Contextualization of Morphological Inflection

Ekaterina Vylomova\textsuperscript{1}  Ryan Cotterell\textsuperscript{2}  
Timothy Baldwin\textsuperscript{1}  Trevor Cohn\textsuperscript{1}  Jason Eisner\textsuperscript{2}

\textsuperscript{1}School of Computing and Information Systems  
The University of Melbourne

\textsuperscript{2}Department of Computer Science  
Johns Hopkins University
Language Modelling

This is Marvin:
Language Modelling

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

Hmmm, let me guess ...

<table>
<thead>
<tr>
<th>noun</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>sitting</td>
<td>3.01 * 10^{-4}</td>
</tr>
<tr>
<td>play</td>
<td>2.87 * 10^{-4}</td>
</tr>
<tr>
<td>running</td>
<td>2.53 * 10^{-4}</td>
</tr>
<tr>
<td>nice</td>
<td>2.32 * 10^{-4}</td>
</tr>
<tr>
<td>lost</td>
<td>1.97 * 10^{-4}</td>
</tr>
<tr>
<td>playing</td>
<td>1.66 * 10^{-4}</td>
</tr>
<tr>
<td>sat</td>
<td>1.54 * 10^{-4}</td>
</tr>
<tr>
<td>plays</td>
<td>1.32 * 10^{-4}</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
</tbody>
</table>
Language Modelling

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

That narrows things down a lot ...

<table>
<thead>
<tr>
<th>Verb</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>sitting</td>
<td>$3.01 \times 10^{-4}$</td>
</tr>
<tr>
<td>play</td>
<td>$2.87 \times 10^{-4}$</td>
</tr>
<tr>
<td>running</td>
<td>$2.53 \times 10^{-4}$</td>
</tr>
<tr>
<td>nice</td>
<td>$2.32 \times 10^{-4}$</td>
</tr>
<tr>
<td>lost</td>
<td>$1.97 \times 10^{-4}$</td>
</tr>
<tr>
<td>playing</td>
<td>$1.66 \times 10^{-4}$</td>
</tr>
<tr>
<td>sat</td>
<td>$1.54 \times 10^{-4}$</td>
</tr>
<tr>
<td>plays</td>
<td>$1.32 \times 10^{-4}$</td>
</tr>
</tbody>
</table>
Hey, this reminds me a bit of .... a wug ... and a second wug:

This is a WUG

Now there is another one.
There are two of them.
There are two _____.

Language Modelling
... as well as the SIGMORPHON morphological inflection task

SIGMORPHON Shared Task 2016–2019

$PLAY + \text{PRESENT PARTICIPLE} \rightarrow \text{playing}$

$played + \text{PRESENT PARTICIPLE} \rightarrow \text{playing}$

<table>
<thead>
<tr>
<th>Lemma</th>
<th>Tag</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>RUN</td>
<td>PAST</td>
<td>ran</td>
</tr>
<tr>
<td>RUN</td>
<td>PRES;1SG</td>
<td>run</td>
</tr>
<tr>
<td>RUN</td>
<td>PRES;2SG</td>
<td>run</td>
</tr>
<tr>
<td>RUN</td>
<td>PRES;3SG</td>
<td>runs</td>
</tr>
<tr>
<td>RUN</td>
<td>PRES;PL</td>
<td>run</td>
</tr>
<tr>
<td>RUN</td>
<td>PART</td>
<td>running</td>
</tr>
</tbody>
</table>

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!

<table>
<thead>
<tr>
<th>SubTask 1</th>
<th>SubTask 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Two</strong> TWO/NUM</td>
<td><strong>Two</strong></td>
</tr>
<tr>
<td>cats CAT/N+PL</td>
<td>cats</td>
</tr>
<tr>
<td>are BE/AUX+PRES+3PL</td>
<td>are</td>
</tr>
<tr>
<td>?? PLAY</td>
<td>?? PLAY</td>
</tr>
<tr>
<td>together TOGETHER/ADV</td>
<td>together</td>
</tr>
</tbody>
</table>
**Contextualization:** The tags must be inferred from the context!

