In Search of Abstract Morphological Structure

Micha Elsner, Ohio State University
2021
Morphology: grammatical information within the word

English: *cat* (SG) ~ *cats* (PL)
Morphology: grammatical information within the word

English: *cat* (SG) ~ *cats* (PL)

But also:

Latin:  
*laudāverītis*  
“You all would have praised”

Murrinhpatha (Forshaw et al 2016):  
*bamngingkardunginthha*  
“The two non-siblings (either ♀♀ or ♀♂) saw me”
A **lexeme** (“a word”) has a number of related **forms**, which make up its **paradigm**.

The set of properties that a form realizes is its **cell**.

<table>
<thead>
<tr>
<th>N.SG</th>
<th>cat</th>
</tr>
</thead>
<tbody>
<tr>
<td>N.PL</td>
<td>cats</td>
</tr>
</tbody>
</table>

Of course, some of these tables are bigger than others...
What goes in the cells?

Language users constantly face data sparsity when trying to remember what goes where (Bybee 2003, Ackerman and Malouf 2013)

Challenges from:

- Unfamiliar lexemes
- Rare cells
- Less-dominant language with lower exposure
Both lexemes and cells are Zipfian

(Lignos and Yang 2016)

**Figure 27.1.** Frequencies of CHILDES Spanish lemmas across inflection categories
Luckily, analogy can fill in some of the gaps

<table>
<thead>
<tr>
<th>Across words</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N.SG</td>
<td>cat</td>
<td>mink</td>
</tr>
<tr>
<td>N.PL</td>
<td>cats</td>
<td>?</td>
</tr>
</tbody>
</table>
Luckily, analogy can fill in some of the gaps

<table>
<thead>
<tr>
<th>N.SG</th>
<th>Across words</th>
<th>N.PL</th>
<th>Across cells</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cat</td>
<td>cats</td>
<td>“You praise”</td>
</tr>
<tr>
<td></td>
<td>mink</td>
<td>?</td>
<td>“You all praise”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“You would praise”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“You all would praise”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“You would have praised”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“You all would have praised”</td>
</tr>
</tbody>
</table>
Abstraction

These analogies copy bits of structure (the -s and -ti-) within the word.

But not all analogies work this way.

<table>
<thead>
<tr>
<th>Latin</th>
<th>PL “woman”</th>
<th>PL “fire”</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM</td>
<td>fēmin-ae</td>
<td>ign-ēs</td>
</tr>
<tr>
<td>GEN</td>
<td>fēmin-ārum</td>
<td>ign-ium</td>
</tr>
<tr>
<td>DAT</td>
<td>fēmin-īs</td>
<td>ign-ibus</td>
</tr>
<tr>
<td>ACC</td>
<td>fēmin-ās</td>
<td>ign-ēs</td>
</tr>
<tr>
<td>ABL</td>
<td>fēmin-īs</td>
<td>ign-ibus</td>
</tr>
</tbody>
</table>
Generalizations about paradigms (Wurzel 1989)

Which cells are the same/different

Which cells are easy/hard to predict

Where and how the morphological marking appears in the word

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>Irish</th>
</tr>
</thead>
<tbody>
<tr>
<td>N.SG</td>
<td>cat</td>
<td>mín</td>
</tr>
<tr>
<td>N.PL</td>
<td>cats</td>
<td>mínte</td>
</tr>
<tr>
<td></td>
<td>ox</td>
<td>gáibéal</td>
</tr>
<tr>
<td></td>
<td>oxen</td>
<td>gáibéil</td>
</tr>
</tbody>
</table>
Paradigm discovery

The paradigm discovery problem: Erdmann, Elsner, Wu, Cotterell and Habash, ACL 2020

Modeling morphological learning, typology, and change: What can the neural sequence-to-sequence framework contribute?: Elsner, Sims, Erdmann, et al, JLM 2019

Alex Erdmann (Ph.D. 2020)
The problem: paradigms from raw text

The cat *watched* me *watching* it.
I *followed* the show but she hadn’t *seen* it.
Let’s *see* who *follows* your logic.

Which types fill the same cells?
Which ones are forms of the same lexeme?
What fills in the unattested cells?
The problem: paradigms from raw text

The cat **watched** me **watching** it.
I **followed** the show but she hadn’t **seen** it.
Let’s **see** who **follows** your logic.

