Neurolinguistics and second language teaching: A view from the crossroads

Stefano Rastelli

University of Greenwich, UK

Abstract

The topic of this article is the link between research on the neurocognition of the teaching–acquisition interface and research on second language teaching. This recent scientific enterprise investigates whether and how different aspects of second language instruction may change both the anatomy and the functioning of an adult learner's brain even in a short period of time. In this article, I analyse how neurolinguists have operationalized three aspects specifically related to second language teaching: (1) learners' proficiency; (2) the between-groups experimental design; (3) the implicit vs. explicit teaching dichotomy. I suggest that the degree of replicability of such neurolinguistics studies can be increased by adopting non-circular operational definitions. Such definitions should not be based on psycholinguistic or neurolinguistic metrics, but on standards that are commonly discussed in the literature on instructed second language acquisition, second language teaching, and assessment. Finally, I suggest that for future research neurolinguists should consider the advantages of welcoming on board more developmental linguists and teachers.

I Topic and motivation

The topic of this article is the link between research on the neurocognition of the teaching–acquisition interface and research on second language (L2) teaching. This recent area of enquiry investigates whether and how different aspects of second language instruction may change both the anatomy and the functioning of an adult learner's brain even in a short period of time. The motivation for this analysis is that increasing numbers of neurolinguists are interested in second language instruction and are dealing with

Corresponding author:
Stefano Rastelli, University of Greenwich, Old Royal Naval College, Park Row, Greenwich, London SE10 9LS, UK.
Email: S.Rastelli@gre.ac.uk

teaching-related variables with which both second language acquisition (SLA) experts and teachers have been familiar for many decades. On the other hand, increasing numbers of SLA experts and teachers are adopting models of L2 neurocognition to explain the findings from behavioural classroom-based experiments. In order to augment the quality of research, researchers need to exchange information. I hope that the current analysis will contribute to this exchange. The article is organized as follows. Section II provides the theoretical background that is necessary for the interpretation of the neurolinguistics studies under analysis. The topic of Section III is how neurolinguists use the variable 'proficiency'. Section IV reviews problems that arise from the between-groups experimental design. Section V deals with the operationalization of the dichotomous variable 'explicit vs. implicit teaching'. At the end of each section, I suggest a way to increase the degree of replicability of neurolinguistics studies that elaborate upon the teaching-related aspects and variables. Finally, a terminological clarification is needed. In this article, the term 'variable' – without further specification – refers exclusively to an independent, explanatory variable (or covariate), that is, the teaching-related factors that are operationalized in a study and are invoked to explain significant changes in behavioural and physiological data.[1]

II Background

Research on the neurocognition of the teaching–acquisition interface is a quite recent field. To my knowledge, the first explicit connection between brain data and second language teaching can be traced back to the work of Lee Osterhout and colleagues (McLaughlin et al., 2004; Osterhout et al., 2004, 2006, 2008) and to the work of Michael Ullman (Ullman, 2004, 2005). In those studies, L2 teaching was observed to correlate with a pattern of electrophysiological responses and also with changes in brain anatomy (e.g. the density of white matter). Such patterns and changes were taken as cues of environmental adaptation directly following classroom-based second language instruction. In subsequent studies, it was also observed that second language students with high levels of proficiency might show neural profiles (both electrophysiological and relative to brain-oxygenation level) that significantly overlap with those of native speakers, regardless of the age of acquisition (Nickels and Steinhauer, 2018; Nickels et al., 2013; Steinhauer, 2014; White et al., 2012). This pattern, which had already been found with learners of artificial languages (Friederici et al., 2002), led some researchers to conclude that L2 proficiency could be more important than age of acquisition and that the stronger version of the Critical Period Hypothesis should be discarded.

The theoretical background that links together neurolinguistics and L2 teaching research has become available during the 1990s with the Declarative Procedural (DP) model (for a review of the neurocognitive fundamentals of the DP model, see Ullman and Lovelett, 2018). Michael Ullman and Michel Paradis – in different ways – began to systematically use the well-known functional-anatomical distinction between procedural and declarative memory circuits to refer to the different levels at which the items of the second language can be learned, represented, and processed over time (Paradis, 2004, 2009; Ullman, 2004, 2005). In more recent years, the interest of some neurolinguists has gradually turned towards how second languages are taught in the classroom. Some

crucial teaching-related variables were thus operationalized in neuro-functional and neuro-anatomical research (explicit vs. implicit instruction, immersion-based vs. college-based instruction, spaced vs. massed presentation, feedback vs. absence of feedback, etc.) (see Ullman and Lovelett, 2018).

According to the DP model, the learning, representation, and storage of distinct aspects of both the first and the second language are connected to two different memory circuits in the brain. The declarative memory system is the relational memory of facts and events. It is an associative memory system that underlies aspects of the mental lexicon[2] and of chunk learning. It is subserved by medial temporal lobe regions (hippocampal regions, entorhinal and pararhinal cortices, parahippocampal cortex) and parietotemporal neocortical regions. The procedural memory, in contrast, supervises learned behaviours such as stimulus–response habits. It promotes the learning and control of cognitive and motor skills which involve automatic sequences and procedures and is assumed to underlie some important aspects of the mental grammar, such as the combinatorial rules of simple past formation in English (e.g. *walk* + *-ed* = *walked*). The procedural memory system is subserved by a network of different subcortical and neocortical brain structures: the basal ganglia (especially the caudate nucleus, putamen, globus pallidus) and the frontal cortex (especially the supplementary motor area and the posterior region of Broca's area).

In both Paradis' and Ullman's versions of the DP model, assumptions are made about language acquisition. Moreover, some predictions based on the DP model are extended to the effectiveness of L2 teaching (for a review see Morgan-Short and Ullman, 2011). In the DP model, it was posited that part of the language (L1 as well as L2) is memorized as a whole, and part of the language is computed by combinatorial rules. Namely, what is memorized is the mental lexicon (see footnote 2), and what is computed by rule is the mental grammar. Within the DP model framework, classroom activities and instruction were treated as variables for the first time in studies on event-related potential (ERP). Some of these ERP studies have found systematic changes in learners' brains. Electrophysiological responses to stimuli are attested even in classroom-based L2 learners at a range of proficiencies. This change, sometimes referred to as the 'biphasic pattern',[3] is characterized by a shift between N400 and P600 ERP components (for a short overview on ERP methodology, see Nickels and Steinhauer, 2018; Roberts et al., 2016 and this issue; Steinhauer, 2014). According to this pattern, during the early stages of the acquisition of verb morphology, L2 learners seem to be sensitive only to statistical rules (e.g. transition probabilities) that keep together whole-word sequences (chunks and formulas). Grammatical violations at this stage elicit mainly N400-like brainwaves which are associated with declarative memory circuits. At a certain point, the same learners may go beyond statistically based patterns in the input and may use productive, combinatorial rules to process the same phenomena. Grammatical violations at this stage elicit mainly P600-like brainwaves that are linked to procedural memory circuits (McLaughlin et al., 2010: 138–142; Osterhout et al., 2004, 2006).[4] To give an example of the N400–P600 shift, in Osterhout et al. (2008), the same wrong sequence in French (e.g. **Tu adorez*) after one month of instruction was found to elicit an N400-like effect in 14 L1 English, L2 French initial learners. In four months, the effect was replaced by a P600 component, which was even larger at the third session (after 80 hours of instruction) and at that point comparable to native controls. Osterhout et al. (2008) claim that

the N400–P600 biphasic pattern is a cue of a sudden change in the neural source of SLA of morphosyntax. McLaughlin et al. (2010: 142) also concluded that 'there are qualitative changes in the neurocognitive mechanism underlying language processing during the first year of instruction.' The first step of such qualitative and physiological change is that 'learners initially learned about words, but not rules' (Osterhout et al., 2004: 290; for a commentary about the N400–P600 shift, see Roberts et al., 2016 and this issue).