---

**SIGMORPHON Shared Task 2018 Task 2**

<table>
<thead>
<tr>
<th>SubTask 1</th>
<th>SubTask 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Two</em></td>
<td><em>Two</em></td>
</tr>
<tr>
<td><strong>TWO/NUM</strong></td>
<td><strong>TWO/NUM</strong></td>
</tr>
<tr>
<td><em>cats</em></td>
<td><em>cats</em></td>
</tr>
<tr>
<td><strong>CAT/N+PL</strong></td>
<td><strong>CAT/N+PL</strong></td>
</tr>
<tr>
<td><em>are</em></td>
<td><em>are</em></td>
</tr>
<tr>
<td><strong>BE/AUX+PRES+3PL</strong></td>
<td><strong>BE/AUX+PRES+3PL</strong></td>
</tr>
<tr>
<td><em>playing</em></td>
<td><em>playing</em></td>
</tr>
<tr>
<td><strong>PLAY</strong></td>
<td><strong>PLAY</strong></td>
</tr>
<tr>
<td><em>together</em></td>
<td><em>together</em></td>
</tr>
<tr>
<td><strong>TOGETHER/ADV</strong></td>
<td><strong>TOGETHER/ADV</strong></td>
</tr>
</tbody>
</table>
A Hybrid (Structured–unstructured) Model

Let’s predict both tags and forms!
... or, in other words,  \( p(w, m \mid \ell) = (\prod_{i=1}^{n} p(w_i \mid \ell_i, m_i)) p(m \mid \ell) \)
... or, in other words, $p(w, m \mid \ell) = (\prod_{i=1}^{n} p(w_i \mid \ell_i, m_i)) \ p(m \mid \ell)$
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:

<table>
<thead>
<tr>
<th>Language</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulgarian (bg), Slavic</td>
<td></td>
</tr>
<tr>
<td>English (en), Germanic</td>
<td></td>
</tr>
<tr>
<td>Basque (eu), Isolate</td>
<td></td>
</tr>
<tr>
<td>Finnish (fi), Uralic</td>
<td></td>
</tr>
<tr>
<td>Gaelic (ga), Celtic</td>
<td></td>
</tr>
<tr>
<td>Hindi (hi), Indic</td>
<td></td>
</tr>
<tr>
<td>Italian (it), Romance</td>
<td></td>
</tr>
<tr>
<td>Latin (la), Romance</td>
<td></td>
</tr>
<tr>
<td>Polish (pl), Slavic</td>
<td></td>
</tr>
<tr>
<td>Swedish (sv), Germanic</td>
<td></td>
</tr>
</tbody>
</table>

Diagram showing the distribution of morphosyntactic features across different languages.
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

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

English:

Kim gives Sandy an interesting book
Subject IObject DObject

Polish:

Jenia daje Maszy ciekawą książkę
Nom Dat Acc.Fem.Sg Acc.Sg

Maszy daje Jenia ciekawą książkę
==

ciekawą książkę daje Jenia Maszy
==

Maszy daje Jenie ciekawą książkę
!=

Vylomova, Cotterell, Baldwin, Cohn, Eisner
Contextualization of Morphological Inflection
Experiments

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

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

Do linguistic features enhance performance?
Experiments

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?
Experiments

**Data:** Universal Dependencies v1.2

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

**SM**: biLSTM encoder–decoder with context window of size 2

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

**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

Kementchedjiev et al., 2018
Experiments

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

![Bar chart showing accuracy for different languages]
Experiments

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

![Bar chart showing the comparison between 1.Direct and 2.SM for different languages: BG, EN, EU, FI, GA, HI, IT, LA, PL, SV. The chart displays the percentage of correct predictions for each language and comparison type.](image-url)
Experiments

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

![Graph](Image)

- BG
- EN
- EU
- FI
- GA
- HI
- IT
- LA
- PL
- SV

Options:
- 1.Direct
- 2.SM
- 3.CPH
Finally, use neural CRF to predict tag sequence and hard monotonic attention model for forms:
Experiments

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!
Discussion

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?
Discussion

Q5: Which language is least affected by lemmatization?

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 DAWAĆ’ MASZA CIEKAWY KSIĄŻKA*

Why English?
Discussion

Q5: Which language is least affected by lemmatization?

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 \(\rightarrow\) KIM GIVE SANDY AN INTERESTING BOOK
PL: Jenia daje Maszy ciekawą książkę \(\rightarrow\) JENIA DAWAĆ’ MASZA CIEKAWY KSIĄŻKA

Why English?

SVO/Roles are still there

Flexible/Roles are partially lost
Future Directions

Evaluation of grammaticality
How well do neural models model grammaticality?

Data de-biasing (e.g., En–Ru)

smart student $\rightarrow$ umnyj.Nom.Masc.Sg student.Nom.Sg

augment with:

smart student $\rightarrow$ umnaja.Nom.Fem.Sg studentka.Nom.Fem.Sg
Thank you! Questions?