HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

University of Colorado **Boulder**

An Encoder-Decoder Approach to the Paradigm Cell-Filling Problem

Miikka Silfverberg
University of Helsinki
University of Colorado

Mans Hulden
University of Colorado

Paradigm Cell-Filling Problem

Partial inflection table

V;INF	speak
V;PRES	?
V;PAST	?
V;PAST;PCPLE	spoken
V;PRES;PCPLE	?

Farrell Ackerman, James P. Blevins, and Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In James P. Blevins and Juliette Blevins, editors, *Analogy in Grammar*, pages 54–82. Oxford University Press.

Paradigm Cell-Filling Problem

Partial inflection table			Completed inflection table	
V;INF	speak		V;INF	speak
V;PRES	?		V;PRES	speaks
V;PAST	?	$\Longrightarrow$	V;PAST	spoke
V;PAST;PCPLE	spoken		V;PAST;PCPLE	spoken
V;PRES;PCPLE	?		V;PRES;PCPLE	speaking

Farrell Ackerman, James P. Blevins, and Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In James P. Blevins and Juliette Blevins, editors, *Analogy in Grammar*, pages 54–82. Oxford University Press.

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

University of Colorado Boulder

Paradigm Cell-Filling Problem

https://herrkramski.wordpress.com/2014/11/06/finnish-is-not-a-hard-language-pt-1/

Finnish
“dog”

Contribution

1. We investigate PCFP in three different settings.
2. We present two neural models for the PCFP task.
3. We present new data sets for PCFP.

Three Settings

Three Settings

n > 1

walk	INF
walks	PRES
?	PAST
?	PCPLE
?	INF
smiles	PRES
?	PAST
smiling	PCPLE

Three Settings

n > 1		n = 1	
walk	INF	?	INF
walks	PRES	?	PRES
?	PAST	walked	PAST
?	PCPLE	?	PCPLE
?	INF	?	INF
smiles	PRES	smiles	PRES
?	PAST	?	PAST
smiling	PCPLE	?	PCPLE

Three Settings

n > 1

walk	INF
walks	PRES
?	PAST
?	PCPLE
?	INF
smiles	PRES
?	PAST
smiling	PCPLE

n = 1

?	INF
?	PRES
walked	PAST
?	PCPLE
?	INF
smiles	PRES
?	PAST
?	PCPLE

by frequency

walk	INF
walks	PRES
?	PAST
walking	PCPLE
smile	INF
smiles	PRES
?	PAST
?	PCPLE

General Approach

- In the n > 1 and “frequent words” settings we train a LSTM encoder-decoder model with attention.

General Approach

- In the n > 1 and “frequent words” settings we train a LSTM encoder-decoder model with attention.

General Approach

- In the n > 1 and “frequent words” settings we train a LSTM encoder-decoder model with attention.

- In the n = 1 setting, we apply adaptive character dropout and then train an encoder-decoder.