The existence of this N400–P600 biphasic pattern has so far been investigated for the following areas: verb inflection, adjective declension, gender and number agreement in the NP, verb–subject agreement in the VP, word order, and verb order in complement-clause constructions. Classroom teaching is a key factor in these studies because it is possible that the declarative knowledge of grammatical forms can eventually be sided by the procedural knowledge of the same forms through practice (Ullman, 2005: 160). This does not mean that declarative knowledge is transformed into procedural knowledge, but only that lexical forms and constructions stored in the declarative memory may provide a database from which grammatical rules 'can gradually and implicitly be abstracted by the procedural memory system' (Rastelli, 2014; Ullman, 2004: 247; Ullman and Lovelett, 2018). Classroom practice can in fact indirectly increase performance in procedural memory: 'in some cases explicit knowledge of the rules themselves may help guide processing, perhaps enhancing the procedural rule acquisition' (Ullman, 2004: 247).[5] There are of course also studies that question both the existence of the N400–P600 biphasic pattern and its relevance for L2 acquisition and teaching. Some of these studies point out that P600 effects have been observed in L2 learners of low proficiency processing violations of syntactic rules common to the L2 and the L1 (e.g. Tokowicz and MacWhinney, 2005). Others reveal that the P600 could not be found in very proficient learners, especially when L2 features were not realized in the L1 (e.g. Sabourin and Haverkort, 2003; Sabourin and Stowe, 2008). Finally, other studies show that both advanced learners and native speakers – when processing morphosyntactic violations – can be either N400- or P600-dominant, depending on factors such as gender, handedness, and kind of linguistic cues (Tanner, 2013; Tanner and van Hell, 2014; Tanner et al., 2013, 2014).

III Handling the variable 'L2 proficiency'

1 Defining the problem

Learners' proficiency is a crucial aspect, which must be dealt with by any research on the neurocognition of the teaching–acquisition interface. In neurolinguistics studies, proficiency is often utilized as a variable in order to explain changes in the brain following instruction. It can be generally agreed that the term proficiency refers to the extent to which L2 learners master the second language at definite points in time. Deep differences and also inconsistencies emerge as soon as one moves away from the generic definition and tries to operationalize proficiency by explicitly defining its indicators. Looking at the current standards of language teaching and assessment, learners' proficiency is often linked to learners' performance of communicative skills displayed in everyday-life scenarios. For instance, the American Council on the Teaching of Foreign Languages

(ACTFL) provides a detailed list of skills that define proficiency depending on learners' ability to present their own ideas and thoughts, to interact with people, and to understand others' ideas (although a distinction is made between 'performance' and 'proficiency').[6] Another example of a skill-based definition of proficiency can be found on the website of the US Department of State.[7] Here, 'professional proficiency' is defined in terms of skills across five stages. At the 'elementary proficiency' stage, learners are 'able to satisfy routine travel needs and minimum courtesy requirements' and 'able to read some personal and place names, street signs, office and shop designations, numbers and isolated words and phrases'. At the 'full professional proficiency stage', learners are able 'to use the language fluently and accurately on all levels pertinent to professional needs' and 'to read all styles and forms of the language pertinent to professional needs'.

Cognitive linguists have observed that skill-based proficiency scales (such as those proposed in the Common European Framework of Reference) are not supported by any proper theory of language. For instance, skill-based proficiency scales do not make the necessary distinction between function and meaning and do not propose how to evaluate those formal-functional elements of language that are not related to meaning and do not impact the effectiveness of communicative performance (Hulstijn, 2009). Moreover, these scales do not take into account the distinction between the components of language knowledge and the components of language processing (De Jong et al., 2012). Operationalizing language proficiency is a major issue in SLA studies as well. Hulstijn (2012) reviewed the way in which language proficiency was measured in a corpus of 140 empirical articles published in volumes 1–14 (1998–2011) of the journal *Bilingualism: Language and Cognition*. He found that in 55% of these articles proficiency was not measured via an objective test.

In this section, eight neurolinguistics studies will be analysed. In all these studies, L2 proficiency is compared to other variables, such as 'kind of instruction', 'learner's age', 'item complexity', and 'development' (typical or atypical). The purpose of this section is to investigate what is meant by proficiency in neurolinguistic studies. Handling the construct proficiency consistently should be seen as crucial in research on the neurocognition of the teaching–acquisition interface. In fact, the appearance of the P600 component in the N400–P600 biphasic pattern has been taken as a hallmark of native-like sentence processing. When learners have attained high levels of proficiency, they may show ERP responses that are indistinguishable from those of native speakers (Steinhauer et al., 2009: 30). SLA researchers and L2 teachers would expect evidence demonstrating that L2 learners showing the P600-like signatures are actually more proficient than learners showing only the N400 component. They would also expect that in those studies, it is clearly specified in what sense those learners showing the P600 are more proficient than the others. Evidence of this superior proficiency should not be neuro-functional or neuro-anatomical, otherwise the argumentation would be circular. Therefore, the crucial question is: Who is a proficient L2 learner and what can they do?

2 Analysis

Steinhauer et al. (2009) present a meta-analysis of 19 different ERP studies. Participants in those studies are divided into novice, very low, low to intermediate, intermediate,

intermediate to high/near-native like and very high/native-like proficiency. These stages are subsequently analysed in terms of their neurolinguistic indicators. For each stage of proficiency, the neurolinguistic profile of learners who are at that stage is described in detail. In the same meta-analysis, one could also expect a summary of how proficiency across these stages has been operationalized (in non-neurolinguistic terms). Unfortunately, this indication is missing. Some neurolinguistics studies measure proficiency by combining various tests with self-reported evaluations and questionnaires. In their classical study, Weber-Fox and Neville (1996) measured learners' proficiency via standardized tests of English grammar, self-reports, and accuracy of acceptability judgements. Steinhauer et al. (2006) examined late French and Chinese learners of English at two different proficiency levels (high vs. low). The high/low proficiency grouping was based on performance in a sentence completion test (cloze test), with 'high proficiency' requiring at least 90% correct completions, and on self-reported evaluations. White et al. (2012) utilized a cloze test of English proficiency and acceptability judgements. The cloze test consisted of a one-page passage with approximately every seventh word missing, 30 in total. Participants were required to read the text and fill in the missing words by selecting a word from among four multiple-choice options. The L2 proficiency measure that was eventually used for a correlation analysis (as 'behavioural performance') was the grammaticality judgement score and not the cloze test score, which served only to assess the baseline for evaluating progress over time. Moreover, in each session, participants self-rated their L2 abilities on a 7-point scale along the following six dimensions: listening, reading, pronunciation, fluency, vocabulary, and grammar. Finally, the participants completed language background questionnaires that provided information about their previous and current English experiences.

