

# Contextualization of Morphological Inflection

Ekaterina Vylomova<sup>1</sup> Ryan Cotterell<sup>2</sup>  
Timothy Baldwin<sup>1</sup> Trevor Cohn<sup>1</sup> Jason Eisner<sup>2</sup>

<sup>1</sup>School of Computing and Information Systems  
The University of Melbourne

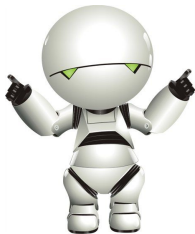
<sup>2</sup>Department of Computer Science  
Johns Hopkins University



THE UNIVERSITY OF  
MELBOURNE



JOHNS HOPKINS  
UNIVERSITY



This is Marvin:

OK, Marvin, which word comes next: *Two cats are \_\_\_\_\_*

Hmmm, let me guess ...



<i>sitting</i>	3:01	10	4
<i>play</i>	2:87	10	4
<i>running</i>	2:53	10	4
<i>nice</i>	2:32	10	4
<i>lost</i>	1:97	10	4
<i>playing</i>	1:66	10	4
<i>sat</i>	1:54	10	4
<i>plays</i>	1:32	10	4

: :

Let's add a constraint by providing a lemma: *Two cats are [PLAY]*

That narrows things down a lot ...

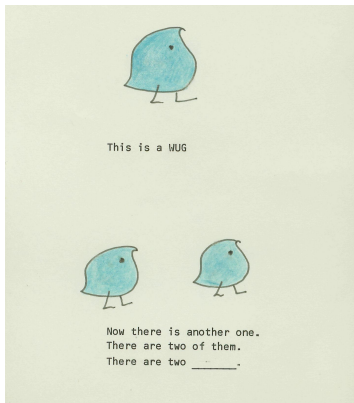


<i>sitting</i>	3:01	10	4
<i>play</i>	2:87	10	4
<i>running</i>	2:53	10	4
<i>nice</i>	2:32	10	4
<i>lost</i>	1:97	10	4
<i>playing</i>	1:66	10	4
<i>sat</i>	1:54	10	4
<i>plays</i>	1:32	10	4

:

:

Hey, this reminds me a bit of .... a wug ... and a second wug:



... as well as the SIGMORPHON morphological inflection task



## SIGMORPHON Shared Task 2016–2019

*PLAY* + PRESENT PARTICIPLE ! *playing*

*played* + PRESENT PARTICIPLE ! *playing*

Lemma	Tag	Form
<i>RUN</i>	PAST	<i>ran</i>
<i>RUN</i>	PRES;1SG	<i>run</i>
<i>RUN</i>	PRES;2SG	<i>run</i>
<i>RUN</i>	PRES;3SG	<i>runs</i>
<i>RUN</i>	PRES;PL	<i>run</i>
<i>RUN</i>	PART	<i>running</i>

2018 : 96% accuracy on avg.

in high-resource setting

Contextualization: But why choose PRESENT PARTICIPLE? Context!



## SIGMORPHON Shared Task 2016–2019

*PLAY* + PRESENT PARTICIPLE ! *playing*  
*played* + PRESENT PARTICIPLE ! *playing*

Contextualization: The tags must be inferred from the context!



## SIGMORPHON Shared Task 2018 Task 2

### SubTask 1

<i>Two</i>	<i>cats</i>	<i>are</i>	<i>???</i>	<i>together</i>
<i>TWO/NUM</i>	<i>CAT/N+PL</i>	<i>BE/AUX+PRES+3PL</i>	<i>PLAY</i>	<i>TOGETHER/ADV</i>

### SubTask 2

<i>Two</i>	<i>cats</i>	<i>are</i>	<i>???</i>	<i>together</i>
			<i>PLAY</i>	



Contextualization: The tags must be inferred from the context!



## SIGMORPHON Shared Task 2018 Task 2

### SubTask 1

<i>Two</i>	<i>cats</i>	<i>are</i>	<i>playing</i>	<i>together</i>
<i>TWO/NUM</i>	<i>CAT/N+PL</i>	<i>BE/AUX+PRES+3PL</i>	<i>PLAY</i>	<i>TOGETHER/ADV</i>

### SubTask 2

<i>Two</i>	<i>cats</i>	<i>are</i>	<i>playing</i>	<i>together</i>
			<i>PLAY</i>	

# A Hybrid (Structured-unstructured) Model

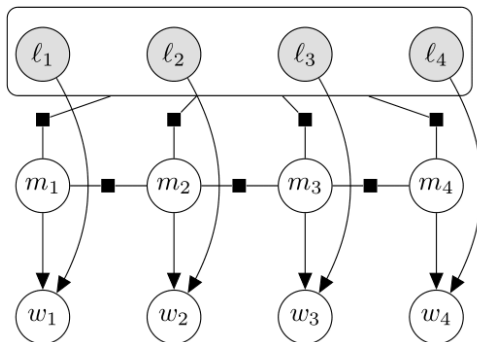
Let's predict both tags and forms!



