

# *Effects of interactivity of written practice on incidental vocabulary acquisition*

Serge Bibauw<sup>1,2,3</sup>    Thomas François<sup>2</sup>    Piet Desmet<sup>1</sup>

<sup>1</sup> ITEC, imec research group at KU Leuven

<sup>2</sup> CENTAL, UCLouvain

<sup>3</sup> Universidad Central del Ecuador

Vocab@Leuven International Conference  
Leuven, 1 July 2019

imec

KU LEUVEN

UCLouvain



# Effects of interactivity of written practice on incidental vocabulary acquisition

Interactive practice *with a dialogue system*, in this case, integrated within an educational video game.



Randomized controlled experiment with young Flemish learners of French (n=209)

# Effects of interactivity of written practice on incidental vocabulary acquisition



## **Interaction-based vocabulary acquisition**

How do output and interaction help lexical development?

## **Dialogue systems for SLA research**

Definition, summarized typology, research questions

## **Randomized controlled experiment**

Groups, learners' profiles, instruments, intervention

## **Results and discussion**

Mixed-effects models, effects of encounters in input/output

# Effects of interactivity of written practice on incidental vocabulary acquisition



Let's begin with some theory...

## Interaction-based vocabulary acquisition

How do output and interaction help lexical development?

## Dialogue systems for SLA research

Definition, summarized typology, research questions

## Randomized controlled experiment

Groups, learners' profiles, instruments, intervention

## Results and discussion

Mixed-effects models, effects of encounters in input/output

Theoretical background

## **Incidental vocabulary acquisition**

**Incidental vs. intentional learning** (e.g. Eysenck, 1982)

exposure to material, no test announcement

**Incidental vocabulary learning** (Schmidt, 1994)

within a meaning-focused task

**Contextual vocabulary learning** (Elgort, Brysbaert, Stevens & Van Assche, 2018)

# Theoretical background

## Interactive tasks and vocabulary acquisition

### Task-induced involvement (Laufer & Hulstijn, 2001)

factors for word retention in a task: *need, search, evaluation*

more effective tasks:

- **Negotiation of meaning (+)** (Newton, 1995)
- **Word use for completing the task (++)** (Newton, 1995)
- **Interactionally modified output (++)** (Ellis & He, 1999)

→ Many reasons to study incidental vocabulary learning beyond reading

Theoretical background

## Effectiveness on vocabulary learning: experimental evidence

### Meta-analysis of incidental word learning from spoken input

(de Vos, Schriefers, Nivard & Lemhöfer, 2018)

$d = 1.05$  (large effect)

**interactive tasks ( $d +0.73$ )** > non-interactive tasks ( $d +0.10$ ) > input only

population: university ( $d +0.92$ ) > high school ( $d +0.74$ ) > elementary school

outcome: recognition ( $d +0.42$ ) > recall

# Theoretical background

## Effectiveness on vocabulary learning: broader evidence

### Studies of out-of-school exposure to English:

Peters, Noreillie, Heylen, Bulté & Desmet, 2019: **significant SEM parameters for vocabulary size (n=138):**

**online/games ( $\beta = .39^{***}$ )**

n.s. parameters: use (friends/family), audiovisual input, written media

De Wilde, Brysbaert & Eyckmans, 2019: **best predictors of receptive vocabulary (n=747):**

**using social medial in L2 ( $\beta = .29^{***}$ )**

**playing games in L2 ( $\beta = .18^{***}$ )**

**speaking L2 ( $\beta = .13^{***}$ )**

compare with watching TV ( $\beta = -.07$  n.s.), music ( $\beta = -.08^*$ )

→ **all profoundly interactive sources of exposure**

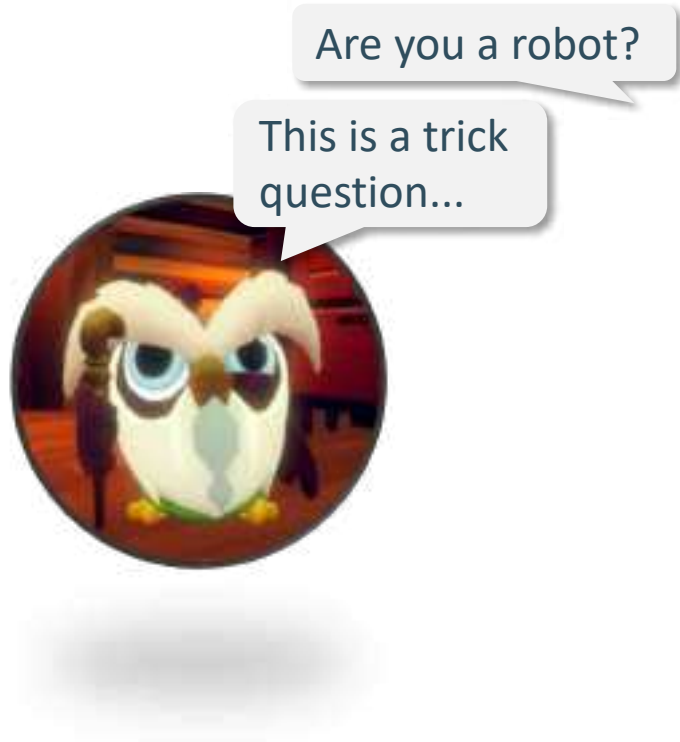


# Towards our research question...

Interactive tasks seem to be among the most beneficial for incidental vocabulary acquisition, but **what is it that makes them so effective?**

Attention to form within meaning-focused task? Need to use?  
Negotiated input? Negotiated output? Multiple exposures?

# Effects of interactivity of written practice on incidental vocabulary acquisition



## Interaction-based vocabulary acquisition

How do output and interaction help lexical development?

