Contextualization of Morphological Inflection

> ¹School of Computing and Information Systems The University of Melbourne

> > ²Department of Computer Science Johns Hopkins University







This is Marvin:

э

イロト イポト イヨト イヨ

OK, Marvin, which word comes next: Two cats are

Hmmm, let me guess ...

$3.01 * 10^{-4}$
$2.87 * 10^{-4}$
$2.53 * 10^{-4}$
$2.32 * 10^{-4}$
$1.97 * 10^{-4}$
$1.66 * 10^{-4}$
$1.54 * 10^{-4}$
$1.32 * 10^{-4}$



Image: A mathematical states and a mathem

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

That narrows things down a lot ...

sitting	$3.01 * 10^{-4}$
play	$2.87 * 10^{-4}$
running	$2.53 * 10^{-4}$
nice	$2.32 * 10^{-4}$
lost	$1.97 * 10^{-4}$
playing	$1.66 * 10^{-4}$
sat	$1.54 * 10^{-4}$
plays	$1.32 * 10^{-4}$



Vylomova, Cotterell, Baldwin, Cohn, Eisner

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





Vylomova, Cotterell, Baldwin, Cohn, Eisner

Contextualization of Morphological Inflection

э

イロト 不同ト 不同ト 不同ト

... as well as the SIGMORPHON morphological inflection task

SIGMORPHON Shared Task 2016–2019

PLAY + PRESENT PARTICIPLE \rightarrow playing played + PRESENT PARTICIPLE \rightarrow playing

Lemma	Tag	Form	
RUN	PAST	ran	_
RUN	PRES;1SG	run	
RUN	PRES;2SG	run	2018 :~ 96% accuracy on avg.
RUN	PRES;3SG	runs	in high-resource setting
RUN	PRES;PL	run	In figh-resource setting
RUN	PART	running	 <



Contextualization of Morphological Inflection

Contextualization: But why choose PRESENT PARTICIPLE? Context!



SIGMORPHON Shared Task 2016–2019

PLAY + PRESENT PARTICIPLE \rightarrow playing played + PRESENT PARTICIPLE \rightarrow playing

Contextualization: The tags must be inferred from the context!



SIGMORPHON Shared Task 2018 Task 2

		SubTask 1		
Two cats TWO/NUM CAT/N+PL		<i>cats are</i> <i>CAT</i> /N+PL <i>BE</i> /AUX+PRES+3PL		together TOGETHER/ADV
		SubTask 2		
Two cats		are	??? PLAY	together

Contextualization: The tags must be inferred from the context!



SIGMORPHON Shared Task 2018 Task 2

		SubTask 1		
<i>Two cats</i> <i>TWO</i> /NUM <i>CAT</i> /N+PL		<i>are</i> PL <i>BE</i> /AUX+PRES+3PL		together TOGETHER/ADV
		SubTask 2		
Two cats		are	playing PLAY	together

A Hybrid (Structured-unstructured) Model

Let's predict both tags and forms!





< 47 ►

A Hybrid (Structured-unstructured) Model

... or, in other words, $p(\mathbf{w}, \mathbf{m} \mid \boldsymbol{\ell}) = (\prod_{i=1}^{n} p(w_i \mid \ell_i, m_i)) p(\mathbf{m} \mid \boldsymbol{\ell})$





▶ ∢ ⊒

A Hybrid (Structured-unstructured) Model

... or, in other words, $p(\mathbf{w}, \mathbf{m} \mid \boldsymbol{\ell}) = (\prod_{i=1}^{n} p(w_i \mid \ell_i, m_i)) p(\mathbf{m} \mid \boldsymbol{\ell})$





Vylomova, Cotterell, Baldwin, Cohn, Eisner

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

							-		1
19 ⁶ 0		4	2	3	3	4	3		
er	0	2	2	3	3	2	3	2	
ev	4	15	2	3	2		3	0	
\$	0	16	0	2	0		3	2	
5 %	0	4	2	1	2		3	3	
Ŵ	2	2	0	0	2		3		
*	0	0	2	2	2		3	4	
12	2	7	0	3	3	2	3	5	
è		7	0	2	3		3	3	
Ş	0	3	2	3	3		0	2	
	spect	Case	finite	egree	ender	mber	erson	Fense	
	A		De	D	Ğ	n	ď		

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 ciekawą książkę

Languages and Grammar Categories

Kim

Subject

gives

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

Sandy

IObject

English:

Polish:

	Jen ia	daje	Maszy	ciekaw ą	książk ę
	Nom		Dat	Acc.Fem.Sg	Acc.Sg
==	Maszy	daje	Jen ia	ciekaw ą	książk ę
==	ciekaw ą	książk ę	daje	Jen ia	Masz y
!=	Jen ie	daje	<i>Masza</i>	ciekaw ą	książk ę
		5		2	

an

interesting

book

DObject





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



Image: A math a math

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)





Vylomova, Cotterell, Baldwin, Cohn, Eisner

Contextualization of Morphological Inflection

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?

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?

・ロト ・ 一下・ ・ ヨト・



Q5: Which language is least affected by lemmatization?





A B > A B > A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A



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:







Q5: Which language is least affected by lemmatization?

Why English?

(日) (同) (日) (日)

Word Order vs. Morphology

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



Evaluation of grammaticality

How well do neural models model grammaticality?

Data de-biasing (e.g., En-Ru)

 $\textit{smart student} \rightarrow \textit{umnyj.Nom.Masc.Sg student.Nom.Sg}$

augment with:

smart student \rightarrow umnaja.Nom.Fem.Sg studentka.Nom.Fem.Sg



Thank you! Questions?





Vylomova, Cotterell, Baldwin, Cohn, Eisner

Contextualization of Morphological Inflection

・ロト ・四ト ・ヨト ・ヨ