L2 development and use of English grammatical morphemes: Insights from learner corpora

Akira Murakami
University of Birmingham
a.murakami@bham.ac.uk

Learner Corpora in L2 Research

- Until the early 2010s, not much research had been carried out at the intersection between learner corpus research and second language acquisition (SLA).

- Recent years have seen the publication of special issues (Fuchs & Werner, 2018; Hasko, 2013; Rebuschat et al., 2017), books (Le Bruyn & Paquot, 2021; Lu, 2022; Tracy-Ventra & Paquot, 2021), and papers in flagship journals in SLA (e.g., Murakami, 2016; Paquot, 2019) at the interface.

- As the use of learner corpora is becoming increasingly common in L2 research (e.g., Paquot & Plonsky, 2017), it comes to be important for us to understand the nature and characteristics of learner corpus data, as well as how to analyze the relatively new data properly.

Learner Corpora as Data in SLA

- From the perspective of SLA, learner corpora are a type of free production data, which are not uncommon.
    - Even earliest studies in SLA employed free production data (Dulay & Burt, 1973; Hakuta, 1974), and Dulay and Burt (1974) indeed called their data “corpuses”.
- The ESF (European Science Foundation Second Language) Database (Feldweg, 1991; Klein & Perdue, 1992; Perdue, 1993) can be considered as a precursor to modern learner corpora.

Learner Corpora as Data in SLA

- Many methodological issues and analytical techniques are shared between learner corpora and other types of free production data.
    - With rich metadata, it is possible to perform e.g., conversation analyses with a learner corpus (e.g., *no* VERB vs *don't* VERB in Eskildsen, 2012).

- That said, a prominent feature of modern learner corpora is their scale, and researchers often exploit the scale in conducting L2 research based on learner corpora.

Implications

- Larger data size does not mean that findings are generalizable.
- Larger data size means
  1. Analyses have higher statistical power.
     - Small differences can be reliably identified.
     - Infrequent features can also be investigated with a reasonable power.
  2. Multifactorial analyses are possible.
     - Larger scope and finer granularity
- The quantity of data in learner corpora allows us to answer the research questions that we would not have been able to without them.

Analysis of Large-Scale Learner Corpora

- A trend in learner corpus research is from monofactorial studies to multifactorial studies.

  - monofactorial: single predictor (e.g., lexical aspect)

  - multifactorial: multiple predictors (e.g., lexical aspect + word frequency + L1 + proficiency + task effects + . . .)

Multifactorial Analyses are Desirable

- Multifactorial analyses are preferred in learner corpus research (see Gries & Wulff, 2013).
- Since practically any phenomenon in L2 research is influenced by a number of factors, multifactorial analyses are generally more desirable than monofactorial analyses.
    - In particular, in corpus analysis, variables (e.g., tasks, L1, proficiency, etc.) often need to be controlled for in a post-hoc manner through statistical analysis.

Multifactorial Analyses

- Multifactorial analyses lead to

    1. larger scope because they incorporate multiple predictors
    2. finer granularity because they disentangle different sources of variability in data

- Based on International Corpus of English (ICE), Rautionaho and Deshors (2018) showed that genres and L2 varieties interact with other variables (e.g., lexical aspect) in the use of progressive *-ing*.

Illustration

Grammatical Morphemes

- Grammatical morphemes are known to be notoriously difficult for L2 learners to acquire.

- Morpheme development is a slow, gradual process (e.g., Jia & Fuse, 2007).

- Not all inflected forms are as easily acquired.

    - Some morphemes are acquired earlier than others (e.g., morpheme order studies)

    - Within each morpheme, certain forms are acquired earlier than others (e.g., Aspect Hypothesis)

Grammatical Morphemes

- Grammatical morphemes are known to be notoriously difficult for L2 learners to acquire.

- Morpheme development is a slow, gradual process (e.g., Jia & Fuse, 2007).

- Not all inflected forms are as easily acquired.

  - Some morphemes are acquired earlier than others (e.g., morpheme order studies)

  - Within each morpheme, certain forms are acquired earlier than others (e.g., Aspect Hypothesis)

Grammatica

- Grammatical morphemes are
for L2 learners to acquire.

Murakami, A., & Alexopoulou, T. (2016). L1 influence on the acquisition order of English grammatical morphemes: A learner corpus study. *Studies in Second Language Acquisition, 38*(3), 365–401. http://doi.org/10.1017/S0272263115000352

- Some morphemes are acq
morpheme order studies)

- Within each morpheme, ce
than others (e.g., Aspect H

Studies in Second Language Acquisition, 2016, **38**, 365–401.
doi:10.1017/S0272263115000352

L1 INFLUENCE ON THE ACQUISITION ORDER OF ENGLISH GRAMMATICAL MORPHEMES

A Learner Corpus Study

Akira Murakami
University of Birmingham

Theodora Alexopoulou
University of Cambridge

We revisit morpheme studies to evaluate the long-standing claim for a universal order of acquisition. We investigate the L2 acquisition order of six English grammatical morphemes by learners from seven L1 groups across five proficiency levels. Data are drawn from approximately 10,000 written exam scripts from the Cambridge Learner Corpus. The study establishes clear L1 influence on the absolute accuracy of morphemes and their acquisition order, therefore challenging the widely held view that there is a universal order of acquisition of L2 morphemes. Moreover, we find that L1 influence is morpheme specific, with morphemes encoding language-specific concepts most vulnerable to L1 influence.

Background and Aim

- It has been believed in SLA that the L2 acquisition order of English grammatical morphemes is universal, irrespective of learners' L1s (Krashen, 1977; Ortega, 2009).

- Natural Order (Krashen, 1977):

-ing *plural -s* *copula be*	→	auxiliary *be* articles	→	irregular past tense	→	regular past tense third person *-s* possessive *'s*

- Luk and Shirai (2009) suggested that L1 affects the order.

- The aim of our study was to empirically test the universality of the order.

Data

- A subcorpus of the Cambridge Learner Corpus
- Exam scripts of Cambridge English Main Suite Exams aligned with Common European Framework of Reference levels (KET, PET, FCE, CAE, CPE)
- Manually error-tagged
- Seven L1 groups: Japanese, Korean, Spanish, Russian, Turkish, German, and French
- Approximately 12,000 essays
- Six morphemes: articles, past tense *-ed*, plural *-s*, possessive *'s*, progressive *-ing*, and third person *-s*

Accuracy Measure

- We calculated the TLU (Target-Like Use; Pica, 1983) of each morpheme in each L1 group in each proficiency level based on the error annotation.

$$\text{TLU score} = \frac{\text{number of correct suppliance}}{\text{number of obligatory contexts} + \text{number of incorrect suppliance}}$$

PET (B1)

	L1 Japanese	L1 Korean	L1 Spanish	L1 Russian	L1 Turkish	L1 German	L1 French
1	past tense *-ed*	past tense *-ed*	articles	past tense *-ed*	past tense *-ed*	articles	articles
	progressive *-ing*	progressive *-ing*	plural *-s*	plural *-s*	progressive *-ing*	plural *-s*	
			progressive *-ing*	progressive *-ing*	third person *-s*		
				~~third person *-s*~~			
2	plural *-s*	plural *-s*	past tense *-ed*	articles	plural *-s*	past tense *-ed*	past tense *-ed*
	third person *-s*	~~possessive *'s*~~		possessive *'s*			plural *-s*
		third person *-s*					progressive *-ing*
3	articles	articles	possessive *'s*		~~possessive *'s*~~	progressive *-ing*	third person *-s*
	~~possessive *'s*~~		third person *-s*		articles	third person *-s*	
4						~~possessive *'s*~~	possessive *'s*

PET (B1)