## Dialogue systems for SLA research

Definition, summarized typology, research questions

## Randomized controlled experiment

Groups, learners' profiles, instruments, intervention

## Results and discussion

Mixed-effects models, effects of encounters in input/output

# Dialogue systems for language learning

(Bibauw, François & Desmet, 2019)

Any application or system allowing

to maintain a **dialogue**

[ immediate, synchronous interaction ]

[ written or spoken ]

with an **automated agent**

[ chatbot, talking robot, automated personal assistant, conversational agent, non-player character in a video game... ]

[ tutorial CALL (≠ computer-mediated communication) ]

for **language learning** purposes.

# Dialogue systems for language learning

## Types of systems (Bibauw, François & Desmet, 2019)



### Form-focused systems



CALL-SLT (Baur, Rayner & Tsourakis, 2014)

### Goal-oriented systems



SPELL (Morton, Gunson & Jack, 2012)

# Dialogue systems for language learning

## Types of systems (Bibauw, François & Desmet, 2019)



### Form-focused systems

**Explicit constraints on meaning:**  
gap-filling, predetermined answers

**Focus of forms**

**Limited interactivity:**  
mostly corrective feedback

**No dialogue management:**  
pre-scripted dialogue

### Goal-oriented systems

**Contextual constraints on meaning:**  
interactional task and context

**Focus on meaning/form**

**High interactivity:**  
conversation influenced by user

**Advanced dialogue management:**  
→ high-level NLP required

# Dialogue systems for SLA research

## Research questions

1. Dialogue-based CALL systems exhibit large variation in terms of **interactivity and freedom** vs. **constraints** of the learner within the dialogue: what does it change?
2. Technologically, it is **considerably easier to “fake” the interaction** by restraining the learner: it is worth it pedagogically to implement advanced natural language understanding and dialogue management?
3. Traditionally, in large-classes school contexts, dialogue is often practiced through written “**Dialogue Completion Tasks**”: it is similar in terms of perception, performance and learning?

# Effects of interactivity of written practice on incidental vocabulary acquisition



## Interaction-based vocabulary acquisition

How do output and interaction help lexical development?

## Dialogue systems for SLA research

Definition, summarized typology, research questions

## Randomized controlled experiment

Groups, learners' profiles, instruments, intervention

## Results and discussion

Mixed-effects models, effects of encounters in input/output

Intervention · Dialogue system

## *LanguageHero*, dialogue-based game for young learners

Developed by Leuven-based start-up *Linguineo*.

(Main) target audience: teenagers (10-14).

Prototype developed for French for Dutch-speaking learners.

Task-based free conversational written interaction.



# Language Hero

## Conversations:

Structured around goal-oriented conversations

Conversation 1: After the storm - Meet Sansai and find out what happened and where you are.

Meilleur score: 828

Conversation 2: Meet Baldog - Meet Baldog and ask him for help.

Meilleur score: 0

Conversation 3: The snails - Vincent - Get to know the snails family

Meilleur score: 426

Conversation 4: The snails - Angélique - Get to know the mother of the snails family

Meilleur score: 0

Conversation 5: The snails - Claudette - Get to know one of the triplets of the snails family

Meilleur score: 0

Conversation 6: Return to Baldog - Go back to Baldog and tell him his problem is solved.



Contextualization

▶ He: *Bien le bonjour! Comment t'appelles-tu?*

✔ You: *bonjour je m'appelle Marco*

▶ He: *Enchanté de faire ta connaissance, Rlnc! Rlnc. Rlnc. Rlnc. Ne t'en fais pas, je ne suis pas fou. C'est juste que je répète ton nom pour ne pas l'oublier.*

You: *Comment tu t'appelle?*

He does not seem to have heard you...

You: *Tu t'appelle comment?*

He does not seem to have heard you...

✔ You: *Tu t'appelle comment?*

Correction: *appelle - Vérifiez l'accord entre le pronom « Tu » et le verbe « appelle ».*

Task accomplished: *Good, That was what we were wondering about.*

▶ He: *Moi, c'est Vincent. Elle, là-bas, c'est Angélique. Ça, c'est Delphine. Puis on a Georges dans le coin. Et évidemment, on ne peut pas oublier les triplées : Lisette, Claudette et Yvette. Oh! Et puis le petit là-bas, c'est Louis.*

Corrective feedback

Score: 405 ?  
Friendship lvl0: Acquaintance

Gamification

Current task (2/30):  
Say it is nice to meet them.

Microtasks to guide the conversation

Type or say your answer:

Type text:

Free written input

➔ Send your reply	🎤 Record your answer
❓ Disable help	⊗ End conversation

We can give you suggestions you can use to come up with an answer:

Scaffolding



# Intervention · Conditions

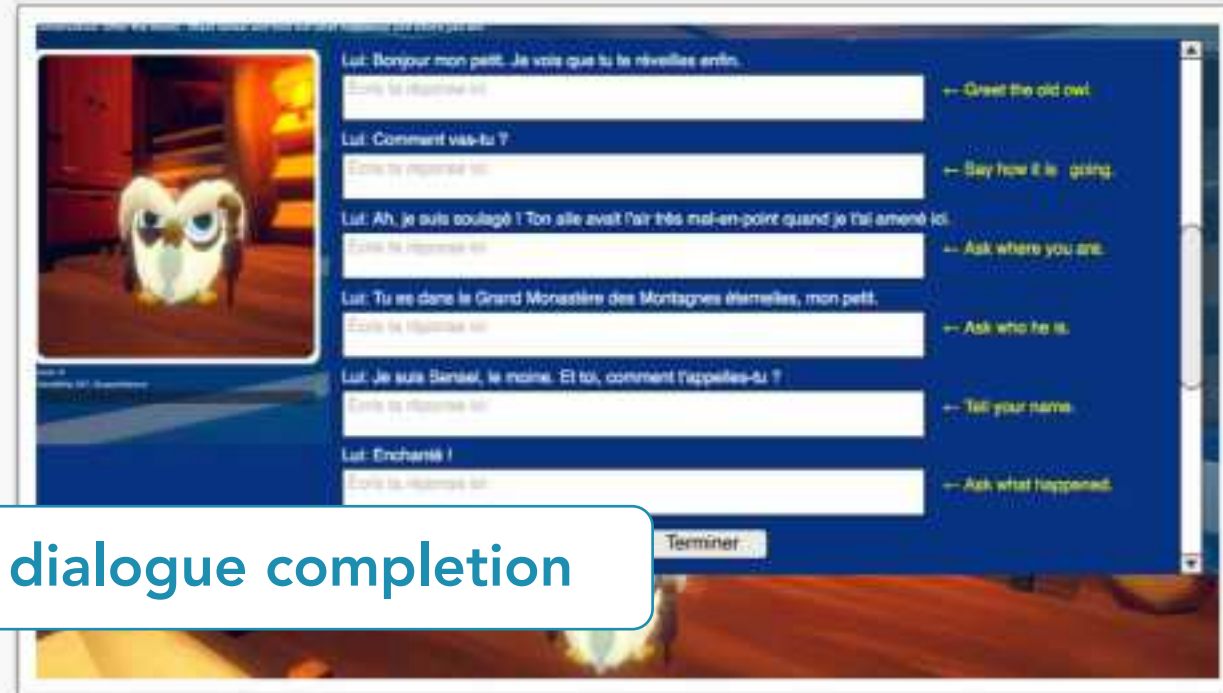
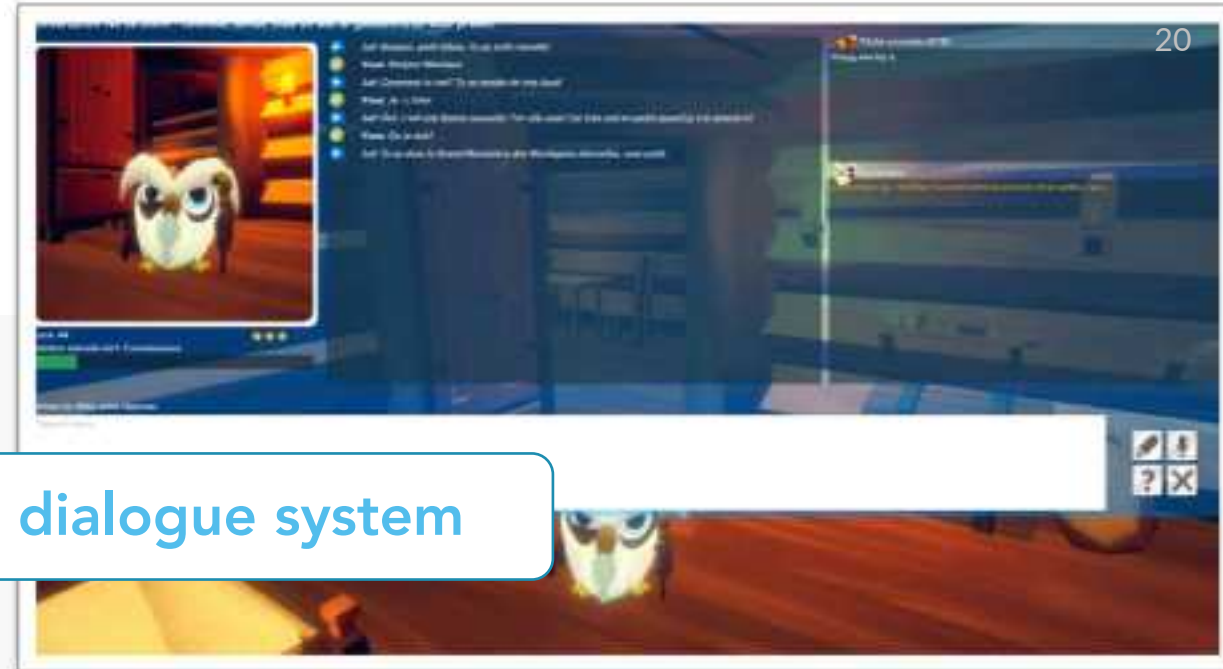
## Interactive vs. static dialogue

Compare:

(A) fully interactive,  
immediate/synchronous  
**dialogue system**

(B) classic, asynchronous  
**dialogue completion task**

Conditions with identical tasks,  
input, output opportunities,  
feedback and scaffolding.



# Methods

## Population and group assignment

4 schools volunteered to participate, with 2-3 classes each:

$$N_{\text{clusters}} = 11$$

$$N_{\text{participants}} = 215 \text{ (208 complete cases)}$$

Random assignment of classes to 3 conditions (distributed equally across schools):

- **Dialogue System** (experimental):  $n_{\text{D.Sys.}} = 81$
- **Dialogue Completion** ('baseline'):  $n_{\text{D.Compl}} = 79$
- **Control** ('business-as-usual'):  $n_{\text{control}} = 49$

Flemish 2<sup>nd</sup> year secondary school learners of French ( $M_{\text{age}} = 13.4$  y.o.)

