

This is a repository copy of *L1 explicit instruction can improve L2 online and offline performance*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/99267/>

Version: Accepted Version

---

**Article:**

McManus, Kevin [orcid.org/0000-0002-7855-6733](https://orcid.org/0000-0002-7855-6733) and Marsden, Emma [orcid.org/0000-0003-4086-5765](https://orcid.org/0000-0003-4086-5765) (2017) L1 explicit instruction can improve L2 online and offline performance. *Studies in Second Language Acquisition*. pp. 459-492. ISSN 1470-1545

<https://doi.org/10.1017/S027226311600022X>

---

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

L1 explicit instruction can improve L2 online and offline performance. An exploratory study

Kevin McManus  
The Pennsylvania State University, USA

Emma Marsden  
University of York, UK

Address for correspondence: Kevin McManus, Department of Applied Linguistics, The  
Pennsylvania State University, University Park, PA 16802, USA. Email: [kmcmanus@psu.edu](mailto:kmcmanus@psu.edu)

#### Acknowledgements

This research was supported by a British Academy Postdoctoral Fellowship (PF130001) awarded to Kevin McManus, with Emma Marsden as Mentor. Many thanks to the learners who participated in this study, and to Laurence Richard for help with recruitment. We would also like to thank the Editors and two anonymous *SSLA* reviewers for their valuable comments, and Oriane Boulay, Margaret Coleman, and Sylvie Goodlow for their research assistance.

Revised manuscript submitted April 30, 2016

Running head: L1 explicit instruction and L2 online and offline performance

L1 explicit instruction can improve L2 online and offline performance. An exploratory study

### Abstract

This study investigated the effectiveness of providing L1 explicit information with practice for making more accurate and faster interpretations of L2 French *Imparfait*. Two treatments were investigated: (i) ‘L2-only’, providing explicit information (EI) about the L2 with L2 interpretation practice and (ii) ‘L2+L1’, providing the exact same L2-only treatment and including EI about the L1 (English) with practice interpreting L1 features that are equivalent to the *Imparfait*. 50 L2 French learners were randomly assigned to either L2-only, L2+L1, or a Control group. Online (self-paced reading) and offline (context-sentence matching) measures from Pretest, Posttest and Delayed Posttests showed that providing additional L1 EI and practice improved not only offline L2 accuracy, but also the speed of online L2 processing. To our knowledge, this makes an original and significant contribution about the nature of EI with practice, the role of L1 (Tolentino & Tokowicz, 2014), and extends a recent line of research examining EI effects in online sentence processing (Andringa & Curcic, 2015).

Key words: instructed second language acquisition, explicit knowledge, L1 influence, online processing

## Explicit information and practice in L2 learning

The usefulness of providing L2 learners with explicit information (EI) about a target feature and subsequent practice in processing the input is not fully understood. As noted by Henry, Culman and VanPatten (2009, p.573) “not all EI is the same, not all structures are the same, and the interaction of EI, structure, and processing problem may yield different results in different studies”. The current study investigates the effects of EI with practice in the L2 and also, in light of research documenting persistent difficulty when the L1 and L2 express the same meaning differently (Izquierdo & Collins, 2008; McManus, 2013, 2015; Roberts & Liszka, 2013), whether additional EI about the L1 with L1 practice can help a specific processing problem – interpreting the habitual *versus* ongoing meanings of L2 French *Imparfait* for L1 English learners. We tested whether making this conceptual distinction explicit, via EI and meaning-based practice in both L1 and L2, would aid form-meaning mapping. First, we briefly discuss research into L2 EI and practice, before justifying the investigation of a role for *L1* EI and practice, and then move on to discuss why EI and practice (in L2 and L1) may have an effect on *online* processing.

### L2 EI and Practice

EI about the L2 is useful for learning, according to information processing and skill acquisition theories, because some declarative information can become proceduralised via practice and automatized, resulting in automatised declarative knowledge and/or knowledge that appears indistinguishable from implicit knowledge (DeKeyser, 2015). 'Weaker' accounts suggest that learners can use EI to segment or parse the input (Terrell, 1991), notice features (Schmidt, 1990), understand a rule and help production (Leow, 2015), and arrive at correct interpretations with fewer practice items (Henry et al., 2009). The effectiveness of EI is likely

to depend on several factors, including its precise nature – the type of information conveyed and the feature in focus. There is a considerable body of research into feature difficulty and amenability to different kinds of instruction (e.g. DeKeyser, 2012). VanPatten & Rothman (2015) suggest that features most likely to benefit might be those that must be learned directly from the input, such as representations of inflectional verb morphology, as EI may make them better noticed in the input. In addition, the selection of *which features* could benefit from *which kinds* of EI and practice could partially be informed by the nature of L1-L2 differences.

### **L1 EI and practice**

Long-term difficulties have been documented for learning L2 features that share some similarity across the L1-L2, and yet have different form-meaning mappings (Murakami & Alexopoulou, 2016; Spada, Lightbown & White, 2005). Yet, as Ellis and Shintani (2014) noted, "there is almost no research that has investigated the actual effects of the classroom use of the L1 on L2 learning" (p.247). A few classroom studies have shown that raising learners' awareness about L1-L2 differences benefits learning of L2 lexis (Laufer & Girsai, 2008; White & Horst, 2012) and grammar (Horst, White & Bell, 2010; Kupferberg, 1999; Spada, Lightbown & White, 2005), as measured during the actual learning events or immediately afterwards via offline vocabulary and writing tests. Evidence also suggests that providing EI about L1-L2 differences correlated positively with learners' performance on untimed tests that allowed access to that awareness, i.e. grammaticality judgments and sentence construction tasks (Ammar, Lightbown & Spada, 2010). One of the few intervention studies to address the amenability of L1-L2 differences to instruction was conducted by Tolentino and Tokowicz (2014). Grammaticality judgement tests showed that providing EI, input flood, and repetition practice of features that exist in both languages but are realized

differently did not in fact benefit from instruction (although benefits were found for features unique to the L2). However, as that study provided no EI about the L1 or any meaning-based practice, the effectiveness of instruction for cross-linguistically different features remains to be explored further.

To our knowledge, the current study addresses three significant gaps in this agenda. First, learning has not yet been documented using delayed posttests. Second, research has not yet investigated the benefits of *systematic L1 practice* as an intentional component of instruction. Third, research has not yet examined the effects of L1 EI and practice on *online* sentence processing, discussed next.

### **Online effects of EI and practice**

Studies showing benefits for EI with (or without) practice have almost exclusively used offline tests that allow access to explicit knowledge, with very little use of online measures (as noted by DeKeyser & Prieto Botana, 2015; VanPatten & Rothman, 2015). Online measures can provide “fine-grained information about moment-by-moment sentence comprehension [... to] examine what happens at precise points in a sentence” (Keating & Jegerski, 2015, p.2). For example, longer reaction times relative to comparison items may indicate a processing cost brought about by ungrammaticality, ambiguity or complexity (Roberts, 2016). Some theorists argue for an even stronger, causal relation between online processing and learning (e.g. O'Grady, 2005, 2015; see Phillips & Ehrenhofer, 2015). That is, online processing may be a mechanism by which learning is driven and constrained, and processing difficulty (e.g. complexity or cost) a key factor in learning. To this end, VanPatten and Rothman (2015) recommended:

"moving away from knowledge-testing more generally and more into the interface

between knowledge and processing via techniques such as [...] self-paced listening/reading [...] Currently, these are used largely to understand the processing outcomes of acquisition. We think they can be used to study acquisition-as-processing itself."

(VanPatten and Rothman, 2015, p. 113)

To our knowledge, the relationship between online L2 processing and explicit instruction has been investigated in two published studies to date<sup>1</sup>. First, Andringa and Curcic (2015) provided half their participants with brief EI explaining that a preposition predicts the animacy of the upcoming direct object in a novel language. Learners in both conditions ( $\pm$ EI) were then exposed to 104 sentences, 52 of which provided exposure to the direct object marker rule. Although the +EI group performed better offline in a grammaticality judgement test, there was no evidence that EI developed into knowledge that was beneficial online, as measured by predictive eye movements. Second, again in an artificial language study, Marsden, Williams and Liu (2013) investigated the effects of task-essential practice with yes/no feedback on interpreting inflections for tense and number (Experiment 3). Again, findings were that offline measures (accuracy of lexical decisions) demonstrated learning, but online measures (reaction times in a cross-modal priming task) provided no evidence that training the learners to orient their attention to the meaning of the inflection had affected online processing of cross-modal representations.

Related to the knowledge gap about the effects of L2 EI and practice on online performance is whether L1 EI and practice can affect L2 online processing. This is of particular relevance for features with cross-linguistic differences in processing routines. Of course, some theories foreground a role for the L1 in critical aspects of input processing, such as: attention allocation being entrenched by the L1 (Ellis, 2006); processing routines being influenced by the L1 when a feature is not unique to the L2 (MacWhinney, 2005); or L2 processing routines

being difficult to learn in cases where adopting related L1 routines would require fewer processing resources (O'Grady, 2005, 2015). There is also growing evidence of L1 co-activation/influence during online L2 sentence processing (e.g. Tolentino & Tokowicz, 2011). Tokowicz and Warren's (2010) self-paced reading study with beginner learners reported slower reading times at morphosyntactic violations for L2 Spanish features that were cross-linguistically similar (verb aspect licensing), but not for those that were entirely unique to the L2 (determiner–noun gender agreement). Similarly, Roberts and Liszka (2013) with advanced learners found slower reading times at morphosyntactic (aspectual) violations in L2 English when L1 and L2 both grammaticalized aspect (L1 French), but not when the means of expressing aspect was unique to the L2 (L1 German). L1-L2 morphosyntactic co-activation has also been documented amongst bilinguals, for example, during comprehension (Sanoudaki & Thierry, 2014), production (Runnqvist, Gollan, Costa, & Ferreira, 2013), and cross-linguistic priming (Hartsuiker & Pickering, 2008). However, there is surprisingly little evidence about the role of the L1 in online processing amongst *learners*, rather than bilinguals/near natives. As VanPatten (2015) noted, "the question is open as to whether and to what degree there is L1 influence in basic Input Processing, and whether that influence is an actual processing procedure or lexical influence" (p.120). Critically, as noted above, there is very little, if any, research on the influence of L1 *explicit information and practice* on online processing that could inform us about the development of sentence processing in instructed contexts.

### **Rationale for the current study**

In sum, the above lines of research inform our understanding about the roles of EI with practice in L2 offline knowledge and about the existence of L1 effects in L2 online processing. But there is no research, to our knowledge, that investigates a) the potential



benefits of EI with practice (L1 or L2) on *online* processing for learning a natural language over time and b) the benefits of EI about the *L1* with practice in processing it (on offline or online performance).