Training for n > 1 Given Forms

walk	INF
walks	**PRES**
?	PAST
?	PCPLE
?	INF
smiles	PRES
?	PAST
smiling	PCPLE

Input Output

Training for n > 1 Given Forms

walk	INF
walks	**PRES**
?	PAST
?	PCPLE
?	INF
smiles	PRES
?	PAST
smiling	PCPLE

Input Output

Training for n > 1 Given Forms

walk	INF
walks	**PRES**
?	PAST
?	PCPLE

?	INF
smiles	PRES
?	PAST
smiling	PCPLE

Input	Output
walk	

Training for n > 1 Given Forms

walk	INF
walks	**PRES**
?	PAST
?	PCPLE
?	INF
smiles	PRES
?	PAST
smiling	PCPLE

Input	Output
`walk+INF`	

Training for n > 1 Given Forms

walk	INF
walks	PRES
?	PAST
?	PCPLE

?	INF
smiles	PRES
?	PAST
smiling	PCPLE

Input	Output
`walk+INF>PRES`	

Training for n > 1 Given Forms

walk	INF
walks	PRES
?	PAST
?	PCPLE
?	INF
smiles	PRES
?	PAST
smiling	PCPLE

Input	Output
`walk+INF>PRES`	`walks`

Training for n > 1 Given Forms

walk	INF
walks	PRES
?	PAST
?	PCPLE
?	INF
smiles	PRES
?	PAST
smiling	PCPLE

Input	Output
walk+INF>PRES	walks
walks+PRES>INF	walk

Training for n > 1 Given Forms

walk	INF
walks	PRES
?	PAST
?	PCPLE

?	INF
smiles	PRES
?	PAST
smiling	PCPLE

Input	Output
walk+INF>PRES	walks
walks+PRES>INF	walk
smiles+PRES>PCPLE	smiling
smiling+PCPLE>PRES	smiles

Training for n > 1 Given Forms

walk	INF
walks	PRES
?	PAST
?	PCPLE

?	INF
smiles	PRES
?	**PAST**
smiling	PCPLE

Input	Output
walk+INF>PRES	walks
walks+PRES>INF	walk
smiles+PRES>PCPLE	smiling
smiling+PCPLE>PRES	smiles

Training for n > 1 Given Forms

walk	INF
walks	PRES
?	PAST
?	PCPLE

?	INF
smiles	PRES
?	**PAST**
smiling	PCPLE

Input	Output
walk+INF>PRES	walks
walks+PRES>INF	walk
smiles+PRES>PCPLE	smiling
smiling+PCPLE>PRES	smiles
smiles+PRES>PAST	?
smiling+PCPLE>PAST	?

Training for n > 1 Given Forms

walk	INF
walks	PRES
?	PAST
?	PCPLE

?	INF
smiles	PRES
?	**PAST**
smiling	PCPLE

Input	Output
walk+INF>PRES	walks
walks+PRES>INF	walk
smiles+PRES>PCPLE	smiling
smiling+PCPLE>PRES	smiles
smiles+PRES>PAST	smiled
smiling+PCPLE>PAST	smiled

Training for n > 1 Given Forms

walk	INF
walks	PRES
?	PAST
?	PCPLE

?	INF
smiles	PRES
smiled	**PAST**
smiling	PCPLE

Input	Output
walk+INF>PRES	walks
walks+PRES>INF	walk
smiles+PRES>PCPLE	smiling
smiling+PCPLE>PRES	smiles
smiles+PRES>PAST	smiled
smiling+PCPLE>PAST	smiled

↓

smiled

Training for n > 1 Given Forms

walk	INF
walks	PRES
?	PAST
?	PCPLE

?	INF
smiles	PRES
smiled	**PAST**
smiling	PCPLE

Input	Output
walk+INF>PRES	walks
walks+PRES>INF	walk
smiles+PRES>PCPLE	smiling
smiling+PCPLE>PRES	smiles

smiles+PRES>PAST	smiled
smiling+PCPLE>PAST	smiled

↓

smiled

Katharina Kann and Hinrich Schütze. 2016. Single-model encoder-decoder with explicit morphological representation for reinflection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 555–560, Berlin, Germany. Association for Computational Linguistics.

Case n = 1

?	INF
?	PRES
walked	PAST
?	PCPLE

?	INF
smiles	PRES
?	PAST
?	PCPLE

Input Output

Case n = 1

?	INF
?	PRES
walked	PAST
?	PCPLE

?	INF
smiles	PRES
?	PAST
?	PCPLE

Input	Output
walked+PAST>PAST	walked
smiles+PRES>PRES	smiles

Case n = 1

?	INF
?	PRES
walked	PAST
?	PCPLE

?	INF
smiles	PRES
?	PAST
?	PCPLE

Input	Output
walked+PAST>PAST	walked
smiles+PRES>PRES	smiles

walked+PAST>PRES	?

Case n = 1

?	INF
?	PRES
walked	PAST
?	PCPLE

?	INF
smiles	PRES
?	PAST
?	PCPLE

Input	Output
~~walked+PAST>PAST~~	~~walked~~
~~smiles+PRES>PRES~~	~~smiles~~

walked+PAST>PRES	?

What about Stemming?

walked+PAST>PAST	walked
smiles+PRES>PRES	smiles

What about Stemming?