L1 = 95,3 % Dutch

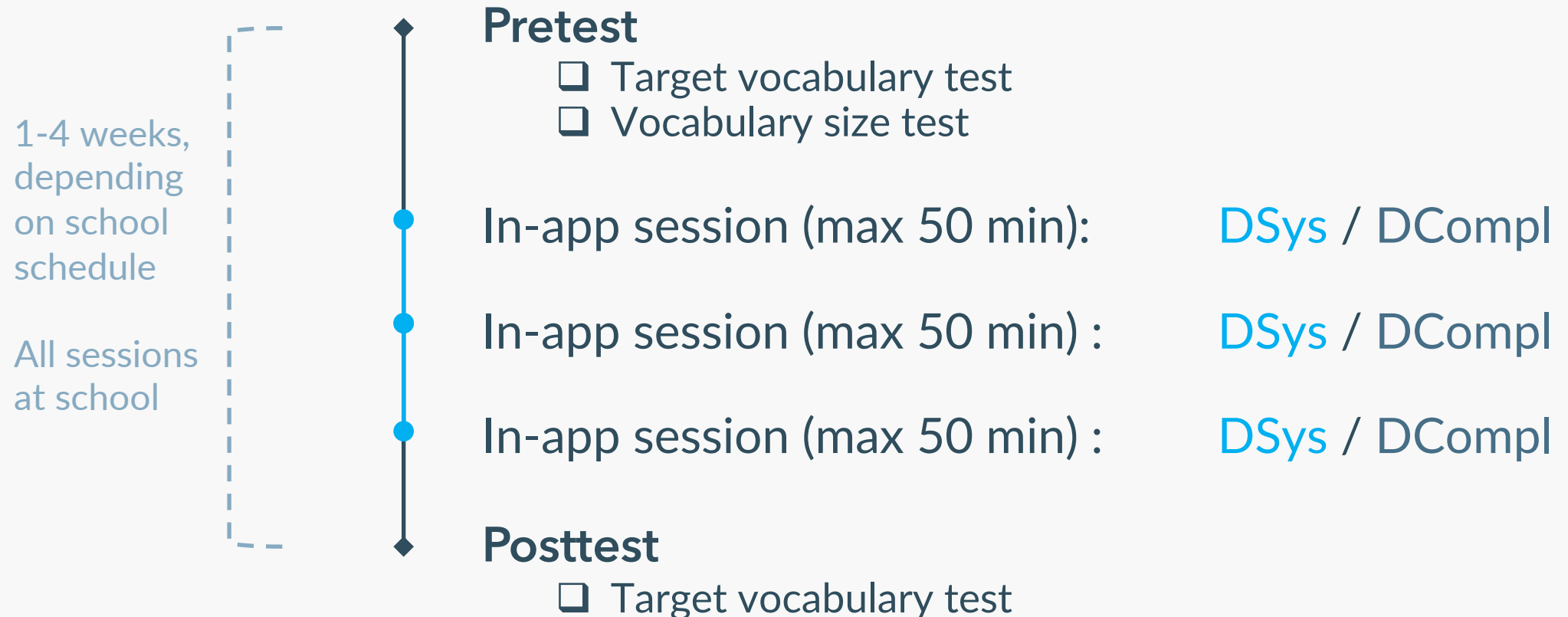
L2 = French = first L2,  $M = 3,1$  years of instruction, mostly at **A1** level

( $M_{\text{score}}$  in productive vocabulary size test = 3.6/30 in 1K frequency band)

10 (near-)native speakers of French excluded (final  $N = 198$ )

# Methods

## Procedure



## Target Vocabulary Test (1)

**“Target” words and sequences** seen and potentially produced inside the intervention: based on frequency of exposure across whole available content, selecting the most frequent lemmas and the most frequent formulaic sequences.

But no explicit target of instruction (no specific feedback, nor glossing, nor systematic presentation)  
⇒ **Incidental learning only**

At pre- and post-test (identical, randomized order)

## Methods · Instruments

# Target Vocabulary Test (2)

- Receptive** part (meaning recognition): 25 items  
 translation, as multiple choice  
 e.g., Potager:  soep  moestuin  vriend  potaarde  *Ik weet het niet*  
 soup  vegetable garden  friend  potting soil  I don't know
- Productive** part (in-context form recall): 25 items  
 gap-filling (L2 only) on formulaic sequences  
 e.g., Cet auteur a vraiment \_ \_ \_ \_ \_ d'imagination : ses livres sont très originaux !  
*This author really has a lot of imagination: his books are really special!*

## Methods

# Advanced in-game logging

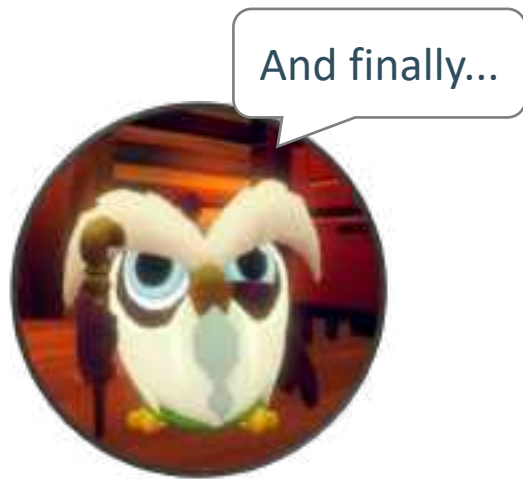
Full logging of all messages read and written in the system

Total: 48 353 messages, 577 494 tokens

Keystroke logging (in ms) for writing fluency measurement and repair behaviours analysis.



# Effects of interactivity of written practice on incidental vocabulary acquisition



## Interaction-based vocabulary acquisition

How do output and interaction help lexical development?