Other studies utilize different tests to define participants' proficiency. Using fMRI, Yusa et al. (2011) tested two groups (one instructed, one control) of adult Japanese learners of English. The two groups did not differ in terms of overall English proficiency, as estimated by the Test of English for International Communication (TOEIC). The instructed group acquired near-native command of the subject–verb inversion following negative adverbials after one month of instruction. Participants in the instruction group met twice a week for one month (eight classes in total), with one training session lasting an hour in addition to their regular classes, and were required to hand in assignments based on the training sessions. Instruction consisted of repeated practice of simplex negative inversion sentences (e.g. *I will never eat sushi* → *Never will I eat sushi*). The authors eventually ruled out baseline (before training) L2 proficiency as a cause of changes in brain signatures and declared instruction to be the cause. The test used to assess baseline proficiency – the TOEIC – comprises a listening comprehension section and a reading comprehension section. In the former, the test takers look at one photograph and then listen to four short declarative sentences, one after the other. They have to choose the sentence that best describes the photograph. In the other part, the test takers listen to a question and then listen to three possible responses. They have to choose the correct response.[8] Test takers are not involved in any form of interaction, nor are they evaluated on anything resembling a communicative, functional task. Much of the TOEIC test is similar to the sentence-picture matching tasks usually administered in psycholinguistic labs. The only difference is that reaction times (recorded by the computerized version of the test) do not enter the evaluation.

Tanner et al. (2014) determined the L2 English proficiency of 24 learners (L1 Spanish) by means of self-rating scores and a paper-and-pencil test consisting of 50 questions selected from the Michigan Examination for the Certificate of Proficiency in English (ECPE). According to the authors of this test,[9] the purpose of the ECPE is to certify advanced English language proficiency. In the writing session, test takers write an essay based on one of two topic choices. In the listening part, a short recorded conversation is accompanied by three printed statements. Test takers choose either the statement that conveys the same meaning as what was heard or the statement that is true based on the conversation. The grammar part presents an incomplete sentence which is followed by a choice of words or phrases to complete it. Only one choice is grammatically correct. Crucially, the speaking part of this test was not mentioned in the article and possibly was not considered by Tanner et al. (2014); see also below.

Some other neurolinguistics studies define proficiency by using psycholinguistic measures. Consonni et al. (2012) is a study that evaluates two groups of highly proficient Italian–Friulian bilinguals only differing in age of acquisition. The authors wanted to demonstrate that – when proficiency and exposure are kept constant – noun and verb production recruit the same neural network. To assess proficiency in the two languages, the authors used the Bilingual Aphasia Test (BAT) in Friulian and Italian. The BAT assesses language comprehension, naming, and metalinguistic abilities by means of: (1) pointing, (2) simple and complex commands, (3) verbal auditory discrimination, reading and listening comprehension for words and sentences, grammaticality judgements, and semantic acceptability tasks.[10] It is worth mentioning that what is meant by 'sentence comprehension' is actually an offline translation exercise (apparently without any time constraints). So a great deal of the examination has to do with offline translation. The examiner asks the participant to translate sentences from Italian to Friulian, and vice versa.[11] In their (2011) ERP study, Pakulak and Neville tested 36 L1 German learners of English (Pakulak and Neville, 2011). L2 proficiency was assessed through the Speaking/ Grammar section of the Test of Adolescent and Adult Language (TOAL-3). In this test, participants must exactly repeat sentences spoken by the examiner. The sentences increase in syntactic difficulty over time. The grammar part of the test requires that participants determine, out of three sentences presented aurally, which two sentences have similar meaning.[12] It must be stressed that this test is often used to help identify individuals who may have a language disorder and to help determine in what area of the brain the dysfunction is located.

Finally, Bowden et al. acknowledge that many ERP studies have varied widely in their choice of proficiency measures (Bowden et al., 2013: 2495). Their study was designed to avoid this methodological weakness. All 32 L2 Spanish learners in the study were therefore given a standardized proficiency test, the SOPI (Simulated Oral Proficiency Interview), which is based on the speaking proficiency guidelines of the American Council on the Teaching of Foreign Languages (ACTFL). According to these guidelines (National Standards in Foreign Language Education Project, 1999), the oral interview session of a standardized test should elicit learners' ability to engage in a conversation on a given topic. The administration procedure of the SOPI is as follows: Each student listens to five recorded questions. Each question is asked twice and is followed by an acoustic signal. After the signal, the student has 20 seconds to answer before another signal cuts the

answer off. Evaluation criteria are: (1) the student answers properly and fluently, and (2) the student uses all of the time at his or her disposal. If the answer is too short, the score drops. Questions students are expected to answer resemble the following:

- Which was your favourite course last year? Why did you like it?
- How would your ideal university be?
- How do you think your life will be in ten years?

In the SOPI test, there is no real interlocutor, since it is a monologue, and the evaluation criteria do not reflect any of the conversational features that a proficient L2 learner is expected to master (the turn-taking system, the partly unpredictable interactional moves by the interlocutor, the necessity of scaffolding, etc.).

In the eight neurolinguistics studies described above, very different measures of L2 proficiency have been adopted. In a couple of these studies, L2 competence was assessed by means of the same tools that are used for impaired speakers. Apparently, none of the examined studies tested communicative functions or language skills involved in real-world language uses and interactions. Even when the authors chose to utilize an oral interview – like in the case of Bowden et al. (2013) – the interview was only simulated and the interlocutors were not real. The discussed measures of proficiency are not consistent and do not reflect most of the criteria of second language competence as outlined by the Common European Framework of Reference (CEFR) or the American Council on the Teaching of Foreign Languages (ACTFL). It is a pity that Tanner et al. (2014) seem to have excluded the speaking part of the ECPE in their experiment. Although one cannot be sure that interaction-based proficiency measures are necessarily more revealing than other measures, this is the part of the test that could be more easily agreed upon as being useful (for assessing proficiency) by L2 teachers and SLA experts. In fact, in the oral section of the ECPE, test takers participate in a decision-making task in pairs. Each test taker is given descriptions of two different options. Test takers collaborate to decide on, present, and defend a single option. For example, in the paired format, each test taker might be given a description of two people who have applied for a particular job. The test takers decide which person should be offered the job. The skills that are elicited during this session are: asking and answering questions, orally summarizing information, providing suggestions or recommendations, explaining opinions and decisions, negotiating and justifying a decision.[13] It is an easy prediction that – had Tanner et al. (2014) included this section of the test – proficiency scores in their experiment would probably have been closer to the scores the same participants would have obtained if assessed by any official proficiency test worldwide. Establishing a correlation between the cognitive and the neurophysiological measures of proficiency on the one hand and the interaction-based measures of proficiency on the other hand would be important to link neurolinguistics and language teaching research.

To sum up: most neurolinguistics studies use proficiency as a variable. In many of these studies, it is debated whether proficiency rather than age of acquisition modulates L2 learners' native-like attainment. But what is meant by 'proficiency' in those studies often differs to such extent that the comparability of the results is seriously undermined. Finally, the fact that interaction-based proficiency is never taken into account in ERP

studies contributes to make ERP data less suitable and interpretable in both SLA and language teaching research.