It remains an empirical question as to whether providing EI about and practice in the different processing routines in learners' L1 can benefit L2 learning. Benefits may be found if making a processing routine explicit could create some declarative knowledge that might serve processing and learning in a variety of ways. For example, clarifying and rehearsing non-transparent, conceptual distinctions in the L1 (e.g. polyfunctionality of the English '-ed' verbal morpheme, McManus, 2015) may facilitate accurate mapping of those concepts to L2 forms, compatible with views that assume a role for EI and explicit rehearsal in language development (DeKeyser, 2015). Thus, L1 and L2 EI with meaning-based practice may lead to new or more efficient L2 processing routines. Initially represented declaratively, these routines may gradually become proceduralised and automatized through practice (DeKeyser, 2015). Combined L1 and L2 EI and practice may also provide data about morphosyntactic distributions of features in the input that could, according to some theories (e.g. Ellis, 2006, 2008; see Andringa & Curic 2015), interface with the developing language system and promote L2 form-meaning re-mapping.

### **A potential testing ground for instruction in a cross-linguistic processing problem: The French *Imparfait***

The *Imparfait* (IMP) is well documented to be late acquired, even after considerable naturalistic exposure (Bartning & Schlyter, 2004), and its habitual meaning in particular has been shown to be influenced by L1 background (Howard, 2005; Izquierdo & Collins, 2008; McManus, 2013, 2015). Although French and English both express past habituality and

ongoingness with verbal morphology, these are mapped differently. For English-speaking learners of French, this is illustrated in 1-2.

(1) Past ongoing and habitual meanings can be expressed by one morpheme in French (a),

but not in English (b & c):

a) *Il jouait*<sup>IMP\_ongoing\_i/habit\_j</sup> *au foot quand j'ai*<sup>-i</sup> *appelé/ quand nous étions*<sup>-j</sup> *petits*

“He play<sup>IMP\_ongoing\_i/habit\_j</sup> football when I <sup>-PERF\_i</sup> called / when we were<sup>-IMP\_j</sup> little”

b) He was<sup>ongoing\_i</sup> playing football

c) He played<sup>-habit\_j</sup> football

(2) Past perfective and habitual meaning can be expressed by one morpheme in English

(a), but not in French (b & c):

a) He played<sup>PERF\_i/IMP\_habit\_j</sup> football once last year<sup>-i</sup> / every Saturday<sup>-j</sup>

b) *Il jouait*<sup>-habit\_j</sup> *au foot quand nous étions*<sup>-j</sup> *petits*

“He play<sup>IMP\_habit\_j</sup> football when we were<sup>-IMP\_j</sup> small

c) *Il a*<sup>PERF\_i</sup> *joué au foot*

“He <sup>PERF</sup>play football”

These sentences illustrate that one French inflectional verb morpheme alone does not disambiguate habitual from ongoing meaning in the past. Nor do lexical phrases *reliably* co-occur with IMP to distinguish between ongoing and habitual ('la semaine dernière' ('last week')) co-occurs with morphology other than IMP). However, *morphosyntactic* information in the discourse context *is* a reliable cue to meaning (de Swart, 1998; Smith, 1997). That is, to resolve the aspectual ambiguity inherent in IMP, the discourse context either: provides an 'interruption' to the event via the past perfective, *Passé Composé* (PC), thus coercing an ongoing meaning of IMP (sentence 1a<sup>-i</sup>); or it provides concurrent information with another IMP, coercing a habitual meaning (sentence 1a<sup>-j</sup> and 2b<sup>-j</sup>) (Comrie, 1976). The IMP can be

before or after its disambiguating verb, and not necessarily in the same sentence or speech turn.

English, on the other hand, does not require inter-clausal morphosyntax to disambiguate past habituality from ongoingness. Instead, this is done within a clause via verb morphology and/or lexical means, leaving no ambiguity in need of resolution e.g. I was walking = past ongoingness; I walked/used to walk/would walk = habituality.

If learners tend to rely on L1 English processing routines they would not reliably use morphological information in the discourse to disambiguate past habituality from ongoingness (which arguably demands more processing resources than immediate disambiguation within the verb phrase (O'Grady, 2005, 2015)). This could result in non-optimum (less accurate and slower) interpretations of IMP. It seems possible that EI with practice that renders explicit the mapping procedures required in L1 could facilitate the re-mapping of procedures for interpreting the L2.

### **Task-essential form-meaning mapping practice for aspectual distinctions**

A large body of research has demonstrated the learning benefits of focusing learners' attention on making form-meaning mappings from the input (Loschky & Bley-Vroman, 1993; VanPatten, 1996, 2002). This has included presenting stimuli without temporal adverbs, thus forcing attention on the temporal meaning of verb inflections (e.g. Benati, 2005; Marsden, 2006; Marsden & Chen, 2011; Sagarra & Ellis, 2013), without overt subjects, thus forcing attention on person and number meanings (Marsden, 2006; Marsden et al. 2013), or without lexical phrases for doubt and certainty, forcing attention on subjunctive versus indicative inflections (Fernández, 2008). Task-essential form-meaning input mapping practice has been found to lead to more learning than input activities with equal numbers of

target features that focus attention on verb semantics or sentential meaning (Marsden, 2006; Marsden & Chen, 2011).

To date, two studies have examined task-essential activities involving IMP (Benati, Lee & Laval, 2008 and Lee, Benati, Aguilar-Sánchez & McNulty, 2007<sup>2</sup>). These studies removed aspectuo-temporal adverbs (e.g. *tous les jours* ‘every day’) that can sometimes co-occur with IMP and may render it less likely to be attended (VanPatten, 2002). However, no research has yet investigated how to make task-essential the two different *aspectual* meanings of IMP - ongoingness and habituality. As described above, this can only be determined reliably by *morphosyntactic* information in the discourse context. To date, we do not know whether task-essential practice can help learning that requires co-indexation with *morphosyntax in another clause* to ascertain the correct form-meaning mapping.

### **Current study**

The current study begins to address the gaps identified above by: i) investigating a hitherto neglected target feature - two aspectual dimensions of the French IMP that require inter-clausal morphosyntactic cues (past habituality and ongoingness); ii) measuring both the speed of online processing and offline interpretation to investigate learning; iii) investigating the role of L1 EI and L1 task-essential practice in L2 learning.

Based on previous research, we expected that L2 EI and task-essential practice would result in gains in offline measures, whereas the Control condition (tests only) would not. As little/no research has examined specifically a) the *online* effects of EI with task-essential practice (L1 or L2) or b) the role of *L1* EI with practice, we could not adopt strong expectations for these two dimensions of the study. However, for a), based on research suggesting links between processing and learning, we thought that online measures might show increased sensitivity to correct/incorrect use of IMP following both L2-only and L2+L1 treatments. This is partly

because our treatments included extensive form-meaning mapping practice, unlike previous studies that have not found online effects for EI or practice. For b), based on research suggesting L1-L2 co-activation during processing and evidence of potential benefits of making explicit and rehearsing cross-linguistically complex form-meaning mappings, it was thought that, compared to Control, the L2+L1 treatment would lead to larger and more consistent on- and offline effects than the L2-only treatment compared to Control.

## **Method**

### **Participants**

Participants were 50 (42 females, 8 males) English-speaking learners of French as a foreign language in semester two of a four-year BA Honours degree in French at a large university in England. We required every participant to be a native speaker of English, have completed A-level (A2) French (equivalent to CEFR level B2, normally after about 700-800 hours of instruction), and not to have spent more than six weeks in a French speaking country. We collected background information via a questionnaire and excluded six people based on these criteria. Participants' mean age was 19, and the mean time spent in a French speaking country was 3.8 weeks. 39 participants declared knowledge of other romance languages (Spanish=29, Portuguese=5, Italian=4), twelve declared knowledge of German, and twelve declared knowledge of other languages (two each for Greek, Latin, Welsh, and one each for Arabic, BSL, Japanese, Mandarin, Polish, Russian).

### **Design**

The study had three between-subjects groups (Control, L2-only, L2+L1) and three within-subject tests (Pretest in week 1, Posttest in week 5, Delayed Posttest in week 12). All

tests and treatments were administered one-to-one with laptops using E-Prime 2.0 (Schneider, Eschman & Zuccolotto, 2012). We assigned participants to a group using matched randomization<sup>3</sup>, resulting in 17 participants in the L2+L1 group, 17 in the L2-only group, and 16 in the Control group (tests only). By contemporary standards these are small numbers, due to the amount of time and relatively long timescale required by the study, and we acknowledge, therefore, that this constitutes an exploratory study. The treatments were delivered in four 45-minute sessions over three weeks, totalling 3.5 hours. Sessions 1 and 2 were delivered in Week 2, Session 3 in Week 3, and Session 4 in Week 4. The Control group completed all tests and did not receive any intervention treatment, but continued normal instruction along with all other participants between Pre- and Posttests. Due to the vacation between Post- and Delayed Posttests, none of the participants received any instruction (either as part of their university programme or our experiment). This increased the likelihood that any effects found at Delayed Posttest were due to our intervention. In the university programme from which all participants were drawn, explicit grammar instruction only took place in Semester 1, i.e. prior to the study, corroborated by interviews with university tutors. The entire study was piloted in a condensed timescale with 10 English-speaking learners of French at another university.

### **Target structure: French IMP**

All exemplars of IMP were third-person singular forms. This included 25 regular (e.g. *jouait*) and 23 irregular (e.g. *finissait*) verb types. Regulars and irregulars were included because 1) the study's focus was inflectional morphology that remains orthographically and phonologically constant in the IMP across regular and irregular stems (L'Huillier, 1999), 2) the tests were receptive and so production of irregular stems was not measured, 3) there is some evidence that learning the two functions of IMP (habitual and ongoing) may relate to

lexical verb type (e.g. activities – *manger*, achievements – *arriver*) and as we wished to counter-balance lexical types this entailed inclusion of frequent irregular verbs (Andersen & Shirai, 1994)<sup>4</sup>; 5) the study's ecological validity was increased by including both regular and irregulars; 6) any potential effect of verb type was experienced by all three participant groups equally, as verb types were counterbalanced across test versions, and test versions counter-balanced across conditions and test phases.

### **Instructional treatments**

The L2+L1 and L2-only treatments included an identical core of EI and practice in interpreting French IMP<sup>5</sup>. We first describe this common core, before describing the L1 treatment received by the L2+L1 group (see also Supplementary Materials: Treatment).

**EI about the L2.** EI was provided in two ways: a) Pre-practice, approximately 5 minutes at the start of each session; and b) during the task-essential practice activities following only incorrect answers, which, as Supplementary Materials: Treatment shows, was infrequent and occurred in almost identical amounts in both treatments. The pre-practice EI depicted conceptual-semantic information via a short video, image or sound file of events. Then the appropriate aural and written forms were presented, and information given about how to interpret their meaning.

**Practice in interpreting the L2.** The short pre-practice EI was followed by task-essential, form-meaning mapping practice, listening and reading in equal amounts, focussing attention on meaning contrasts expressed by different forms. In line with other task-essential activities, this was done via learners choosing the meaning of a stimulus from fixed options (e.g. Marsden, 2006; Marsden & Chen, 2011; Sanz & Morgan-Short, 2004; VanPatten, 2002).

