Automating elicited imitation for spoken practice in German L2: task design, speech recognition, and language models

Frederik Cornillie (University of Leuven & imec)
Dirk De Hertog (University of Leuven & imec)
Kristof Baten (Ghent University)
Spoken practice: what and why?

spoken activities in a L2
that focus on specific linguistic constructions
and that involve a considerable amount of
recycling, feedback, and often time pressure,
with the goal of developing explicit knowledge
about these constructions
as well as skills in the L2

All you need is input

vs.

Output practice and feedback can aid noticing
and automatization

the Krashen school

the interactionist school
The relative effects of input and output practice

- Inconsistent findings:
  - Effects on comprehension:
    - Input practice ~ output practice (Morgan-Short & Bowden, 2006; Nagata, 1998; Salaberry, 1997; Toth, 2006)
    - Input practice > output practice (Benati, 2001; 2005; DeKeyser & Sokalski, 1996)
  - Effects on production:
    - Input practice ~ output practice (Benati, 2001; 2005)
    - Output practice > input practice (Dekeyser & Sokalski, 1996; Morgan-Short & Bowden, 2006; Nagata, 1998; Toth, 2006)

- Limitations:
  - (very) short treatments (1-6 hours) over short periods of time (1-7 days)
  - Only accuracy rates considered

→ No evidence of relative effects on automatization: transfer to communicative tasks?
CALL to the rescue? (a call from the past)

Research on practice [must be] very fine-grained, to allow for tracking of stimuli and responses in milliseconds […] while being longitudinal in nature […] Perhaps new technology can solve this problem by allowing for massive data collection and sophisticated analysis at the fine-grained level and longitudinally, from many learners, without losing sight of the importance of individual differences.

Robert DeKeyser

Data collection today

in everyday apps

- longitudinal and massive
- uncontrolled environments
- updated and analyzed continuously
- valorized (e.g. for personalization)

in SLA research

- typically no longer than a couple of weeks
- controlled environments
- write once, analyze once
- typically not valorized in learning environments
But … big data is gaining traction in CALL

---

[Message]

From: Stephen Fansrali <sfansrali@gmail.com>
Subject: [edm-announce] Jobs @ Duolingo: Multiple R&D Roles
To: edm-announce@freelists.org

[on behalf of Burr Settles]

Duolingo is hiring several research and data scientist positions. We already have a strong group of interdisciplinary scholars focused on novel applied research at the intersection of machine learning, computational linguistics, and cognitive science, and we are expanding quickly. https://www.duolingo.com

Duolingo is the world’s largest online language learning service with more than 150 million users worldwide. Our apps have been selected as Apple’s iPhone App of the Year and Google’s Best of the Best for Android multiple times. Our scientists split their time between primary research on new and unique problems (publication is encouraged), and translating these research findings into production systems that improve learning and engagement outcomes for millions.

Candidates interested in user modeling, educational data mining, applied machine learning, NLP and/or speech processing should apply!

See the full advertisements for two positions below.

Cheers,
ORAL ELICITED IMITATION
Oral elicited imitation: the basic task

stimulus
relatively short and simple sentences

response
repeat as exactly as possible
Oral elicited imitation: cognitive processes

**stimulus**
- relatively short and simple sentences
- (target-language-like or deviating)

**response**
- repeat and reconstruct
- insight in the learner’s interlanguage system

**SEMANTIC PROCESSING**
- erases memory of the form (Erlam, 2006)

**SYNTACTIC PROCESSING**
Oral elicited imitation in L2 assessment

- OEI can measure
  - oral proficiency (Tracy-Ventura, McManus, Norris, & Ortega, 2014)
  - implicit knowledge (e.g. Erlam, 2009)
  - automatized explicit knowledge (Suzuki & DeKeyser, 2015)

- The assessment task can be automated with speech recognition
  - (Cook, Mcghee, & Lonsdale, 2011; Graham, Lonsdale, Kennington, Johnson, & McGhee, 2008)
Oral elicited imitation for output practice: issues for CALL

meaningful language processing or mechanical parroting?

corrective feedback in order to stimulate noticing

speech recognition technology & language models for error diagnosis
EMPIRICAL STUDY ON GERMAN L2
The current study

Goal

prepare task design, materials and technology

for a study on the relative effects of output practice in German L2

Research questions:

1. Does the design of the OEL task focus learners’ attention on meaning?
   → task design

2. How accurately does state-of-the-art speech recognition transcribe
   the learners’ production?
   → speech recognition

3. What was the nature of linguistic variation in the learners’ production?
   → language models
Materials: target constructions

- **transitives** – e.g. [The dog chases the man]
  
  Der Hund verfolgt den Mann.
  
  *Der Hund verfolgt der Mann.
  
  Den Mann verfolgt der Hund.
  
  *Der Mann verfolgt der Hund.

- **ditransitives** – e.g. [The teacher gives the headmaster flowers]
  
  Die Lehrerin schenkt dem Direktor die Blumen.
  
  *Die Lehrerin schenkt der Direktor die Blumen.
  
  Dem Direktor schenkt die Lehrerin die Blumen.
  
  *Der Direktor schenkt die Lehrerin die Blumen.

- **prepositional phrases** – e.g. [The man walks through NP]
  
  Der Mann spaziert durch den Tunnel.
  
