Computational simulations of second language construction learning

Yevgen Matusevych
Afra Alishahi
Ad Backus

Tilburg University, the Netherlands
A usage-based account of SLA

• Nick Ellis: a usage-based account of SLA. Language is a complex adaptive system (Beckner et al., 2009; Ellis & Larsen-Freeman, 2006).

• Experimental and corpus studies (e.g., Ellis, O’Donnell, & Römer, 2013; Ellis & Ferreira-Junior, 2009).

• Computational modeling is very simplistic (Ellis & Larsen-Freeman, 2009).
Why computational modeling?

- Lots of variables involved into SLA process!
Why computational modeling?

• Lots of variables involved into SLA process!
Why computational modeling?

- Lots of variables involved into SLA process!
Why computational modeling?

- Lots of variables involved into SLA process!
Why computational modeling?

- Lots of variables involved into SLA process!
Computational modeling in language acquisition

Child language acquisition:

• word segmentation (ten Bosch, Van Hamme, Boves, & Moore, 2009); &

• learning words and their meanings (Yu & Ballard, 2007; Frank, Goodman, & Tenenbaum, 2009; Fazly, Alishahi, & Stevenson, 2010);

• learning language structure (Elman, 1990);

• etc.

Second language acquisition:

• Very few models, often do not model existing L1 knowledge.

• Mostly phonetics and lexis, not language structure.
Our project

Task: learning argument structure constructions (VACs)

Adele Goldberg: “Argument structure constructions are a special subclass of constructions that provides the basic means of clausal expression in a language.”

We find these constructions in nearly every utterance. Therefore their knowledge can be a good estimate of learners' overall language proficiency.
An existing Bayesian model (Alishahi & Stevenson, 2008).

Conceptually frames are utterance–scene pairs.

Utterance: Mom put toys in boxes.
Scene: $\text{Put}_{\text{cause,move}}(\text{MOM}_{\text{agent}}, \text{TOYS}_{\text{theme}}, \text{In}_{\text{In}}(\text{BOXES}_{\text{destination}}))_{\text{destination}}$
Frame for the utterance *Mom put toys in boxes*:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head verb</td>
<td>put</td>
</tr>
<tr>
<td>Verb semantic primitives</td>
<td>&lt;cause, move&gt;</td>
</tr>
<tr>
<td>Number of arguments</td>
<td>3</td>
</tr>
<tr>
<td>Arguments</td>
<td>mom, toys, boxes</td>
</tr>
<tr>
<td>Argument 1 semantics</td>
<td>&lt;parent, person, female, ... &gt;</td>
</tr>
<tr>
<td>Argument 1 event properties</td>
<td>&lt;volitional, cause-location, ... &gt;</td>
</tr>
<tr>
<td>Syntactic pattern</td>
<td>arg1 verb arg2 arg3</td>
</tr>
</tbody>
</table>
Constructions in this model are groups of frames:
Model of early argument structure acquisition (4)

\( \text{BestConstruction}(F) = \arg\max_k P(k \mid F) \)

\[
P(k \mid F) = \frac{P(k)P(F \mid k)}{P(F)} \sim P(k)P(F \mid k)
\]

\[
P(0) = \frac{1}{N + 1}
\]

\[
P(k) = \frac{N_k}{N + 1}
\]

\[
P(F \mid k) = \prod_{i \in \text{Features}(F)} P_i(v \mid k)
\]

\(F \text{ – frame, } k \text{ – construction, } v \text{ – feature value}\)
Evaluation is based on language use and can include various comprehension/production tasks (predicting missing features in a frame).

Sentence production task: predict the syntactic pattern for a given frame.
Naturalistic data

The Flensburg English Classroom corpus

English teachers’ speech addressed to German children

6 frequent verbs: go, come, read, show, look, put
Naturalistic data

The Flensburg **English** Classroom corpus

English teachers’ speech addressed to German children

6 frequent verbs: *go, come, read, show, look, put*

CHILDES: 3 **German** children

Child-directed speech

6 frequent verbs: *machen, kommen, gucken, gehen, sehen, geben*
The Flensburg **English** Classroom corpus

English teachers’ speech addressed to German children

6 frequent verbs: *go, come, read, show, look, put*

---

CHILDES: 3 **German** children

Child-directed speech

6 frequent verbs: *machen, kommen, gucken, gehen, sehen, geben*

---

Extracting 100 verb instances, annotating argument structure
Results

General L2 development
General L2 development: baseline

Manipulating the input using different scenarios. For this case:

- L1 training: 20 frames
- L1 testing: 20 frames
- L1+L2 training: Ratio 3:1, 20 frames
- L1 & L2 testing: 20 frames

25 times 20 = 500 frames
General L2 development: baseline

Manipulating the input using different scenarios. For this case:

- **L1 training**: 20 frames
- **L1 testing**: 20 frames
- **L1+L2 training**: Ratio 3:1, 20 frames
- **L1 & L2 testing**: 20 frames

25 times 20 = 500 frames
Manipulating the input using different scenarios. For this case:

1: \( p < .05 \)

2: \( p < .05 \)
General L2 development: baseline

Manipulating the input using different scenarios. For this case:

1. p < .05
2. p < .05

L2 proficiency differs from L1 proficiency

L1 training 20 frames ➔ L1 testing 20 frames ➔ L1+L2 training Ratio 3:1 20 frames ➔ L1 & L2 testing 20 frames ➔ ...