The numbers of French exemplars are shown in Table 1. The practice drew on 48 lexical verb types: each one occurred eight times with IMP (n=384): four for reading (two habitual, two ongoing/interrupted); four for listening (two habitual, two ongoing/interrupted). The lexical semantic properties of verb types were counterbalanced across listening/reading and ongoing/habitual items: 12 states (e.g. be happy), 12 activities (e.g. swim in the sea), 12 accomplishments (e.g. walk to the shop) and 12 achievements (e.g. arrive home). Verb type frequency was balanced across the four lexical semantic classes using Lonsdale and Le Bras's (2009) frequency dictionary of French. Aural stimuli were recorded by two native French speakers. The French sentences were verified for authenticity by 26 native French speakers: all were rated as 100% acceptable, with the meanings (ongoing/ habitual, present/past) as intended by the researchers.

[Table 1 here]

**The L2+L1 treatment.** The L2+L1 group received the exact same treatment as above, with no additional French L2 exemplars. The L2+L1 treatment additionally included EI about English and task-essential form-meaning mapping practice in English. Table 2 shows the numbers of tokens of English exemplars (all third-person singular). See Supplementary Materials: Treatment for full descriptions and example activities and stimuli.

[Table 2 here]

### **Outcome measures**

Two versions of each outcome measure were administered in a split-block design. The versions alternated the lexical items carrying IMP and PC inflections and the order of items.



**Context-matching tests (listening and reading).** All participants took two context-matching tests: first a listening (CMT-Listen), then a reading (CMT-Read), each with 24 target and 8 filler trials. Each trial consisted of two parts: (i) the English context: two sentences describing either a habitual or an ongoing activity written in English and (ii) the French stimulus: a two-clause French sentence that either matched ( $k=12$ ) or mismatched ( $k=12$ ) the meaning of the English context. Critically, the French stimuli and English contexts were never translations of each other; rather, the context gave a fuller description of an event in which either a habitual or ongoing function of IMP would be required, on different lexical items, in the shorter stimulus sentence. In this way, we were not eliciting direct translations between context and stimulus, but specific functions of IMP. For example:

MATCHED TRIAL in CMT-READ

**Context** (ongoing): Yesterday, Patrick was expecting his wife to come back from work any minute. Just as he was on his way out, she appeared in the driveway.

**Stimulus** (ongoing): *Quand Patrick quittait la maison, il a vu sa femme*

“When Patrick was leaving the house, he saw his wife”

MISMATCHED TRIAL in CMT-READ

**Context** (ongoing): Yesterday, Patrick was expecting his wife to come back from work any minute. Just as he was on his way out, she appeared in the driveway.

**Stimulus** (habitual): *Quand Patrick quittait la maison, il voyait sa femme*

“When Patrick left the house, he used to see his wife”

In both CMTs the English context appeared on screen for 10 seconds. Then, the French stimulus appeared orally (CMT-Listen) or in writing (CMT-Read). Participants were instructed to rate how good the match was between the meaning of the French stimulus and

English context by pressing a number on the keyboard from 1 (“very good”), 2 (“good”), 3 (“neither good nor bad”), 4 (“poor”), 5 (“very poor”), with a separate option for “I don’t know” (9). The written French stimulus remained on screen until a number was pressed, and then participants could not change their answer. The task was untimed and took approximately 20-25 minutes.

The CMTs drew on 24 of the 48 lexical verbs (3<sup>rd</sup> person singular) from the intervention<sup>6</sup>.

Items were counterbalanced across the match and mismatch conditions for:

ongoingness/habituality, verb frequency, lexical aspect class, verb regularity, and clause ordering (main->subordinate / subordinate-> main).

In addition to the pilot study, we checked the English contexts with three native speakers of English, the French stimuli with 26 native speakers of French, and the match and mismatch combinations with three L1 English very advanced learners of French.

**Self-paced reading test.** The self-paced reading (SPR) test was administered after the CMTs and used 16 items from the CMT-Listen<sup>7</sup>, with eight context-stimulus matches and eight mismatches, counterbalanced as described above for the CMT tests. Half the items were followed by Yes/No comprehension questions to increase the likelihood that participants focused on meaning (see Keating & Jegerski, 2015)<sup>8</sup>. The answers to the questions only depended on a lexico-semantic feature (but not verb stems) and not on inflectional morphology. For each trial, the English context appeared for 10 seconds before an X appeared in the centre of the screen. A spacebar press brought up the first and then each subsequent word of the French stimuli. After the last word, the next screen displayed “END”. Participants were instructed to read as quickly as possible. Reaction times were collected from each word according to the noncumulative moving-window procedure (Marinis et al.,

2005). The font was 18-point Courier New, displayed in the centre of a white background, without line breaks.

### **Data scoring and analysis**

For the CMTs, responses were coded following standard protocols for judgement tasks (see Mackey and Gass, 2013): 5 points for each correct response (i.e. pressing 1 or 2 for match trials, and 4 or 5 for mismatch trials); 3 points for midway responses (pressing 3 for match and mismatch trials); 1 point for each incorrect response (pressing 4 or 5 for match trials, and 1 or 2 for mismatch trials). Cronbach's alphas were: CMT-Listen version A ( $\alpha=.72$ ), CMT-Listen version B ( $\alpha=.74$ ), CMT-Read version A ( $\alpha=.72$ ), CMT-Read version B ( $\alpha=.74$ ).

In the SPR, reaction times in each French stimulus were on the critical word, underlined for illustrative purposes in Table 3 - the verb in the co-ordinating clause (either IMP or PC) that disambiguated the main clause verb's meaning (habitual *versus* ongoing). Mean sentence length was 10 words (SD=1.5, range 9 - 14). Mean length of the critical word was 2.3 syllables for IMP and 2.9 for PC (all auxiliaries were one syllable, past participle mean = 1.9), with bi- and tri-syllabic words counterbalanced across test versions. Reaction times (RT) for the critical word were calculated from the onset of the critical word to the onset of the next word<sup>9</sup>. We also analysed whole sentence RTs calculated as the time taken to read from the onset of the first word to the onset of the 'END' screen. We analysed the raw RT data, which we trimmed in line with recommendations for SPR (Keating & Jegerski, 2015), removing critical word RTs less than 150ms and greater than 2000ms (three (0.5%, habitual match) and four (0.7%, ongoing match) data points across 50 participants).

Slower RTs are usually interpreted as indications of a processing burden, and, conversely, faster RTs as indications of relative processing ease (e.g. Leung & Williams, 2014; Marinis et

al., 2005). We therefore analysed RTs for changes over time in different training conditions. We also compared RTs in matched *versus* mismatched trials, to detect changes in sensitivity to violations in the use of IMP. If learners became more sensitive to the different functions of IMP following training, it was expected that differences in their RTs between match and mismatch trials would become (more) apparent. For example, following a habitual context, if learners were sensitive to a context-stimulus anomaly, an IMP+PC stimulus (mismatch) would cause a slower RT compared to an IMP+IMP (match).

[Table 3 here]

As none of the data were normally distributed (according to Shapiro-Wilks tests, all datasets  $p < .05$ , see Supplementary Materials: Statistics, Table 5), we present the results of nonparametric tests (Field, 2013; Norris, Plonsky, Ross & Schoonen, 2015). Nevertheless, for parity with other studies, we note that the patterns of findings did not differ when parametric tests were used (i.e. mixed design ANOVAs with planned contrasts).

First, between-group differences at Pretest were checked using Kruskal-Wallis H tests.

Second, Friedman tests were used to compare Pretest, Posttest, and Delayed Posttest scores within each group, and, if a significant difference was found, then within-subject comparisons were made using Wilcoxon Signed-Rank tests with bonferroni corrected alpha levels (equivalent to post-hoc analyses) between pairs of test results: Pre-Post, Pre-Delayed, and Post-Delayed. Finally, we compared SPR performance between matched and mismatched trials, in each group, with Wilcoxon Signed-Rank tests<sup>10</sup>.

Following recent discussion on decreasing the probability of Type II errors, including the observation that p-values can be strongly influenced by sample size (Plonsky, 2015; Plonsky & Oswald, 2014) and that low stakes outcomes should entail setting higher alpha levels (Norris, 2015), for the Kruskal-Wallis H and Friedman tests the alpha level was set at  $0.10^{11}$ .

For the post-hoc Wilcoxon Signed-Rank tests, we used a Bonferroni adjustment making the revised alpha value  $0.10/3 = .033$  because three comparisons were carried out for each significant omnibus test result (Field, 2013). For interpreting the magnitude of change, we present Cohen's  $d$  effect sizes for all paired comparisons. Effect sizes between tests, i.e. within-subjects, were calculated in relation to the mean and standard deviation of the Pretest as a baseline (and the Posttest for effect sizes at Delayed Posttest). As within-subject effect sizes tend to be larger than between-subject, we also give effect sizes compared to the Control group using the mean and standard deviation of the Control group, Tables 5 and 7, with both the raw effect size and an adjusted effect size corrected for baseline differences (even though all were non-statistically significant differences, Plonsky & Oswald 2014). Following Plonsky and Oswald (2014), Cohen's  $d$  field-specific benchmarks are used for interpretation:  $d=.40$  (small),  $.70$  (medium) and  $.10$  (large). Confidence intervals for effect sizes are in the Supplementary Materials: Statistics (Tables 2 and 3).

## Results

Results are presented first for habitual contexts, then for ongoing contexts, separately for matched and mismatched trials. After establishing baseline parity at Pretest, we present change over time within each group on each test (CMT-Read, CMT-Listen, SPR critical word and whole sentence) and effect sizes compared to the Control group on each test. For the SPR, we present comparisons of matched versus mismatched trials in each group.

## Habitual contexts

The accuracy of judgements (CMTs) and reaction times (SPR) are presented in Table 4. All participants achieved 100% accuracy in SPR comprehension questions. Clause ordering comparisons showed no significant effects (see Table 1A, Supplementary Materials: Statistics).

[Table 4 here]

**Matched trials.** Kruskal-Wallis tests revealed no statistically significant between-group differences at Pretest in CMT-Listen ( $X^2(2)=.358$ ,  $p=.836$ ), CMT-Read ( $X^2(2)=.481$ ,  $p=.786$ ), SPR critical word ( $X^2(2)=.316$ ,  $p=.854$ ), and SPR whole sentence ( $X^2(2)=.151$ ,  $p=.927$ ).

For the Control group, in all three tests, there were no statistically significant changes over time: CMT-Read,  $X^2(2)=2.032$ ,  $p=.362$  (Pre-Post,  $d=.33$ ; Pre-Delayed,  $d=-.65$ ; Post-Delayed,  $d=-.26$ ); CMT-Listen,  $X^2(2)=1.348$ ,  $p=.510$  (Pre-Post,  $d=-.13$ ; Pre-Delayed,  $d=-.13$ ; Post-Delayed,  $d=.00$ ); SPR critical word,  $X^2(2)=2.625$ ,  $p=.269$  (Pre-Post,  $d=-.25$ ; Pre-Delayed,  $d=-.29$ ; Post-Delayed,  $d=-.05$ ); SPR whole sentence,  $X^2(2)=3.875$ ,  $p=.144$  (Pre-Post,  $d=-.63$ ; Pre-Delayed,  $d=-.67$ ; Post-Delayed,  $d=.03$ ).