`walked+PAST>PAST`	`walked`
`smiles+PRES>PRES`	`smiles`

What about Stemming?

`walked+PAST>PAST`	`walked`
`smiles+PRES>PRES`	`smiles`

`walk>PAST`	`walked`
`smile>PRES`	`smiles`

Probabilistic Stemming Using a Language Model

PAST forms:

```
walked
slept
heard
missed
loved
ate
taped
```

Probabilistic Stemming Using a Language Model

PAST forms:

```
walked
slept
heard
missed
loved
ate
taped
```

Probabilistic Stemming Using a Language Model

PAST forms:

```
walked
slept
heard
missed
loved
ate
taped
```

- We train a character language model for each label (INF, PRES, PAST,…)

Probabilistic Stemming Using a Language Model

PAST forms:

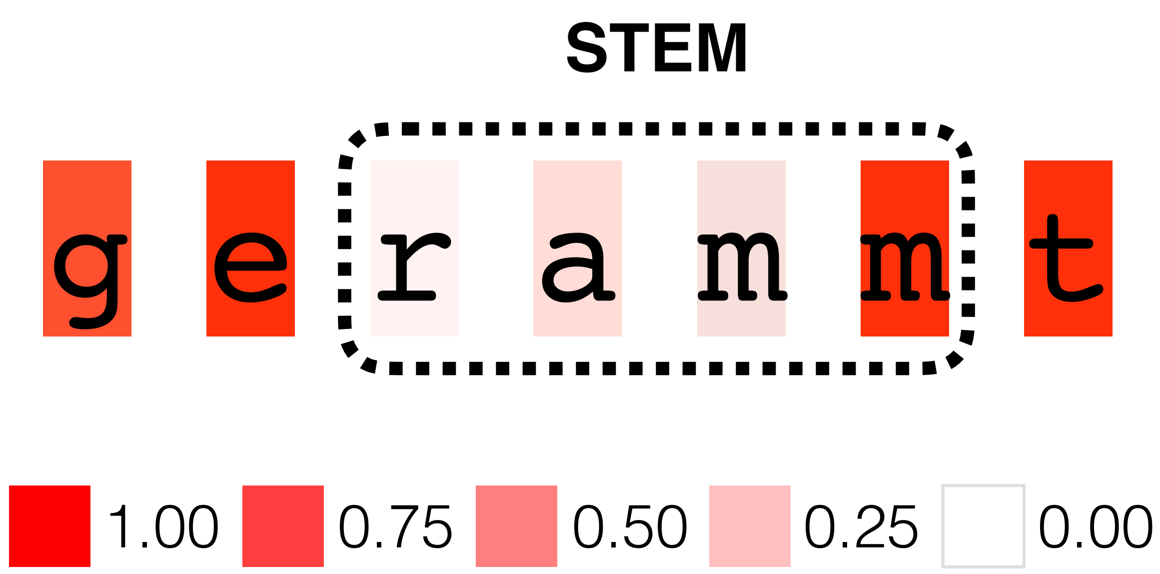
```
walked
slept
heard
missed
loved
ate
taped
```

- We train a character language model for each label (INF, PRES, PAST,…)

- Language model confidence is used for identifying word stems and affixes.