	L1 Japanese	L1 Korean	L1 Spanish	L1 Russian	L1 Turkish	L1 German	L1 French
	past tense *-ed*	past tense *-ed*	articles	past tense *-ed*	past tense *-ed*	articles	articles
	progressive *-ing*	progressive *-ing*	plural *-s*	plural *-s*	progressive *-ing*	plural *-s*	
1			progressive *-ing*	progressive *-ing*	third person *-s*		
				~~third person *-s*~~			
	plural *-s*	plural *-s*	past tense *-ed*	articles	plural *-s*	past tense *-ed*	past tense *-ed*
	third person *-s*	~~possessive *'s*~~		possessi *'s*			plural *-s*
2		third person *-s*					progressive *-ing*
	articles	articles	possessive *'s*		~~possessive *'s*~~	progressive *-ing*	third person *-s*
3	~~possessive *'s*~~		third person *-s*		articles	third person *-s*	
4						~~possessive *'s*~~	possessive *'s*

PET (B1)

	L1 Japanese	L1 Korean	L1 Spanish	L1 Russian	L1 Turkish	L1 German	L1 French
1	past tense *-ed*	past tense *-ed*	articles	past tense *-ed*	past tense *-ed*	articles	articles
	progressive *-ing*	progressive *-ing*	plural *-s*	plural *-s*	progressive *-ing*	plural *-s*	
			progressive *-ing*	progressive *-ing*	third person *-s*		
				~~third person *-s*~~			
2	plural *-s*	plural *-s*	past tense *-ed*	articles	plural *-s*	past tense *-ed*	past tense *-ed*
	third person *-s*	~~possessive 's~~		possessive *'s*			plural *-s*
		third person *-s*					progressive *-ing*
3	articles	articles	possessive *'s*		~~possessive 's~~	progressive *-ing*	third person *-s*
	~~possessive 's~~		third person *-s*		articles	third person *-s*	
4						~~possessive 's~~	possessive *'s*

PET (B1)

	L1 Japanese	L1 Korean	L1 Spanish	L1 Russian	L1 Turkish	L1 German	L1 French
1	past tense *-ed*	past tense *-ed*	articles	past tense *-ed*	past tense *-ed*	articles	articles
	progressive *-ing*	progressive *-ing*	plural *-s*	plural *-s*	progressive *-ing*	plural *-s*	
			progressive *-ing*	progressive *-ing*	third person *-s*		
				~~third person *-s*~~			
2	plural *-s*	plural *-s*	past tense *-ed*	articles	plural *-s*	past tense *-ed*	past tense *-ed*
	third person *-s*	~~possessive 's~~		possessive *'s*			plural *-s*
		third person *-s*					progressive *-ing*
3	articles	articles	possessive *'s*		~~possessive 's~~	progressive *-ing*	third person *-s*
	~~possessive 's~~		third person *-s*		articles	third person *-s*	
4						~~possessive 's~~	possessive *'s*

PET (B1)

	L1 Japanese	L1 Korean	L1 Spanish	L1 Russian	L1 Turkish	L1 German	L1 French
1	past tense *-ed* progressive *-ing*	past tense *-ed* progressive *-ing*	articles plural *-s* progressive *-ing*	past tense *-ed* plural *-s* progressive *-ing* ~~third person *-s*~~	past tense *-ed* progressive *-ing* third person *-s*	articles plural *-s*	articles
2	plural *-s* third person *-s*	plural *-s* ~~possessive *'s*~~ third person *-s*	past tense *-ed*	articles possessive *'s*	plural *-s*	past tense *-ed*	past tense *-ed* plural *-s* progressive *-ing*
3	articles ~~possessive *'s*~~	articles	possessive *'s* third person *-s*		~~possessive *'s*~~ articles	progressive *-ing* third person *-s*	third person *-s*
4						~~possessive *'s*~~	possessive *'s*

CPE (C2)

	L1 Japanese	L1 Korean	L1 Spanish	L1 Russian	L1 Turkish	L1 German	L1 French
1	past tense *-ed* progressive *-ing*	past tense *-ed* progressive *-ing*	articles past tense *-ed* plural *-s* progressive *-ing* third person *-s*	past tense *-ed* plural *-s* progressive *-ing* third person *-s*	past tense *-ed* plural *-s* progressive *-ing*	articles past tense *-ed* plural *-s* third person *-s*	articles past tense *-ed* plural *-s* third person *-s*
2	plural *-s* possessive *'s* third person *-s*	plural *-s* possessive *'s* third person *-s*	possessive *'s*	possessive *'s*	articles third person *-s*	possessive *'s* progressive *-ing*	progressive *-ing*
3	articles	articles		articles	possessive *'s*		possessive *'s*
4							

CPE (C2)

	L1 Japanese	L1 Korean	L1 Spanish	L1 Russian	L1 Turkish	L1 German	L1 French
1	past tense *-ed* progressive *-ing*	past tense *-ed* progressive *-ing*	articles past tense *-ed* plural *-s* progressive *-ing* third person *-s*	past tense *-ed* plural *-s* progressive *-ing* third person *-s*	past tense *-ed* plural *-s* progressive *-ing*	articles past tense *-ed* plural *-s* third person *-s*	articles past tense *-ed* plural *-s* third person *-s*
2	plural *-s* possessive *'s* third person *-s*	plural *-s* possessive *'s* third person *-s*	possessive *'s*	possessive *'s*	articles third person *-s*	possessive *'s* progressive *-ing*	progressive *-ing*
3	articles	articles	articles		possessive *'s*		possessive *'s*
4							

QPE (C2)

	L1 Japanese	L1 Korean	L1 Spanish	L1 Russian	L1 Turkish	L1 German	L1 French
1	past tense *-ed*	past tense *-ed*	articles	past tense *-ed*	past tense *-ed*	articles	articles
	progressive *-ing*	progressive *-ing*	past tense *-ed*	plural *-s*	plural *-s*	past tense *-ed*	past tense *-ed*
			plural *-s*	progressive *-ing*	progressive *-ing*	plural *-s*	plural *-s*
			progressive *-ing*	third person *-s*		third person *-s*	third person *-s*
			third person *-s*				
2	plural *-s*	plural *-s*	possessive *'s*	possessive *'s*	articles	possessive *'s*	progressive *-ing*
	possessive *'s*	possessive *'s*			third person *-s*	progressive *-ing*	
	third person *-s*	third person *-s*					
3	articles	articles		articles	possessive *'s*		possessive *'s*
4							

CPE (C2)

	L1 Japanese	L1 Korean	L1 Spanish	L1 Russian	L1 Turkish	L1 German	L1 French
1	past tense *-ed* progressive *-ing*	past tense *-ed* progressive *-ing*	articles past tense *-ed* plural *-s* progressive *-ing* third person *-s*	past tense *-ed* plural *-s* progressive *-ing* third person *-s*	past tense *-ed* plural *-s* progressive *-ing*	articles past tense *-ed* plural *-s* third person *-s*	articles past tense *-ed* plural *-s* third person *-s*
2	plural *-s* possessive *'s* third person *-s*	plural *-s* possessive *'s* third person *-s*	possessive *'s*	possessive *'s*	articles third person *-s*	possessive *'s* progressive *-ing*	progressive *-ing*
3	articles	articles		articles	possessive *'s*		possessive *'s*
4							

OPE (C2)

	L1 Japanese	L1 Korean	L1 Spanish	L1 Russian	L1 Turkish	L1 German	L1 French
1	past tense *-ed* progressive *-ing*	past tense *-ed* progressive *-ing*	articles past tense *-ed* plural *-s* progressive *-ing* third person *-s*	past tense *-ed* plural *-s* progressive *-ing* third person *-s*	past tense *-ed* plural *-s* progressive *-ing*	articles past tense *-ed* plural *-s* third person *-s*	articles past tense *-ed* plural *-s* third person *-s*
2	plural *-s* possessive *'s* third person *-s*	plural *-s* possessive *'s* third person *-s*	possessive *'s*	possessive *'s*	articles third person *-s*	possessive *'s* progressive *-ing*	progressive *-ing*
3	articles	articles		articles	possessive *'s*		possessive *'s*
4							

Between-L1 Differences of Orders

Level	Articles	Past tense *-ed*	Plural *-s*	Possessive *'s*	Progressive *-ing*	Third person *-s*
CPE	SGF > JKRT T > JKR			JKR > STF	JKRT > GF	
CAE	SGF > JKRT	JKT > S		K > JSRTGF	JKSRT > GF	JRTGF > K
FCE	SGF > JKRT	JKTF > SR G > S	SR > JKTF	JK > STGF	JKSRTF > G	RT > SF
PET	SGF > JKRT	JK > SG TG > S	G > JKT S > T		JKST > G	T > SGF J > SF

J = L1 Japanese; K = L1 Korean; S = L1 Spanish; R = L1 Russian; T = L1 Turkish; G = L1 German; F = L1 French

L1 Differences of Orders

In L1 Spanish, German, and French learners of English, the rank of the articles is higher than in L1 Japanese, Korean, Russian, and Turkish learners of English.