For the L2-only group, accuracy in CMT-Listen increased over time ( $X^2(2)=7.276$ ,  $p=.026$ ), due to Pre-Post changes ( $Z=-2.163$ ,  $p=.031$ ,  $d=.69$ ), but not Pre-Delayed ( $Z=-.512$ ,  $p=.609$ ,  $d=.51$ ) or Post-Delayed ( $Z=-1.179$ ,  $p=.238$ ,  $d=.00$ ). No changes over time were found for CMT-Read ( $X^2(2)=.250$ ,  $p=.882$ : Pre-Post,  $d=-.21$ ; Pre-Delayed,  $d=.43$ ; Post-Delayed,  $d=.51$ ). In SPR, although whole sentence reading got faster over time ( $X^2(2)=12.118$ ,  $p=.002$ ), Pre-Post ( $Z=-2.343$ ,  $p=.019$ ,  $d=-.55$ ) and Pre-Delayed ( $Z=-2.722$ ,  $p=.006$ ,  $d=-.68$ ), with minimal change Post-Delayed ( $Z=-1.160$ ,  $p=.246$ ,  $d=-.16$ ), RTs for the critical word

were not significantly different, though effect sizes suggest possible trends ( $X^2(2)=1.529$ ,  $p=.465$ : Pre-Post,  $d=-.63$  ; Pre-Delayed,  $d=-.67$ : Post-Delayed,  $d=-.03$ ).

For the L2+L1 group, we found significant changes over time on all measures. Accuracy improved over time in both CMT-Read ( $X^2(2)=9.254$ ,  $p=.010$ ) and CMT-Listen ( $X^2(2)=7.356$ ,  $p=.025$ ), with small effects Pre-Post (CMT-Read,  $Z=-1.484$ ,  $p=.138$ ,  $d=.41$ ; CMT-Listen,  $Z=-.908$ ,  $p=.364$ ,  $d=.35$ ), large for Pre-Delayed (CMT-Read,  $Z=2.692$ ,  $p=.007$ ,  $d=1.67$ ; CMT-Listen,  $Z=-2.843$ ,  $p=.004$ ,  $d=1.13$ ), and with small-to-medium effects Post-Delayed (CMT-Read,  $Z=-2.275$ ,  $p=.023$ ,  $d=.58$ ; CMT-Listen,  $Z=-1.616$ ,  $p=.106$ ,  $d=.74$ ). In the SPR, critical word RTs decreased significantly ( $X^2(2)=17.765$ ,  $p=.000$ ), Pre-Post ( $Z=-3.432$ ,  $p=.001$ ,  $d=-1.20$ ) and Pre-Delayed ( $Z=-3.290$ ,  $p=.001$ ,  $d=-1.14$ ), but not Post-Delayed ( $Z=-.592$ ,  $p=.554$ ,  $d=.02$ ), with larger effect sizes than the L2-only group. Whole sentence RTs also decreased significantly ( $X^2(2)=12.824$ ,  $p=.002$ ) for Pre-Post ( $Z=-3.243$ ,  $p=.001$ ,  $d=-1.06$ ) and Pre-Delayed ( $Z=-3.195$ ,  $p=.001$ ,  $d=-.90$ ), with no differences Post-Delayed ( $Z=-1.160$ ,  $p=.246$ ,  $d=.03$ ).

***Effect sizes compared to Control.*** As shown in Table 5, negligible effects were found between L2-only and Control at Post and Delayed in CMT-Read, SPR whole sentence and critical word. In CMT-Listen, negligible effects were found at Post and a small effect at Delayed. In contrast, larger effects were found between L2+L1 and Control: in CMT-Listen, medium effects at Post and large at Delayed; in CMT-Read, small at Post and large at Delayed; in SPR, for the critical word, large at Post and medium at Delayed, and for the whole sentence small at Post and Delayed.

**Mismatched trials.** Kruskal-Wallis tests revealed no statistically significant between-group differences at Pretest in CMT-Listen ( $X^2(2)=1.216$ ,  $p=.544$ , CMT-Read ( $X^2(2)=.005$ ,

$p=.998$ ), SPR whole sentence ( $X^2(2)=1.292$ ,  $p=.524$ ), and SPR critical word ( $X^2(2)=.189$ ,  $p=.910$ ).

For the Control group, there were no changes over time in CMT-Read ( $X^2(2)=.933$ ,  $p=.627$ : Pre-Post,  $d=.12$ ; Pre-Delayed,  $d=.10$ ; Post-Delayed,  $d=-.02$ ); SPR critical word ( $X^2(2)=.000$ ,  $p=1.000$ : Pre-Post,  $d=-.18$ ; Pre-Delayed,  $d=-.18$ ; Post-Delayed,  $d=-.06$ ); SPR whole sentence ( $X^2(2)=3.875$ ,  $p=.144$ : Pre-Post,  $d=-.49$  and Pre-Delayed,  $d=-.53$ ; Post-Delayed,  $d=-.11$ ). In CMT-Listen, accuracy scores deteriorated ( $X^2(2)=6.136$ ,  $p=.047$ ) Pre-Delayed ( $Z=-2.632$ ,  $p=.018$ ,  $d=-.84$ ) and Post-Delayed ( $Z=-2.300$ ,  $p=.021$ ,  $d=-.63$ ), but not Pre-Post ( $Z=-.789$ ,  $p=.430$ ,  $d=-.28$ ).

For the L2-only group, performance did not change over time in CMT-Listen,  $X^2(2)=1.581$ ,  $p=.454$  (Pre-Post,  $d=-.18$ ; Pre-Delayed,  $d=.18$ ; Post-Delayed,  $d=.37$ ), CMT-Read,  $X^2(2)=.892$ ,  $p=.892$  (Pre-Post,  $d=.34$ ; Pre-Delayed,  $d=.20$ ; Post-Delayed,  $d=-.10$ ), SPR whole sentence (though with a medium effect size Pre-Post) ( $X^2(2)=4.353$ ,  $p=.113$ : Pre-Post,  $d=-.78$ ; Pre-Delayed,  $d=-.59$ ; Post-Delayed,  $d=-.15$ ), or critical word ( $X^2(2)=1.412$ ,  $p=.494$ : Pre-Post,  $d=-.11$ ; Pre-Delayed,  $d=-.16$ ; Post-Delayed,  $d=-.09$ ).

For the L2+L1 group, changes were observed in all tests. There were significant improvements in accuracy for both CMT-Read ( $X^2(2)=9.270$ ,  $p=0.10$ ) and CMT-Listen ( $X^2(2)=12.406$ ,  $p=.002$ ), with large effects Pre-Delayed (CMT-Read  $Z=-3.004$ ,  $p=.003$ ,  $d=1.65$ ; CMT-Listen  $Z=-2.819$ ,  $p=.005$ ,  $d=1.22$ ), and medium-to-large effects Post-Delayed (CMT-Read,  $Z=-2.296$ ,  $p=.022$ ,  $d=.91$ ; CMT-Listen,  $Z=-2.872$ ,  $p=.004$ ,  $d=1.02$ ), notable given that no participant received instruction Post-Delayed. Negligible effects were found Pre-Post (CMT-Read,  $Z=-1.086$ ,  $p=.277$ ,  $d=.37$ ; CMT-Listen,  $Z=-.959$ ,  $p=.338$ ,  $d=.34$ ). In the SPR, there was no significant change in RT for the whole sentence ( $X^2(2)=3.294$ ,  $p=.193$ : Pre-Post,  $d=-.56$ ; Pre-Delayed,  $d=-.26$ ; Post-Delayed,  $d=.26$ ), but critical word reading got faster with medium-large effect sizes ( $X^2(2)=19.882$ ,  $p=.000$ ) Pre-Post ( $Z=-3.432$ ,  $p=.001$ ,



$d=-.85$ ), Pre-Delayed ( $Z=-3.574$ ,  $p=.000$ ,  $d=-.1.09$ ), but not Post-Delayed ( $Z=-.497$ ,  $p=.619$ ,  $d=-.34$ ).

***Effect sizes compared to Control.*** (See Table 5). For the L2-only group, effects were: in CMT-Listen, negligible at Post but large at Delayed; in CMT-Read, effects were negligible; in SPR (critical word and whole sentence), effects were also negligible. For the L2+L1 group, effects sizes were: in CMT-Listen, large at Post and Delayed; in CMT-Read small at Post and large at Delayed; in SPR, effects on the critical word were small at Post and Delayed, and for the whole sentence, small at Post, but negligible at Delayed.

The results for habitual trials (matched and mismatched) suggest patterns of greater accuracy and faster processing speeds following the L2+L1 training (corroborated by confidence intervals in Supplementary Materials: Statistics, Table 2). In terms of SPR results, we also note a) larger effects for critical word RT in L2+L1 than for whole sentence RT, b) effects for both critical word and whole sentence are larger for L2+L1 than for L2-only and Control, suggesting a larger effect for the L2+L1 treatment on processing in general and c) any Pre-Delayed effects for L2-only and Control found for whole sentence RT are larger than for critical word RT. Lastly, we do not see speed of processing effects after Post.

***Reading times in matched versus mismatched trials.*** At Pretest, no group performed significantly differently across different trial types on the critical item (Control,  $Z=-1.655$ ,  $p=.098$ ,  $d=-.47$ ; L2-only,  $Z=-1.444$ ,  $p=.149$ ,  $d=-.41$ ; L2+L1,  $Z=-1.302$ ,  $p=.193$ ,  $d=-.37$ ) or whole sentence (Control,  $Z=-.259$ ,  $p=.796$ ,  $d=-.08$ ; L2-only,  $Z=-1.302$ ,  $p=.193$ ,  $d=-.37$ ; L2+L1,  $Z=-1.538$ ,  $p=.124$ ,  $d=-.45$ ). However, the L2+L1 group's RTs were significantly slower in mismatched compared to matched trials at both Post and Delayed for critical word (Post,  $Z=-2.722$ ,  $p=.006$ ,  $d=-.82$ ; Delayed,  $Z=-2.390$ ,  $p=.017$ ,  $d=-.82$ ) and whole sentence (Post,  $Z=-1.965$ ,  $p=.049$ ,  $d=-.58$ ; Delayed,  $Z=-2.012$ ,  $p=.044$ ,  $d=-.58$ ). In contrast, we found

no between-trial differences in the L2-only group for the critical word (Post,  $Z=-.024$ ,  $p=.981$ ,  $d=-.01$ ; Delayed,  $Z=-.402$ ,  $p=.687$ ,  $d=-.12$ ) or whole sentence (Post,  $Z=-.308$ ,  $p=.758$ ,  $d=-.08$ ; Delayed,  $Z=-1.254$ ,  $p=.210$ ,  $d=-.37$ ), nor in the Control group: critical word (Post,  $Z=-.672$ ,  $p=.501$ ,  $d=-.19$ ; Delayed,  $Z=-.571$ ,  $p=.605$ ,  $d=-.16$ ) or whole sentence (Post,  $Z=-1.183$ ,  $p=.255$ ,  $d=-.35$ ; Delayed,  $Z=-.155$ ,  $p=.877$ ,  $d=-.04$ ).