Which types fill the same cells?
Which ones are forms of the same lexeme?
What fills in the unattested cells?

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>see</td>
</tr>
<tr>
<td>C2</td>
<td>follows</td>
</tr>
<tr>
<td>C3</td>
<td>watching</td>
</tr>
<tr>
<td>C4</td>
<td>watched, followed</td>
</tr>
<tr>
<td>C5</td>
<td>seen</td>
</tr>
</tbody>
</table>
The problem: paradigms from raw text

The cat **watched** me **watching** it.
I **followed** the show but she hadn’t **seen** it.
Let’s **see** who **follows** your logic.

<table>
<thead>
<tr>
<th>“watch”</th>
<th>“follow”</th>
<th>“see”</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>?</td>
<td>see</td>
</tr>
<tr>
<td>?</td>
<td>follows</td>
<td>?</td>
</tr>
<tr>
<td>watching</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>watched</td>
<td>followed</td>
<td>?</td>
</tr>
</tbody>
</table>

Which types fill the same cells?
Which ones are forms of the same lexeme?
What fills in the unattested cells?
The problem: paradigms from raw text

The cat **watched** me **watching** it.
I **followed** the show but she hadn’t **seen** it.
Let’s **see** who **follows** your logic.

Which types fill the same cells?

Which ones are forms of the same lexeme?

What fills in the unattested cells?

<table>
<thead>
<tr>
<th>“watch”</th>
<th>“follow”</th>
<th>“see”</th>
</tr>
</thead>
<tbody>
<tr>
<td>watch</td>
<td>follow</td>
<td>see</td>
</tr>
<tr>
<td>watches</td>
<td>follows</td>
<td>sees</td>
</tr>
<tr>
<td>watching</td>
<td>following</td>
<td>seeing</td>
</tr>
<tr>
<td>watched</td>
<td>followed</td>
<td>saw</td>
</tr>
<tr>
<td>watched</td>
<td>followed</td>
<td>seen</td>
</tr>
</tbody>
</table>
Step 1: finding cells

Represent the word types with FastText (Bojanowski et al 2017), a Word2Vec-like method that incorporates substring information.

Ideally, things with similar distribution and string markers will be embedded close together.

Use $k$-means to extract hard clusters.
Results: clustering into cells is challenging

<table>
<thead>
<tr>
<th>Language</th>
<th>F-score of paradigm cell mates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic nouns</td>
<td>39.9</td>
</tr>
<tr>
<td>German nouns</td>
<td>35.2</td>
</tr>
<tr>
<td>English verbs</td>
<td>64.0</td>
</tr>
<tr>
<td>Latin nouns</td>
<td>38.8</td>
</tr>
<tr>
<td>Russian nouns</td>
<td>44.5</td>
</tr>
</tbody>
</table>
Step 2: finding lexemes

Basic idea: forms of the same lexeme *usually* consist of some shared content (the **theme**) and some cell-specific marking (**distinguishers**)
Building lexemes incrementally

Consider adding a new form to the lexeme…

Adding “watches” adds 5 theme characters and 2 unexplained characters.

Adding “water” adds 3 theme characters and 6 unexplained characters.
After each pass, re-estimate the distinguishers
Results: clustering into lexemes is not too bad

<table>
<thead>
<tr>
<th>Language</th>
<th>F-score of lexeme mates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic nouns</td>
<td>48.5</td>
</tr>
<tr>
<td>German nouns</td>
<td>59.4</td>
</tr>
<tr>
<td>English verbs</td>
<td>80.1</td>
</tr>
<tr>
<td>Latin nouns</td>
<td>73.2</td>
</tr>
<tr>
<td>Russian nouns</td>
<td>72.2</td>
</tr>
</tbody>
</table>
Filling the gaps

