Starting from the very beginning: Unraveling Third Language (L3) Development with Longitudinal Data from Artificial Language Learning and Related Epistemology

Jorge González Alonso[1,2] *, Pablo Bernabeu[2], Gabriella Silva[1], Vincent DeLuca[2], Claudia Poch[1], Iva Ivanova[3] & Jason Rothman[2,1,4]

1. Nebrija University | 2. UiT The Arctic University of Norway | 3. University of Texas at El Paso | 4. Lancaster University

* Corresponding author: jorge.gonzalez.alonso@uit.no | jgonzalezal@nebrija.es

Abstract

The burgeoning field of third language (L3) acquisition has increasingly focused on intermediate stages of language development, aiming to establish the groundwork for comprehensive models of L3 learning that encompass the entire developmental sequence. This article underscores the importance of a robust epistemological foundation, advocating for incremental knowledge building through longitudinal research. In the study presented here, we use artificial languages to investigate L3 acquisition from initial exposure with complete input control, factoring in individual differences in executive functions and history of bilingual exposure/engagement to assess the role of these variables in shaping learning trajectories and modulating cross-linguistic influence (CLI). This approach not only advances our understanding of L3 development under controlled conditions, but also links L3 acquisition research to broader cognitive science inquiries.

INTRODUCTION

Third language acquisition and linguistic transfer

The dedicated, independent study of how adults acquire languages beyond a second one has been justified on many grounds: social, educational, cognitive, theoretical, epistemological, methodological and otherwise (see Cabrelli et al., 2023; González Alonso, 2023; Rothman, González Alonso & Puig-Mayenco, 2019, for a summary of arguments). From an acquisition/processing perspective alone, empirical work asking and testing whether all non-native acquisition can/should be treated as if the same—i.e., is additive third or more language

(L3/L*n*) distinct from additive second language (L2) learning—decisively shows they are not. Within acquisition/processing approaches, the most compelling arguments to date underscoring the need to treat additive multilingualism separately revolve around the observation that L3/L*n* acquisition has fundamentally different initial conditions to those of both first (L1) and L2 acquisition, which also differ significantly from one another. In particular, L3/L*n* learners have, by definition, previous linguistic knowledge/representations organized in *more than one* grammatical system as well as previous experience resolving dual language competition at the very onset of L3 learning (Rothman et al., 2019).

Research in L2 acquisition (SLA) has long considered (e.g., Weinreich, 1953), and by the 1980s established, that prior linguistic knowledge plays a significantly deterministic role in L2 development, and from the very onset of L2 interlanguage (see, e.g., Kellerman, 1977; Odlin, 1989; Gass, 1984, 1988; Schwartz & Sprouse, 1996; Treffers-Daller & Sakel, 2012). The structure of the other language an L2 learner already speaks matters for L2 development, and does so systematically. Obviously, in the case of L2 acquisition, specific language influence/transfer can only come from the native L1—the only specific language an L2 learner has experience with. Is the same true for L3/L*n,* or will they have recourse to transfer from their L2 or perhaps both/all their previously acquired languages? Seminal studies around the turn of the century examining L3 acquisition specifically (e.g., Flynn, Foley & Vinnitskaya, 2004; Leung, 1998, 2001, 2002) showed that the L1 was not the sole language of potential influence in L3/L*n* interlanguage development. From this launching pad, it was clear that bespoke theoretical models needed to be developed to account for prior language influence when more than one previously acquired grammar was at play.[1]

Insofar as there is no privileged status for L1 transfer in L3/L*n* and yet it is clear that transfer effects apply, the fact that L1 and L2 representations will often not align in their configurations for any given grammatical property—for example, the L1 might have phonologically null subjects, while the L2 does not—makes it non-trivial to determine where the L3/L*n* grammar starts from (the L3/L*n* initial state, or stages). This is so even under theoretical approaches that assume very early, full transfer of the grammatical representations of a single previous language into the L3/L*n*. In L3 acquisition, having stronger (or sole) representational transfer from the L1 vs. the L2 would make all the difference in estimating the developmental path for any given linguistic property. It would also be important in comparing

[1] These studies are not strictly speaking the seminal papers on the acquisition of an L3 grammar (see, e.g., Klein, 1995; Mägiste, 1979, 1986), but they were the first ones to study this process systematically, from which one can trace the provenance of the burgeoning field of L3/L*n* from a cognitive perspective that exists today.

how complex that developmental path would be for native speakers of the same L1 who are acquiring the same target language, but as an L2 instead of an L3. Let us illustrate this point with an example.

