"Oye Siri, aprendo un idioma hablando contigo?"
Meta-análisis de IA conversacional para idiomas

Serge Bibauw[1,2,3], con Wim Van den Noortgate[2], Thomas François[3] y Piet Desmet[2]

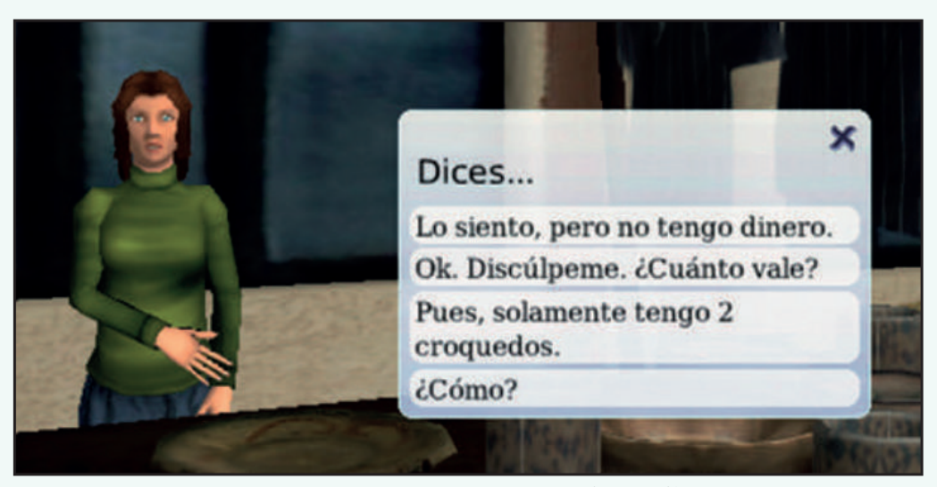
[1] Fac. de Filosofía, Carrera de Ingles, Universidad Central del Ecuador [2] ITEC, KU Leuven [3] Cental, UCLouvain

Dialogue-based CALL

Dialogue-based CALL systems involve (Bibauw, François & Desmet, *2019*)

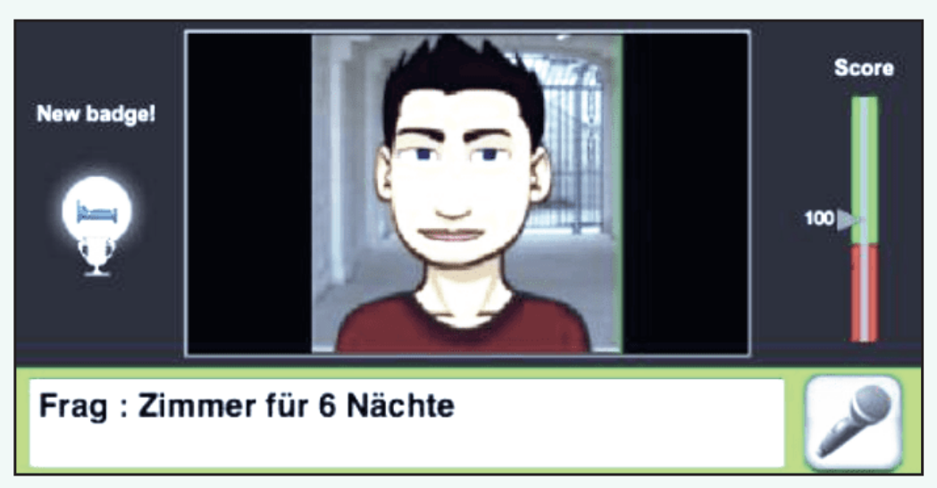
- a dialogue (i.e., sequence of conversational turns)
- with an automated agent (chatbot, robot, voice assistant, non player character...)
- as a language learning task (≠scaffolding).

Narrative systems


Dices...
Lo siento, pero no tengo dinero.
Ok. Discúlpeme. ¿Cuánto vale?
Pues, solamente tengo 2 croquedos.
¿Cómo?