25 times 20 = 500 frames

Accuracy

L1
L2
Steps

1: p < .05
2: p < .05

L2 proficiency differs from L1 proficiency
Manipulating the input using different scenarios. For this case:

- L1 training: 20 frames
- L1 testing: 20 frames
- L1+L2 training: Ratio 3:1, 20 frames
- L1 & L2 testing: 20 frames

25 times 20 = 500 frames

Accuracy

1: p < .05
2: p < .05

L2 proficiency differs from L1 proficiency
General L2 development: equal AoE

The learner starts receiving L2 input together with L1 input.

L1+L2 training
Ratio 3:1
20 frames

L1 & L2 testing
20 frames

... 25 times 20 = 500 frames
General L2 development: equal AoE

The learner starts receiving L2 input together with L1 input.

L1+L2 training
Ratio 3:1
20 frames

L1 & L2 testing
20 frames

25 times 20 = 500 frames

1: p < .05
2: p < .05

L2 proficiency differs from L1 proficiency given equal AoE but different ratio
General L2 development: equal ratio

The amount of L1 and L2 exposure is equal, but AoE is different.

L1 training 20 frames → L1 testing 20 frames → L1+L2 training Ratio 1:1 20 frames → L1 & L2 testing 20 frames → ...

25 times 20 = 500 frames
The amount of L1 and L2 exposure is equal, but AoE is different.

L2 proficiency differs from L1 proficiency given equal ratio but different AoE.
General L2 development: all equal

Equal learning conditions for L1 and L2:

- L1+L2 training
  - Ratio 1:1
  - 20 frames
- L1 & L2 testing
  - 20 frames
- ...  

25 times 20 = 500 frames
General L2 development: all equal

Equal learning conditions for L1 and L2:

L1+L2 training
Ratio 1:1
20 frames

L1 & L2 testing
20 frames

…

25 times 20 = 500 frames

1: p > .05
2: p < .05

L2 proficiency may differ from L1 proficiency given equal learning conditions
Results

Structural priming in L2
Structural priming effects in L2

- Structural priming: speakers tend to reproduce a recently encountered linguistic structure.

- Observed both in L1 and L2 learners.
Structural priming effects in L2

• Structural priming: speakers tend to reproduce a recently encountered linguistic structure.

• Observed both in L1 and L2 learners.

• Experiment by Gries & Wulff (2005):
  1. Show a prime sentence:
     P. *The racing driver showed the torn uniform to the manager.*
     D. *The racing driver showed the mechanic the damaged tyre.*
  2. Ask to continue a test sentence:
     *The racing driver showed…*
Structural priming effects in L2

Two priming frames:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head verb</td>
<td>show</td>
</tr>
<tr>
<td>Verb semantic primitives</td>
<td>&lt;act, cause, perceive&gt;</td>
</tr>
<tr>
<td>Number of arguments</td>
<td>3</td>
</tr>
<tr>
<td>Arguments and their properties</td>
<td>vary</td>
</tr>
<tr>
<td>Syntactic pattern: <strong>P-prime</strong></td>
<td>AGENT show THEME to BENEFICIARY</td>
</tr>
<tr>
<td>Syntactic pattern: <strong>D-prime</strong></td>
<td>AGENT show BENEFICIARY THEME</td>
</tr>
</tbody>
</table>
Structural priming effects in L2

- **L1 training**: 250 frames
- **L1+L2 training**: 10 frames
- **L2 priming (OR NOT)**: 1 frame
- **L2 testing**: 1 frame

25 times 10 = 250 frames
Structural priming effects in L2

L1 training
250 frames

L1+L2 training
Ratio 3:1
10 frames

L2 priming
(OR NOT)
1 frame

L2 testing
1 frame

25 times 10 = 250 frames

P-patterns production during test

D-patterns production during test
Structural priming effects in L2

We could simulate the structural priming effect only at earlier stages; after a certain point the effect is leveled off.

P-patterns production during test

D-patterns production during test
Effect of skewed input frequencies
Facilitatory effect of skewed input frequencies

- General idea: language learners are sensitive to manipulating various frequencies in the input.

- In the natural input, certain verbs occur in a certain construction much more often than other verbs (Ellis & Ferreira-Junior, 2009).

- Skewed distributions may have facilitatory effect on learning.

- The effect has been shown for learning novel (artificial) L2 verbs (Boyd & Goldberg, 2009), but no results for natural L2 yet.
Facilitatory effect of skewed input frequencies

Year and Gordon (2009):

- **Participants**: Korean learners of English
- **Materials**: 5 English verbs in a ditransitive construction
- **Conditions**: skewed (24-4-4-4-4) or balanced (8-8-8-8-8)
- **Result**: Found no facilitatory effect
Facilitatory effect of skewed input frequencies

Three conditions:

1. **Naturalistic data.**
2. **Balanced data.**
3. **Skewed data (go and show).**

The same scenario:

- **L1 training**
  - 500 frames

- **L1+L2 training**
  - Ratio 3:1
  - 20 frames

- **L2 testing**
  - 20 frames

...  

25 times 20 = 500 frames
Facilitatory effect of skewed input frequencies
Conclusions

• We replicated computationally general patterns of L2 learning and the priming effect, although with some limitations.

• We could observe the facilitatory effect of skewed frequency on L2 learning, which has not been shown for L2 yet.

• This can be a good framework for studying different L2 phenomena, because we can easily manipulate all the variables, which is not the case with human learners.
Future work

• Careful consideration of each phenomenon is needed.

• A study of verbs/constructions distribution in input/output data (similar to Ellis & Ferreira-Junior, 2009) would allow us to evaluate the model in the light of the existing theory of construction learning (Ellis, O’Donnell, & Römer, 2013).

• Improvements in the model, in the clustering structure particularly. Multilevel organization of constructicon.
Questions?

THANK YOU
FOR YOUR ATTENTION!