Rothman and Cabrelli Amaro (2010) showed that not all English—non-null subject language—speakers acquiring non-native French—also a non-null subject language—are the same. Their study revealed that those acquiring French as an L2 showed little issues from the very outset with the obligatory nature of overt syntactic subjects in French. This was in sharp contrast to their L3 counterparts, who had previously acquired Spanish—a null-subject language—as an L2, whose initial French interlanguage grammars displayed all the hallmarks of having transferred (Spanish) null-subjects. This was the case despite both groups being under the exact same conditions for learning French. Rothman and Cabrelli Amaro also reported the same L2/L3 pattern (L1 English transfer for L2 and L2 Spanish transfer for the L3) for two additional groups who were acquiring Italian—a null-subject language—as the L3 target. These patterns mean that the learning tasks are distinct across the groups: French L2 learners need no reconfiguration, French L3 learners need to somehow unlearn the transferred null subjects and acquire anew the obligatory nature of overt syntactic subject arguments. The L2 Italian learners must do the same as the L3 French group whereas the L3 Italian learners need no reconfiguration. From this data set alone, it cannot be determined if the results favor a privileged status for L2 transfer (Bardel & Falk, 2007) or if there is another factor at play, namely a deterministic role for typological/structural similarity in multilingual transfer selection (Rothman, 2010, 2011, 2015; Westergaard, Mitrofanova, Mykhaylyk & Rodina, 2017). Suffice it to say for the immediate expository aim, what is clear from this early study is that it documents how the learning task of L2 and L3 can be significantly distinct even when an L2 and an L3 learner share the same native L1.

What is at stake in having one or the other language transfer grammatical knowledge, then, is the degree of facilitation and/or, indeed, the degree of complication for the subsequent (developmental) learning tasks within the L3 acquisition process. Given the relevance of this phenomenon, a (very) significant part of the theoretical and empirical work on L3 acquisition/processing since the early 2000s has been dedicated to investigating when, from where (L1, L2 or both?), how, how much and what role other interacting factors play in conditioning previous language influence in L3/L*n* acquisition. Several theoretical models have been developed (see González Alonso, 2023, for an updated overview), and while there are non-trivial divergence points between them, a collective reading of the theoretical and empirical L3/L*n* literature reveals some areas of consensus. So far, we are more or less in agreement about the following:

1) Without knowing something about the nature of the implicated languages in the pairings (and the context of L3 acquisition), the task of predicting and accounting for patterns of apparent L1 or L2 transfer is nearly impossible (e.g., Rothman et al., 2019; Puig-Mayenco, González Alonso & Rothman, 2020).[2]
2) Throughout the process of L3 acquisition, both languages are likely to influence development, although their influence might not be equally strong at all stages—if at all (e.g., Cabrelli & Iverson, 2023; Westergaard, 2021).

Until around 2015, most studies focused on—or otherwise attempted to make claims about—**early** L3 acquisition. This was partly due to the SLA-inherited concern about the initial conditions (*initial state*, in the specialized SLA literature) of the non-native acquisition process. In fact, one of the theories developed from the early 2010s, the Typological Primacy Model (TPM; Rothman, 2010, 2011, 2015), has sought explicitly (in fact, exclusively) to model the first instances of independent L3/L*n* interlanguage grammars. The TPM argues that the initial L3/L*n* stages are the locus of significant theoretical, empirical and practical importance. The model advocates for full (complete) transfer—in short, a copy of an existing underlying grammar—argued to take place as early as possible after first exposure to the L3/L*n* and where selection between the L1 or the L2 obtains on the basis of the parser's assessment of overall typological/structural overlap to the L3 (see Rothman, 2015; Rothman et al., 2019 for the implicational hierarchy of cues argued to guide the parser).