Croquelandia (Sykes, 2013)

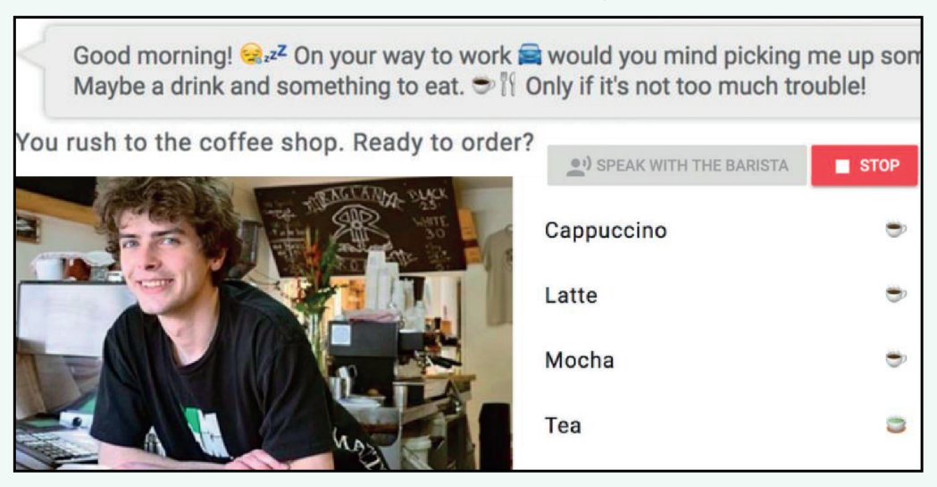
Form-focused systems



Frag : Zimmer für 6 Nächte

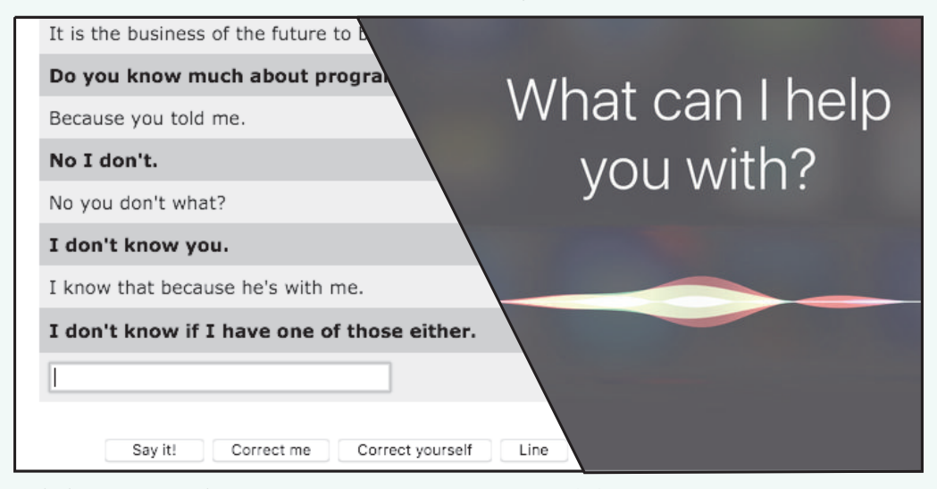
CALL-SLIT (Baur *et al*, 2014)

Goal-oriented systems


Good morning! On your way to work would you mind picking me up som
Maybe a drink and something to eat. Only if it's not too much trouble!
You rush to the coffee shop. Ready to order?

HALEF/ETS (Sydorenko *et al*, 2018)

Reactive systems


It is the business of the future to b
Do you know much about progra
Because you told me.
No I don't.
No you don't what?
I don't know you.
I know that because he's with me.
I don't know if I have one of those either.

