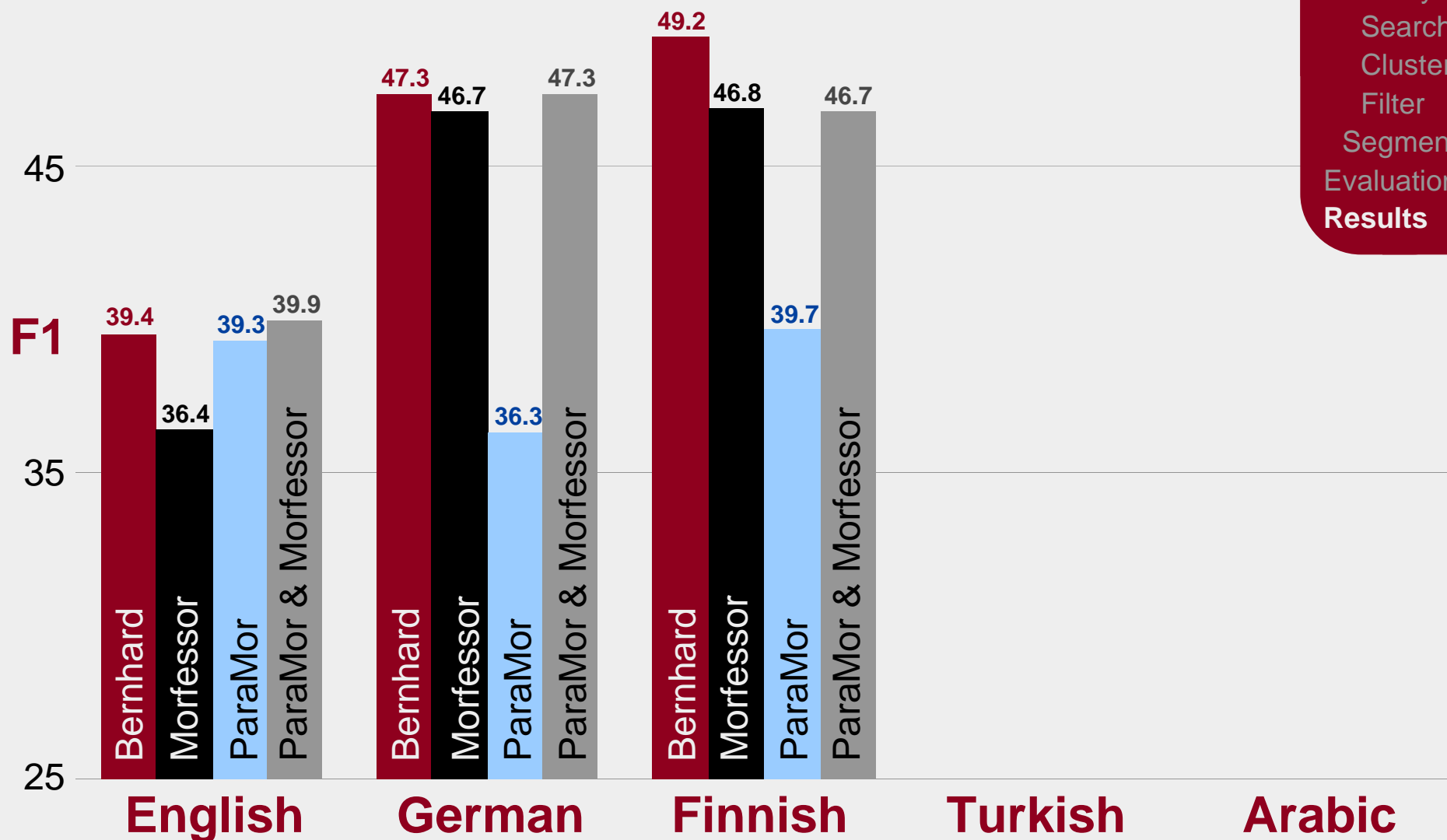
ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**