Language Model Confidences

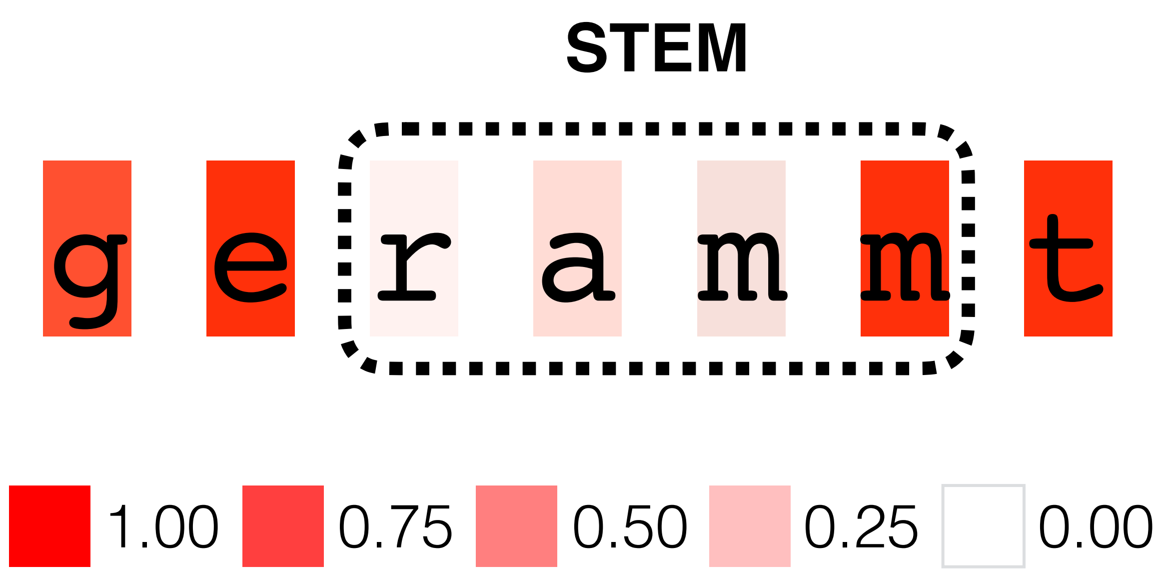
Confidence for predicting character x_t based on $x_1, \ldots, x_{t-1}$:



German
“rammed”

Language Model Confidences

Confidence for predicting character x_t based on $x_1, ..., x_{t-1}$:



German
"rammed"

Characters belonging to the word stem typically have **low** language model confidence.

Adaptive Character Dropout

During training we drop characters from the input form based on $\mathrm{LM}_{\mathrm{PAST}}$ confidence:

`walked+PAST`

Adaptive Character Dropout

During training we drop characters from the input form based on $\mathrm{LM}_{\mathrm{PAST}}$ confidence:

`walk`[redacted]`+PAST`

Adaptive Character Dropout

During training we drop characters from the input form based on LM_{PAST} confidence:

`walk█+PAST`

Because it is problematic to determine which characters belong to the stem, we drop characters probabilistically.

Training Examples Before Character Dropout

```
résigner>INF -> résigner
résigner>INF -> résigner
résigner>INF -> résigner
résigner>INF -> résigner
résigner>INF -> résigner
résigner>INF -> résigner
résigner>INF -> résigner
résigner>INF -> résigner
résigner>INF -> résigner
résigner>INF -> résigner
```

French "resign"