3 Proposed solution: handling 'proficiency' as a relational notion

Steinhauer et al. (2009) and White et al. (2012) proposed that 'structure-specific proficiency' – rather than general proficiency – is a more appropriate indicator of the neurocognitive mechanism underlying grammar processing by L2 learners (see also Nickels and Steinhauer, 2018). These authors maintain that proficiency in neurolinguistic studies should be factorized only in relation to the morphosyntactic structure that is investigated. I suspect that the exclusive adoption of such restricted measures of proficiency will increase – rather than reduce – the gap between neurolinguistics and research on L2 teaching. One may agree with Steinhauer et al. (2009) that L2 proficiency could be treated as a relational notion in future neurolinguistics research, provided that the term 'relational' means that the different proficiency scores are based not only on the syntactic structure that is investigated, but also on learners' communicative performance. Using such heterogeneous relational measures as breakdown points to assess L2 proficiency (so that continuous neural measures can be regressed onto it) would mean that neurolinguists also need to refer explicitly to some aspects of language usage.

Research on language teaching and language assessment can provide some useful relational proficiency metrics based on communicative skills. For instance, the CEFR grid for assessing speaking[14] suggests three main domains of assessment. The first one is competence (linguistic, sociolinguistic, pragmatic, etc.), the second one is skills (e.g. being capable of speaking rather than writing or talking on the phone, etc.), and the last one is 'can do' statements (e.g. being capable of performing language functions such as expressing opinions, disagreeing, thanking someone, etc.).

The adoption of such a differentiated baseline for evaluating participants' L2 proficiency in neurolinguistics studies – together with structure-specific metrics – would greatly benefit the growing enterprise of the neurocognition of L2 teaching–acquisition interface, for at least three reasons. First, the comparison between the two kinds of L2 proficiency (morphosyntactic and interaction-based) could help paint a bigger and more realistic picture of a learners' competence. The neural correlates of second language communicative skills would be evaluated, not just learners' capacity of reasoning about the language (as reflected by acceptability/grammaticality judgements). Second, defining a skill-based grid to evaluate L2 proficiency would require a closer collaboration between SLA experts, L2 teachers, and neurolinguists. This would be a positive outcome in itself. Third, by adopting shared criteria, neurolinguistics classroom studies would be much more replicable than they are now.

If a common baseline for evaluating participants' L2 proficiency in neurolinguistics studies were eventually adopted, we could be in a position to evaluate – for instance – whether neurological markers of L2 acquisition (for example, the N400–P600 biphasic pattern) are likely to precede or follow (cause or are caused by) pragmatic, sociocultural or communicative markers of L2 acquisition over the course of the developmental path. For example, a learner's capacity of engaging in real, unplanned interactions with native speakers or their capacity of comprehending monodirectional, spoken discourse (without

interaction) are two different markers of pragmatic competence. One could propose the hypothesis that different neural correlates (neural markers) exist for each one of these pragmatic markers of acquisition. As to the lexis, a learner's capacity of using multi-word expressions or their capacity of comprehending and using single words are two different markers of lexical competence. One could offer the hypothesis that different neural correlates (neural markers) exist for each of these capacities. At the moment, the literature on L2 neurocognition is lacking in indications of such a cross-disciplinary nature.

IV Handling the between-groups experimental design

1 Defining the problem

Classroom studies in SLA typically use a between-groups design where different variables (such as 'kind of training') represent the factor that differentiates the experimental and the control group. In many of these studies, a particular subset of non-probability sampling procedures is used, which Dörnyei (2007: 98) refers to as 'convenience or opportunity sampling'. Participants of such kind are in fact those easily accessible in the researcher's own institution, typically students in a language classroom. Although students in real classrooms represent the most obvious instance of a convenience sample, it is common in SLA studies that – from the students attending a class – a number of participants can be further selected, depending on certain key characteristics that are related to the purpose of the investigation. Experience with the language, kind of exposure, sex, age, working memory and aptitude are typical additional grouping variables that re-shape the original convenience sample across different, more purposeful dimensions. The identification of the contribution of each variable is then left to the choice of a proper statistical method (and *post-hoc* tests).

In L2 neurocognitive studies using ERP and involving participants who learn languages in formal settings, the issues of sampling and experimental design are important because the crucial ERP components can only emerge from the comparison of large numbers of participants.[15] However, this technical constraint risks to eclipse individual (between- participant) differences which are relevant for SLA research and for L2 teaching research (see Faretta-Stutenberg and Morgan-Short, 2018). In fact, if neurolinguists want to observe the neurocognitive effects of training or of other dichotomic variables (as is often the case in studies on L2 teaching), then they must use groups of learners who undergo different forms of training or have different degrees of exposure to the target language, etc. On the other hand, if neurolinguists decide to use such a between-group design (whether classroom-based or more fine-tuned), between-participant variability over time cannot be easily observed. In other terms, on one hand all statistical models would require that the differences between groups (possibly due to teaching-related factors) are greater than the individual ERP differences within each group. On the other hand, individual differences in ERP profiles could be of much interest for the purposes of research on language teaching and assessment.

Another complication is that the crucial variables (e.g. 'proficiency scores' and 'group') are often nested. This means for instance that good learners and bad learners – as they get sorted by test scores (such as acceptability judgements) – can be distributed

unevenly across different sub-groups. As a consequence, the factor 'group' (which is decisive in training studies) becomes even more unreliable, which makes it difficult to single out the main factors (responsible for the changing brain signatures) from other intervening, moderating, or confounding variables.

2 Analysis

Osterhout et al. (2006) proposed a longitudinal paradigm in order to overcome the difficulty of handling individual differences in between-group designs when group comparisons originate from a convenience sample (such as real language classrooms). In this paradigm, neurocognitive measures of L2 proficiency are recorded over time from the same groups of participants, typically students enrolled in university language courses or in study-abroad courses. Longitudinal studies with such within-participant designs are meant to minimize between-participant variability because participants in these groups act as their own controls (Osterhout et al., 2006: 200–210).[16] Ideally, one could combine the between-group design and the longitudinal, within-participant design in order to observe whether neurocognitive changes over time are an artefact of the experimental conditions under scrutiny. But in order for these changes to be unequivocally linked to one factor rather than another, all conditions must be carefully controlled – not only between-groups and within-participant, but also between-participants factors. This limitation is important and inherent in neurolinguistic studies: to use ERP, you need groups, but to see whether teaching (rather than something else) affects ERPs, you need to group in different ways the individual profiles within each group. To my knowledge, neither Osterhout et al. (2006) nor anyone else so far have suggested how this problem can be fixed.

3 Proposed solution: Experimental design should allow a three-layered analysis

Morgan-Short and Ullman (2011: 292) observed that the adoption of longitudinal design is enough to neutralize between-participant variability. But another source of variability that must be controlled for is between- participants (with final -*s*) variability. This refers to what might differentiate participants within the same group, that is, all kinds of moderating variables (e.g. length of exposure, aptitude, motivation,[17] working memory, etc.) that shape in different ways the group and that may interact with the variables under observation (e.g. kind of training). Unlike between-participant variability, between-participants variability cannot be neutralized by longitudinal design because the same learner can react differently to different variables over time (see Faretta-Stutenberg and Morgan-Short, 2018).