Use a small Transformer network (Wu and Cotterell 2020) to fill the empty slots

<table>
<thead>
<tr>
<th>“watch”</th>
<th>“follow”</th>
<th>“see”</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>?</td>
<td>see</td>
</tr>
<tr>
<td>?</td>
<td>follows</td>
<td>?</td>
</tr>
<tr>
<td>watching</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>watched</td>
<td>followed</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>seen</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>“watch”</th>
<th>“follow”</th>
<th>“see”</th>
</tr>
</thead>
<tbody>
<tr>
<td>watch</td>
<td>follow</td>
<td>see</td>
</tr>
<tr>
<td>watches</td>
<td>follows</td>
<td>sees</td>
</tr>
<tr>
<td>watching</td>
<td>following</td>
<td>seeing</td>
</tr>
<tr>
<td>watched</td>
<td>followed</td>
<td>saw</td>
</tr>
<tr>
<td>watched</td>
<td>followed</td>
<td>seen</td>
</tr>
</tbody>
</table>
Results: filling empty slots is also difficult

<table>
<thead>
<tr>
<th></th>
<th>% missing forms predicted (anywhere)</th>
<th>Supervised</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic nouns</td>
<td>49.5</td>
<td>87.0</td>
</tr>
<tr>
<td>German nouns</td>
<td>56.7</td>
<td>74.9</td>
</tr>
<tr>
<td>English verbs</td>
<td>67.5</td>
<td>80.7</td>
</tr>
<tr>
<td>Latin nouns</td>
<td>72.9</td>
<td>88.0</td>
</tr>
<tr>
<td>Russian nouns</td>
<td>86.2</td>
<td>96.8</td>
</tr>
</tbody>
</table>
Why is this so tough?

Things that look the same don’t always belong together:

<table>
<thead>
<tr>
<th>NOM</th>
<th>GEN</th>
<th>DAT</th>
<th>ACC</th>
<th>ABL</th>
<th>NOM</th>
<th>GEN</th>
<th>DAT</th>
<th>ACC</th>
<th>ABL</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>serv-us</td>
<td>i</td>
<td>o</td>
<td>um</td>
<td>o</td>
<td>i</td>
<td>orum</td>
<td>is</td>
<td>os</td>
<td>is</td>
<td>“slave.M”</td>
</tr>
<tr>
<td>serv-a</td>
<td>ae</td>
<td>ae</td>
<td>am</td>
<td>a</td>
<td>ae</td>
<td>arum</td>
<td>is</td>
<td>as</td>
<td>is</td>
<td>“slave.F”</td>
</tr>
<tr>
<td>frat-er</td>
<td>ris</td>
<td>ri</td>
<td>rem</td>
<td>re</td>
<td>res</td>
<td>rum</td>
<td>ribus</td>
<td>res</td>
<td>ribus</td>
<td>“brother”</td>
</tr>
</tbody>
</table>

This can cause cascading errors. If we decide *servum* and *servorum* belong in the same cell, we must assign them to different lexemes.

If we put *fratrum* and *servum* in the same cell, we hide the real correspondence *fratrum* ~ *servorum*. 
What’s missing

More sensitive use of context (but this isn’t a cure-all, especially in languages with relatively free word order)

Better use of paradigm structure; use internal consistency of the tables to avoid being misled by surface overlap.

Better integration of the transformer: can feedback from the neural learner help to exclude implausible guesses?
Work is still ongoing

A SigMorphon 2021 shared task (Wiemerslage et al) investigates a similar problem.

(Bayesian) clustering-based systems are still state-of-the-art (McCurdy et al)

And scores are still comparatively low...
Looking for paradigm structure

Formalizing Inflectional Paradigm Shape with Information Theory: Lefevre, Elsner and Sims, SCiL 2021

Grace LeFevre
(B.A. 2021, now Ph.D. program at Northwestern)
"Paradigm shape" in Spanish