Level	Articles	Past tense *-ed*	Plural *-s*	Possessive *'s*	Progressive *-ing*	Third person *-s*
CPE	SGF > JKRT T > JKR			JKR > STF	JKRT > GF	
CAE	SGF > JKRT	JKT > S		K > JSRTGF	JKSRT > GF	JRTGF > K
FCE	SGF > JKRT	JKTF > SR G > S	SR > JKTF	JK > STGF	JKSRTF > G	RT > SF
PET	SGF > JKRT	JK > SG TG > S	G > JKT S > T		JKST > G	T > SGF J > SF

J = L1 Japanese; K = L1 Korean; S = L1 Spanish; R = L1 Russian; T = L1 Turkish; G = L1 German; F = L1 French

Within-L1 Differences of Orders

L1	Articles	Past tense *-ed*	Plural *-s*	Possessive *'s*	Progressive *-ing*	Third person *-s*
Japanese						
Korean						
Spanish						
Russian						
Turkish						P > Ca
German						
French	P > F				FP > CpCa	CpCa > FP

Cp = CPE; Ca = CAE; F = FCE; P = PET

Within-L1 Differences of Orders

L1	Articles	Past tense *-ed*	Plural *-s*	Possessive *'s*	Progressive *-ing*	Third person *-s*
Japanese						
Korean						
Spanish						
Russian						
Turkish						P > Ca
German						
French		P > F			FP > CpCa	CpCa > FP

Cp = CPE; Ca = CAE; F = FCE; P = PET

Large-scale corpus data allowed us to compare between-L1 and within-L1 differences (Jarvis, 2000, 2010), thereby supporting our claim that what we observe is indeed L1 influence.

“Natural” Order

1	plural *-s*
	progressive *-ing*
2	articles
3	third person *-s*
	possessive *'s*
	past tense *-ed*

Differences Between the Observed Order and the Natural Order

Level	Articles	Past tense *-ed*	Plural *-s*	Possessive *'s*	Progressive *-ing*	Third person *-s*
CPE	NO > JKRT	JKT > NO	NO > JK		NO > GF	
CAE	NO > JKRT	JKT > NO	NO > K		NO > GF	
FCE	NO > JKRT	JKTGF > NO	NO > JKTF		NO > G	
PET	NO > JKRT GF > NO	JKTG > NO	NO > JKT		NO > G	JT > NO

J = L1 Japanese; K = L1 Korean; R = L1 Russian; T = L1 Turkish; G = L1 German; F = L1 French; NO = Natural Order

Differences Between the Observed Order and the Natural Order

Level	Articles	Past tense *-ed*	Plural *-s*	Possessive *'s*	Progressive *-ing*	Third person *-s*
CPE	NO > JKRT	JKT > NO	NO > JK		NO > GF	
CAE	NO > JKRT	JKT > NO	NO > K		NO > GF	
FCE	NO > JKRT	JKTGF > NO	NO > JKTF		NO > G	
PET	NO > JKRT GF > NO	JKTG > NO	NO > JKT		NO > G	JT > NO

J = L1 Japanese; K = L1 Korean; R = L1 Russian; T = L1 Turkish; G = L1 German; F = L1 French; NO = Natural Order

- Few differences with the Natural Order in possessive *'s* and third person *-s*
  - Surprising in the case of possessive *'s*, as Luk and Shirai (2009) found a robust effect of L1 on its acquisition

Differences Between the Observed Order and the Natural Order

Level	Articles	Past tense *-ed*	Plural *-s*	Possessive *'s*	Progressive *-ing*	Third person *-s*
CPE	NO > JKRT	JKT > NO	NO > JK		NO > GF	
CAE	NO > JKRT	JKT > NO	NO > K		NO > GF	
FCE	NO > JKRT	JKTGF > NO	NO > JKTF		NO > G	
PET	NO > JKRT GF > NO	JKTG > NO	NO > JKT		NO > G	JT > NO

J = L1 Japanese; K = L1 Korean; R = L1 Russian; T = L1 Turkish; G = L1 German; F = L1 French; NO = Natural Order

- Few differences with the Natural Order in possessive *'s* and third person *-s*
  - Surprising in the case of possessive *'s*, as Luk and Shirai (2009) found a robust effect of L1 on its acquisition

- The accuracy order of L1 Spanish learners of English does not deviate from the Natural Order.
  - Supports Luk and Shirai's (2009) hypothesis that the "natural" order is a mere reflection of the order of acquisition by Spanish learners of English

Summary and Comments

- Despite the numerous claims for the “natural” order, there does not seem to be a fixed accuracy order in L2 English grammatical morphemes.

- The pieces of evidence provided in this study should be more than sufficient to cast a strong doubt on the universality of the accuracy order.

- The use of a large-scale learner corpus allowed us to perform a study with a relatively large scope (i.e., six morphemes in seven L1 groups in five proficiency levels)

Grammatical Morphemes

- Grammatical morphemes are known to be notoriously difficult for L2 learners to acquire.

- Morpheme development is a slow, gradual process (e.g., Jia & Fuse, 2007).

- Not all inflected forms are as easily acquired.

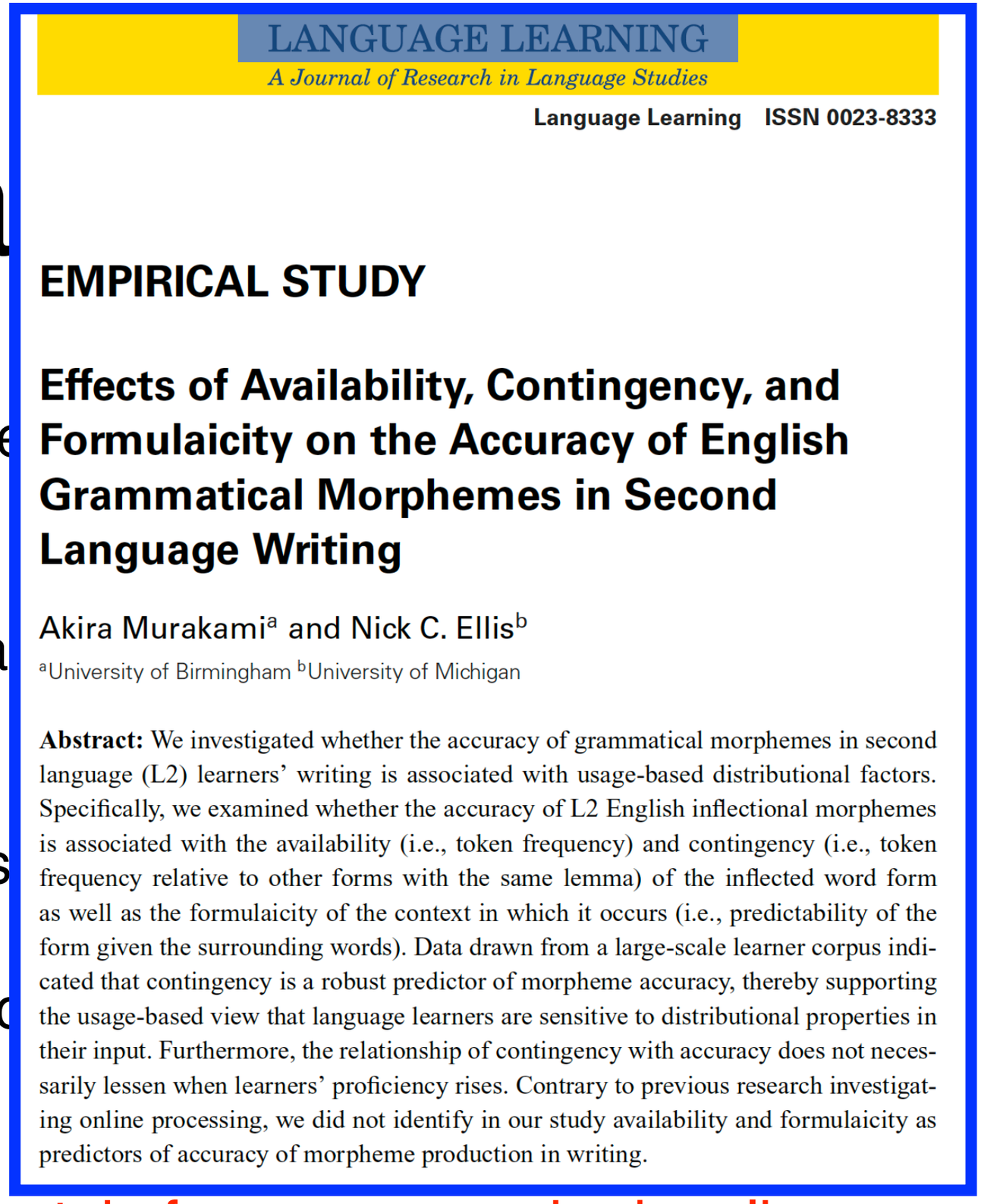
  - Some morphemes are acquired earlier than others (e.g., morpheme order studies)