## Dialogue systems for SLA research

Definition, summarized typology, research questions

## Randomized controlled experiment

Groups, learners' profiles, instruments, intervention

## Results and discussion

Mixed-effects models, effects of encounters in input/output

# Results

## Receptive vocabulary

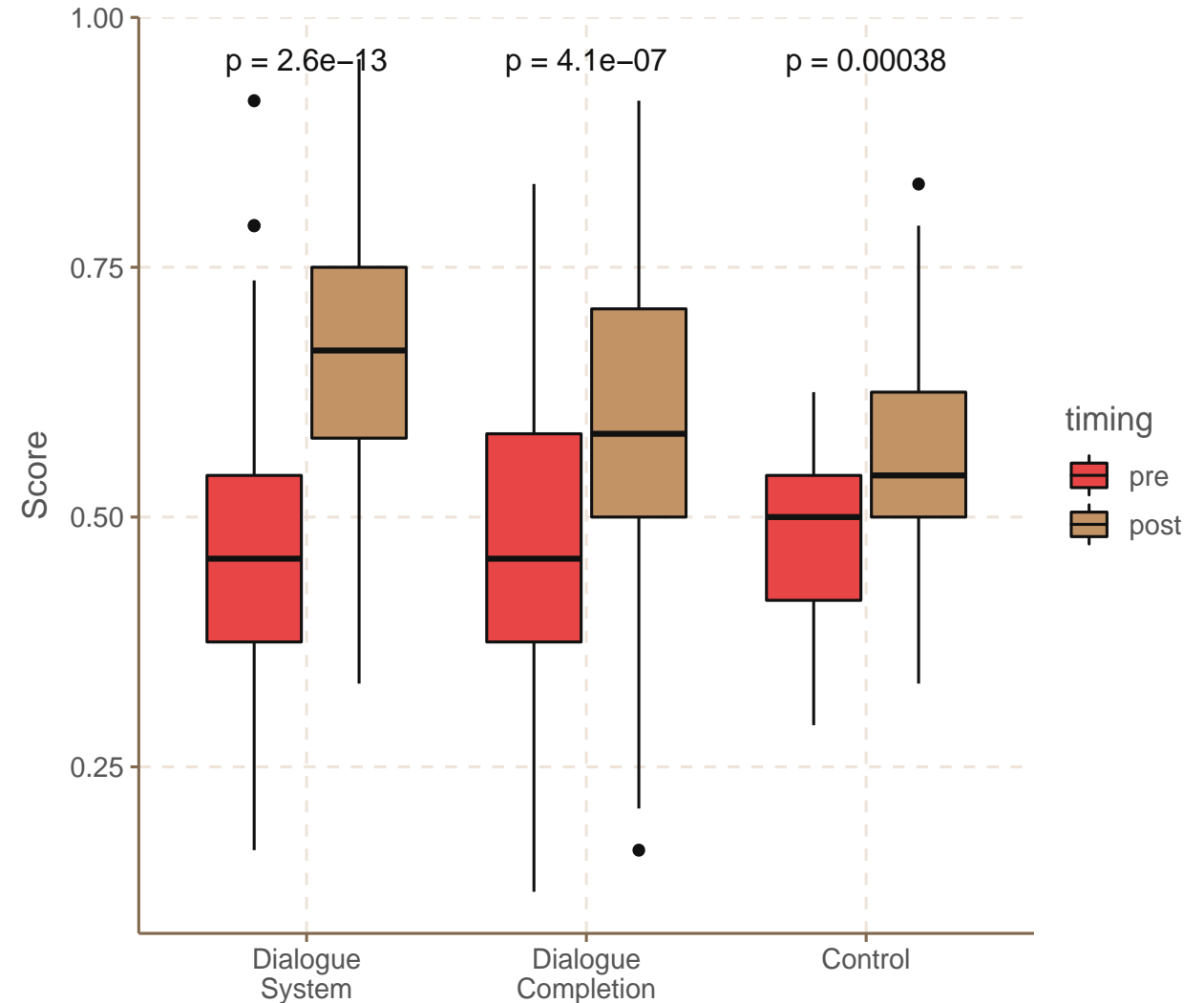
Very significant increase.

$$d_{\text{DSystem}} = 1.17^{***}$$

$$d_{\text{DCompletion}} = 0.80^{***}$$

$$d_{\text{DControl}} = 0.67^{***}$$

Considering the short treatment (2h), clear difference between conditions.



# Results

## Productive vocabulary

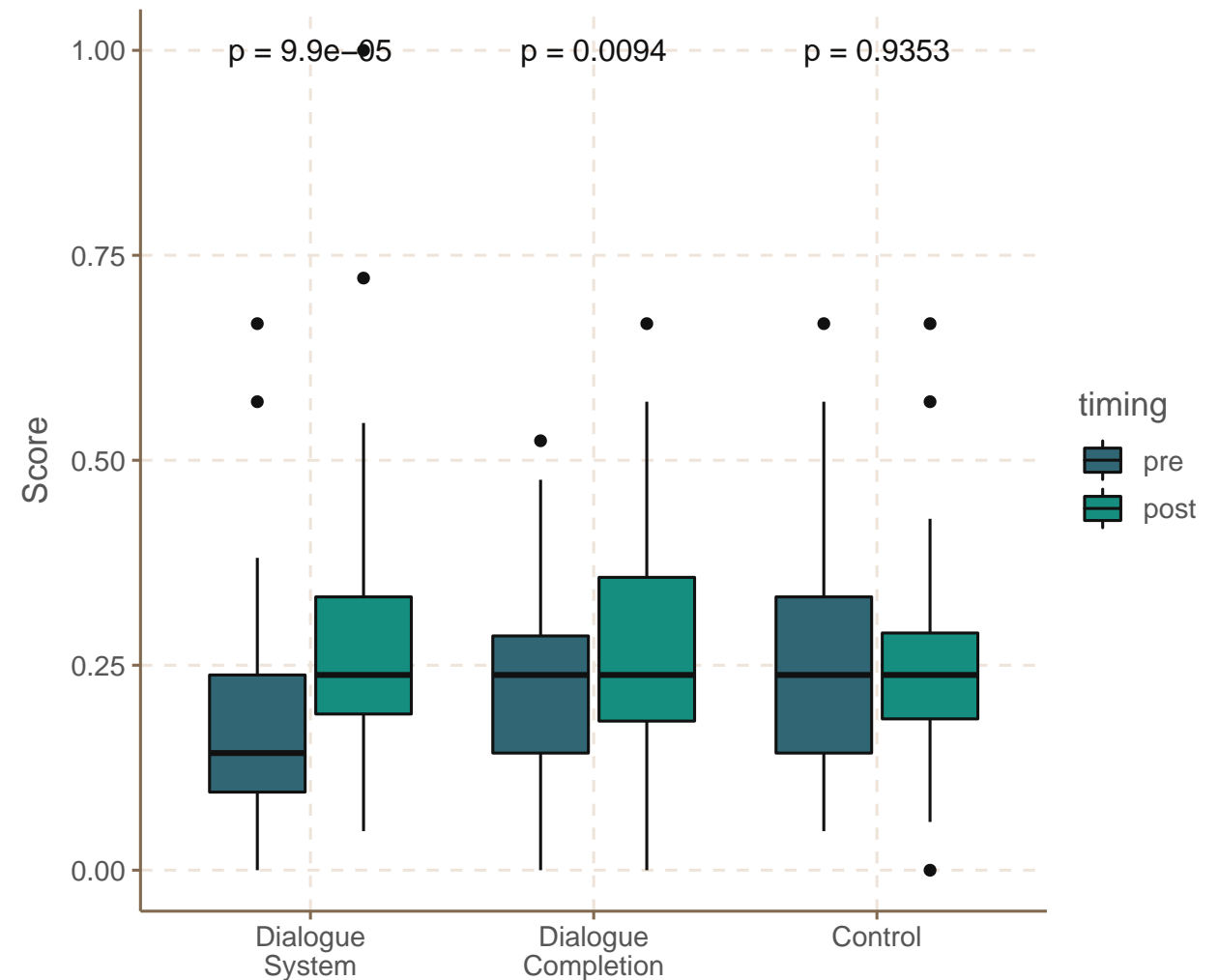
Less marked increase,  
and much more difficult test.