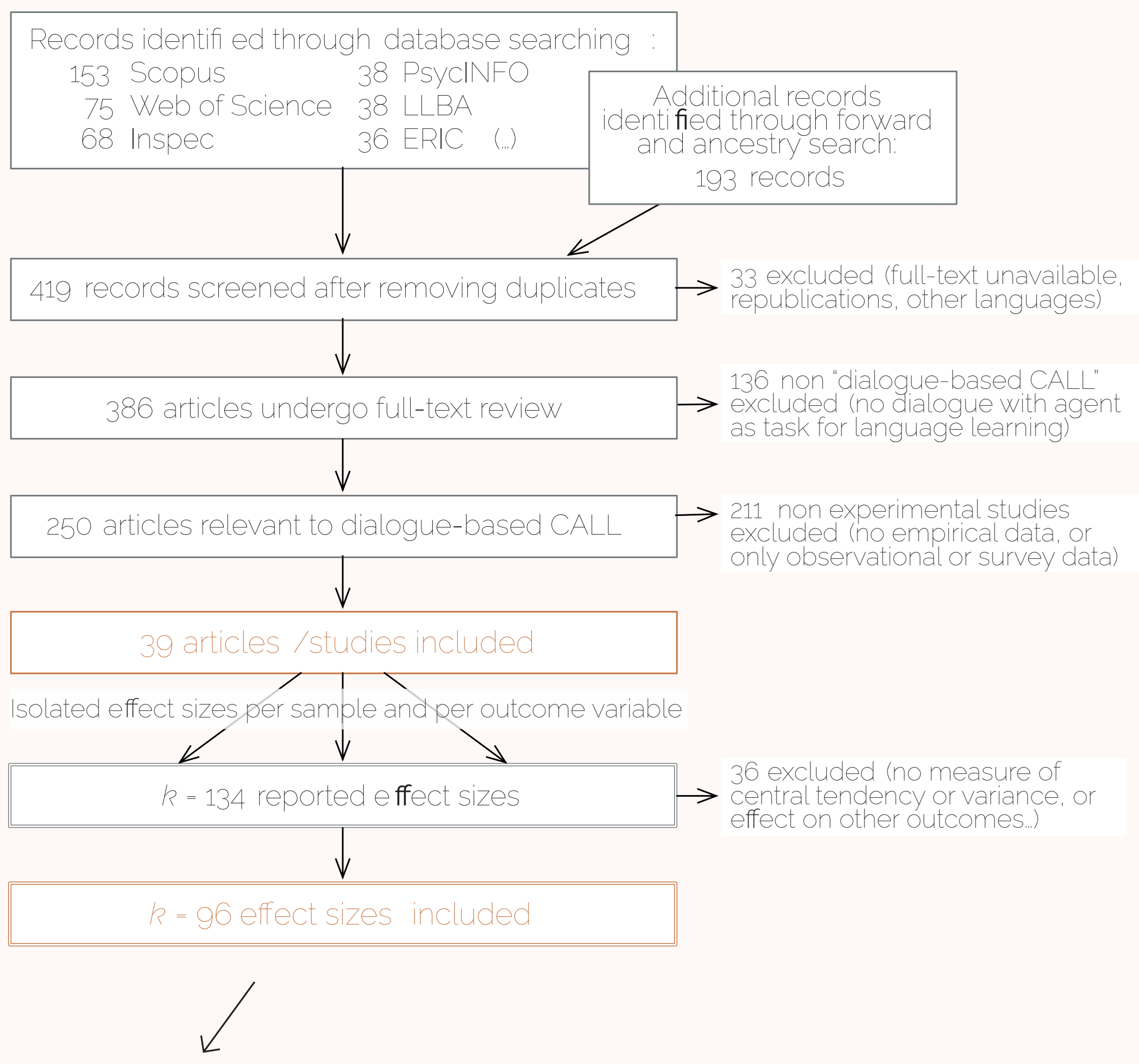
What can I help you with?

Jabberwacky (Fryer *et al*, 2017); *Siri* (Underwood, 2017)

RQ1 How effective is dialogue-based CALL in general for L2 development?