  - Within each morpheme, certain forms are acquired earlier than others (e.g., Aspect Hypothesis)

Grammatica

- Grammatical morphemes are
- Not all inflected forms are as
  - Some morphemes are acc morpheme order studies)
  - Within each morpheme, certain forms are acquired earlier than others (e.g., Aspect Hypothesis)

Murakami, A., & Ellis, N. C. (2022). Effects of availability, contingency, and formulaicity on the accuracy of English grammatical morphemes in second language writing. *Language Learning*. http://doi.org/10.1111/lang.12500



LANGUAGE LEARNING
A Journal of Research in Language Studies
Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Effects of Availability, Contingency, and Formulaicity on the Accuracy of English Grammatical Morphemes in Second Language Writing

Akira Murakami[a] and Nick C. Ellis[b]
[a]University of Birmingham [b]University of Michigan

Abstract: We investigated whether the accuracy of grammatical morphemes in second language (L2) learners' writing is associated with usage-based distributional factors. Specifically, we examined whether the accuracy of L2 English inflectional morphemes is associated with the availability (i.e., token frequency) and contingency (i.e., token frequency relative to other forms with the same lemma) of the inflected word form as well as the formulaicity of the context in which it occurs (i.e., predictability of the form given the surrounding words). Data drawn from a large-scale learner corpus indicated that contingency is a robust predictor of morpheme accuracy, thereby supporting the usage-based view that language learners are sensitive to distributional properties in their input. Furthermore, the relationship of contingency with accuracy does not necessarily lessen when learners' proficiency rises. Contrary to previous research investigating online processing, we did not identify in our study availability and formulaicity as predictors of accuracy of morpheme production in writing.

Distributional Properties in Input

- From the perspective of usage-based theories, distributional properties in input affect the ease of processing and acquisition, as well as the accuracy of use.
- What kind of distributional characteristics influence the accurate vs erroneous (non-)use of inflectional morphemes?
  - Availability
  - Contingency
  - Formulaicity

Availability

- Availability: How often learners experience a particular form in their input

- One straightforward measure of availability is surface-form frequency.
  - If *wanted* is more frequent than *graduated*, then the accuracy of past tense *-ed* would be higher in *wanted* than in *graduated*.

- It has been widely demonstrated to impact the processing, acquisition, and use in both L1 (e.g., Ambridge, Kidd, Rowland, & Theakston, 2015) and L2 (e.g., Ellis, 2002; Ellis et al., 2016).

- In L1 acquisition, research has shown that high frequency surface forms are less prone to errors (Braine et al. 1990; Marchman, 1997; Aguado-Orea, 2004; Finley, 2018).

Contingency

- Contingency refers to a probabilistic association between cue and outcome.
- In the present study, the cue is the lemma and the outcome is the target inflected form.
- One measure of contingency is reliability.
- Reliability = P(inflected form | lemma)

Surface form	Frequency	Ratio (reliability)
arrive	22,619	0.23
arrived	56,489	0.56
arrives	10,106	0.10
arriving	10,876	0.11
Total	100,090	1.00

Contingency

- Contingency refers to a probabilistic association between cue and outcome.
- In the present study, the cue is the lemma and the outcome is the target inflected form.
- One measure of contingency is reliability.
- Reliability = P(inflected form | lemma)
- Contingency influences the associative learning between a cue and the outcome (Shanks, 1995).
- The reliability of the association has been shown to affect L1 and L2 processing, acquisition, and use (e.g., Hay, 2001; Matthews & Theakston, 2006, Sugaya & Shirai, 2009, Ellis et al., 2014; Tatsumi et al, 2018).

Formulaicity

- Formulaicity: The extent to which a given word sequence is a fixed, prefabricated, or memorised expression.

- More formulaic language is processed faster (e.g., Tremblay et al., 2011) and acquired earlier (e.g., Bannard & Matthews, 2008).

- Typically operationalised via either frequency or association strength.

- But using frequency alone can be problematic (Ellis, 2012).

- For instance, the expression *a lot of Americans* is frequent as a sequence of four words. But this is only because the sequence of the first three words (*a lot of*) is frequent and not because there is an inherent association between *a lot of* and *Americans*.

- What is important here is the association between *a lot of* and *Americans*.
  → ΔP

ΔP

- ΔP (Ellis, 2006; Gries, 2013) is a metric that quantifies the association strength between cue and outcome.

- One characteristic of ΔP is that it is unidirectional.

- In quantifying collocational strength of *of course*, for example, the ΔP value with *of* as a cue and *course* as its outcome is different from the ΔP value with *course* as a cue and *of* as its outcome (Gries, 2013).

- In the present study, the context surrounding the target word serves as the cue of the target inflected form (outcome).

Guo and Ellis (2021)

- Examined the effects of availability, reliability, and formulaicity through elicited imitation tasks.
- Participants: Chinese learners of English
- Target morphemes:
  - past tense *-ed*
  - progressive *-ing*
  - third person *-s*
  - plural *-s*
- Task: Participants listened to a sentence including the above morphemes and typed it on a computer.
- Availability, contingency, and formulaicity (operationalised as the frequency of four-word sequences) were calculated based on the Corpus of Contemporary American English (COCA; Davies, 2008-).

Guo and Ellis (2021)

- Results
  - Availability and contingency influenced the accuracy of morpheme provision in third person *-s* and plural *-s*.
  - The effect of contingency was particularly strong in lower proficiency learners.
  - Formulaicity was associated with accuracy as well.

- Due to the experimental nature of the work, its scope was limited.

- Also, while the elicited imitation task can experimentally control for a variety of factors (e.g., linguistic contexts), it potentially threatens ecological validity.

Elicited Imitation Task

- Implicit linguistic knowledge (Bowles, 2011; Erlam, 2006; Ellis et al., 2006; Spada et al., 2015) or automatised explicit knowledge (Suzuki & DeKeyser, 2015)

- The elicited imitation task involves perceiving words, linking them syntactically, and interpreting the sentence. Input factors such as availability, contingency, and formulaicity are relevant to each of these stages.

- More pronounced frequency effects in online processing tasks than in production tasks.

Free Writing

- Explicit processing (Elder & Ellis, 2009), which might mitigate the effect of input factors stemming from implicit processing

- Learners can pay attention to forms and/or edit writing.

Our Study

- Aim: Conceptually replicate and extend Guo and Ellis's (2021) experimental study by drawing data from a large-scale learner corpus of L2 writing.