In order to cope with this further complication, I propose to systematically apply a three-layered analysis. In longitudinal studies, the outcome of different statistical analyses that calculate the impact of different variables (within the same group) should be compared in a regression model. For instance, if two different groups of learners undergo two different modes of training (e.g. implicit vs. explicit) in a longitudinal design, the same measures can be taken at different moments in time. By doing so, the between-participant analysis integrates the between-groups analysis. This combination can be further enhanced

by adding a between-participants analysis, where the final *-s* means that different grouping factors are analysed within each group. In order to check to what extent learners' individual differences affect the results, one needs to go beyond the grouping factor 'mode of teaching'. This can be done by analysing both between-participants and between participant variability, for example, with linear mixed-effects modeling (Baayen et al., 2008) in order to verify if other factors rather than teaching explain a significant degree of the variance in the data in the same participant. For instance, if motivation explains a great deal of the between-participants variance, then it is possible that the importance of the teaching factor has to be scaled down or modulated. In sum, while keeping the longitudinal design combined with between-groups comparison would allow researchers to track within-participant neural changes over time, an additional between-participants comparison would help narrow the range of competing (or interacting) variables. Linear mixed-effects statistics would allow correlating the individual differences in ERP profiles with all variables (not just with the between-group differentiating factor/s). This would make researchers more confident that teaching is actually the main factor for brain changes.

To my knowledge, this three-layered design has not been adopted yet in L2 neurocognitive studies.[18] Tanner et al. (Tanner and van Hell, 2014; Tanner et al., 2014) for instance study individual differences in ERP profiles of L2 learners through regression-based statistics in order to capture the continuous nature of individual variation. They found that gender and handedness modulate the ERP responses, with female being more 'P600 dominant' than male. Unfortunately, the design of these studies is cross-sectional. White et al. (2012) utilize a nine-week longitudinal design together with between-groups comparisons (the differentiating factor being the L1) and also a post-hoc between-participants analysis. This latter analysis is carried out in order to determine whether the individuals' ability to discriminate grammatical forms from ungrammatical ones is reflected by ERP measures. In fact, a correlation is found between d-prime scores in acceptability judgement tasks and P600 amplitude measured at the end of the course. According to White et al. (2012), this demonstrates that L2 proficiency – and not the L1 – determines the presence and magnitude of the P600. White et al. (2012) is not a training study though and the between-group factor is only the L1. Moreover, in their study, L2 proficiency is taken to be a function of acceptability judgements, while individual differences that may affect individual ERP profiles are not investigated further.

Although the three-layered experimental design I am proposing would initially complicate the design of studies such as White et al. (2012) and similar, I think it is a necessary step if one wants to study how teaching-related factors affect learners' neural profiles. As a matter of fact, whether or not L2 learners are taught in groups, they always react to teaching as individuals. This should be accounted for properly in the experimental design.

V Handling the variable 'implicit vs. explicit' teaching

1 Defining the problem

In this section, I deal with how the dichotomous variable 'implicit' vs. 'explicit' teaching is operationalized in two seminal neurolinguistics studies. The word 'explicit' refers to

all situations where language teachers drive learner's attention on the functioning of the target-language via the explanation of grammar rules. The word 'implicit' in this article covers two different features at the same time. In fact, acquisition in the implicit condition must be both unconscious and incidental (Ellis, 2005; Rebuschat, 2013). 'Unconscious' relates to the fact that attentional processes are not involved in acquisition (meant as the internalization of linguistic knowledge). The term 'incidental' refers to the situation when learners' attention in the classroom has been totally diverted from the language and has been directed exclusively towards the accomplishment of communicative tasks. These two conditions ('unconscious' and 'incidental') can be dealt with separately and operationalized differently in a training experiment. As an instance, according to some definitions, teaching is implicit just because learners do not receive information concerning rules underlying the input (Hulstijn, 2005: 130). Of course language teachers can keep learners 'implicitly' focused on the language even when they are not explicitly teaching the rules of grammar. In this case, teaching may well be unconscious, but it is not incidental.

The issue of the effectiveness of explicit vs. implicit teaching is among the most quoted and debated in the SLA literature of the last 25 years at least. This time span can be further extended backwards if one considers that the debate on the utility of the teaching of grammar rules in language classrooms traces back to the 1970s (for a review, see Benati, 2015; VanPatten and Benati, 2015). Since then, the work of three generations of SLA experts has been revolving around the following basic question: Is the second language learned better (i.e. more quickly and stably) when learners focus their attention on language forms (as they are organized and presented in a syllabus), or when they just use the language and are prompted (by communicative tasks) to detect language regularities only incidentally, by relying on unconscious, implicit devices?

On the one hand, those who promote the effectiveness of implicit teaching (as opposed to explicit teaching) are interested in verifying whether learners who do not receive any explicit information are capable of detecting patterns of covariations between elements in the rule-governed stimuli they were previously exposed to. These researchers are also interested in knowing whether such learners can use the outcome of this inductive process productively for novel, not previously trained sentences (for a collection of studies, see Rebuschat, 2015). Proponents of implicit L2 teaching rely on the evidence that – in a steady-state, native language competence – being capable of using a rule productively does not equal being aware of its existence or functioning (this distinction echoes the distinction between 'epistemic awareness' and 'functional awareness' proposed by Reber, 1993). Proponents of explicit L2 teaching, on the other hand, stress the importance of noticing and of other attentional cognitive processes, especially in adult SLA. For instance, Ellis (1994) points out that learners who are taught explicitly are more inclined to utilize conscious operations of formulating and testing hypotheses in a search for structure. Proponents of explicit teaching can also rely on the favourable results of systematic meta-analyses conducted in the last 20 years. Such analyses would demonstrate that L2 instruction is more beneficial than implicit instruction, especially in the long term (Goo et al., 2015).

2 Analysis

Morgan-Short et al. (2010, 2012) are the first neurolinguistics studies that face this issue in an upfront, clear, and comprehensive manner, which makes them must-read articles for SLA experts and L2 teachers interested in neurolinguistics. In their seminal works, these authors use an artificial language learning paradigm to see how explicit and implicit training conditions, respectively, affect the behavioural and electrophysiological (ERP) measures of syntactic processing of word-order violations and inflectional morphology (noun–adjective and noun–article agreement). In the case of word-order violations, the results show that only learners in the implicit training conditions display a native-like pattern consisting of an anterior negativity, followed by a P600 that is accompanied by a late anterior negativity. In the case of agreement, the results are less telling and the differences between groups are subtler. At a low proficiency level (i.e. towards the beginning of training), the P600 component is absent. The only difference between groups is that the implicitly trained learners show an N400 in both agreement conditions, while the explicitly trained learners exhibit the N400 only in the noun–adjective agreement condition. Moreover, the N400 component in explicitly trained learners is delayed. However, at the end of training, explicit and implicit groups respond similarly. All learners show a P600 for noun–article agreement and an N400 for noun–adjective agreement. Morgan-Short et al. (2012) characterize explicit training as the one that approximates 'traditional grammar-focused classroom settings' and implicit training as the one that approximates immersion settings.