Other L3 models, perhaps most notably the Cumulative Enhancement Model (CEM; Flynn et al., 2004), the Linguistic Proximity Model (LPM; Westergaard et al., 2017; Westergaard, 2021) and the Scalpel Model (SM; Slabakova, 2017), disagree with the idea that the early stages of L3/L*n* acquisition should be given any special consideration. As these models argue that cross-linguistic influence effects are primarily the result of dual activation of previously acquired languages and assume this to be a constant throughout development, they postulate no urgent need to focus on the initial stages of L3/L*n* acquisition when investigating cross-linguistic effects. Partially as a consequence of these models, a theoretical

---

[2] This does *not* mean that being an L1 or an L2 is inconsequential (e.g., Cabrelli & Iverson, 2023), only that, while relevant, L1 or L2 status—as it has been called—is not a priori deterministic without consideration of many other conspiring variables.

landscape has been carved out for work on L3 acquisition focusing on groups of learners with intermediate and advanced proficiencies both with and without any (empirical) attempt to chart, or directly consider, progress from early L3/L*n* development—i.e., without data from these learners at the earliest stages. Following Rothman et al.'s (2019) argumentation that the best approach to L3/L*n* development would be to model it in full consideration of its initial stages conditions, L3 developmental models whose predictions for L3 development are predicated on the basis of what has been previously documented in initial stages transfer studies are beginning to emerge (e.g., Cabrelli & Iverson, 2023; Cabrelli, Iverson, Giancaspro & Gonzalez, 2020; Puig-Mayenco, 2023; Puig-Mayenco, Tubau & Rothman, 2022).

Tracking development from the beginning

Arguments for undertaking L3/L*n* developmental research that is anchored in a solid understanding/documenting of the initial stages are methodological and epistemological in nature, oriented toward scientific progress for the field as a whole. Methodologically, the early stages of L3/L*n* acquisition constitute the only real overlap in scope between all models that otherwise focus on somewhat different mechanisms. If a model of cross-linguistic influence in L3/L*n* acquisition does not ascribe any particular significance to the initial stages (e.g., the CEM, the LPM or the Scalpel Model), because the mechanisms posited there are assumed to operate in similar or identical ways throughout the whole developmental sequence, then testing learners at the beginning where their predictions still apply and where other models (e.g., the TPM, or the Abbreviated Grappling Period Model; Sprouse & Schwartz, 2023) make explicit predictions, should be the best way to adjudicate between these competing accounts. Epistemologically, starting at the very beginning is on the whole a safer bet for best practice. This is so since it is particularly difficult, if not impossible, to pinpoint the provenance of grammatical representations beyond a certain point in (L3/L*n*) development. In other words, one cannot distinguish (targetlike) transferred representations from those acquired without mediation of previously acquired languages—or even from more complex situations where the L3 grammar has landed on target after initial non-facilitative transfer from one previous language and reconfiguration with or without help from the other.

This last epistemological argument connects with a methodological, practical issue in non-native language acquisition research: precisely because it is difficult to tell how a learner has arrived at a targetlike representation, *errors* are the most informative pieces of information for the researcher interested in understanding language development, at all instances (Rothman et al., 2019): it is non-targetlike language that provides a window into the hidden mechanisms of language acquisition, especially when we are focusing on tracking sources of cross-linguistic influence. Errors, however, become less and less frequent the further down the developmental sequence one looks. If we want to determine the sources of transfer in L3/L*n* morphosyntax, we are likely to have more material to work with at early stages as opposed to later ones.