- The analysis of a large-scale corpus allows us to target a larger number and range of words and learners, leading to a study with a larger scope and a more fine-grained picture of the effects of relevant factors.

- The study examines
  - whether the use of grammatical morphemes is more accurate in more available and more reliable words, as well as in more formulaic contexts
  - whether their effects interact with other factors such as learners' proficiency

Corpus

- EF-Cambridge Open Language Database (EFCAMDAT; Geertzen et al., 2014)
  - Available at https://philarion.mml.cam.ac.uk/

- The writings included in EFCAMDAT come from Englishtown, the online school formerly run by the company called EF Education First.

- The course in Englishtown consisted of 16 levels, each of which included 8 (or 6) units.

- At each unit, each student was asked to respond to a free writing task on a variety of topics (Alexopoulou et al., 2017; Michel et al., 2019).

Corpus

- EFCAMDAT includes 1.2 million writings by 175,000 learners.

- 783,000 (66%) of the writings come with teacher corrections, which were used as error annotation to calculate accuracy.

- Only the error-tagged writings were used in the analysis.

Target Morphemes

- Same as Guo and Ellis (2021): past tense *-ed*, progressive *-ing*, third person *-s*, plural *-s*

- We only examined omission errors and misformation errors (e.g., using third person *-s* instead of past tense *-ed*).

- We did not target overgeneralisation errors because the expected direction of the effect of the three distributional factors differs between omission errors and overgeneralisation errors.

    - For example, if high-frequency inflected forms tend to be used in those inflected forms (as opposed to other forms with the same lemma), then learners are less likely to make omission errors in high-frequency forms but more likely to make overgeneralisation errors.

Target Nationality Groups

- 11 nationality groups included in Shatz's (2020) cleaned subcorpus of EFCAMDAT.

- In past tense *-ed*,
  - Brazilian: 46.8%
  - Mexican: 8.2%
  - Chinese: 8.2%
  - German: 7.0%
  - Russian: 6.6%
  - Italian: 6.1%
  - French: 4.8%
  - Saudi Arabian: 4.2%
  - Taiwanese: 3.9%
  - Japanese: 2.9%
  - Turkish: 1.4%

- The other morphemes show similar distributional patterns.

- The distribution is skewed, with Brazilian learners occupying 43-48% of observations for each morpheme, followed by Chinese, Mexican, and German learners each contributing 6-13%.

Target Words

- In verbal morphemes, we only targeted words with five or more occurrences in each of the CEFR levels A1 through C1.
    - In plural *-s*, we employed different criteria to reduce its otherwise unmanageably large data size.
    - Specifically, we only targeted the learners who contributed 20 or more writings with at least one error or accurate use of plural *-s*, and among them, we included the word forms with a minimum of 20 occurrences at each of the CEFR levels A1 through C1.

Morpheme	Learner	Writing	Word	Obligatory Context
Past tense *-ed*	30,955	83,001	87	151,979
Progressive *-ing*	39,744	88,010	94	123,869
Third person *-s*	39,961	78,667	124	113,298
Plural *-s*	2,633	59,759	89	136,601

Availability and Reliability

- Availability, reliability, and formulaicity were calculated based on COCA.

- Availability: log-transformed surface-form frequency of the inflected form used as a verb (in verbal morphemes) or a noun (in plural *-s*)
    - High availability: *used, getting, says, years*
    - Low availability: *requested, raining, prefers, supermarkets*

- Reliability: surface-form frequency divided by the frequency of the corresponding lemma
    - High reliability: *decided*, *trying*, *depends*, *inhabitants*
    - Low reliability: *liked*, *thinking*, *thinks*, *nights*

Calculation of Formulaicity

- ΔP = P(outcome | cue) - P(outcome | ¬cue)
- For instance, in the trigram *what happened yesterday*, we consider the contingency table of the frequency counts of *what* ____ *yesterday* and ____ *happened* ____.

Cue	Target Inflected Form (*happened*)	Non-Target Word
Present (*what* ____ *yesterday*)	318 (*what happened yesterday*)	62
Absent	162,971	763,861,708

$$\Delta P = \frac{318}{318 + 62} - \frac{162971}{162971 + 763861708} = 0.84$$

- ΔP takes the value between -1 and +1, and the larger value indicates a stronger association between cue and outcome.
- ΔP of 0.84 suggests that we can predict *happened* from *what* ____ *yesterday* fairly well.

Target N-grams

N-gram	L4	L3	L2	L1	Target word	R1	R2	R3	R4	ΔP	Standardised log-transformed ΔP
3-gram											
			with	all	kinds					0.01	
				all	kinds	of				0.19	1.51
					kinds	of	people			0.01	
4-gram											
		deal	with	all	kinds					0.03	
			with	all	kinds	of				0.33	
				all	kinds	of	people			0.54	1.61
					kinds	of	people	and		0.02	
5-gram											
	can	deal	with	all	kinds					0.05	
		deal	with	all	kinds	of				0.38	
			with	all	kinds	of	people			0.43	
				all	kinds	of	people	and		0.49	1.54
					kinds	of	people	and	i	0.02	

Expressions with High and Low Formulaicity

- High formulaicity: since i graduated from college, and time is running out, practice makes perfect, ladies and gentlemen

- Low formulaicity: wanted a lot of, going is not, says do not, in the years of

Statistical Analysis

- We built a Bayesian mixed-effects binary logistic regression model for each morpheme.
- Dependent variable: accurate use versus error of each (non-)occurrence of each morpheme
- Fixed-effects
  - the mean Englishtown level of the learner representing the learner's L2 proficiency
  - the learner's writing number representing longitudinal development in an ordinal manner (e.g., 1 = first writing, 2 = second writing)
  - L1 type indicating whether an equivalent feature to the target morpheme was obligatory in the language predominantly spoken in the country or region of the learner's nationality
  - Log-transformed surface-form frequency (availability)
  - Reliability (contingency)
  - ΔP (formulaicity)
  - Two-way interactions of the above variables
- All the quantitative variables have been standardised to a mean of 0 and an SD of 1.
- Weakly informative priors were used.

Random Effects

- Learner
  - writing number
  - frequency
  - reliability
  - ΔP

- Nationality
  - mean Englishtown level
  - writing number
  - frequency
  - reliability
  - ΔP