It is important to examine in detail how the explicit and implicit teaching conditions are differentiated. During the training phases, both groups (implicit and explicit) are provided with meaningful examples of an artificial language called BROCANTO 2. Each sentence describes a move in a chess-like game. For instance, in sentence (1) we find a first subject NP comprising the feminine noun *blom* followed by the feminine adjective *neimo* (*-o* is the feminine ending) and by the feminine article *lu*. A masculine object NP (*neep li*) and the uninflected verb *praz* follow.

(1) Blom   neimo   lu   neep   li   praz
     Blom-F   square-F   the-F   neep-M   the-M   switch
     'The square blom-piece switches with the neep-piece.'

The sentences are accompanied by the corresponding game constellations and moves. The implicit groups listen to 127 such sentences overall. The explicit groups listen to only 33 sentences, but they additionally receive aurally presented explicit metalinguistic information about how the agreement rules work in BROCANTO 2. A sample of the explicit instructions reads as follows:

> there is only one article in BROCANTO 2, which has two forms. *Li* is the masculine form, *lu* is the feminine form. You should remember two points. First, articles always come after nouns. Second, articles must agree with the gender of the nouns. In other words, if a noun is masculine, the masculine form of the article must be used. (Morgan-Short et al., 2012: 192)

After the training period, participants in both conditions practise both comprehension and production skills in BROCANTO 2 by playing the computer-based board game. To

practise comprehension, they listen to the description of the move and are asked to make the stated move. To practise production, they watch a move displayed on the computer screen and have to describe it with a single BROCANTO 2 sentence. In this practice phase, all participants are exposed to a huge number (880) of items of the language over the course of three sessions (training + practice). ERP and behavioural measures are recorded after session 1 and 3. I now focus on how the sentences are presented to the participants during the training phase. Below are the first four items in the sequence.

pleck li ('the pleck')

pleck troise li ('the round pleck')

pleck li ('the pleck')

pleck neime li ('the square pleck')

……

Since these aurally presented items are clearly patterned, the authors' intention could have been to drive participants to detect the regularities in the input and to abstract the target rule (gender agreement). This is the general idea behind the so-called 'variation sets' practice in language teaching and also behind statistical learning in general[19] (Onnis et al., 2008a, 2008b). Therefore, even though the presentation of the sentences was meant to elicit implicit knowledge, the result might have been to draw participants' attention to both the meaning of sentences (the game moves) and the grammatical regularities behind them. This methodological choice has relevant consequences. The way language input is presented in this study differs from the way language input is presented in SLA studies on implicit teaching. In the latter, the input is not structured to support the presentation of a grammar rule. Implicit teaching in SLA research is meant to focus on meaning, not on rules. Learners must in fact pay attention to language form only incidentally, to overcome form–meaning-related problems that have arisen during meaningful interactions.

Let us now analyse the role of the computer game for which both groups were trained. Many SLA experts and teachers would agree that this can be dubbed a real 'language task' which, as such, would perfectly suit a Task-Based Language Teaching (TBLT) pedagogy (Long, 2015). This pedagogy has been revolving around the elaboration of the construct 'implicit teaching' for thirty years. It is very likely that the chess-like game elicited implicit knowledge because the form of any message conveyed to participants (in both comprehension and production) was functional to meaning (the game move) and to a language-related performance (carrying out an instruction). Since both groups were massively exposed to this kind of implicit teaching, it is possible that the effects of explicit instructions were balanced or even overshadowed and that this could at least partially explain the mixed results in Morgan-Short et al. (2010).

3 Proposed solution: 'implicit teaching' should mean more than simply 'not teaching grammar'

The main problem for experiment designers is finding activities that operationalize implicit teaching properly. Implicit teaching is such only if it is both unconscious and

incidental (section 5.1). It is possible that the game activity in Morgan-Short et al. (2010, 2012) – even though it was not designed to do so – promoted genuine incidental learning of BROCANTO 2 because this activity was not focused on the language. Instead, the presentation of sentences – even if no grammar rule was explicitly presented – could not promote implicit learning because those sentences were highly structured. In fact, sentences consisted of 'variation sets', that is, repetitions of partially identical structures with little, clearly detectable variations across items. During this activity, learners' attention was not diverted away from the structure of the language. Yet, learners' efforts were aimed at detecting what changed and what remained identical across sentences. Since learners' attention was not directed to a communicative task, teaching in this experimental condition cannot be said incidental and teaching cannot be properly defined as 'implicit'.

It would be important to replicate studies such as Morgan-Short et al. (2010, 2012) while taking into account the equation 'implicit = incidental' (section 5.1) and while using a natural language as input rather than an artificial one. Giving up the experimental advantages of the latter[20] would be compensated by increased control over the delivery of the crucial variable according to what has been elaborated in SLA theory for at least two decades. The teaching would in fact be truly 'implicit' only if participants could engage in meaningful tasks that divert their attention away from the rules of the language, allowing incidental learning. For instance, participants could resort to the vocabulary, forms, and structures they can draw from statistically controlled (but not rule-motivated) samples of input that they are exposed to in the instructional phase of the experiment, but only for the purpose of accomplishing the communicative task in which they are engaged.

VI Shortcomings of the analysis and conclusions: More L2 teachers and developmental linguists should get on board

In this article, I reviewed some studies in a quite recent area of cross-disciplinary research, dubbed the neurocognition of the teaching–acquisition interface. I analysed how these studies utilize three teaching-related aspects: learners' proficiency; the between-groups experimental design, and implicit vs. explicit teaching. Eventually, I suggested how these aspects and variables should be handled in order to increase the degree of replicability of these studies. The current analysis has of course shortcomings. First, one can always find teaching-related variables that – due to their elusive nature – cannot be factored into experimental studies. Second, not very much is known yet about the causal relationship (not just about the correlation) between environmental factors (such as teaching) and patterns of brain adaptation in adulthood. Another limitation is that all studies quoted in this article deal with the teaching–acquisition of L2 morphosyntax, and nothing is said about the teaching–acquisition of the L2 lexicon. Finally, the majority of the evidence quoted in this article comes from ERP studies. This technique has great advantages, but also well-known inherent limitations (Luck and Kappenman, 2013).

There also exist important teaching-related variables that are not dealt with in this article, such as kind of exposure (immersion vs. college-based instruction), type of feedback (recasts vs. explicit feedback), or type of presentation (spaced vs. massed, see Ullman and Lovelett, 2018). I suggest that for the research to come, neurolinguists should consider the advantages of welcoming on board more developmental linguists, SLA experts, and

teachers. Although this might slightly slow down the workflow at the initial stages of the research process, the whole scientific community could benefit in the long term and the payback for the field of the neurocognition of teaching–acquisition interface would be guaranteed.

Notes

1. In Section IV4.3, the expression 'moderating variable' is used as opposed to simply 'variable'. Elsewhere in the text, it is always specified when the word 'variable' is used to refer to the outcome (dependent) variable.