<table>
<thead>
<tr>
<th>LEXEME</th>
<th>GLOSS</th>
<th>PRS.1SG</th>
<th>PRS.2SG</th>
<th>PRS.3SG</th>
<th>PRS.1PL</th>
<th>PRS.2PL</th>
<th>PRS.3PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>CANTAR</td>
<td>‘sing’</td>
<td>canto</td>
<td>cantas</td>
<td>canta</td>
<td>cantamos</td>
<td>cantáis</td>
<td>cantan</td>
</tr>
<tr>
<td>SUBIR</td>
<td>‘rise’</td>
<td>subo</td>
<td>subes</td>
<td>sube</td>
<td>subimos</td>
<td>subís</td>
<td>suben</td>
</tr>
<tr>
<td>SENTIR</td>
<td>‘feel’</td>
<td>siento</td>
<td>sientes</td>
<td>siente</td>
<td>sentimos</td>
<td>sentís</td>
<td>sienten</td>
</tr>
<tr>
<td>PENSAR</td>
<td>‘think’</td>
<td>pienso</td>
<td>piensas</td>
<td>piensa</td>
<td>pensamos</td>
<td>pensáis</td>
<td>piensan</td>
</tr>
<tr>
<td>MOVER</td>
<td>‘move’</td>
<td>muevo</td>
<td>mueses</td>
<td>mueve</td>
<td>movemos</td>
<td>movéis</td>
<td>mueven</td>
</tr>
</tbody>
</table>

Traditional analysis of the Spanish verbs, based on the vowels in suffixes.
**“Paradigm shape” in Spanish**

<table>
<thead>
<tr>
<th>LEXEME</th>
<th>GLOSS</th>
<th>PRS.1SG</th>
<th>PRS.2SG</th>
<th>PRS.3SG</th>
<th>PRS.1PL</th>
<th>PRS.2PL</th>
<th>PRS.3PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>CANTAR</td>
<td>‘sing’</td>
<td>canto</td>
<td>cantas</td>
<td>canta</td>
<td>cantamos</td>
<td>cantáis</td>
<td>cantan</td>
</tr>
<tr>
<td>SUBIR</td>
<td>‘rise’</td>
<td>subo</td>
<td>subes</td>
<td>sube</td>
<td>subimos</td>
<td>subís</td>
<td>suben</td>
</tr>
<tr>
<td>SENTIR</td>
<td>‘feel’</td>
<td>siento</td>
<td>sientes</td>
<td>siente</td>
<td>sentimos</td>
<td>sentís</td>
<td>sienten</td>
</tr>
<tr>
<td>PENSAR</td>
<td>‘think’</td>
<td>pienso</td>
<td>piensas</td>
<td>piensa</td>
<td>pensamos</td>
<td>pensáis</td>
<td>piensan</td>
</tr>
<tr>
<td>MOVER</td>
<td>‘move’</td>
<td>muevo</td>
<td>mueves</td>
<td>mueve</td>
<td>movemos</td>
<td>movéis</td>
<td>mueven</td>
</tr>
</tbody>
</table>

But it’s also well-known that some verbs have stem alternations.
*This distribution is what Maiden (2005) terms the N-pattern*
Measure similarity in a flexible way

What is the underlying structure of the verb classes?

*Cantar* and *pensar* share their stem vowel, but not their alternation pattern.

Do the verbs form real “clusters”, and if so, which organizing principles matter?
"Confusion sets"

<table>
<thead>
<tr>
<th>PRS.1SG</th>
<th>PRS.2SG</th>
<th>theme</th>
<th>distinguishers</th>
</tr>
</thead>
<tbody>
<tr>
<td>canto</td>
<td>cantas</td>
<td>cant</td>
<td>o, as</td>
</tr>
<tr>
<td>subo</td>
<td>subes</td>
<td>sub</td>
<td>o, es</td>
</tr>
<tr>
<td>siento</td>
<td>sientes</td>
<td>sient</td>
<td>o, es</td>
</tr>
<tr>
<td>pienso</td>
<td>piensas</td>
<td>piens</td>
<td>o, as</td>
</tr>
<tr>
<td>muevo</td>
<td>mueves</td>
<td>muev</td>
<td>o, es</td>
</tr>
</tbody>
</table>

Alignment-based segmentation, following Beniamine et al. (2017)
Segmenting Stem Alternations

<table>
<thead>
<tr>
<th>PRS.1SG</th>
<th>PRS.1PL</th>
<th>theme</th>
<th>distinguishers</th>
</tr>
</thead>
<tbody>
<tr>
<td>subo</td>
<td>subimos</td>
<td>subo</td>
<td>--, ims</td>
</tr>
<tr>
<td>nuevo</td>
<td>movemos</td>
<td>mvo</td>
<td>ue, oems</td>
</tr>
</tbody>
</table>
Method sketch

Generate matrix of entropy values quantifying relations in the inflectional system

Segment locally for subsets of cells

Divide lexemes into “confusion sets” with identical distinguishers

Calculate entropy values: larger confusion sets get higher values

Use these entropies as coordinates in an embedding space

Verbs are similar for a set of columns if their forms are similarly easy or hard to predict
t-SNE Visualization

- Allomorphic groupings based on inflectional exponents
- Distributional groupings based on Maiden’s (2005) stem alternations
Follow-up “deidentified” analysis

Replacing individual characters with abstract identifiers focuses more on which cells contain alternations.