$$d_{\text{DSystem}} = 0.56^{***}$$

$$d_{\text{DCompletion}} = 0.39^{***}$$

$$d_{\text{DControl}} = 0.02 \text{ n.s.}$$

But here, no improvement in control  
group and benefits of practice are  
much clearer.



## Results

## Linear regression with mixed effects modelling

	$\beta$	SE	F	df	p
(Intercept) ***	-0.36	0.08	-4.54	72	0.000
Dialogue System <i>n.s.</i>	-0.03	0.04	-0.77	366	0.440
Dialogue (vs. Control) <i>n.s.</i>	0.05	0.05	1.19	328	0.234
<b>Score at pretest ***</b>	<b>0.39</b>	0.01	35.87	7195	0.000
<b>Vocabulary Size Score ***</b>	<b>0.10</b>	0.01	7.10	171	0.000
<b>Test: Receptive (vs. Productive) ***</b>	<b>0.41</b>	0.10	4.14	54	0.000
Gender: F °	0.05	0.03	1.91	171	0.058
Dialogue × Receptive °	0.08	0.05	1.66	7191	0.098
<b>Dialogue System × Receptive **</b>	<b>0.12</b>	0.04	3.28	7212	0.001
(1 Subject)	0.14				
(1 Item)	0.31				
(Residual)	0.70				

$R^2 = 0.45$

## Results & discussion

### Effects on vocabulary acquisition

Effect is, obviously defined by pretest score and voc. size, and higher for receptive vocabulary.

No direct distinction of effect between Dial.System and Dial.Completion, because it does not affect recall results that much:

**sig. differentiated effect on receptive voc. acquisition:**

Dial. system > Dial. completion > 'business-as-usual'

# Results

## Modelling number of encounters with words

### Frequency counts (lemma) of

- Potential encounters:
  - probable **input** (from bot's utterances)
  - opportunities for **output** (computed from model responses)
- Actual encounters
  - **input**: bot's utterances (but noticing?)
  - **output**: user's messages

# Results

## Modelling actual and potential encounters with words

	$\beta$	SE	F	df	p
(Intercept)	-0,53	0,21	-2,53	721	0,012
Dialogue System ***	-0,27	0,07	-3,86	424	0,000
Dialogue (vs. Control) *	0,45	0,21	2,18	1083	0,030
<b>Score at pretest</b>	<b>0,36</b>	0,02	16,58	1442	0,000
<b>Vocabuly size score</b>	<b>0,11</b>	0,03	3,90	141	0,000
<b>Test: Receptive (vs. Productive)</b>	<b>0,30</b>	0,32	0,94	1240	0,348
Gender: F	0,01	0,05	0,25	121	0,804
<b>Potential Encounters: Input *</b>	<b>0,12</b>	0,06	2,11	24	0,045
Potential Encounters: Output	-0,03	0,07	-0,39	28	0,699
Actual Encounters: Input	-0,03	0,02	-1,59	1356	0,112
<b>Actual Encounters: Output ***</b>	<b>0,10</b>	0,02	4,15	1614	0,000
Dialogue × Receptive	0,24	0,31	0,78	1782	0,435
<b>Dialogue System × Receptive **</b>	<b>0,23</b>	0,08	2,97	1727	0,003
(1 Participant)	0,19				
(1 Item)	0,20				
(Residual)	0,73				

$R^2 = 0.46$

## Results

# Modelling actual and potential encounters with words

Major predictor of incidental vocabulary learning is not input, but **output** (actual number of uses).

Is it different for receptive vs. productive learning?