Training Examples After Character Dropout

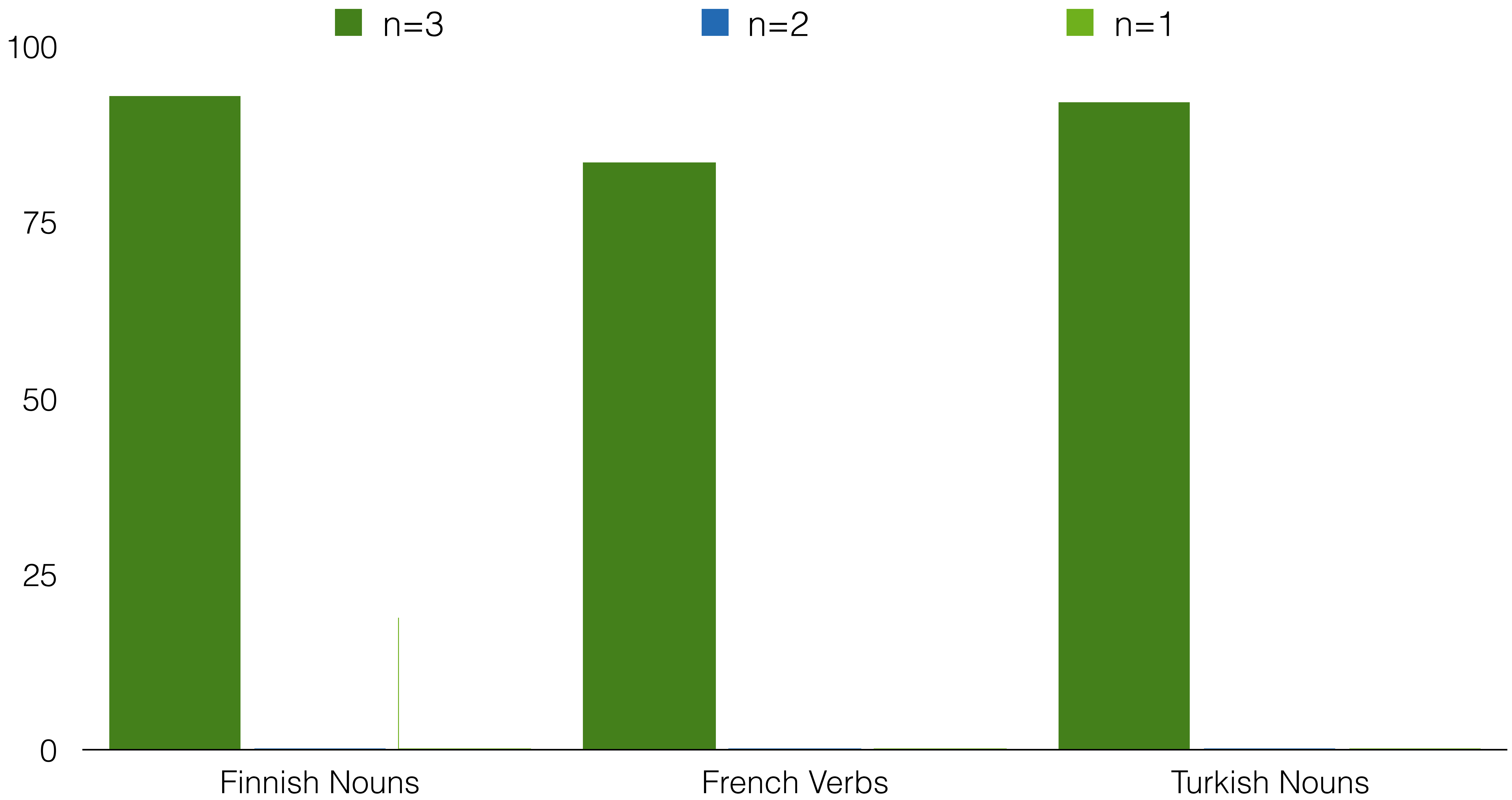
ré[redacted]i[redacted]>INF -> résigner
r[redacted]s[redacted]>INF -> résigner
rés[redacted]g[redacted]>INF -> résigner
résign[redacted]>INF -> résigner
résig[redacted]>INF -> résigner
r[redacted]si[redacted]>INF -> résigner
rési[redacted]>INF -> résigner
r[redacted]si[redacted]>INF -> résigner
rés[redacted]g[redacted]e>INF -> résigner
résig[redacted]>INF -> résigner

French
“resign”

Data Used in Experiments

We present experiments for noun and verb tables for: Finnish, French, Georgian, German, Latin, Latvian, Spanish and Turkish.