Circled clusters have an additional vowel change in the preterite.
Lessons

Spanish verbs have a cross-cutting organization (Baerman 2010); they are clustered differently according to suffix and stem alternation.

This complicates the clustering landscape of verb types.

But there is still a lot of information there for the learner.
Generalizing across languages

What transfers in morphological inflection? Experiments with analogical models: Elsner, SigMorphon 2021

Analogical modeling of morphology for L1 effects in language contact: Elsner and Sims, AIMM5 (presentation only) 2021

Andrea D. Sims (Associate Professor, OSU)
Language transfer

Lots of previous work (McCarthy et al 2019, Kann 2020 ...) on using knowledge from one language to learn another

Important in low-resource settings; also forces model to focus on *abstraction* rather than memorizing specific markers

How to encourage models to do this?
An old idea: analogical learning

Train a model to solve four-part proportional analogy problems like this one (to produce a Maori passive, from waiata “sing”)

waiata karanga:karangatia waiatatia
source exemplar : exemplar form prediction target

*also investigated by Liu and Hulden 2020: “cross-table examples”
Issue: choosing exemplars

Good exemplar (matches target)

waiata karanga:karangatia waiatatia

Bad exemplar (inflects differently from target)

waiata kaukau:kaukauria waiatatia
Two basic options

**Random exemplars:** Easy. Training matches test.

**Similar exemplars:** Pick training exemplar that uses the same alignment-based edit rule (for example, $+tia$). Training mismatches test.

Always use **random exemplars** at test time.
Setup

Transformer (Wu and Cotterell 2020) as learner again

Staged training:

- Learn to copy
- Multilingual model
- Fine-tuning

SigMorphon 2020 development languages as training set
Supervised results

<table>
<thead>
<tr>
<th></th>
<th>Random exemplars</th>
<th>Similar exemplars</th>
<th>Non-analogical transformer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>89</td>
<td>57</td>
<td>90</td>
</tr>
</tbody>
</table>

Test-train mismatch makes “similar” quite poor

“Random” underperforms where inflection class structure is important
One-shot mode

No fine-tuning; apply trained multilingual model to an unseen language

Provide an exemplar from the target language at test time
One-shot results

<table>
<thead>
<tr>
<th></th>
<th>Random exemplars</th>
<th>Similar exemplars</th>
<th>Fine-tuned Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Known family</td>
<td>29</td>
<td>38</td>
<td>80</td>
</tr>
<tr>
<td>Novel family</td>
<td>11</td>
<td>28</td>
<td>92</td>
</tr>
<tr>
<td>Overall</td>
<td>14</td>
<td>30</td>
<td>90</td>
</tr>
</tbody>
</table>

Similar exemplars performs much better; much more faithful copying from exemplars
### One-shot results: Catalan

<table>
<thead>
<tr>
<th>Lemma</th>
<th>Exemplar</th>
<th>Rand. sys</th>
<th>Sim. sys</th>
<th>Correct form</th>
</tr>
</thead>
<tbody>
<tr>
<td>arrissar</td>
<td>posar : posarien</td>
<td>arrissaren</td>
<td>arrissarien</td>
<td>arrissarien</td>
</tr>
<tr>
<td>disputar</td>
<td>descriure : descriuria</td>
<td>disputarta</td>
<td>disputaria</td>
<td>disputaria</td>
</tr>
<tr>
<td>repetir</td>
<td>cremar : creo</td>
<td>repetirer</td>
<td>repetio</td>
<td>repeteixo</td>
</tr>
<tr>
<td>engolir</td>
<td>forjar : forjava</td>
<td>engolire</td>
<td>engoliva</td>
<td>engolia</td>
</tr>
<tr>
<td>llevar-se</td>
<td>terminar : termino</td>
<td>llevar-se</td>
<td>llevar-se</td>
<td>llevo</td>
</tr>
</tbody>
</table>

Model seems to “know” an abstract suffixation rule
What transfers?