### Ongoing contexts

Accuracy of judgement (CMTs) and reaction times (SPR) are presented in Table 6. All groups achieved 100% accuracy in SPR comprehension questions. Clause ordering comparisons showed no significant effects (Supplementary Materials: Statistics Table 1A).

[Table 6 here]

**Matched trials.** There were no significant between-group differences at Pretest on any measure (CMT-Listen,  $X^2(2)=.074$ ,  $p=.964$ ; CMT-Read,  $X^2(2)=.809$ ,  $p=.667$ ; SPR whole sentence,  $X^2(2)=.011$ ,  $p=.994$ ; SPR critical word  $X^2(2)=.036$ ,  $p=.982$ ).

For the Control group, there was no significant change over time on any measure: CMT-Listen,  $X^2(2)=3.049$ ,  $p=.218$  (Pre-Post,  $d=.20$ ; Pre-Delayed,  $d=-.22$ ; Post-Delayed,  $d=-.44$ ), CMT-Read,  $X^2(2)=.255$ ,  $p=.880$  (Pre-Post,  $d=.20$ ; Pre-Delayed,  $d=.16$ ; Post-Delayed,  $d=-.05$ ); SPR critical word,  $X^2(2)=3.875$ ,  $p=.144$  (Pre-Post,  $d=-.63$ ; Pre-Delayed,  $d=-.47$ ; Post-Delayed,  $d=.02$ ); SPR whole sentence,  $X^2(2)=.500$ ,  $p=.779$  (Pre-Post,  $d=-.56$ ; Pre-Delayed,  $d=-.54$ ; Post-Delayed,  $d=.09$ ).

For the L2-only group, in CMT-Read ( $X^2(2)=12.000$ ,  $p=.002$ ) accuracy improved Pre-Post ( $Z=-2.762$ ,  $p=.006$ ,  $d=1.02$ ) and Pre-Delayed ( $Z=-2.773$ ,  $p=.006$ ,  $d=1.08$ ), not Post-Delayed ( $Z=-.085$ ,  $p=.932$ ,  $d=.00$ ). In CMT-Listen ( $X^2(2)=3.180$ ,  $p=.204$ ), accuracy improved Pre-

Post ( $Z=-2.303$ ,  $p=.021$ ,  $d=.75$ ), but not Pre-Delayed ( $Z=-.882$ ,  $p=.378$ ,  $d=.37$ ) or Post-Delayed ( $Z=-.701$ ,  $p=.483$ ,  $d=-.24$ ). In SPR, we found faster processing of the critical word ( $X^2(2)=10.706$ ,  $p=.005$ ), Pre-Post ( $Z=-2.722$ ,  $p=.006$ ,  $d=-.98$ ), but not Pre-Delayed ( $Z=-1.965$ ,  $p=.049$ ,  $d=-.34$ ), with no change Post-Delayed ( $Z=-1.337$ ,  $p=.181$ ,  $d=-.54$ ). For the whole sentence, RTs did not change over time ( $X^2(2)=4.235$ ,  $p=.120$ : Pre-Post,  $d=-.19$ ; Pre-Delayed,  $d=-.53$ ; Post-Delayed,  $d=-.39$ ).

For the L2+L1 group, there was improved accuracy in the CMT-Read ( $X^2(2)=27.098$ ,  $p=.000$ ), Pre-Post ( $Z=-3.413$ ,  $p=.001$ ,  $d=1.67$ ) and Pre-Delayed ( $Z=-3.415$ ,  $p=.001$ ,  $d=1.89$ ), but not Post-Delayed ( $Z=-1.294$ ,  $p=.196$ ,  $d=.35$ ). For CMT-Listen ( $X^2(2)=11.828$ ,  $p=.003$ ), small gains were found Pre-Post ( $Z=-1.741$ ,  $p=.082$ ,  $d=.54$ ), large gains Pre-Delayed ( $Z=-3.250$ ,  $p=.001$ ,  $d=1.72$ ), with some gains Post-Delayed ( $Z=-1.888$ ,  $p=.059$ ,  $d=.72$ ). For SPR, increases in speed were found on the critical word ( $X^2(2)=16.328$ ,  $p=.000$ ) and whole sentence ( $X^2(2)=15.176$ ,  $p=.001$ ), Pre-Post (critical word,  $Z=-3.479$ ,  $p=.001$ ,  $d=-1.37$ ; whole sentence,  $Z=-3.195$ ,  $p=.001$ ,  $d=-1.22$ ) and Pre-Delayed (critical word,  $Z=-3.103$ ,  $p=.002$ ,  $d=-1.20$ ; whole sentence,  $Z=-3.337$ ,  $p=.001$ ,  $d=-1.04$ ), but not Post-Delayed (critical word  $Z=-3.355$ ,  $p=.723$ ,  $d=.15$ ; whole sentence  $Z=-1.065$ ,  $p=.287$ ,  $d=.09$ ).

***Effect sizes compared to Control.*** As shown in Table 7, for L2-only in CMT-Listen effects were small at Post and Delayed; in CMT-Read, medium at Post and large at Delayed. In SPR, there were negligible effects at Post and Delayed for critical word and whole sentence RTs. In contrast, for the L2+L1 group we found larger effects: in CMT-Listen, effects were small at Post and large at Delayed; in CMT-Read, large at Post and Delayed; in SPR, effects were large for the critical word at Post and Delayed and medium for whole sentence at both Post and Delayed.

**Mismatched trials.** There were no significant between-group differences at Pretest (CMT-Listen,  $X^2(2)=.053$ ,  $p=.974$ ; CMT-Read,  $X^2(2)=1.138$ ,  $p=.566$ ; SPR critical word,  $X^2(2)=.182$ ,  $p=.913$ ; SPR whole sentence,  $X^2(2)=1.034$ ,  $p=.596$ ).

For the Control group, performance did not change over time in CMT-Read,  $X^2(2)=.448$ ,  $p=.799$  (Pre-Post,  $d=.39$ ; Pre-Delayed,  $d=.26$ ; Post-Delayed,  $d=-.09$ ); deteriorated in CMT-Listen ( $X^2(2)=7.172$ ,  $p=.028$ ), Pre-Delayed ( $Z=-2.659$ ,  $p=.008$ ,  $d=-.71$ ) but not Pre-Post ( $Z=-.910$ ,  $p=-.363$ ,  $d=-.29$ ) or Post-Delayed ( $Z=-.511$ ,  $p=.609$ ,  $d=-.40$ ); and did not change in SPR critical word,  $X^2(2)=.875$ ,  $p=.646$  (Pre-Post,  $d=-.12$ ; Pre-Delayed,  $d=.10$ ; Post-Delayed,  $d=.18$ ) or whole sentence,  $X^2(2)=3.500$ ,  $p=.174$  (Pre-Post,  $d=-.28$ ; Pre-Delayed,  $d=-.52$ ; Post-Delayed,  $d=-.17$ ).

For the L2-only group, we found no change over time on any measure: CMT-Read,  $X^2(2)=.426$ ,  $p=.808$  (Pre-Post,  $d=.19$ ; Pre-Delayed,  $d=.31$ ; Post-Delayed,  $d=.09$ ), CMT-Listen,  $X^2(2)=2.710$ ,  $p=.258$  (Pre-Post,  $d=.47$ ; Pre-Delayed,  $d=.14$ ; Post-Delayed,  $d=-.36$ ), SPR critical word,  $X^2(2)=1.882$ ,  $p=.390$  (Pre-Post,  $d=-.20$ ; Pre-Delayed  $d=-.19$ ; Post-Delayed,  $d=-.01$ ), or whole sentence,  $X^2(2)=.353$ ,  $p=.838$  (Pre-Post,  $d=-.52$ ; Pre-Delayed,  $d=-.40$ ; Post-Delayed  $d=.10$ ).

For the L2+L1 group, we found significant change over time in all tests, with medium effects in CMT-Read ( $X^2(2)=.14.302$ ,  $p=.001$ ), Pre-Post ( $Z=-2.883$ ,  $p=.004$ ,  $d=.93$ ) and Pre-Delayed ( $Z=-3.242$ ,  $p=.001$ ,  $d=1.46$ ), with small Post-Delayed differences ( $Z=-1.961$ ,  $p=.050$ ,  $d=.43$ ). In CMT-Listen ( $X^2(2)=.11.375$ ,  $p=.003$ ), accuracy increased Pre-Delayed ( $Z=-2.962$ ,  $p=.003$ ,  $d=1.27$ ), but not Pre-Post ( $Z=-1.223$ ,  $p=.221$ ,  $d=.42$ ), with a medium effect Post-Delayed ( $Z=-1.615$ ,  $p=.106$ ,  $d=.71$ ). In SPR, we found large effects for faster processing of the critical word ( $X^2(2)=.22.588$ ,  $p=.000$ ), Pre-Post ( $Z=-3.574$ ,  $p=.000$ ,  $d=-1.01$ ) and Pre-Delayed ( $Z=-3.621$ ,  $p=.000$ ,  $d=-.98$ ), but not Post-Delayed (SPR,  $Z=-.639$ ,  $p=.523$ ,  $d=.04$ ). RTs for the

whole sentence did not change over time ( $X^2(2)=4.235$ ,  $p=.120$ : Pre-Post,  $d=-.49$ ; Pre-Delayed,  $d=-.28$ ; Post-Delayed,  $d=.17$ ).

***Effect sizes compared to Control.*** (See Table 7). For the L2-only group, medium effect sizes were found in CMT-Listen at both Post and Delayed, with negligible effects found in CMT-Read; for SPR, there were no effects for both critical word and whole sentence processing. For the L2+L1 group, in CMT-Listen effect sizes at Post were small and large at Delayed; in CMT-Read, effects were negligible at Post, but medium at Delayed; in SPR, effect sizes for the critical word were medium at both Post and Delayed, and effects were negligible for the whole sentence.

In these comparisons to Control, in matched and mismatched trials, effects seemed larger in the L2+L1 group for accuracy and processing speed. Effect sizes for RTs on the critical word seemed particularly affected, and were larger than for the whole sentence, whereas the L2-only group's effects on both critical word and whole sentence were negligible, suggesting an across the board increase in processing speed due to test-familiarity cannot adequately explain the results. (See also confidence intervals in Supplementary Materials: Statistics, Table 3). For example, for L2+L1, Pre-Delayed effect sizes are larger for critical word RT than for whole sentence, and SPR effects sizes for L2+L1 are consistently larger than for both L2-only and Control, suggesting an advantage for the L2+L1 treatment. Lastly, we do not see speed of processing effects after Posttest (except for L2-only in ongoing match).