# Results

## Exposure to words $\Rightarrow$ Receptive test

	$\beta$	SE	F	df	p
(Intercept)	-0,38	0,28	-1,35	476	0,178
Dialogue System	-0,05	0,08	-0,63	207	0,529
<b>Dialogue (vs. Ctrl) *</b>	<b>0,65</b>	0,28	2,28	469	0,023
<b>Score at pretest ***</b>	<b>0,37</b>	0,03	12,19	692	0,000
<b>Vocabuly size score ***</b>	<b>0,12</b>	0,04	3,34	141	0,001
Potential Encounters: Input <i>n.s.</i>	0,10	0,09	1,17	16	0,257
Potential Encounters: Output <i>n.s.</i>	-0,03	0,08	-0,33	14	0,748
Actual Encounters: Input <i>n.s.</i>	0,02	0,04	0,53	919	0,594
<b>Actua Encounters: Output **</b>	<b>0,08</b>	0,03	3,07	936	0,002
(1 Participant)	0,24				
(1 Item)	0,22				
(Residual)	0,75				

$R^2 = 0.33$

# Results

## Exposure to words $\Rightarrow$ Productive test

	$\beta$	SE	F	df	p
(Intercept)	-0,34	0,23	-1,46	362	0,144
Dialogue System ***	-0,32	0,08	-4,13	163	0,000
<b>Dialogue (vs. Ctrl) *</b>	0,55	0,23	2,42	496	0,016
<b>Score at pretest ***</b>	0,37	0,03	11,39	765	0,000
<b>Vocabuly size score **</b>	0,11	0,04	2,73	141	0,007
Potential Encounters: Input °	0,17	0,09	1,98	8	0,081
Potential Encounters: Output <i>n.s.</i>	-0,03	0,12	-0,20	11	0,842
<b>Actual Encounters: Input *</b>	<b>-0,07</b>	0,03	-1,98	631	0,048
<b>Actual Encounters: Output ***</b>	<b>0,14</b>	0,04	3,38	730	0,001
(1 Participant)	0,19				
(1 Item)	0,21				
(Residual)	0,82				

**R<sup>2</sup> = 0.33**

## Results & discussion

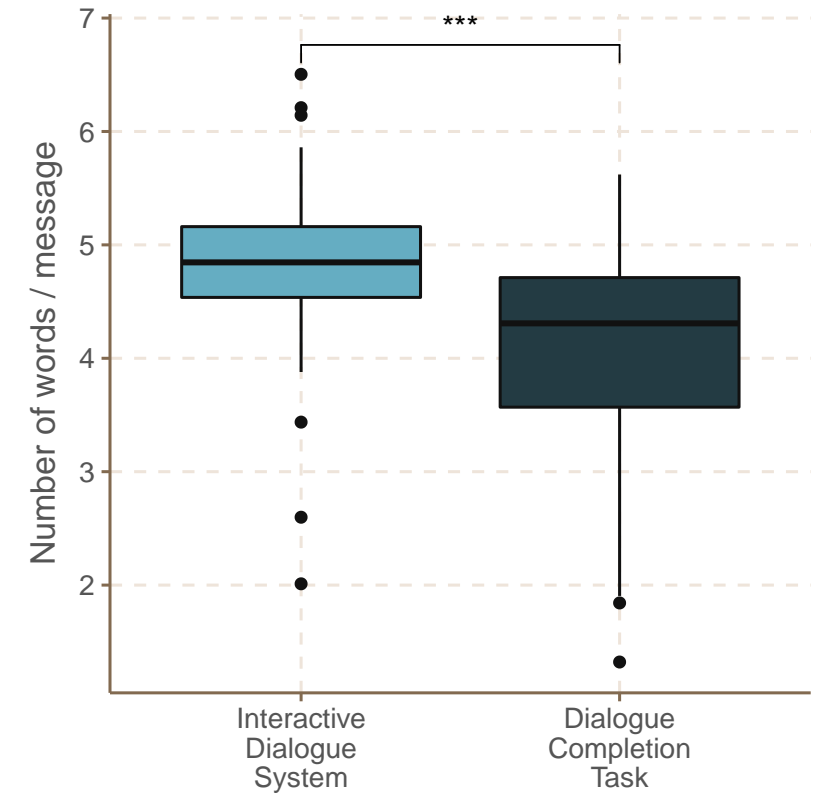
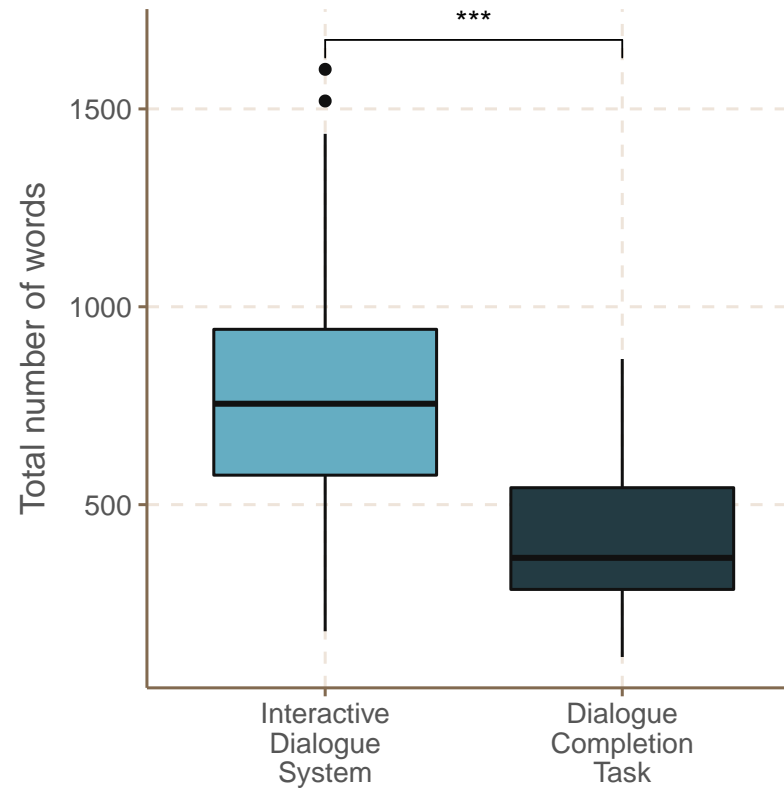
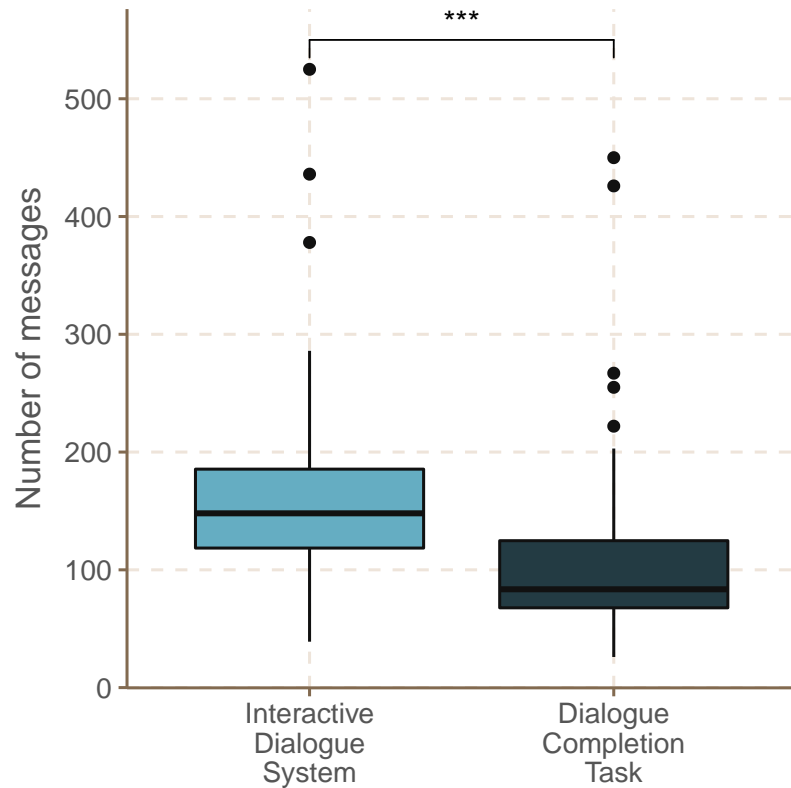
# Output encounters with words