Use synthetic data to probe what the model has learned about specific languages and processes:

\[ \text{modi} \quad \text{gobu} : \text{gogobu} \quad ??? \]

When example is labeled as Tagalog, output is: \text{momodi}

But when labeled as Swahili, output is: \text{gomodi}
Probe tasks

Modeled on real morphological processes with varying representation in the SigMorphon dataset

<table>
<thead>
<tr>
<th>Process</th>
<th>What families (in data?)</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prefixing</td>
<td>Niger-Congo, Austronesian</td>
<td>semet ~ igo semet</td>
</tr>
<tr>
<td>Suffixing</td>
<td>Germanic, Uralic, some others</td>
<td>semet ~ semet igo</td>
</tr>
<tr>
<td>Reduplication</td>
<td>Austronesian</td>
<td>semet ~ sesemet</td>
</tr>
<tr>
<td>Gemination</td>
<td>None</td>
<td>semet ~ semmet</td>
</tr>
</tbody>
</table>
## Results (similar exemplars)

<table>
<thead>
<tr>
<th>Language (Family)</th>
<th>Prefix</th>
<th>Suffix</th>
<th>Reduplicate</th>
<th>Geminate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tagalog (Austronesian)</td>
<td>30</td>
<td>75</td>
<td>88</td>
<td>0</td>
</tr>
<tr>
<td>German (Germanic)</td>
<td>86</td>
<td>99</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Swahili (Niger-Congo)</td>
<td>99</td>
<td>88</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mezquital Otomi (Oto-Manguean)</td>
<td>96</td>
<td>59</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Finnish (Uralic)</td>
<td>52</td>
<td>98</td>
<td>0</td>
<td>12</td>
</tr>
</tbody>
</table>
Future work on language transfer

Currently: making model less sensitive to how exemplars are chosen

Teach the model simulated L1, then look at acquisition trajectories in L2

Also model frequency effects; how are exemplars really selected?
Matching human ratings of “wug” words

2021 SigMorphon shared task
Conclusions

Morphology is about more than just what affixes go where…

There are all sorts of abstract structures that help speakers predict:

Patterns of which cells are the same/different (Latin nouns)
Commonalities in what is predictable/shared across cells (Spanish verbs)
Systemic generalizations about how to mark morphological information (Catalan suffixation)
And more…
Current models learn surface structure well

But we’re still learning how to measure their knowledge of abstract structures

And how to encourage them to learn more and generalize better

Thank you!
## A closer look: reduplication

<table>
<thead>
<tr>
<th>Language</th>
<th>Reduplicated momodi</th>
<th>Prefixed gomodi</th>
<th>Infixed mogodi</th>
<th>Unaltered modi</th>
<th>Other moodi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tagalog (Austronesian)</td>
<td>87.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12.5</td>
</tr>
<tr>
<td>German (Germanic)</td>
<td>0</td>
<td>0</td>
<td>7.5</td>
<td>10</td>
<td>82.5</td>
</tr>
<tr>
<td>Swahili (Niger-Congo)</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mezquital Otomi (Oto-Manguean)</td>
<td>5</td>
<td>87.5</td>
<td>0</td>
<td>2.5</td>
<td>5</td>
</tr>
<tr>
<td>Finnish (Uralic)</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>2.5</td>
<td>92.5</td>
</tr>
</tbody>
</table>
### A closer look: suffixation

<table>
<thead>
<tr>
<th>Language</th>
<th>Suffixed semet-igo</th>
<th>All but first semet-go</th>
<th>All but last semet-ig</th>
<th>Unaltered semet</th>
<th>Other semet-g</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tagalog (Austronesian)</td>
<td>75</td>
<td>16</td>
<td>~0</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>German (Germanic)</td>
<td>100</td>
<td>~0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Swahili (Niger-Congo)</td>
<td>88</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Mezquital Otomi (Oto-Manguean)</td>
<td>59</td>
<td>5</td>
<td>8</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>Finnish (Uralic)</td>
<td>98</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>