RQ2 How different implementations of dialogue-based CALL, distinguished by characteristics of instructional and system design, compare to each other in terms of effectiveness on diverse language learning outcomes?

Métodos


Records identifi ed through database searching :
153 Scopus 38 PsycINFO
75 Web of Science 38 LLBA
68 Inspec 36 ERIC (...)

Additional records identified through forward and ancestry search: 193 records

419 records screened after removing duplicates → 33 excluded (full-text unavailable, republications, other languages)

386 articles undergo full-text review → 136 non "dialogue-based CALL" excluded (no dialogue with agent as task for language learning)

250 articles relevant to dialogue-based CALL → 211 non experimental studies excluded (no empirical data, or only observational or survey data)

39 articles /studies included

Isolated effect sizes per sample and per outcome variable

k = 134 reported e ffect sizes → 36 excluded (no measure of central tendency or variance, or effect on other outcomes...)

k = 96 effect sizes included

Standardized Mean Difference (*d*) computed with single raw metric (Morris & DeShon, 2002):

$$d_{\mathrm{PP,raw}} = c(df_{\mathrm{PP}})\left(\frac{M_{\mathrm{post,E}} - M_{\mathrm{pre,E}}}{SD_{\mathrm{pre,E}}}\right)$$

$$d_{\mathrm{ECPP,raw}} = c(df_{\mathrm{ECPP}})\left(\frac{M_{\mathrm{post,E}} - M_{\mathrm{pre,E}}}{SD_{\mathrm{pre,E}}} - \frac{M_{\mathrm{post,C}} - M_{\mathrm{pre,C}}}{SD_{\mathrm{pre,C}}}\right)$$

Meta-analysis: statistical summary of studies, aggregating all compatible effects to compute a summary effect.

Multivel meta-analysis

- every measurement of effect on each outcome variable for each sample is included as a single effect size;
- lack of independence between effects from the same study taken into account by layer of random variation at the study level;
- allows high granularity in study of moderator variables.

(see Van den Noortgate et al 2012)

Level of aggregation	Items/clusters	Remaining variation
Study	$k_{\mathrm{studies}} = 17$	Variation between-studies
Effect size	$k = 96$	Variation between-subjects
Subject	$n = 803$	Random sampling variance

Mixed-effects model:

- random between-studies effect
- random between-subjects effect
- fixed effects for covariates and moderator variables