Lemmatized  
Sequence

Predicted Tag  
Sequence

Predicted Form  
Sequence



# A Hybrid (Structured-unstructured) Model

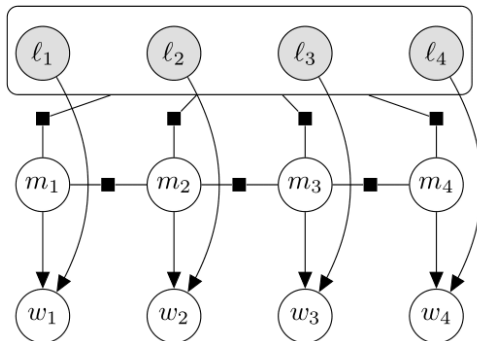
... or, in other words,  $p(w; m, j) = \left( \prod_{i=1}^n p(w_i, j_i; m_i) \right) p(m, j)$



Lemmatized  
Sequence

Predicted Tag  
Sequence

Predicted Form  
Sequence



# A Hybrid (Structured-unstructured) Model

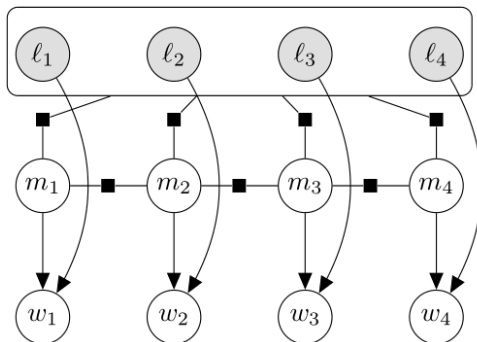
... or, in other words,  $p(w; m, j) = \left( \prod_{i=1}^n p(w_i, j_i; m_i) \right) p(m, j)$



Lemmatized  
Sequence

Predicted Tag  
Sequence

Predicted Form  
Sequence



Lample et al., 2016

$p(m, j)$

Neural CRF

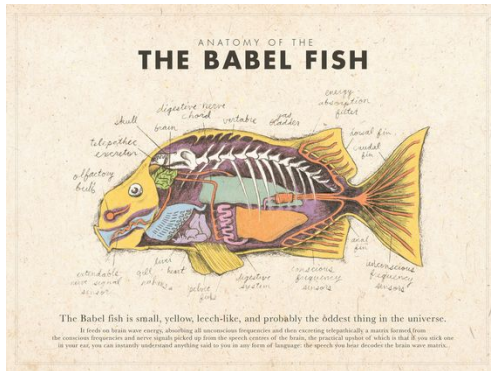
Aharoni et al., 2017

$p(w_i, j_i; m_i)$

Hard Monotonic Attention

# Languages and Grammar Categories

Let's test the model on a wide variety of languages!



# Languages and Grammar Categories

Languages differ in what is explicitly morphosyntactically marked, and how:



Bulgarian (bg), Slavic  
English (en), Germanic  
Basque (eu), Isolate  
Finnish (fi), Uralic  
Gaelic (ga), Celtic  
Hindi (hi), Indic  
Italian (it), Romance  
Latin (la), Romance  
Polish (pl), Slavic  
Swedish (sv), Germanic

<i>bg</i>	2	4	2	3	3	4	3	3
<i>en</i>	0	2	2	3	3	2	3	2
<i>eu</i>	4	15	2	3	2	2	3	0
<i>fi</i>	0	16	0	2	0	2	3	2
<i>ga</i>	0	4	2	1	2	2	3	3
<i>hi</i>	2	2	0	0	2	2	3	3
<i>it</i>	0	0	2	2	2	2	3	4
<i>la</i>	2	7	0	3	3	2	3	5
<i>pl</i>	2	7	0	2	3	2	3	3
<i>sv</i>	0	3	2	3	3	2	0	2
Aspect								
Case								
Definite								
Degree								
Gender								
Number								
Person								
Tense								

# Languages and Grammar Categories

Some languages use word order to express relations between words, while others use morphosyntactic marking:



English:

*Kim gives Sandy an interesting book*

Polish:

*Jenia daje Maszy ciekaw ksi k*

# Languages and Grammar Categories

Some languages use word order to express relations between words, while others use morphosyntactic marking:



English:

<i>Kim</i>	<i>gives</i>	<i>Sandy</i>	<i>an interesting</i>	<i>book</i>
Subject		IObject		DObject

Polish:

	<i>Jenia</i>	<i>daje</i>	<i>Maszy</i>	<i>ciekawą</i>	<i>ksi</i> <i>kę</i>
	Nom		Dat	Acc.Fem.Sg	Acc.Sg
==	<i>Maszy</i>	<i>daje</i>	<i>Jenia</i>	<i>ciekawą</i>	<i>ksi</i> <i>kę</i>
==	<i>ciekawą</i>	<i>ksi</i> <i>kę</i>	<i>daje</i>	<i>Jenia</i>	<i>Maszy</i>
!=	<i>Jenie</i>	<i>daje</i>	<i>Masza</i>	<i>ciekawą</i>	<i>ksi</i> <i>kę</i>



How well can such categories and corresponding forms be predicted in each language?



How well can such categories and corresponding forms be predicted in each language?

Do linguistic features enhance performance?



How well can such categories and corresponding forms be predicted in each language?

Do linguistic features enhance performance?

Does morphological complexity impact on empirical performance?



Nivre et al., 2016

Data: Universal Dependencies v1.2

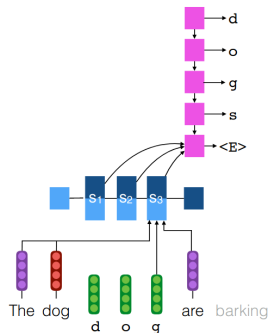
Baselines: the baseline of the SIGMORPHON 2018 shared task as well as the best performing system of that year



Cotterell et al., 2018

SM: biLSTM encoder-decoder with context window of size 2

- input = concat (left+right forms, lemma, tags, char-level center lemma)



Kementchedjhieva et al., 2018

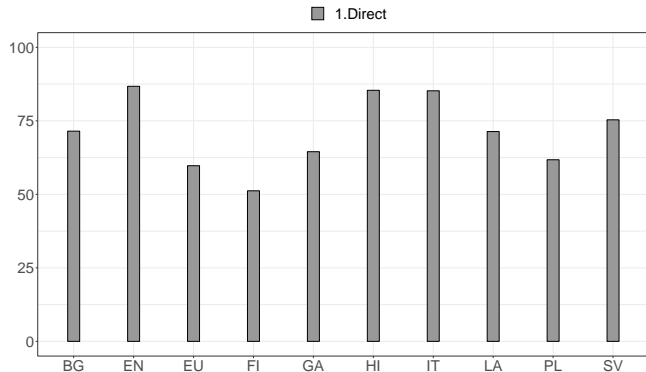
CPH: biLSTM encoder–decoder with no context window size restrictions

- input = concat (full context, lemma, tags, char-level center lemma)
- also predicts target tags as an auxiliary task

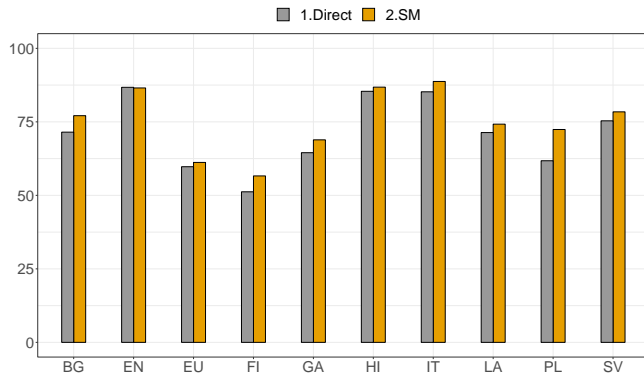
Direct: more basic model that relies only on forms and lemmas



Let's condition only on contextual forms and lemmas (1-best accuracy for form prediction):

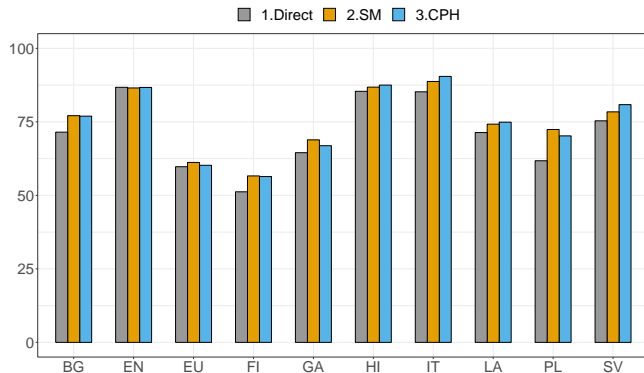


Now also supply contextual tag information, still predicting forms only:

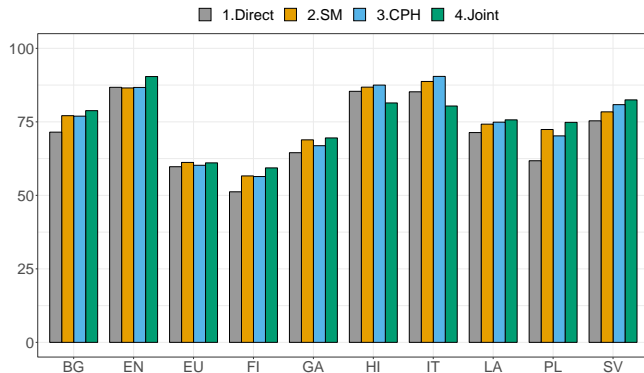




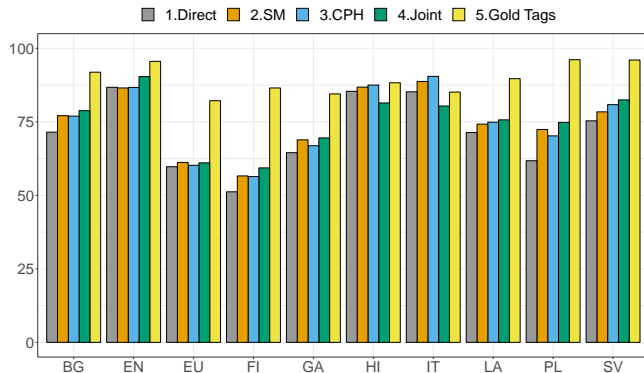
Now use a wider context and predict tags as an auxiliary task:



Finally, use neural CRF to predict tag sequence and hard monotonic attention model for forms:



How far are we from the results for forms predicted from gold tag sequence?





Q1: Do linguistic features help?

Yes, they do!

Most systems that make use of morphological tags outperform the “Direct” baseline on most languages

Joint prediction of tags and forms further improves the results





Q2: Does morphological complexity impact empirical performance?

Yes, it does!

Performance drops in languages with rich case systems such as Slavic and Uralic

The model needs to learn which grammatical categories should be in agreement





Q3: How well is agreement captured?

Adjective – Noun (AMod)

is captured quite well

Verb – Noun (Subject – Verb)

is more challenging, since agreement categories can vary depending on tense

General-purpose inference of agreement categories is still a challenging task!



Q4: Where does most uncertainty come from?

## Inherent and Contextual Morphological Categories

Contextual categories participate in agreement: adjective number, case, gender, verbal gender, etc.

Inherent express the speaker's intentions: noun number, verbal tense

Q4: Where does most uncertainty come from?

## Inherent and Contextual Morphological Categories

Contextual categories participate in agreement: adjective number, case, gender, verbal gender, etc.

Inherent express the speaker's intentions: noun number, verbal tense

Most uncertainty comes from inherent categories!



Q4: Where does most uncertainty come from?

## Inherent and Contextual Morphological Categories

Contextual categories participate in agreement: adjective number, case, gender, verbal gender, etc.

Inherent express the speaker's intentions: noun number, verbal tense

Most uncertainty comes from inherent categories!

Often such categories must be inferred

Q5: Which language is least affected by lemmatization?

Q5: Which language is least affected by lemmatization?

Q5: Which language is least affected by lemmatization?

Why English?

## Word Order vs. Morphology

Most information on roles and dependencies is expressed non-morphologically, e.g. in word order or by prepositions:

EN: Kim gives Sandy an interesting book KIM GIVE SANDY AN INTERESTING BOOK

PL: Jenia daje Maszy ciekawą książkę JENIA DAWA, MASZA CIEKAWY KSIĄŻKA

Q5: Which language is least affected by lemmatization?

Why English?

## Word Order vs. Morphology

Most information on roles and dependencies is expressed non-morphologically, e.g. in word order or by prepositions:

EN: Kim gives Sandy an interesting book KIM GIVE SANDY AN INTERESTING BOOK

SVO/Roles are still there

PL: Jenia daje Maszy ciekawą książkę JENIA DAWA MASZY CIEKAWY KSIĄŻKA

Flexible/Roles are partially lost

## Evaluation of grammaticality

How well do neural models model grammaticality?

## Data de-biasing (e.g., En–Ru )

*smart student ! umnyj.Nom.Masc.Sg student.Nom.Sg*

augment with:

*smart student ! umnaja.Nom.Fem.Sg studentka.Nom.Fem.Sg*



# Thank you! Questions?