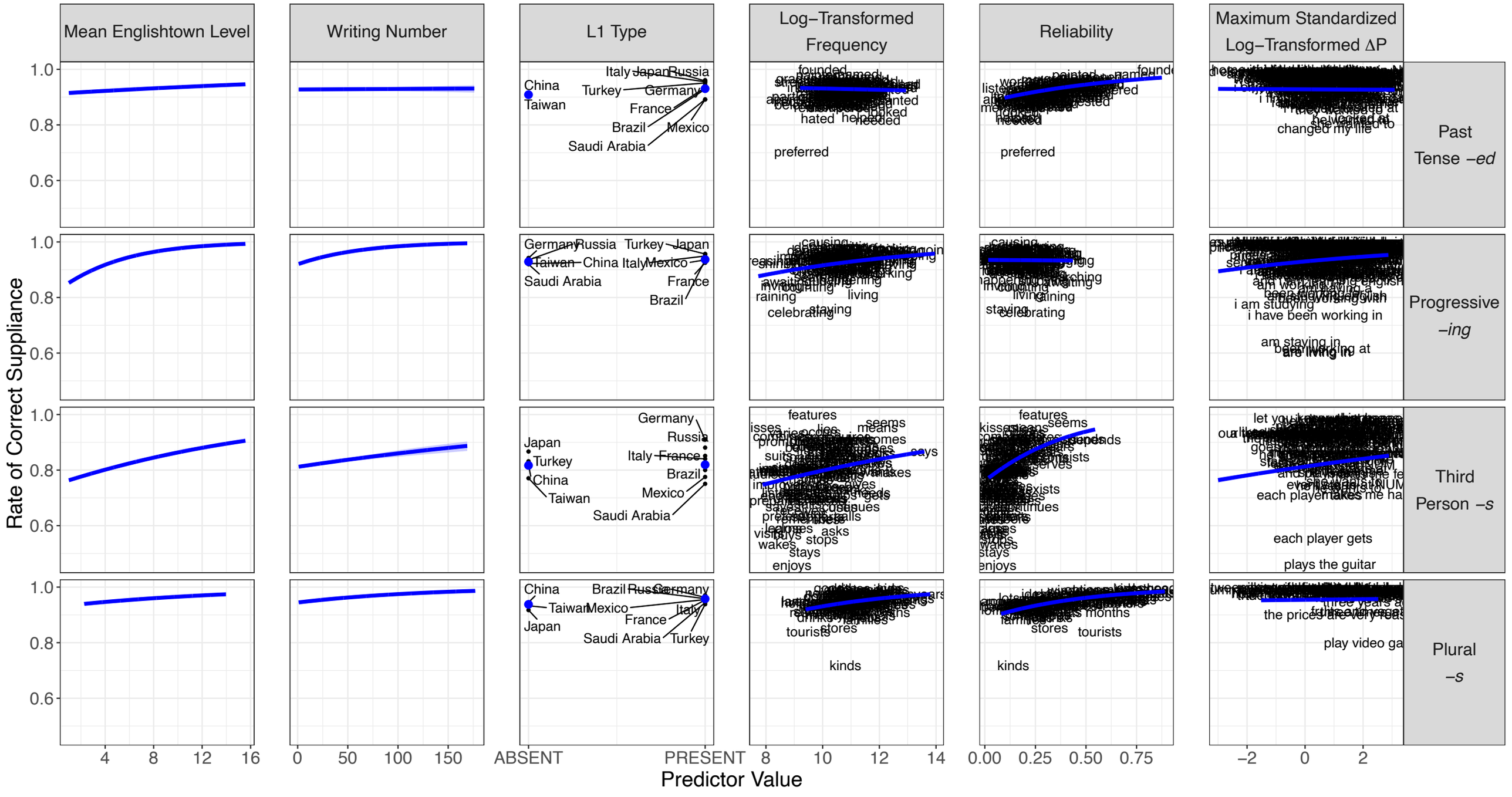
- Inflected form
  - mean Englishtown level
  - writing number
  - ΔP

- Topic of writing
  - L1 type
  - mean Englishtown level
  - writing number
  - frequency
  - reliability
  - ΔP

- Interaction between inflected forms and topics of writing

- This is close to the maximal model (Barr et al., 2013), except that interaction terms were not included as random slopes.

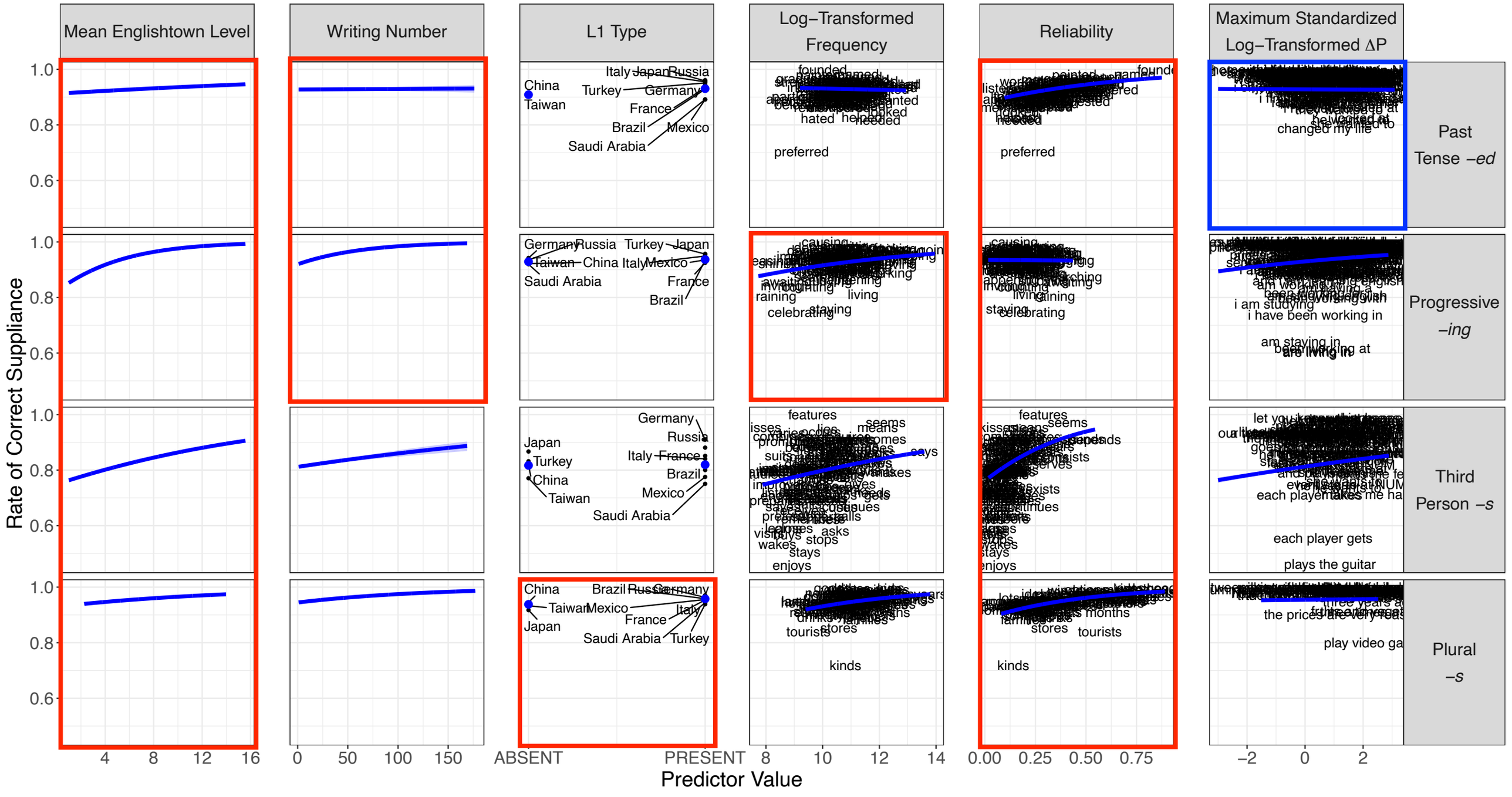
Results


Mean Englishtown Level
Writing Number
L1 Type
Log−Transformed Frequency
Reliability
Maximum Standardized Log−Transformed ΔP
Past Tense –ed
Progressive –ing
Third Person –s
Plural –s
Rate of Correct Suppliance
1.0
0.8
0.6
China
Taiwan
Turkey
Italy
Japan
Russia
Germany
France
Brazil
Mexico
Saudi Arabia
preferred
hated
needed
changed my life
i am studying
i have been working in
am staying in
are living in
each player gets
plays the guitar
kinds
tourists
stores
months
4
8
12
16
0
50
100
150
ABSENT
PRESENT
8
10
12
14
0.00
0.25
0.50
0.75
−2
0
2
Predictor Value