***Reading times in matched versus mismatched trials.*** At Pretest, Wilcoxon Signed Rank tests revealed no significant between-trial RT differences in any group for either the critical word (Control,  $Z=-.362$ ,  $p=.717$ ,  $d=-.10$ ; L2-only,  $Z=-.166$ ,  $p=.868$ ,  $d=-.04$ ; L2+L1,  $Z=-.592$ ,  $p=.554$ ,  $d=-.16$ ) or the whole sentence (Control,  $Z=-.259$ ,  $p=.769$ ,  $d=-.08$ ; L2-only,  $Z=-.876$ ,  $p=.381$ ,  $d=-.24$ ; L2+L1,  $Z=-1.444$ ,  $p=.149$ ,  $d=-.41$ ). However, after the intervention, the

L2+L1 group's processing was significantly slower in mismatched than in matched trials at both Post (critical word,  $Z=-2.402$ ,  $p=.016$ ,  $d=-.72$ ; whole sentence,  $Z=-2.485$ ,  $p=.013$ ,  $d=-.75$ ) and Delayed (critical word,  $Z=-2.485$ ,  $p=.013$ ,  $d=-.75$ ; whole sentence,  $Z=-2.580$ ,  $p=.010$ ,  $d=-.77$ ). In contrast, no between-trial type differences were found for the L2-only group (Post: critical word,  $Z=-1.160$ ,  $p=.246$ ,  $d=-.32$ ; whole sentence,  $Z=-.497$ ,  $p=.619$ ,  $d=-.14$ ; Delayed: critical word,  $Z=-1.112$ ,  $p=.266$ ,  $d=-.32$ ; whole sentence,  $Z=-1.349$ ,  $p=.177$ ,  $d=-.39$ ) or Control group (Post: critical word,  $Z=-.776$ ,  $p=.438$ ,  $d=-.24$ ; whole sentence,  $Z=-.310$ ,  $p=.756$ ,  $d=-.08$ ; Delayed: critical word,  $Z=-.155$ ,  $p=.877$ ,  $d=-.24$ ; whole sentence,  $Z=-.207$ ,  $p=.836$ ,  $d=-.06$ ).

[Table 7 here]

## Discussion

We examined the extent to which EI with task-essential, form-meaning practice influenced learners' online processing and offline interpretation of L2 French morphosyntax for past habituality and ongoingness. An L2-only group received L2 EI with task-essential, form-meaning practice. An L2+L1 group received the same, with additional L1 EI with task-essential, form-meaning practice.

### Summary of findings

Our expectation that L2 EI with practice would result in learning gains whereas the Control condition would not was partially supported. The Control group showed no significant improvement over time (and performed worse in the CMT-Listen Pre-Delayed, mismatched, habitual and ongoing). For the L2-only group, some improvements were found,

although these were mostly in offline tasks and limited: development up to Delayed Posttest was maintained only in CMT-Read (ongoing, matched); short-term gains made in CMT-Listen (habitual, matched) and SPR (ongoing, matched) were lost by Delayed Posttest; negligible-to-no gains were made in mismatched contexts in all tests. This is somewhat inconsistent with previous research, which has shown clear benefits for task-essential form-meaning mapping practice (e.g. Marsden, 2006; Marsden & Chen, 2011; VanPatten, 2002). This discrepancy may relate to the fact that the SPR online measure has not been used in previous form-meaning practice studies. It could be that the inter-clausal nature of the processing problem, and/or the specific cross-linguistic properties, moderated the benefits of this instructional treatment.

Our hunch that the L2+L1 EI and task-essential practice would result in development was supported, in all outcome measures at six weeks after the intervention. Learners' interpretation of IMP had improved at Delayed Posttest for both habitual and ongoing contexts, in matched and mismatched trials, and in reading and listening CMTs. They also increased the speed of distinguishing habitual and ongoing meanings of IMP between Pre-Post, maintained at Delayed, with larger and longer lasting effect sizes than the L2-only group. Additionally, the L2+L1 group demonstrated increased sensitivity to incorrect (mismatch) compared to correct (match) usage of IMP. At Pretest, all groups' processing speeds were similar in matched and mismatched trials, indicating a lack of online sensitivity. Only the L2+L1 group's processing in mismatched trials became significantly slower relative to matched trials at Post and Delayed Posttest. In contrast, there continued to be no between-trial differences for L2-only and Control. These results are in line with findings about sensitivity to the processing cost of anomalous aspectual distinctions (Roberts & Liszka, 2013) and other violations (Leung & Williams, 2014; Marinis et al., 2005; Tokowicz & Warren, 2010).

The likelihood that these observations were largely ascribable to the Pre-Post phase is increased by the fact that participants received no instruction at all Post-Delayed.

### **Learning mechanisms potentially at play**

Providing learners with EI about and practice in L1 form-meaning mappings may have helped establish, or strengthen, *conceptual* representations of habituality and ongoingness. In turn, this may have facilitated the strengthening, via L2 instruction, of L2 mappings for these concepts. By practising these (re)mappings, declarative knowledge of new processing routines may have been proceduralised and automatized, reflected in faster online processing (DeKeyser, 2015). That is, the L1 EI clarified concepts and form-meaning mappings; the L1 practice reinforced these concepts and mappings; the L2 EI and interspersed practice strengthened the mappings between these (now better represented) concepts and French forms. Our data largely supports evidence of difficulties created by cross-linguistic form-meaning mapping differences (Izquierdo & Collins, 2008; McManus, 2013, 2015).

### **Observation of online effects**

Our findings from SPR may be accounted for in various ways. The L2+L1 treatment could have reduced reliance on L1-based processing strategies for interpreting aspect – which do *not* require co-indexation with inter-clausal morphosyntax - and routinized L2 processing strategies that *do* require inter-clausal co-indexation. Indeed, the interweaving of the English and French practice items may have promoted some co-activation of French morphosyntax when, at test, English contexts were read. Thus, after reading the English, the L2+L1 group



were more ready to make a faster decision about the expected French morphosyntax. Clearer habituality and ongoing conceptual distinctions could have given anticipatory benefits for accessing the appropriate French in the stimuli, e.g. increasing the speed of processing inter-clausal forms required for habituality. These findings are broadly compatible with existing evidence about L1 effects during L2 processing (e.g. Sanoudaki & Thierry, 2014; Tolentino & Tokowicz, 2011). However, further research is required to examine whether a) this co-activation is unique to mixed language tests (our CMTs and SPR both included an L1 context and L2 stimulus) and b) current findings would hold for tests only presented in the L2. To this end, preliminary findings from spoken narrative tests (beyond the scope of the current article) suggest some similar patternings of results in tests that do not provide an L1 context, suggesting that L1-L2 co-activation is not entirely restricted to mixed language testing contexts. Further research should also examine whether comprehension and SPR tests that are only in the L2 would pattern similarly.

However, our observations of online effects contrast with findings from Andringa and Curic (2015) and Marsden et al. (2013). A number of reasons could explain this difference: our provision of L1 EI and practice; our fuller EI about the L2 (compared to Andringa & Curic's brief EI containing two examples, and Marsden et al.'s Yes/No feedback); our longer practice with its task-essential form-meaning mappings; our different outcome measures (SPR here, versus anticipatory eye movements in Andringa & Curic and cross-modal priming in Marsden et al.); finally, our different target features and languages (a more established, natural lexicon here, versus a novel artificial lexicon in Andringa & Curic and Marsden et al.).

In sum, our findings suggest that providing L2+L1 EI and task-essential form-meaning practice for a feature exhibiting complex L1-L2 differences resulted in L2 performance that

appeared to benefit from L1 knowledge, rather than being adversely affected by it.

Compatible with previous studies that explored teaching/knowledge about the L1 for L2 learning (Kupferberg, 1999; Horst et al., 2010; Spada et al., 2005), it seems likely that our L2+L1 treatment was beneficial (and more reliably so than L2-only) because of the specific nature of the learning problem i.e. L1-L2 form-meaning mapping differences. Our evidence therefore supports the points raised by the opening quotation (Henry, Culman and VanPatten, 2009): that the effectiveness of EI seems dependent on the nature of the EI, the target structure, and processing problem.

### **Limitations and future research**

First, we emphasise that our findings and accounts of learning are tentative, given our small sample sizes.

As described above, the English exemplars in the L2+L1 condition were in addition to the L2-only treatment, and this difference in exposure requires interpretation. We consider it highly unlikely that, for English speakers who already have 19 years of exposure to English as L1 speakers, the mere exposure *alone* to the English exemplars would affect L2 French online or offline performance. Furthermore, as described above, English habituality and ongoingness are expressed with entirely different morphology to French (played => *jouait* and *a joué*; was playing => *jouait*; used to play/would play => *jouait*). Thus, additional exposure alone to English is unlikely to change the processing of French. We suspect, therefore, that the combination of the EI about English and practice in interpreting the English and French resulted in stronger conceptualisations, and more accurate and faster (re)mapped interpretations of L2 French *IMP*. However, further investigation, including

replication, is required to isolate the role of the English EI from the English practice, to understand the contribution of each to learning.

This dataset clearly offers opportunities for further analyses that are beyond the scope of this paper, such as detailed statistical analyses on the effects of lexical verb type, and grammatical/viewpoint aspect, in line with research informed by the Aspect Hypothesis (Andersen & Shirai, 1994). We found tentative support for previous evidence that habituality in French seems more difficult than ongoingness for L1 English learners (Howard, 2005; McManus, 2013, 2015), as, descriptively, Pretest scores were slightly lower for habituality. However, we find no clear patterns of more improved performance in habituality versus ongoingness as a result of either treatment. That is, the effect sizes on our interpretation measures did not suggest more benefits for one meaning of IMP than the other, although further research is required.

This study has provided some evidence that EI about the L1 with interpretation practice had benefits on both off and online measures, at least for our feature with cross-linguistic form-meaning differences. However, further research is required, with larger sample sizes, to ascertain whether this type of L2+L1 instruction would be as beneficial for different language features, including syntactic phenomena with and without referential meaning, for different L1-L2 combinations, and for different L2 proficiency levels. In particular, further research should consider a) the influence of different types of EI and practice on processing over time and b) the relationship of the L1-L2 morphosyntax and the extent to which types of EI and practice interface with this, for on- and offline behaviour. It is also important to investigate how changes in on- and offline interpretation behaviour relate to the development of production.