Results

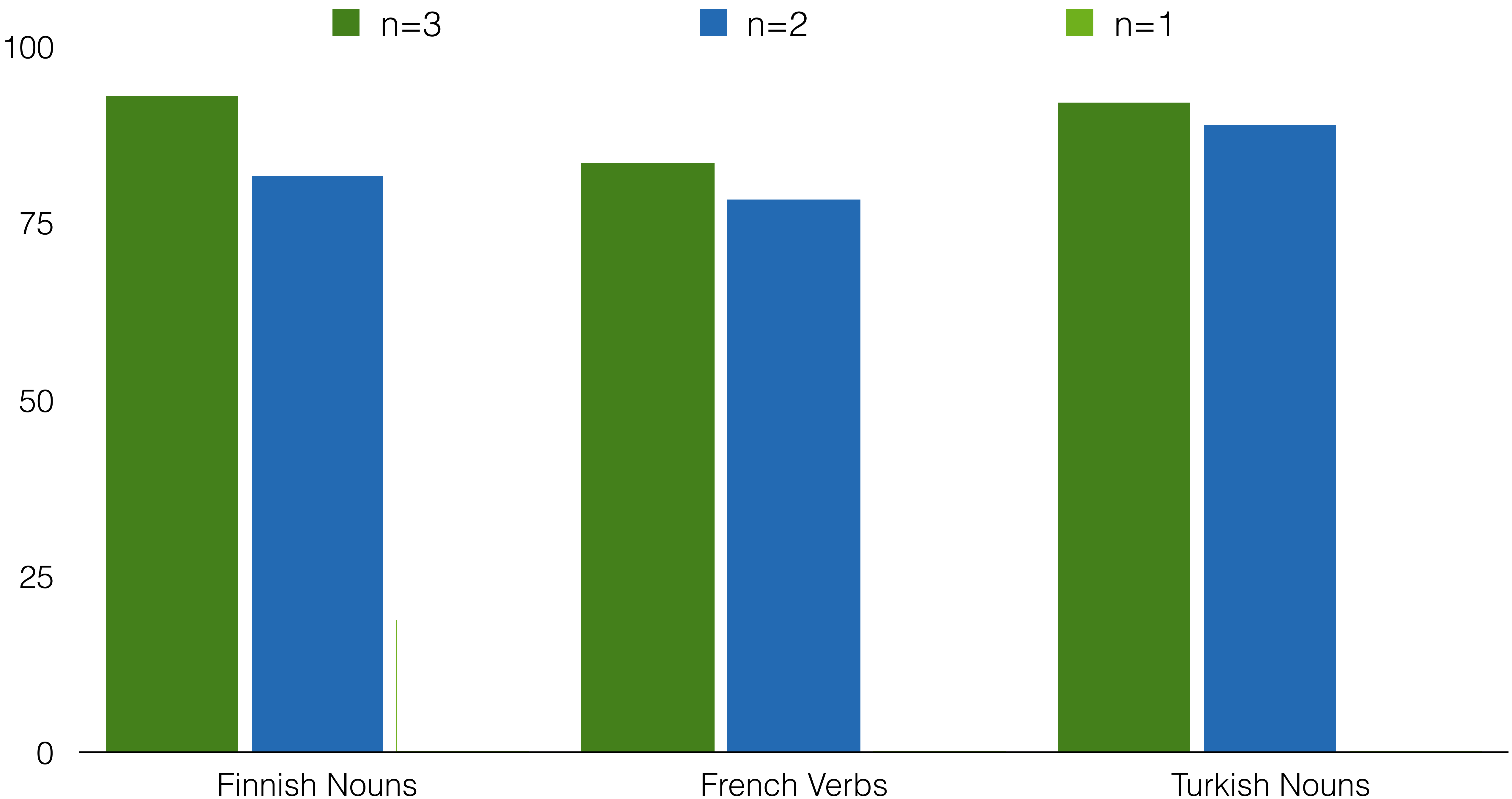


We use 1000 inflection tables for each language.