= the lower bound of 95% CI > 0

= the upper bound of 95% CI < 0

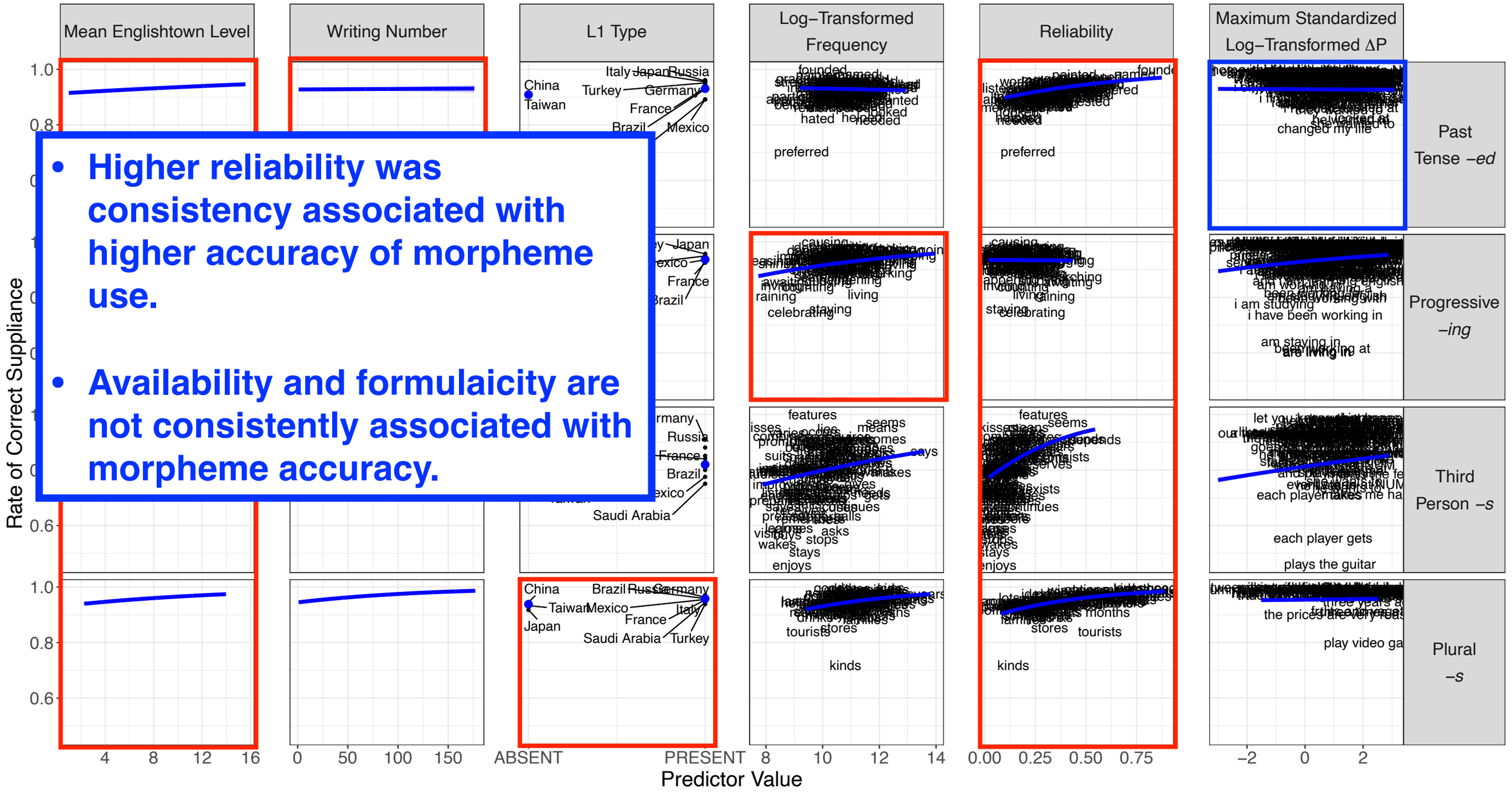
Results



= the lower bound of 95% CI > 0

= the upper bound of 95% CI < 0

Results



- **Higher reliability was consistency associated with higher accuracy of morpheme use.**
- **Availability and formulaicity are not consistently associated with morpheme accuracy.**

Availability

- The lack of noticeable effects of availability and formulaicity is interesting, given their pervasive impact documented in the usage-based literature and more specifically in Guo and Ellis (2021).

- We speculate this is potentially due to (i) the nature of writing, which does not require online processing, (ii) target words, and (iii) the operationsalisation of formulaicity.

- In free writing, learners respond to a prompt with whatever language forms associated with the ideas they come up with.

- Availability means that higher-proficiency learners are able to use a wider range of items than lower-proficiency learners.

- But what we examined is not availability in this sense. What we examined was whether learners correctly inflected the lemmas that came to their mind in that context.

- Also, all of the target words in our study were fairly high in their frequency, and their frequency effects might have approached the ceiling.

Formulaicity

- As mentioned earlier, elicited imitation tasks draw on implicit (or automatised explicit) knowledge, while free writing draws on explicit knowledge.

- Unlike elicited imitation tasks requiring online processing, conscious writing strategies may have overridden the effects of distributional factors offered up from the implicit (or automatised) system.

- Editing can, for example, disrupt what is first offered by the implicit system and moderate reliance on formulaic knowledge.

- There are also issues with the operationalisation of the construct.

    - We did not look at non-adjacent (e.g., verb and its direct object) or abstract patterns (e.g., part-of-speech sequences).

    - There are many other ways to calculate association strength (e.g., MI, n-gram frequency), and ΔP may or may not be the best measure.

Contingency

- The results on contingency were robust.

- The positive association between reliability and the accuracy in morpheme use suggests that the lemma functions as a cue of its inflected forms, and that L2 learners make use of the contingency in processing them.

- Reliability of the link between cue and outcome is important in all kinds of associative learning, and contingency learning plays a central role in the Competition Model (MacWhinney, 1987; MacWhinney & Bates, 1989).

    - In learning the association between cue and outcome, learners start with available cues and then increasingly rely on reliable cues.

Summary and Comments

- Systematic patterns observed in this study indicate that L2 learners are sensitive to certain distributional properties in their input.

- In particular, reliability is strongly associated with morpheme accuracy.

- Methodologically, the study exemplifies multifactorial research by modeling the accuracy of grammatical morphemes as a function of a number of predictors including proficiency, L1, availability, contingency, and formulaicity.

- Distributional patterns are only available through large-scale corpora (e.g., Tracy-Ventura & Medina, 2018).

- It also complements a prior experimental study (Guo & Ellis, 2018) through methodological triangulation.

Wrap Up

- In this talk, I hope I have shown that there are quite a lot of things you can do with large-scale learner corpora.

- Combined with multifactorial analyses, studies based on large-scale learner corpora showed
    - that the L2 acquisition order of English grammatical morphemes differs across the learners' L1 backgrounds and

    - that contingency between inflected forms and their lemmas is robustly associated with morpheme accuracy, suggesting L2 learners are sensitive to the within-lemma distribution of inflected forms in their input.

- Large-scale learner corpora make it possible to perform a multifactorial study that yields more fine-grained findings with a larger scope.

References (1/5)

Alexopoulou, T., Michel, M., Murakami, A., & Meurers, D. (2017). Task effects on linguistic complexity and accuracy: A large-scale learner corpus analysis employing Natural Language Processing techniques. *Language Learning, 67*(S1), 180–208

Aguado-Orea, J. J. (2004). *The acquisition of morpho-syntax in Spanish: Implications for current theories of development*. Unpublished PhD thesis. University of Nottingham, Nottingham, UK.

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. Journal of *Child Language, 42*(2), 239–273.

Andersen, R. W., & Shirai, Y. (1994). Discourse motivations for some cognitive acquisition principles. *Studies in Second Language Acquisition, 16*(2), 133-156

Bannard, C. & Matthews, D. (2008). Stored word sequences in language learning: the effect of familiarity on children's repetition of four-word combinations. *Psychological Science, 19*(3), 241-248

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Bowles, M. (2011). Measuring implicit and explicit linguistic knowledge: what can heritage language learners contribute? *Studies in Second Language Acquisition, 33*(2), 247-271

Braine, M. D. S., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L., & Fisch, S. M. (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Language, 29*(5), 591-610

Dulay, H. C., & Burt, M. K. (1973). Should we teach children syntax? *Language Learning, 23*(2), 245–258.

Dulay, H. C., & Burt, M. K. (1974). Natural sequences in child second language acquisition. *Language Learning, 24*(1), 37–53.

Elder, C., & Ellis, R. (2009). Implicit and explicit knowledge of an L2 and language proficiency. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders. *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 167-193). Bristol, UK: Multilingual Matters.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*(2), 143–188.

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics, 27*(1), 1–24

References (2/5)

Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics, 32*, 17–44.

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2014). The processing of verb-argument constructions is sensitive to form, function, frequency, contingency and prototypicality. *Cognitive Linguistics, 25*(1), 55–98.