2. In Ullman's version of the DP model, the word lexicon corresponds to what Paradis (2009: 15–19) calls vocabulary. Paradis (2004, 2009) in contrast defines the 'lexicon' as the complete, structural set of variously combining items forming a network of combinable stems and affixes. In more recent articles, Paradis seems to agree with Ullman that lexical items are comparable to vocabulary items defined as 'form–meaning pairs', which are stored in declarative memory (see above; Paradis, 2013). In this article, by 'lexicon' I mean the closed set of form–meaning pairs that are stored as unanalysed wholes in the declarative memory system, as in Ullman's version of the DP model (see Ullman and Lovelett, 2016 and this issue).

3. An anonymous reviewer observed that – in the literature on first and second language acquisition – the term 'biphasic pattern' is used very frequently to refer to the presence of both a LAN and a P600 in response to syntactic violations. In fact, some SLA studies also questioned whether or not only high-proficiency learners can elicit a biphasic LAN/P600 pattern that is statistically indistinguishable from that of native speakers (e.g. Steinhauer et al., 2006). However, in many other studies, the expression 'biphasic pattern' is explicitly used to refer to the N400–P600 shift that occurs when learners' proficiency increases (e.g. Erdocia et al., 2014: 818; McLaughlin et al., 2010: 128; Meulman et al., 2014: 1; Mueller et al., 2008: 78; Tanner, 2013: 353; Tanner et al., 2013: 368; Roberts et al., 2016 and this issue). This terminological coincidence – although unfortunate – is not confounding, provided authors always specify which one of the two (biphasic) patterns they refer to.

4. An anonymous reviewer pointed out that – in the neurolinguistics literature – the presence of the P600 component is also linked to a post-processing revision stage and to wrap-up effects. For instance, Steinhauer and Connolly (2008) express the common view that the P600 is very likely comprised of a number of distinct sub-components, each of which may reflect different kinds of cognitive processes; some of these may, moreover, be recruited across different cognitive domains (e.g. in music, mathematics, etc.). However, in all the studies quoted in this section, the P600 component – whatever its underpinnings and domains of application – can be also taken as a neurophysiological signature of L2 development, given that in most studies it can be found at advanced stages of second language acquisition and not at the early stages.

5. According to Michel Paradis, however, declarative knowledge cannot be transformed into procedural knowledge (Paradis, 2004).

6. Performance is the ability to use language that has been learned and practised in an instructional setting. Proficiency is the ability to use language in real-world situations in a spontaneous interaction and non-rehearsed context, and in a manner acceptable and appropriate to native speakers of the language; see: https://www.actfl.org/publications/guidelines-and-manuals/actfl-performance-descriptors-language-learners (accessed November 2016).

7. https://careers.state.gov/gateway/lang_prof_def.html (accessed November 2016).

8. Examples can be found here: http://www.examenglish.com/TOEIC/TOEIC_listening_part2.htm (accessed November 2016).

9. For examples go to: http://www.cambridgemichigan.org/institutions/products-services/tests/proficiency-certification/ecpe (accessed November 2016).

10. For examples go to: https://www.mcgill.ca/linguistics/files/linguistics/Friulian-Italian.pdf (accessed November 2016).

11. Example sentence: 'the maid is always at the window'.

12. Examples can be found here: https://prezi.com/bxrwd3_shdap/toal-3 (accessed November 2016).

13. An inventory of such language functions together with a set of instructions on how they can be used for the purpose of language examination is available from the website of the CEFR at: http://www.coe.int/t/dg4/linguistic/Manuel1_EN.asp (accessed November 2016).

14. http://www.coe.int/t/dg4/linguistic/Manuel1_EN.asp#Contents (accessed November 2016).

15. In ERP studies, learners' individual differences are eclipsed by the grand-averaging procedure. This procedure is necessary in order to increase the signal-to-noise ratio that allows to uncover meaningful patterns in microvolt potentials (Luck and Kappenman, 2013).

16. Davidson (2010) adds the 'learning-based experimental approach' as a sub-case of longitudinal studies.

17. See Dörnyei, 2007.

18. As a reviewer pointed out, the reason could be that this design is unrealistically difficult to implement and can be pursued only in the context of a big funded longitudinal study.

19. Variation sets are partially alignable segments of sentences, that is, repetitions of partially identical structures within a short array of texts.

20. On the one hand it is difficult to generalize findings from laboratory or artificial-language studies to real-life L2 classroom settings (see criticism in Robinson, 2005). On the other hand, dealing with natural languages makes it difficult – if not almost impossible – to control for a number of moderating factors (see discussion in Ettlinger et al., 2015).

References

Baayen RH, Davidson DJ, and Bated DM (2008) Mixed-effects modeling with crossed random effects fro subjects and items. *Journal of Memory and Language* 59: 390–412. doi: 10.1016/j.jml.2007.12.005.

Benati AG (2015) *Key methodological frameworks in second language research*. London: Equinox.

Bowden HW, Steinhauer K, Sanz C, and Ullman MT (2013) Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia* 51: 2492–2511.

Consonni M, Cafiero R, Marin D et al. (2012) Neural convergence for language comprehension and grammatical class production in highly proficient bilinguals is independent of age of acquisition. *Cortex* 49: 1252–58.

Davidson D (2010) Short-term grammatical plasticity in adult language learners. *Language Learning* 60: 109–22.

De Jong NH, Steinel MP, Florijin AF, Schoonen R, and Hulstijn JH (2012) Facets of Speaking Proficiency. *Studies in Second Language Acquisition* 34: 5–34.

Dörnyei Z (2007) *Research methods in applied linguistics*. Oxford: Oxford University Press.

Ellis NC (1994) (ed.) *Implicit and explicit learning of languages*. San Diego, CA: Academic Press.

Ellis NC (2005) At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition* 27: 305–52.

Erdocia K, Zawiszewski I, and Laka I (2014) Word order processing in a second language: From VO to OV. *Journal of Psycholinguistic Research* 43: 815–37.

Ettlinger M, Morgan-Short K, Faretta-Stutenberg M, and Wong PCM (2015) The relationship between artificial and second language learning. *Cognitive Science* 40: 822–47.

Faretta-Stutenberg M and Morgan-Short K (2018) The interplay of individual differences and context of learning in behavioral and neurocognitive second language development. *Second Language Research* 34(1): 67–101. This article appears in the current special issue of *Second Language Research*.

Friederici AD, Steinhauer K, and Pfeifer E (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* 99: 529–34.

Goo J, Grañena G, Yilmaz Y, and Novella M (2015) Implicit and explicit instruction in L2 learning: Norris and Ortega (2000) revisited and updated. In: Rebuschat P (ed) *Implicit and explicit learning of languages*. Amsterdam: John Benjamins, pp. 443–83.

Hulstijn JH (2005) Theoretical and empirical issues in the study of implicit and explicit second-language learning. Introduction. *Studies in Second Language Acquisition* 27: 129–40.

Hulstijn JH (2009) The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal* 91: 663–67.

Hulstijn JH (2012) The construct of language proficiency in the study of bilingualism from a cognitive perspective. *Bilingualism: Language and Cognition* 15: 422–33.

Long MH (2015) *Second language acquisition and task-based language teaching*. Oxford: Wiley-Blackwell.

Luck S and Kappenman ES (eds) (2013) *The Oxford handbook of event-related potential components*. Oxford and New York: Oxford University Press.