Tables

Table 1. Frequency counts of French exemplars used in both treatments

	Session	Listening	Reading	Total
<i>IMP Ongoing/Interrupted</i>	1	24	24	48
	3	36	36	72
	4	36	36	72
<i>IMP Habitual</i>	2	24	24	48
	3	36	36	72
	4	36	36	72
TOTAL IMP		192	192	384
Contrasted with...				
<i>Présent</i>	1	24	24	48
<i>Présent</i>	2	24	24	48
<i>Passé Composé</i>	4	36	36	72
GRAND TOTALS		276	276	552

Table 2. Frequency counts of English exemplars used in L2+L1 treatment

	Session	Listening	Reading	Total
<i>Past Progressive (Ongoing)</i>	1	8	8	16
	3	12	12	24
	4	8	8	16
<i>Past Simple (Habitual)</i>	2	8	8	16
	3	12	12	24
	4	8	8	16
TOTAL ONGOING & HABITUAL		56	56	112
Contrasted with...				
<i>Present Progressive</i>	1	8	8	16
<i>Present Simple</i>	2	8	8	16
<i>Past Simple</i>	4	8	8	16
GRAND TOTALS		80	80	160

Table 3: Critical words analysed in the SPR

English context	(Mis)match of French stimuli	Critical disambiguating word	Example sentence (English gloss, not given to participants)
Habitual	Match	IMP	Il quittait la maison quand il <u>voyait</u> sa femme (‘He would be leaving the house when he saw his wife’)
	Mismatch	<i>Passé Composé</i>	Il quittait la maison quand il <u>a</u> vu <sup>12</sup> sa femme (‘He was leaving the house when he saw his wife’)
Ongoing	Match	<i>Passé Composé</i>	Il quittait la maison quand il <u>a</u> vu sa femme
	Mismatch	IMP	Il quittait la maison quand il <u>voyait</u> sa femme

Table 4. Descriptive results for CMTs (offline) and SPR (online) tasks in habitual contexts

Measure	L2+L1 group				L2-only group				Control group				
	CMT-Listen	CMT-Read	SPR		CMT-Listen	CMT-Read	SPR		CMT-Listen	CMT-Read	SPR		
			Critical word	Whole Sentence			Critical word	Whole Sentence			Critical word	Whole Sentence	
		Accuracy (max = 5)		RT (ms)		Accuracy (max = 5)		RT (ms)		Accuracy (max = 5)		RT (ms)	
<i>Matched contexts (IMP + IMP)</i>													
<i>Pretest</i>	M	3.76	4.04	613.89	24367.12	3.71	4.02	628.97	25949.35	3.79	4.03	629.28	23399.75
	(SD)	(.79)	(.76)	(328.87)	(10789.56)	(.73)	(.63)	(407.88)	(13774.35)	(1.05)	(.61)	(326.41)	(9280.53)
<i>Posttest</i>	M	4.16	4.33	307.29	15106.12	4.24	3.83	541.78	19570.94	3.65	3.75	552.83	18079.31
	(SD)	(1.12)	(.89)	(152.05)	(5971.47)	(.81)	(1.14)	(180.16)	(9101.35)	(1.15)	(1.03)	(282.57)	(7430.99)
<i>Delayed Posttest</i>	M	4.77	4.73	310.49	16181.76	4.24	4.31	496.10	18189.65	3.65	3.48	538.12	18274.38
	(SD)	(.33)	(.41)	(180.74)	(7016.38)	(1.27)	(.71)	(109.84)	(8337.93)	(1.14)	(1.04)	(304.92)	(5707.1)
<i>Mismatched contexts (IMP + PC<sup>#</sup>)</i>													
<i>Pretest</i>	M	3.51	3.08	634.31	22014.94	3.41	2.96	624.32	28353.24	3.35	2.98	625.09	23584.13
	(SD)	(1.07)	(.92)	(205.05)	(11488.66)	(.99)	(.87)	(323.38)	(15386.55)	(.99)	(.98)	(202.38)	(9858.51)
<i>Posttest</i>	M	3.86	3.55	483.39	16747.06	3.24	3.29	594.22	19118.1	3.10	3.10	587.69	19449.13
	(SD)	(.78)	(1.08)	(133.13)	(6633.28)	(.89)	(1.07)	(219.35)	(6659.55)	(.79)	(1.05)	(214.08)	(6697.09)
<i>Delayed Posttest</i>	M	4.53	4.33	434.94	19072.06	3.59	3.17	565.84	20389.24	2.63	3.08	570.09	18596.19
	(SD)	(.51)	(.55)	(146.79)	(10744.34)	(1.01)	(1.25)	(410.54)	(11315.54)	(.71)	(1.11)	(386.25)	(8818.7)

Note. CMT-Listen = Context-matching listening test, CMT-Read = Context-matching reading test, SPR = Self-paced reading test, M = Mean, SD = Standard deviation, Max=

Maximum, RT = Reaction time; <sup>#</sup>reaction times for processing PC are for auxiliary (patterns of results were the same for aux. + participle)

Table 5. Habitual contexts: Effect size (Cohen's *d*) comparisons with Control, and effect size changes with effects adjusted for baseline differences.

	CMT-Listen		CMT-Read		SPR critical word		SPR whole sentence	
	L2+L1 vs Control	L2-only vs Control	L2+L1 vs Control	L2-only vs Control	L2+L1 vs Control	L2-only vs Control	L2+L1 vs Control	L2-only vs Control
<b>MATCHED</b>								
Pretest	-.03	-.09	.01	-.02	-.04	.00	.09	.22
Posttest	.45	.59	.60	.07	-1.08	-.05	-.44	.18
Delayed Posttest	1.34	.49	1.60	.93	-.91	-.18	-.33	-.01
Pre-Post <i>d</i> change	.48	.68	.59	.09	-1.04	-.05	-.53	-.04
Pre-DPT <i>d</i> change	1.37	.58	1.59	.95	-.87	-.18	-.42	-.23
<b>MISMATCHED</b>								
Pretest	.16	.06	.11	-.02	.05	.00	-.15	.37
Posttest	.97	.17	.42	.18	-.59	.03	-.41	-.05
Delayed Posttest	3.09	1.09	1.44	.08	-.46	-.01	.05	.18
Pre-Post <i>d</i> change	.81	.11	.31	.20	-.64	.03	-.26	-.42
Pre-DPT <i>d</i> change	2.93	1.03	1.33	.10	-.51	-.01	.20	-.19

Table 6. Descriptive results for CMTs (offline) and SPR (online) tasks in ongoing contexts

Measure		L2+L1 group				L2-only group				Control group			
		CMT-Listen	CMT-Read	SPR		CMT-Listen	CMT-Read	SPR		CMT-Listen	CMT-Read	SPR	
				Critical word	Whole sentence			Critical word	Whole sentence			Critical word	Whole sentence
		Accuracy (max = 5)		RT (ms)		Accuracy (max = 5)		RT (ms)		Accuracy (max = 5)		RT (ms)	
<i>Matched contexts (IMP + IMP)</i>													
<i>Pretest</i>	M	3.84	3.61	584.31	23778.88	3.83	3.92	585.44	26045.76	3.81	3.88	586.28	24752.69
	(SD)	(.72)	(.88)	(322.42)	(8439.14)	(.74)	(.86)	(167.73)	(14662.91)	(.78)	(.76)	(212.52)	(9981.01)
<i>Posttest</i>	M	4.29	4.73	250.89	14870.82	4.33	4.67	433.65	23534.29	3.96	4.04	478.64	19825.31
	(SD)	(.93)	(.36)	(119.97)	(6006.4)	(.59)	(.58)	(139.61)	(11065.99)	(.73)	(.83)	(111.56)	(7362.70)
<i>Delayed Posttest</i>	M	4.79	4.84	264.89	15828.82	4.14	4.67	524.43	19317.18	3.65	4.00	481.42	20404.13
	(SD)	(.30)	(.27)	(193.55)	(6705.05)	(.94)	(.47)	(186.41)	(10552.78)	(.68)	(.74)	(232.61)	(5426.27)
<i>Mismatched contexts (IMP + PC<sup>#</sup>)</i>													
<i>Pretest</i>	M	2.82	2.98	617.71	22185.53	2.82	2.98	624.97	29038.94	2.81	3.12	612.70	23444.5
	(SD)	(.85)	(.75)	(283.78)	(11133.76)	(1.04)	(.77)	(266.52)	(18250.65)	(1.00)	(.75)	(227.24)	(9694.48)
<i>Posttest</i>	M	3.24	3.75	386.00	17748.06	3.29	3.17	577.29	21221.88	2.50	3.48	579.66	20720.5
	(SD)	(1.12)	(.90)	(158.41)	(6105.99)	(.95)	(1.17)	(197.09)	(11092.51)	(1.14)	(1.08)	(326.59)	(9997.92)
<i>Delayed Posttest</i>	M	3.98	4.12	392.96	19189.71	2.96	3.27	574.24	22502.94	2.04	3.38	643.48	19381.88
	(SD)	(.97)	(.81)	(156.11)	(10194.09)	(.89)	(1.09)	(261.43)	(13926.08)	(1.15)	(1.19)	(383.95)	(5286.05)

Note. CMT-Listen = Context-matching listening test, CMT-Read = Context-matching reading test, SPR = Self-paced reading test, M = Mean, SD = Standard deviation, Max = Maximum, RT = Reaction time



Table 7. Ongoing contexts: Effect size (Cohen's *d*) comparisons with Control, and effect size changes with effects adjusted for baseline differences.

	CMT-Listen		CMT-Read		SPR critical word		SPR whole sentence	
	L2+L1 vs Control	L2-only vs Control	L2+L1 vs Control	L2-only vs Control	L2+L1 vs Control	L2-only vs Control	L2+L1 vs Control	L2-only vs Control
<b>MATCHED</b>								
Pretest	.04	.03	-.33	.05	-.01	.00	-.11	-.10
Posttest	.39	.56	1.08	.88	-1.97	-.37	-.74	.39
Delayed Posttest	2.17	.59	1.51	1.08	-1.01	.20	-.75	-.13
<b>Pre-Post <i>d</i> change</b>	.35	.53	1.41	.83	-1.96	-.37	-.63	.49
<b>Pre-DPT <i>d</i> change</b>	2.13	.56	1.84	1.03	-1.00	.20	-.64	-.03
<b>MISMATCHED</b>								
Pretest	.01	.01	-.19	-.18	.02	.05	-.12	.38
Posttest	.65	.75	.27	-.28	-.75	-.01	-.36	.05
Delayed Posttest	1.82	.89	.73	-.09	-.85	-.21	-.02	.29
<b>Pre-Post <i>d</i> change</b>	.64	.74	.46	-.10	-.77	-.06	-.24	-.33
<b>Pre-DPT <i>d</i> change</b>	1.81	.88	.92	-.09	-.87	-.26	.10	-.09

## References

- Ammar, A., Lightbown, P., & Spada, N. (2010). Awareness of L1/L2 differences: does it matter? *Language Awareness, 19*(2), 129-146.
- Andersen, R., & Shirai, Y. (1994). Discourse motivations for some cognitive acquisition principles. *Studies in Second Language Acquisition, 16*(02), 133–156.
- Andringa, S., & Curcic, M. (2015). How explicit knowledge affects online L2 processing. Evidence from differential object marking acquisition. *Studies in Second Language Acquisition, 37*(3), 237-268.
- Benati, A. (2005). The effects of processing instruction, traditional instruction and meaning—output instruction on the acquisition of the English past simple tense. *Language Teaching Research, 9*(1), 67-93.
- Benati, A., Lee, J., & Laval, C. (2008). From processing instruction on the acquisition of French imparfait to secondary transfer-of-training effects on French subjunctive and to cumulative transfer-of-training effect with French causative constructions. In Benati, A., & Lee, J., (Eds.). *Grammar Acquisition and Processing Instruction: Secondary and Cumulative Effects* (pp 121-157). Bristol: Multilingual Matters.
- Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.
- De Swart, H. (1998). Aspect shift and coercion. *Natural Language & Linguistic Theory, 16*(2), 347-385.
- DeKeyser, R. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning, 62*(Suppl. 2), 189-200.