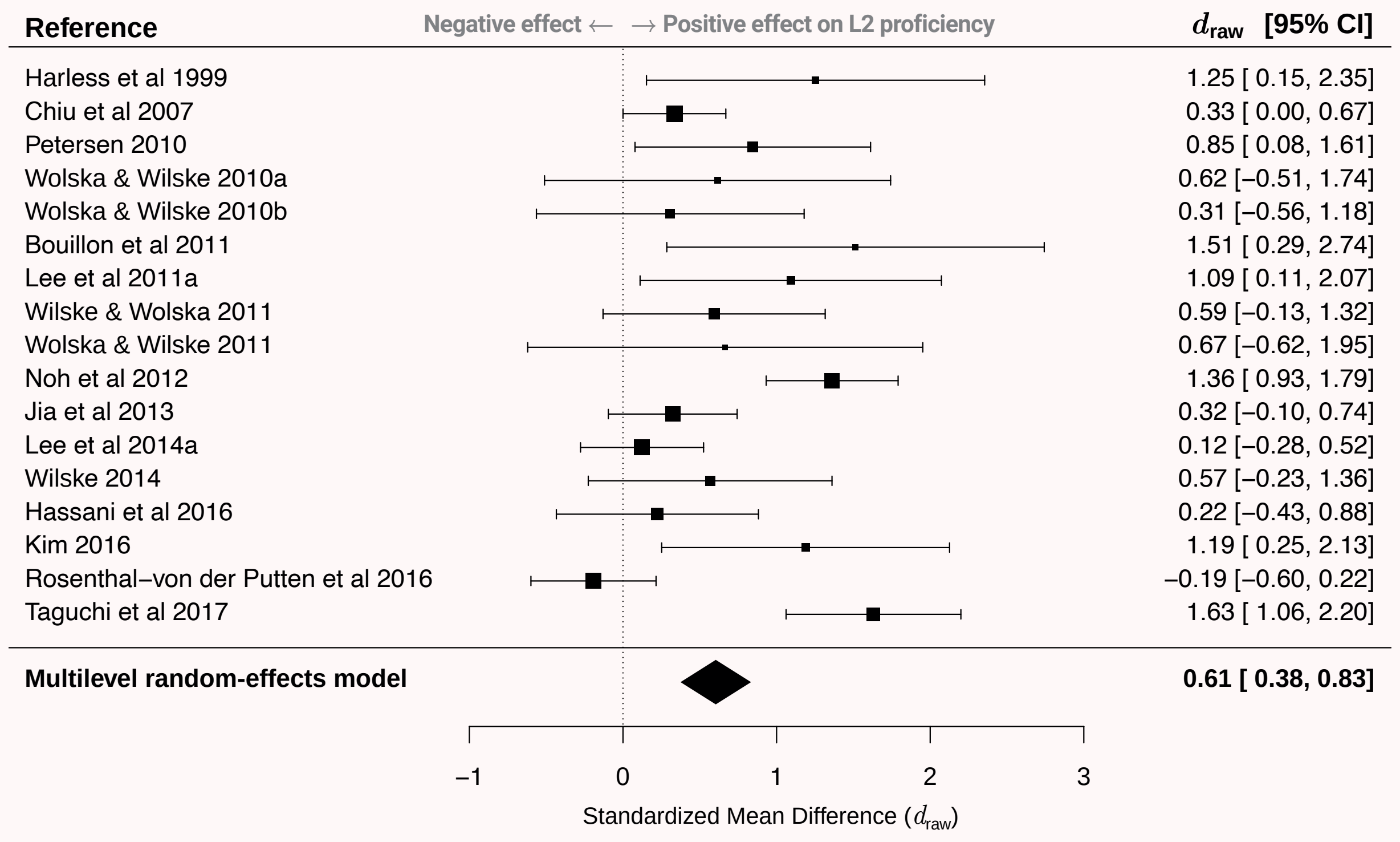
Resultados

Efecto global medio-alto del DBCALL sobre competencia en L2:

$$\boldsymbol{d}_{\mathrm{raw}} = \mathbf{.61}$$

Alta heterogeneidad y potencia estadística limitada en estudios existentes:

$$Q(df = 95) = 301.3$$



Análisis de moderadores

Type	Variable	*df*	*F*	*p*	Values	*k*	*d*	*SE*	CI
Population	L2 proficiency*	4	9.55	.049	intercept		0.69	0.38	[-0.049, 1.436]
					A1	38	0.36	0.21	[-0.056, 0.775]
					A2	89	0.18	0.30	[-0.416, 0.769]
					B1	77	-0.42	0.25	[-0.910, 0.066]
					B2	28	-0.41	0.28	[-0.962, 0.150]
	Context	2	1.03	.599	school	18	0.68	0.23	[0.235, 1.121]**
					university	75	0.54	0.15	[0.259, 0.830]***
					military	3	1.08	0.55	[0.002, 2.160]*
Treatment	Duration*	4	10.29	.036	intercept		0.09	0.20	[-0.300, 0.484]
					+1 hour on task		0.15	0.05	[0.049, 0.256]**
					+1 session		0.31	0.11	[0.094, 0.523]**
					+1 week		-0.19	0.08	[-0.338, -0.037]*
	Type of system	3	1.38	.710	narrative	4	0.31	0.49	[-0.643, 1.261]
					form-focused	15	0.86	0.27	[0.336, 1.392]**
					goal-oriented	71	0.56	0.16	[0.244, 0.877]***
					reactive	6	0.57	0.37	[-0.156, 1.287]
	Type of interaction	2	0.46	.794	goal-oriented	66	0.64	0.14	[0.373, 0.907]***
					open-ended	6	0.56	0.36	[-0.146, 1.276]
					system-guided	4	0.31	0.48	[-0.627, 1.245]
	System modality	1	0.03	.873	spoken	25	0.59	0.17	[0.256, 0.920]***
					written	61	0.63	0.17	[0.293, 0.960]***
	Corrective feedback	2	2.53	.283	explicit	36	0.75	0.16	[0.447, 1.059]***
					implicit	37	0.71	0.15	[0.415, 1.005]***
					none	23	0.37	0.18	[0.013, 0.732]*
Outcome	Test modality	1	1.72	.190	spoken	35	0.74	0.16	[0.427, 1.054]***
					written	61	0.52	0.14	[0.249, 0.799]***
	Matching modality (treatment=test)	1	2.52	.113	true	72	0.68	0.13	[0.428, 0.923]***
					false	24	0.40	0.17	[0.063, 0.745]*
	Outcome type***	2	16.32	<.001	comprehension	4	-0.45	0.33	[-1.095, 0.201]
					production	80	0.76	0.16	[0.453, 1.069]***
					vocabulary	12	0.41	0.25	[-0.083, 0.899]
	Outcome dimension**	6	18.68	.005	holistic proficiency	11	0.76	0.26	[0.263, 1.265]**
					complexity	1	0.68	0.48	[-0.262, 1.614]
					accuracy	49	0.52	0.18	[0.176, 0.871]**
					lexicon	17	0.83	0.23	[0.375, 1.292]***
					fluency	14	0.65	0.23	[0.207, 1.097]**
	Type of test	3	7.75	.051	metaling. judgment	20	0.58	0.20	[0.184, 0.969]**
					selected response	9	0.17	0.23	[-0.280, 0.621]
					constrained resp.	32	0.71	0.18	[0.355, 1.064]***
					free response	35	0.76	0.18	[0.412, 1.109]***
	Temporality of effects	1	0.60	.439	short-term	73	0.62	0.12	[0.388, 0.860]***
					long-term	23	0.52	0.16	[0.202, 0.838]**

Differenciated effects across levels: beginners tend to benefit more

Tentative modelization of effects of treatment duration: Time on task + #Session - Time between sessions

Goal/task-oriented interaction seems to provide more learning oppportunities than open-ended (e.g., small talk) or system-guided interactions

← Spoken and written practice seem to have very similar effects

← ...but effects could be slightly stronger or more visible on speaking

Learning effects are much stronger on production outcomes, and could be close to zero regarding an improvement in comprehension

All 4 CALF dimensions seem to benefit from DBCALL, but the effects seem stronger on vocabulary & fluency (and possibly complexity)

Effects are higher when tested through free or constrained production tasks than in other types of instruments

Resumen

El "CALL basado en diálogo" (dialogue-based CALL) incluye chatbots, agentes conversacionales, asistentes de voz, robots y NPC parlantes para el aprendizaje de idiomas.

Realizamos un meta-análisis multinivel de los estudios de eficacia realizados sobre dichos sistemas (250 artículos), recopilando 96 tamaños de efecto. Se aplicaron fórmulas y modelos estadísticos innovadores para integrar resultados.

El efecto general de la práctica CALL basada en el diálogo sobre el desarrollo de la competencia en L2 es medio-alto (*d* = .61). Es comparable, aunque lógicamente inferior, al efecto de la interacción humano-humano medido por otra meta-análisis (Mackey & Goo, 2007: *d* = .75).

Entre las conclusiones del análisis de moderador: efecto diferenciado entre los niveles de competencia (los principiantes se benefician más que los alumnos avanzados) y efectos más fuertes en las tareas de producción, especialmente en las medidas de vocabulario y fluidez.

Descargar este poster
Lista de referencias

sbibauw@uce.edu.ec