# IR Evaluation



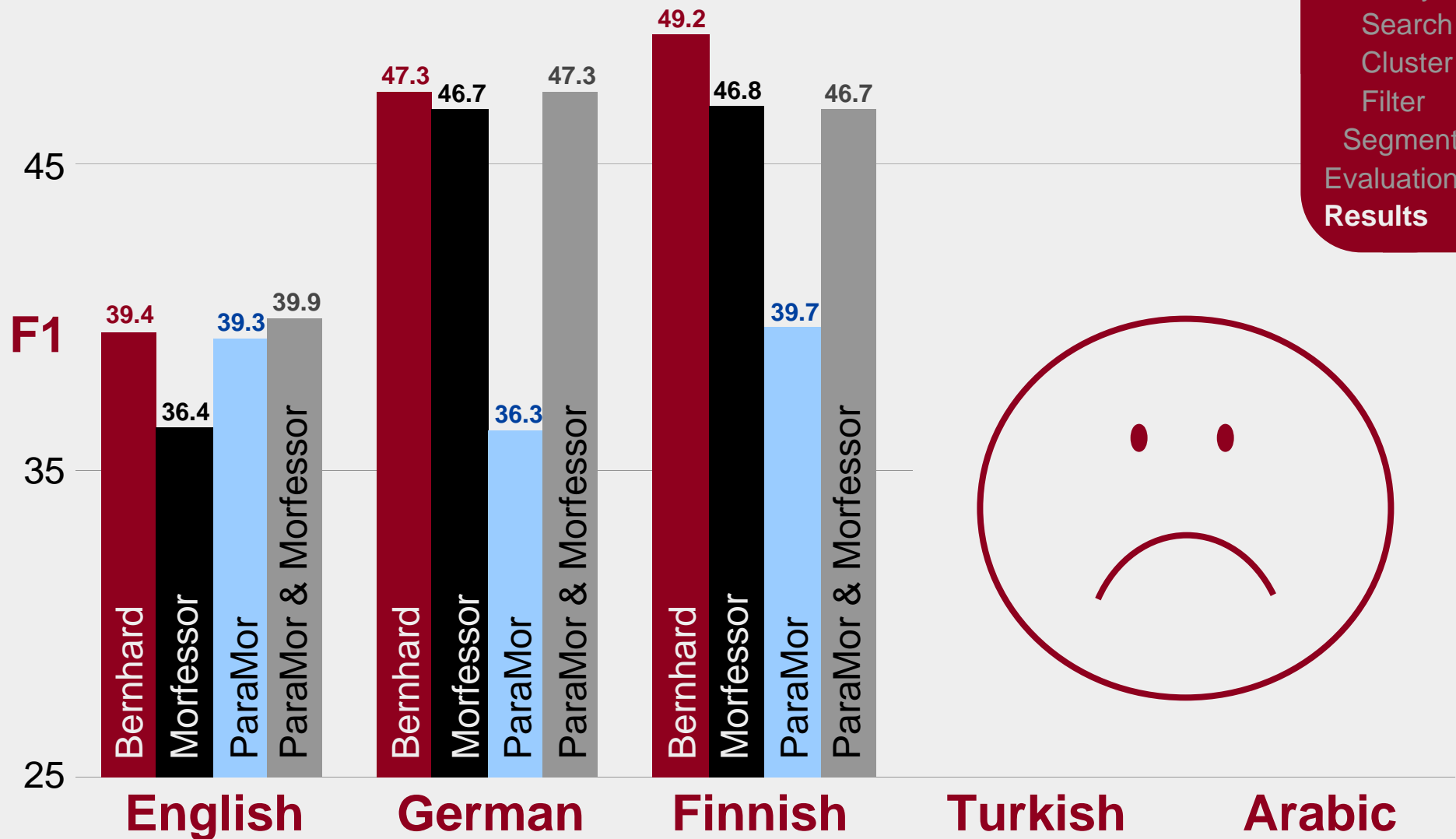
ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**



# IR Evaluation



ParaMor  
Identify  
Search  
Cluster  
Filter  
Segment  
Evaluation  
**Results**



# ParaMor: State-of-the-Art Unsupervised Morphology Induction System

## ParaMor

**I**dentifies paradigms

**T**he organizing **structure** of inflectional morphology

**S**egments words

**A**s discovered paradigms suggest

## Combined with Morfessor

**A**mong the best in Morpho Challenge

**C**onsistent across languages

## Beyond suffixes

**S**traightforward extension to ParaMor for

**P**refixes

**M**ore Challenging

**R**eduplication, **I**nfixation, etc.

## Morphophonology

**I**ncorporate contextual information when clustering

## Improve system combination

**T**rue merging of analyses

**C**ombine more systems