Ellis, N. C., Römer, U., & O'Donnell, M. B. (2016). *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of construction grammar*. Language Learning Monograph Series. Chichester, West Sussex [England]; Malden, MA: Wiley.

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly, 42*(3), 375–396.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition, 28*(2), 339-368

Erlam, R. (2006). elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics, 27*, 464-491

Eskildsen, S. W. (2012). L2 Negation constructions at work. *Language Learning, 62*(2), 335–372.

Finley, S. (2018). Cognitive and linguistic biases in morphology learning. *Wiley Interdisciplinary Reviews: Cognitive Science, 9*(5), e1467

Feldweg, H. (1991). *The European Science Foundation Second Language Database*. Nijmegen.

Fuchs, R., & Werner, V. (Eds.) (2018). Tense and aspect in second language acquisition and learner corpus research [Special Issue]. *International Journal of Learner Corpus Research, 4*(2).

Geertzen, J., Alexopoulou, T., & Korhonen, A. (2014). Automatic linguistic annotation of large scale L2 databases: The EF-Cambridge Open Language Database (EFCAMDAT). In R. T. Millar, K. I. Martin, C. M. Eddington, A. Henery, N. M. Miguel, A. Tseng, … D. Walter (Eds.), *Selected proceedings of the 2012 Second Language Research Forum. Building bridges between disciplines* (pp. 240–254). Cascadilla Proceedings Project.

Gries, S. Th. (2013). 50-something years of work on collocations: What is or should be next …. *International Journal of Corpus Linguistics, 18*(1). 137-165

Gries, S. Th., & Wulff, S. (2013). The genitive alternation in Chinese and German ESL learners: Towards multifactorial notion of context in learner corpus research. *International Journal of Corpus Linguistics, 18*(3), 327–356.

References (3/5)

Guo, R. & Ellis, N. C. (2021). Language usage and second language morphosyntax: Effects of availability, reliability, and formulaicity. *Frontiers in Psychology, 12*. 582259

Hasko, V. (Ed.) (2013). Capturing L2 development through learner corpus analysis [Special Issue]. *Modern Language Journal, 97*(S1).

Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning, 24*(2), 287–297.

Hay, J. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics, 39*(6), 1041–1070

Hiki, M. (1991). *A study of learners' judgments of noun countability*. Indiana University.

Jarvis, S. (2000). Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. *Language Learning, 50*(2), 245–309.

Jarvis, S. (2010). Comparison-based and detection-based approaches to transfer research. In L. Roberts, M. Ó. Howard, M. Laoire, & D. Singleton (Eds.), *EUROSLA Yearbook 10* (pp. 169–192). Amsterdam:Benjamins

Jarvis, S., & Pavlenko, A. (2007). *Crosslinguistic influence in language and cognition*. New York: Routledge.

Jia, G., & Fuse, K. (2007). Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents: Age-related differences. *Journal of Speech, Language, and Hearing Research, 50*(5), 1280–1299.

Klein, W., & Perdue, C. (1992). U*tterance structure: Developing grammars again*. Amsterdam: John Benjamins.

Krashen, S. D. (1977). Some issues relating to the Monitor Model. In H. D. Brown, C. A. Yorio, & R. H. Crymes (Eds.), *On TESOL '77: Teaching and learning English as a second language: Trends in research and practice* (pp. 144–158). Washington D. C.: TESOL.

Le Bruyn, B., & Paquot, M. (2021). *Learner corpus research meets second language acquisition*. Cambridge: Cambridge University Press

Lu, X. (2022). *Corpus linguistics and second language acquisition: Perspectives, issues, and findings*. Routledge

References (4/5)

Luk, Z. P., & Shirai, Y. (2009). Is the acquisition order of grammatical morphemes impervious to L1 knowledge? Evidence from the acquisition of plural -s, articles, and possessive 's. *Language Learning, 59*(4), 721–754.

MacWhinney, B. (1987). Applying the Competition Model to bilingualism. *Applied Psycholinguistics, 8*(4): 315–27

MacWhinney, B., & Bates, E.(Eds.) (1989). *The crosslinguistic study of sentence processing*. Cambridge; New York: Cambridge University Press.

Marchman, V. (1997). Children's productivity in the English past tense: The role of frequency, phonology, and neighborhood structure. *Cognitive Science, 21*(3), 283–303.

Matthews, D. & Theakston, A. L. (2006). Errors of omission in English-speaking children's production of plurals and the past tense: the effects of frequency, phonology, and competition. *Cognitive Science, 30*(6), 1027-1052

Michel, M., Murakami, A., Alexopoulou, T., & Meurers, D. (2019). Effects of task type on morphosyntactic complexity across proficiency: Evidence from a large learner corpus of A1 to C2 writings. *Instructed Second Language Acquisition, 3*(2), 124–152

Murakami, A. (2016). Modeling systematicity and individuality in nonlinear second language development: The case of English grammatical morphemes. *Language Learning, 66*(4), 834–871.

Murakami, A., & Alexopoulou, T. (2016). L1 influence on the acquisition order of English grammatical morphemes: A learner corpus study. *Studies in Second Language Acquisition, 38*(3), 365–401.

Murakami, A., & Ellis, N. C. (2022). Effects of availability, contingency, and formulaicity on the accuracy of English grammatical morphemes in second language writing. *Language Learning*.

Ortega, L. (2009). *Understanding second language acquisition*. London: Hodder Education.

Perdue, C. (1993). Adult language acquisition. Vol 1: Field methods. Cambridge: Cambridge University Press.

Pica, T. (1983). Methods of morpheme quantification: Their effect on the interpretation of second language data. *Studies in Second Language Acquisition, 6*(1), 69–78.

Paquot, M. (2019). The phraseological dimension in interlanguage complexity research. *Second Language Research, 35*(1), 121-145

References (5/5)

Paquot, M., & Plonsky, L. (2017). Quantitative research methods and study quality in learner corpus research. *International Journal of Learner Corpus Research, 3*(1), 61–94.

Rautionaho, P., & Deshors, S. C. (2018). Progressive or not progressive? Modeling the constructional choices of EFL and ESL writers. *International Journal of Learner Corpus Research, 4*(2), 225–252.

Rebuschat, P., Meurers, D., & McEnery, T. (Eds.) (2017). Experimental, Computational, and Corpus-Based Approaches to Language Learning: Evidence and Interpretation [Special Issue]. *Language Learning, 67*(S1).

Shanks, D. R. (1995). *The psychology of associative learning*. Cambridge: Cambridge University Press.

Shatz, I. (2020). Refining and modifying the EFCAMDAT: Lessons from creating a new corpus from an existing large-scale English learner language database. *International Journal of Learner Corpus Research, 6*, 220–236.

Slobin, D. I. (1996). From "thought to language" to "thinking for speaking". In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge: Cambridge University Press.

Spada, N., Shiu, J.L.-J. and Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language Learning, 65*, 723-751

Sugaya, N., & Shirai, Y. (2009). Can L2 learners productively use Japanese tense-aspect markers? A usage-based approach. In *Typological Studies in Language*, R. Corrigan, E. A. Moravcsik, H. Ouali, and K. Wheatley, (Eds.) (pp. 383-423). Amsterdam: John Benjamins Publishing Company

Suzuki, Y., & DeKeyser, R. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning, 65*, 860-895.

Tatsumi, T., Ambridge, B., & Pine, J. M. (2018). Disentangling effects of input frequency and morphophonological complexity on children's acquisition of verb inflection: An elicited production study of Japanese. *Cognitive Science, 42*(May), 555–77

Tracy-Ventura, N., & Medina, J. A. C. (2018). Can native-speaker corpora help explain L2 acquisition of tense and aspect? A study of the "input." *International Journal of Learner Corpus Research, 4*(2), 277–300.

Tracy-Ventura, N., & Paquot, M. (Eds.). (2021). *The Routledge handbook of second language acquisition and corpora*. Oxon: Routledge

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning, 61*(2), 569–613.

Yoon, K. K. (1993). Challenging prototype descriptions: Perception of noun countability and indefinite vs. zero article use. *International Review of Applied Linguistics, 31*(4), 269–289.