McLaughlin J, Osterhout L, and Kim A (2004) Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience* 7: 703–04.

McLaughlin J, Tanner D, Pitkänen I et al. (2010) Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning* 60: 123–50.

Meulman N, Stowe LA, Sprenger S A, Bresser M, and Schmid MS (2014) An ERP study on L2 syntax processing: When do learners fail? *Frontiers in Psychology* 5: 1072.

Morgan-Short K and Ullman MT (2011) The neurocognition of second language. In: Mackey A and Gass S (eds) *The Routledge handbook of second language acquisition*. London: Routledge, pp. 282–99.

Morgan-Short K, Sanz C, Steinhauer K, and Ullman MT (2010) Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning* 60: 154–93.

Morgan-Short K, Steinhauer K, Sanz C, and Ullman MT (2012) Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience* 24: 933–47.

Mueller J, Girgsdies S, and Friederici AD (2008) The impact of semantic-free second-language training on ERPs during case processing. *Neuroscience Letters* 443: 77–81.

National Standards in Foreign Language Education Project (1999) *Standards for foreign language learning in the 21st century*. Lawrence, KS: Allen Press.

Nickels S and Steinhauer K (2018) Prosody–syntax integration in a second language: Contrasting event-related potentials from German and Chinese learners of English using linear mixed effects model. *Second Language Research* 34(1):9–37. This article appears in the current special issue of *Second Language Research*. This article appears in the current special issue of *Second Language Research*.

Nickels S, Opitz B, and Steinhauer K (2013) ERPs show that classroom-instructed late second language learners rely on the same prosodic cues in syntactic parsing as native speakers. *Neuroscience Letters* 557: 107–11.

Onnis L, Waterfall HR, and Edelman S (2008a) Learn locally, act globally: Learning language from variation set cues. *Cognition* 109: 423–30.

Onnis L, Waterfall HR, and Edelman S (2008b) Variation sets facilitate artificial language learning. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*. Available at http://kybele.psych.cornell.edu/~edelman/OnnisWaterfallEdelman-variation-setsCogSci08.pdf.

Osterhout L, McLaughlin J, Kim A, Greenwald R, and Inoue K (2004) Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In: Carreiras M and Clifton Jr C (eds) *The on-line study of sentence comprehension*. New York: Psychology Press, pp. 271–308.

Osterhout L, McLaughlin J, Pitkänen I, Frenck-Mestre C, and Molinaro N (2006) Novice learners, longitudinal designs and event-related potentials: A means for exploring the neurocognition of second language processing. In: Indefrey P and Gullberg M (eds) *The cognitive neuroscience of second language acquisition*. Malden, MA and Oxford: Blackwell, pp. 199–230.

Osterhout L, Poliakov A, Inoue K et al. (2008) Second-language learning and changes in the brain. *Journal of Neurolinguistics* 21: 509–21.

Pakulak E and Neville HJ (2011) Maturational constraints on the recruitment of early processes for syntactic processing. *Journal of Cognitive Neuroscience* 23: 2752–65.

Paradis M (2004) *A neurolinguistic theory of bilingualism*. Amsterdam and Philadelphia, PA: John Benjamins.

Paradis M (2009) *Declarative and procedural determinants of second languages*. Amsterdam and Philadelphia, PA: John Benjamins.

Paradis M (2013) Late-L2 increased reliance on L1 neurocognitive substrates: A comment on Babcock, Stowe, Maloof, Brovetto and Ullman (2012). *Bilingualism: Language and Cognition* 16: 704–07.

Rastelli S (2014) *Discontinuity in second language acquisition: The switch between statistical and grammatical learning*. Bristol: Multilingual Matters.

Reber AS (1993) *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford: Oxford University Press.

Rebuschat P (2013) Measuring implicit and explicit knowledge in second language research. *Language Learning* 63: 595–626.

Rebuschat P (ed.) (2015) *Implicit and explicit learning of languages. Studies in Bilingualism, volume 48*. Amsterdam: John Benjamins.

Roberts L, Alonso JG, Pliatsikas C, and Rothman J (2016) Evidence from neurolinguistic methodologies: Can it actually inform linguistic/language acquisition theories and translate to evidence-based applications? *Second Language Research* 34(1):125–143. This article appears in the current special issue of *Second Language Research*.

Robinson P (2005) Cognitive abilities, chunk-strength and frequency effects in implicit artificial grammar and incidental L2 learning. *Studies in Second Language Acquisition* 27: 235–68.

Sabourin L and Haverkort M (2003) Neural substrates of representation and processing of a second language. In: van Hout R, Hulk A, Kuiken F, and Towell R (eds) *The lexicon–syntax*

interface in second language acquisition. Amsterdam and Philadelphia, PA: John Benjamins, pp. 175–95.

Sabourin L and Stowe LA (2008) Second language processing: When are first and second languages processed similarly? *Second Language Research* 24: 397–430.

Steinhauer K (2014) Event-related potentials (ERPs) in second language research: A brief introduction to the technique, a selected review, and an invitation to reconsider critical periods in L2. *Applied Linguistics* 35: 393–417.

Steinhauer K and Connolly JF (2008) Event-related potentials in the study of language. In: Stemmer B and Whitaker HA (eds) *Handbook of the neuroscience of language*. New York: Elsevier, pp. 91–104.

Steinhauer K, White EJ, Cornell S, Genesee F, and White L (2006) The neural dynamics of second language acquisition: Evidence from event-related potentials. *Journal of Cognitive Neuroscience (supplement)* 99.

Steinhauer K, White EJ, and Drury JE (2009) Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research* 24: 13–41.

Tanner D (2013) Individual differences and streams of processing. *Linguistic Approaches to Bilingualism* 3: 350–56.

Tanner D, McLaughlin J, Herschensohn J, and Osterhout L (2013) Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition* 16: 367–82.

Tanner D and van Hell JG (2014) ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia* 56: 289–301.

Tanner D, Inoue K, and Osterhout L (2014) Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition* 17: 277–93.

Tokowicz N and MacWhinney B (2005) Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potentials investigation. *Studies in Second Language Acquisition* 27: 173–204.

Ullman MT (2004) Contributions of memory circuits to language: The declarative/procedural model. *Cognition* 92: 231–70.

Ullman MT (2005) A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In: Sanz C (ed) *Mind and context in second language acquisition*. Washington, DC: Georgetown University Press, pp. 141–78.

Ullman MT and Lovelett J (2018) Implications of the declarative/procedural model for improving second language learning: The role of memory enhancements techniques. *Second Language Research* 34(1):39–65. This article appears in the current special issue of *Second Language Research*.

VanPatten B and Benati AG (2015) *Second language acquisition: Key terms*. 2nd edition. London: Continuum.

Weber-Fox CM and Neville HJ (1996) Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience* 8: 231–56.

White EJ, Genesee F, and Steinhauer K (2012) Brain responses before and after intensive second language learning: Proficiency based changes and first language background effects in adult learners. *PLoS ONE* 7(12): e52318.

Yusa N, Koizumi M, Kim J et al. (2011) Second language instinct and instruction effects: Nature and nurture in second-language acquisition. *Journal of Cognitive Neuroscience* 23: 2716–30.