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

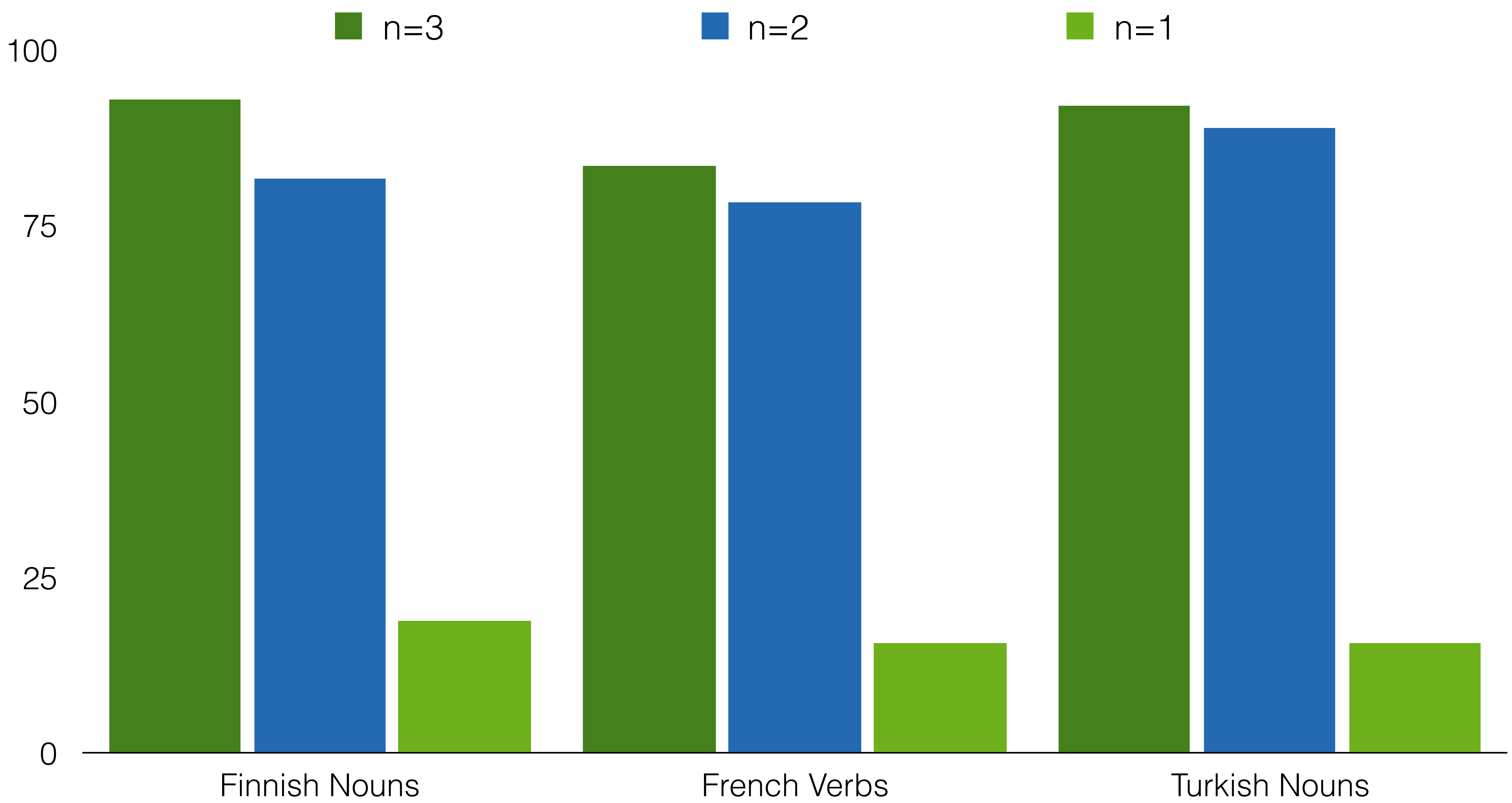
University of Colorado **Boulder**

Results



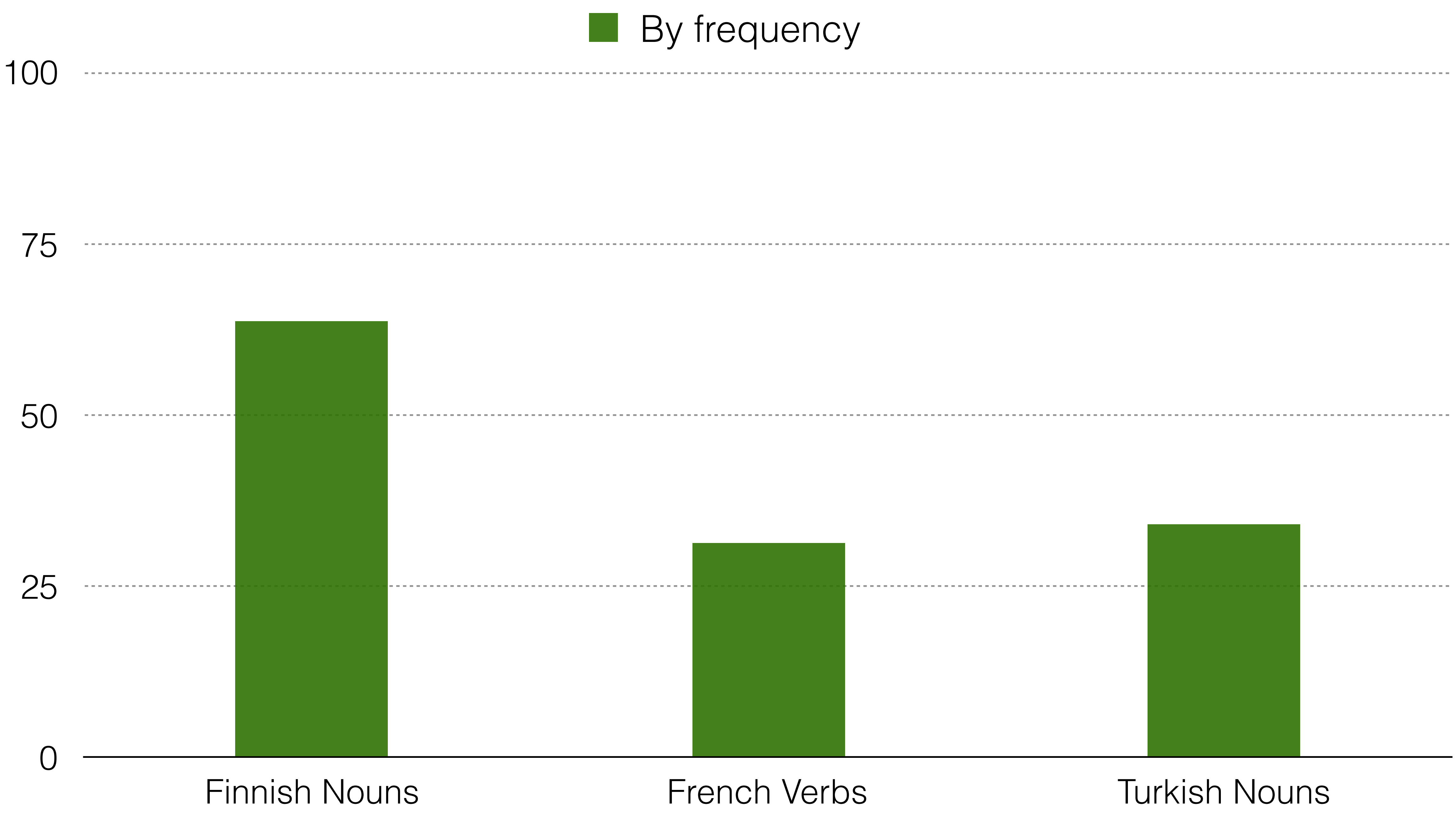
We use 1000 inflection tables for each language.

Results



We use 1000 inflection tables for each language.

Results


By frequency
100
75
50
25
0
Finnish Nouns
French Verbs
Turkish Nouns

Why are Results Low When Top 10k Forms are Given?

Difficult to learn rare forms.

Why are Results Low When Top 10k Forms are Given?

Difficult to learn rare forms.

Major problem: Syncretism

Why are Results Low When Top 10k Forms are Given?

Difficult to learn rare forms.

Major problem: Syncretism

Spießer	N,ACC,PL
Spießer	N,ACC,SG
Spießer	N,DAT,SG
Spießer	N,GEN,PL
	N,DAT,PL
Spießer	N,NOM,PL
Spießer	N,NOM,SG
	N,GEN,SG

German
“philistine”

Why are Results Low When Top 10k Forms are Given?

Difficult to learn rare forms.

Major problem: Syncretism

Spießer	N,ACC,PL
Spießer	N,ACC,SG
Spießer	N,DAT,SG
Spießer	N,GEN,PL
Spießern	N,DAT,PL
Spießer	N,NOM,PL
Spießer	N,NOM,SG
Spießers	N,GEN,SG

German "philistine"

Conclusions & Future Work

- We can learn inflectional morphology even when only given one example per lexeme.

- If every table has two or more forms, accuracy is around 90%.

- Predicting rare forms based on frequent ones is difficult.

- Future work: Need more realistic data set for the L1 learning scenario.

Thank you!