  *Der Mann spaziert durch der Park.
Materials: task design

stimulus

picture matching response

spoken response

Den Mann verfolgt der Hund.
[The dog chases the man]

instruction:
“repeat in as good German as possible”
Participants & data

- participants:
  - Flemish learners of German L2 ($N = 36$)
  - academic programme in Languages and Literature, Ghent University
    - 2nd bachelor ($N=11$)
    - 3rd bachelor ($N=10$)
    - master ($N=15$)
  - 18-23 years old

- data:
  - collected online (item order counterbalanced), using headsets
  - total of 1728 learner-item interactions:
    - 1728 picture-matching responses
    - 1487 spoken responses manually transcribed
Results for task design

Does the design of the task focus learners’ attention on meaning?

Accuracy on picture matching task, by year

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%

Bachelor 2  Bachelor 3  Master

Correct  Incorrect

chance level
Results for task design

Does the design of the task focus learners’ attention on meaning?

Accuracy on picture matching task, by year

difference between groups: $F(2, 33) = 0.88, p = 0.42$
Results for task design

Does the design of the task focus learners’ attention on meaning?

Grammatical accuracy of production (correct picture matching responses only)

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammatical stimuli</td>
<td>36</td>
<td>0.87</td>
<td>1</td>
<td>0.986</td>
<td>.028</td>
</tr>
<tr>
<td>Ungrammatical stimuli</td>
<td>36</td>
<td>0.208</td>
<td>1</td>
<td>0.716</td>
<td>.199</td>
</tr>
</tbody>
</table>

$r = 0.62, p < 0.001, N = 36$

→ reconstructive
## Results for speech recognition

### Tools
- **Google Cloud**
  - easy API
  - black box
  - pay for what you use
- **CMUSphinx**
  - more tricky to set up
  - open source
  - pay for a server

### Implementations
- out of the box
- acoustic model
- language model
- language model & acoustic model

### Evaluation metric

#### Levenshtein edit distance (word level)

<table>
<thead>
<tr>
<th>Google Cloud</th>
<th>CMUSphinx</th>
</tr>
</thead>
<tbody>
<tr>
<td>den Direktor</td>
<td>den Direktor</td>
</tr>
<tr>
<td>schimpfe</td>
<td>schenkt</td>
</tr>
<tr>
<td>Lehrerin</td>
<td>Lehrerin</td>
</tr>
<tr>
<td>die Blumen</td>
<td>die Blumen</td>
</tr>
</tbody>
</table>

→ 3
Results for speech recognition

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Median</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>0</td>
<td>6</td>
<td>0.55</td>
<td>0</td>
<td>1487</td>
</tr>
<tr>
<td>Sphinx</td>
<td>0</td>
<td>14</td>
<td>4.70</td>
<td>5</td>
<td>1412</td>
</tr>
<tr>
<td>Sphinx AM</td>
<td>0</td>
<td>11</td>
<td>2.48</td>
<td>2</td>
<td>1410</td>
</tr>
<tr>
<td>Sphinx LM</td>
<td>0</td>
<td>12</td>
<td>2.23</td>
<td>2</td>
<td>1413</td>
</tr>
<tr>
<td>Sphinx LM+AM</td>
<td>0</td>
<td>13</td>
<td>1.87</td>
<td>1</td>
<td>1413</td>
</tr>
</tbody>
</table>
Results for speech recognition

Some other relevant findings:

- no error correction

  - der Vater zeigt *[den Sohn] die Brille
  - der Mann ist gegen *[dem dem Baum] gefahren
  - der Junge geht *[zu Bäcker]
  - die Lehrerin schenkt dem Direktor *[den Blumen] die Blumen

- possible quick win: improve recognition by prioritizing key vocabulary in the language model

  - der Polizist sucht den Becher (< Bäcker)
  - die Lehrerin schenkt den Jagd aber (< Direktor) die Blumen
Results for language models (work in progress)

What was the nature of linguistic variation in the learners’ production?

- Linguistic variation
  - Semantic
  
  Der Mann ist gegen den Baum gefallen (< gefahren)

  - Morphological
  
  *Die Lehrerin schenkt *den (< dem) Direktor den Blumen

  - Syntactic
  
  Die Lehrerin schenkt dem Direktor die Blumen
  < Dem Direktor schenkt die Lehrerin die Blumen

  - Combinations
  
  Der Vater schenkt der Junge den Junge die Brille
  < Dem Sohn zeigt der Vater die Brille

- Variation due to cognitive processes
  - Self-correction
  
  Das Mädchen kommt aus der Shop - dem Shop

  - Disfluencies
  
  Der Doktor verklauf verkauft dem Clown das Buch

  - Multiple repetitions
  
  Die Frau gibt den Mann den Apfel. Die Frau gibt dem Mann den Apfel.
Discussion and next steps

- OEI as implemented in this study has potential as a practice task
  - Picture matching simulated meaningful language processing
  - Google Cloud speech API handled non-native German speech relatively well

- Limitations:
  - Advanced students > role of working memory?
  - Controlled setting
  - Meaning-focus could be stronger
  - Google Cloud Speech API is a black box

- Next steps:
  - Develop language models for error correction
  - Increase the meaning-focus of the task, e.g. individual sentences form a coherent story
The future of research on CALL practice?

- open data
- open tools and technologies
- real collaboration academics - industry
Thank You!

Acknowledgements

- German native speaker stimuli recorded by Carola Strobl
- Drawings created by Fridl Cuvelier
- Data collected by Wouter Vanacker
- Icons created by Gregor Cresnar and Oksana Latysheva from Noun Project

Frederik Cornillie (University of Leuven & imec)
Dirk De Hertog (University of Leuven & imec)
Kristof Baten (Ghent University)