It confirms the actual uses/output encounters influence both receptive and productive acquisition.

When accounting for output encounters, other coefficients go negative (e.g. Dialogue system) because their specificity is 'absorbed' by the number of uses.

# Results

## Quantity of in-task production



# Discussion

## Limitations

- Somewhat coarse automated processing for detection of encounters: can be refined substantially: include formulaic sequences that are tested in productive part.
- Suboptimal items, esp. in productive test, with very low or very high facility scores.
- Very “unadventurous/passive” behaviour in many participants from both conditions (fulfilling the microtasks like exercises):

Due to school context? age factor? presentation of the instructions?

→ Probably reduced the “interactivity” of the Dialogue system condition a lot.

# Effects of interactivity of written practice on incidental vocabulary acquisition



## Interaction-based vocabulary acquisition

How do output and interaction help lexical development?

## Dialogue systems for SLA research

Definition, summarized typology, research questions

## Randomized controlled experiment

Groups, learners' profiles, instruments, intervention

## Results and discussion

Mixed-effects models, effects of encounters in input/output

## Conclusions

# Interactivity and incidental vocabulary acquisition

- Effectiveness, in particular for **receptive acquisition**:  
Dialogue system > Dial. completion > Business-as-usual
- But effectiveness differences between *Dial. system* and *Dial. completion* not as strong as expected (worth the technological effort?), although it could be explained by the limited involvement of the participants, which could be due to a series of contextual and instructional factors.
- Actual number of uses (encounters) in production are a **highly significant predictor of word learning**, much higher than number of input encounters.

## Perspectives

# Dialogue systems as an SLA research environment

Dialogue systems offer  
**fully controllable and reproducible interaction:**  
opportunities to monitor and to alter infinity of details.

Experimental testing (A/B testing) with different types of tasks, instructions, feedback, exposure, reactions...



Thank you! Dank u!  
Merci! ¡Gracias!

**Serge Bibauw**

[serge.bibauw@kuleuven.be](mailto:serge.bibauw@kuleuven.be)

**Thomas François**

[thomas.françois@uclouvain.be](mailto:thomas.françois@uclouvain.be)

**Piet Desmet**

[piet.desmet@kuleuven.be](mailto:piet.desmet@kuleuven.be)

Any questions?  
Suggestions?  
Ideas?

# References

- Bibauw, S., François, T., & Desmet, P. (2019). Discussing with a computer to practice a foreign language: Research synthesis and conceptual framework of dialogue-based CALL. *Computer Assisted Language Learning*, 0(0), 1–51. <https://doi.org/10.1080/09588221.2018.1535508>
- de Vos, J. F., Schriefers, H., Nivard, M. G., & Lemhöfer, K. (2018). A meta-analysis and meta-regression of incidental second language word learning from spoken input. *Language Learning*, 68(4), 906–941. <https://doi.org/10.1111/lang.12296>
- De Wilde, V., Brysbaert, M., & Eyckmans, J. (2019). Learning English through out-of-school exposure. Which levels of language proficiency are attained and which types of input are important? *Bilingualism: Language and Cognition*, 1–15. <https://doi.org/10.1017/S1366728918001062>
- Elgort, I., Brysbaert, M., Stevens, M., & Assche, E. V. (2018). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*, 40(2), 341–366. <https://doi.org/10.1017/S0272263117000109>
- Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition*, 21(2), 285–301.
- Newton, J. (1995). Task-based interaction and incidental vocabulary learning: A case study. *Second Language Research*, 11(2), 159–176. <https://doi.org/10.1177/026765839501100207>
- Peters, E., Noreillie, A.-S., Heylen, K., Bulté, B., & Desmet, P. (2019). The impact of instruction and out-of-school exposure to FL input on learners' vocabulary knowledge in two languages. *Language Learning*, 69(3). <https://doi.org/10.1111/lang.12351>
- Peters, E., Velghe, T., & Van Rompaey, T. (2019). The VocabLab tests: The development of an English and French vocabulary test. *ITL - International Journal of Applied Linguistics*, 170(1), 53–78. <https://doi.org/10.1075/itl.17029.pet>