- DeKeyser, R. (2015). Skill acquisition theory. In VanPatten, B. & Williams, J. (Eds.), *Theories in Second Language Acquisition. An introduction* (pp. 94-112). London: Routledge.
- DeKeyser, R., & Prieto Botana, G. (2015). The effectiveness of processing instruction in L2 grammar acquisition: A narrative review. *Applied Linguistics*, 36(3), 290-305.
- Dracos, M. (2013). The effects of form-focused training and working memory on the L2 processing and learning of morphological cues. Unpublished Doctoral Dissertation. The Pennsylvania State University. University Park, PA.
- Ellis, N. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, 27(2), 164-194.
- Ellis, N. (2008). Temporal cognition and temporal language the first and second times around. *Language Learning*, 58(s1), 115-121.
- Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. New York: Routledge.
- Fernández, C. (2008). Reexamining the role of explicit information in processing instruction. *Studies in Second Language Acquisition*, 30(3), 277-305.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage.
- Hartsuiker, R., & Pickering, M. (2008). Language integration in bilingual sentence production. *Acta Psychologica*, 128(3), 479-489.

- Henry, N., Culman, H., & VanPatten, B. (2009). More on the effects of explicit information in instructed SLA. A partial replication and a response to Fernández (2008). *Studies in Second Language Acquisition*, 31(4), 559-575.
- Henry, N. (2015). Morphosyntactic processing, cue interaction, and the effects of instruction: An investigation of Processing Instruction and the acquisition of case markings in L2 German. Unpublished Doctoral Dissertation. The Pennsylvania State University. University Park, PA.
- Horst, M., White, J., & Bell, P. (2010). First and second language knowledge in the language classroom. *International Journal of Bilingualism*, 14(3), 331-349.
- Howard, M. (2005). Les contextes prototypiques et marquées de l'emploi de l'imparfait par l'apprenant du français langue étrangère. In Labeau, E., & Larrivée, P. (Eds.). *Nouveaux développements de l'imparfait* (pp 175-197). Amsterdam: Rodopi.
- Izquierdo, J., & Collins, L. (2008). The facilitative role of L1 influence in tense–aspect marking: A comparison of Hispanophone and Anglophone learners of French. *The Modern Language Journal*, 92(3), 350-368.
- Keating, G., & Jegerski, J. (2015). Experimental designs in sentence processing research. *Studies in Second Language Acquisition*, 37(1), 1-32.
- Kupferberg, I. (1999). The cognitive turn of contrastive analysis: Empirical evidence. *Language Awareness*, 8(3-4), 210-222.
- Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, 29(4), 694-716.

- Lee, J., Benati, A., Aguilar-Sánchez, J., & McNulty, E. (2007). Comparing three modes of delivering Processing Instruction on preterite/imperfect distinction and negative informal commands in Spanish. In Lee, J. F., & Benati, A. (Eds.). *Delivering processing instruction in classrooms and virtual contexts. Research and practice* (pp 73-98). London: Equinox.
- Leow, R. (2015). *Explicit learning in the L2 Classroom: A student-centered approach*. New York: Routledge.
- Leung, J., & Williams, J. (2014). Crosslinguistic differences in implicit language learning. *Studies in Second Language Acquisition*, 36(4), 733-755.
- L'Huillier, M. (1999). *Advanced French Grammar*. Cambridge: Cambridge University Press.
- Lonsdale, D., & Le Bras, Y. (2009). *A frequency dictionary of French: core vocabulary for learners*. New York: Routledge.
- Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. In Crookes, G., & Gass, S. (Eds.). *Tasks and Language Learning: Integrating Theory and Practice* (pp 123-167). Clevedon: Multilingual Matters.
- Mackey, A., & Gass, S. (2013). *Second language research: Methodology and design*. New York: Routledge.
- MacWhinney, B. (2005). The emergence of linguistic form in time. *Connection Science*, 17(3-4), 191-211.
- Marsden, E. (2006). Exploring input processing in the classroom: An experimental comparison of processing instruction and enriched input. *Language Learning*, 56(3), 507-566.

- Marsden, E. & Chen, H.-Y. (2011). The roles of structured input activities in Processing Instruction and the kinds of knowledge they promote. *Language Learning*, 61(4), 1058–1098.
- Marsden, E., Williams, J., & Lui, X. (2013). Learning novel morphology: The role of meaning and orientation of attention at initial exposure. *Studies in Second Language Acquisition*, 35(4), 1-36.
- McManus, K. (2013). Prototypical influence in second language acquisition: what now for the Aspect Hypothesis? *International Review of Applied Linguistics in Language Teaching*, 51(3), 299-322.
- McManus, K. (2015). L1-L2 differences in the acquisition of form-meaning pairings: A comparison of English and German learners of French. *Canadian Modern Language Review*, 71(2), 51-77.
- Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, 27(1), 53-78.
- Murakami, A. & Alexopoulou, T. (2016, in press). L1 influence on the acquisition order of English grammatical morphemes. *Studies in Second Language Acquisition*.
- Norris, J. (2015) Statistical significance testing in second language research: Basic problems and suggestions for reform. *Language Learning* 65, S1, 97-126.
- Norris, J. , Plonsky, L., Ross, S. , & Schoonen, R. (2015). Guidelines for reporting quantitative methods and results in primary research. *Language Learning*, 65(2), 470-476.
- O'Grady, W. (2005). *Syntactic carpentry: An emergentist approach to syntax*. New York: Routledge.

- O'Grady, W. (2015). Processing determinism. *Language Learning*, 65(1), 6-32.
- Phillips, C. & Ehrenhofer, L. (2015). The role of language processing in language acquisition. *Linguistic Approaches to Bilingualism*, 5(4), 409-453.
- Plonsky, L. (2015). Statistical power, *p* values, descriptive statistics, and effect sizes: A “back-to-basics” approach to advancing quantitative methods in L2 research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 23-45). New York: Routledge.
- Plonsky, L., & Oswald, F. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878-912.
- Roberts, L. (2016). Self-paced reading and L2 grammatical processing. In Mackey, A. & Marsden, E. (Eds.), *Advancing Methodology and Practice. The IRIS repository of instruments for research into second languages* (pp. 58-72). London: Routledge.
- Roberts, L., & Liszka, S. (2013). Processing tense/aspect-agreement violations on-line in the second language: A self-paced reading study with French and German L2 learners of English. *Second Language Research*, 29(4), 413-439.
- Runnqvist, E., Gollan, T., Costa, A., & Ferreira, V. (2013). A disadvantage in bilingual sentence production modulated by syntactic frequency and similarity across languages. *Cognition*, 129(2), 256–263.
- Sagarra, N., & Ellis, N. (2013). From seeing adverbs to seeing verbal morphology. *Studies in Second Language Acquisition*, 35(2), 261-290.
- Sanoudaki, E., & Thierry, G. (2014). Juggling two grammars. In Thomas, E.M. & Mennen, I. (Eds.), *Advances in the study of bilingualism* (pp. 214–230). Bristol: Multilingual Matters.

- Sanz, C., & Morgan-Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: a computer-assisted study. *Language Learning*, 54(1), 35-78.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 127-158.
- Schneider, W., Eschman, A., and Zuccolotto, A. (2012). *E-Prime Reference Guide*. Pittsburgh: Psychology Software Tools, Inc.
- Smith, C. (1997). *The parameter of aspect*. New York: Springer.
- Spada, N. & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, 60(2), 1–46.
- Spada, N., Lightbown, P., & White, J. (2005). The importance of form/meaning mappings in explicit form-focused instruction. In Housen, A., & Pierrard, M. (Eds.). *Investigations in instructed second language acquisition* (pp 199-234). Berlin: DeGruyter.
- Terrell, T. (1991). The role of grammar instruction in a communicative approach. *The Modern Language Journal*, 75(1), 52-63.
- Tokowicz, N., & Warren, T. (2010). Beginning adult L2 learners' sensitivity to morphosyntactic violations: A self-paced reading study. *European Journal of Cognitive Psychology*, 22(7), 1092-1106.
- Tolentino, L., & Tokowicz, N. (2011). Across languages, space, and time. *Studies in Second Language Acquisition*, 33(1), 91-125.
- Tolentino, L., & Tokowicz, N. (2014). Cross-language similarity modulates effectiveness of second language grammar instruction. *Language Learning*, 64, 279-309.



- VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Norwood, NJ: Ablex.
- VanPatten, B. (2002). Processing instruction: An update. *Language Learning*, 52(4), 755-803.
- VanPatten, B. (2015). Input processing in adult SLA. In VanPatten, B, & Williams, J. (Eds.). *Theories in second language acquisition: An introduction* (pp 113-134). New York: Routledge.
- VanPatten, B., & Rothman, J. (2015). What does current generative theory have to say about the explicit-implicit debate? In Rebuschat, P. (Ed.). *Implicit and explicit learning of languages* (pp 91-116). Amsterdam: John Benjamins.
- White, J., & Horst, M. (2012). Cognate awareness-raising in late childhood: teachable and useful. *Language Awareness*, 21(1-2), 181-196.

---

<sup>1</sup> An anonymous reviewer pointed out that two recent dissertations (Dracos, 2013; Henry, 2015) used online methods to test processing behaviour changes as a result of practice with some type of EI, but they did not manipulate the type of EI with practice on L2 learning and did not investigate L1 EI and practice as in the current study.

<sup>2</sup> Lee et al. primarily investigated the mode of delivery of PI materials (computer versus paper-based).

<sup>3</sup> A cumulative score was created from a vocabulary test and scores from two Pretest measures (CMT-Listen, CMT-Read). The scores were rank ordered and each set of three rank ordered participants (e.g. ranks 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>) were randomly assigned to conditions.

<sup>4</sup> The potential role of lexical semantics has been extensively researched (Andersen & Shirai, 1994) and would involve extensive analysis along a range of variables, including telic\*atelic, regularity\*irregularity, type/token frequency, lexical semantics (4 levels), and is therefore beyond the scope of the current analysis.

<sup>5</sup> Participants did not speak or write during the EI (nor practice).

<sup>6</sup> The same lexical verbs were used in each version, but in different contexts (e.g. he went to the shop vs. she went to the stadium)

<sup>7</sup> So that participants had not read the same stimuli twice in the same test phase.

<sup>8</sup> Judgements about match/mismatches were not requested because explicit judgements in online tasks can lead to unusually slow or careful reading as well as tapping into more explicit resources (see Keating & Jegerski, 2015).

<sup>9</sup> Although we additionally analysed spillover effects in the one, two and three words following the critical word, these are not reported because a) they did not change the patterning of the results b) the syntactic properties of the spillover words varied e.g. determiner, adverb, noun, making interpretation of processing speed difficult c) space constraints do not allow such extensive results to be presented and discussed.

<sup>10</sup> For the RTs for the PC, we took the measurement from the auxiliary (though patterns of results were the same for the entire PC and for the past participle).

<sup>11</sup> We are grateful to Steven Ross and Luke Plonsky for this advice. For parity with other studies, note that the same patterns of significance for these omnibus tests were found whether we used alpha 0.10 or 0.05.

<sup>12</sup> Whether the analysis was carried out on the auxiliary (a), the past participle (vu) or both combined (a + vu) did not make any difference to the pattern of findings. Thus, for parity with the IMP critical words, the raw data we present for the *Passé Composé* are for the auxiliary.