Motivated by the idea of characterizing the initial stages of L3/L*n* acquisition as accurately as possible from the very first moments of exposure, a number of studies have employed semiartificial languages to train groups of bilingual learners on a completely novel L3 (e.g., González Alonso et al., 2020; Mitrofanova, Leivada & Westergaard, 2023; Pereira Soares, Kupisch & Rothman, 2022; Sanz, Park & Lado, 2015; see Grey, Williams & Rebuschat, 2014, for a precursor; see also Grey, 2023, for a primer on the use of artificial languages in multilingual acquisition/processing research). Employing artificial languages as third languages provides four methodological improvements that are crucial to an examination of early L3/L*n* acquisition with significantly less noise than one would encounter in "natural" conditions. The first is total control over exposure: the language is guaranteed to be novel for participants (since it was created *ex novo* for the study), which means that the experimenter knows exactly the quantity and quality of input the learners have received. The method also guarantees that learners cannot be exposed to the language outside the study, keeping all participants equal in terms of input received. Second, these artificial languages are custom-made to have the optimal lexical, phonological/phonotactic (or orthographical/orthotactic) and morphosyntactic properties to test the predictions of any theoretical model(s) guiding our inquiry. While it is often hard to find language combinations where the L1/L2 match and mismatch properties of the L3 grammar in perfectly unambiguous ways, leading to distinct predictions and unconfounded results, this is possible when designing languages at will. Finally, the use of L3 artificial languages with bilingual groups solves a recurring problem of L3/L*n* research from its inception: the small sample sizes researchers have available when looking for homogeneous groups of L3 learners of a particular language combination. Small samples compromise statistical power, and therefore the generalizability of the results. If our task becomes finding homogeneous groups of *bilinguals*, rather than multilinguals, our chances to put together much larger participant numbers are immediately improved. Semiartificial languages, which use known lexicon (from one of the languages of the participants) with novel morphosyntax (typically from another language, either with its own morphological exponents or with novel morphology altogether), provide an extra advantage over purely novel artificial languages: they allow learners to overcome the lexico-semantic bottleneck (e.g., Hopp, 2014) and start processing the grammar almost immediately. Finally, using artificial grammars allows more easily than in the “wild” for systematic and controlled tracking of development, from initial exposure, introducing much needed longitudinal paradigms for theory testing.

Bespoke paradigms for L3 acquisition: EEG and artificial languages

The first paradigm to test L3/L*n* theories was presented in Rothman, Alemán Bañón and González Alonso (2015). This epistemological paper introduced an event-related potential (ERP) methodology to gauge implicit responses to grammatical violations in a novel L3. Since EEG/ERP had been shown to allow for a qualitative distinction of the processes underlying otherwise similar linguistic behavior in L2 learners (e.g., Alemán Bañón, Fiorentino & Gabriele, 2014; Gabriele, Fiorentino, Alemán Bañón & Rossomondo, 2013; McLaughlin et al., 2010), Rothman et al.’s rationale was that a similar method, applied to the early stages of L3/L*n* acquisition, could offer a rubric to distinguish between learner performance stemming from transferred linguistic representations (which should then yield ERP indices of processing similar to those of native grammatical processing, such as the P600 component) and processing based on mechanisms not readily related to syntactic processing (e.g., rhyming or pattern matching, which should manifest in ERP indices such as the N400 component).

González Alonso et al. (2020) was the first study to use this method. Participants were highly proficient L2 speakers of English living in the UK, all with Spanish as their L1. Two miniature artificial languages (ALs), based on the Spanish or English lexicon, were developed. Both ALs shared a nominal morphological agreement paradigm similar to Spanish. Participants, matched for proficiency and other demographic factors, were randomly assigned to one of the two groups and trained in one of the ALs through implicit exposure to the language’s agreement properties for approximately 45 minutes. Following this training, participants underwent an ERP experiment, reading different types of sentences in their assigned AL (grammatical, gender violations, and number violations) while their electrophysiological activity was recorded. Additionally, behavioral responses were collected through a grammatical judgment task presented at the end of each sentence. Rothman et al. (2015) had outlined the predictions for the neural correlates of processing that each of the dominant models at the time should have expected in each condition. These predictions involved biphasic (N400)-P600 or P600 components, depending on the condition and the language combination—i.e., the AL. The behavioral data showed no significant effects in accuracy or response time, and the ERP findings did not confirm any of the original predictions for any of the models. González Alonso et al. (2020) observed that the short exposure time might have been insufficient to allow the learners' linguistic parser to show any linguistic specific transfer effects.[3] Results, however, were not completely null. A P300-like effect was observed as a response to gender violations only in the group exposed to the Spanish-based AL. This was interpreted as evidence of an early allocation of attentional resources (the P300 component is a marker of attentional resource allocation; Datta et al., 2007; Herrmann & Knight, 2001; Patel & Azzam, 2005) to help parse the incoming input. The presence of this finding solely in the Mini-Spanish subgroup suggested that the difference might be due to the biases produced by the Spanish- and English-based lexicons, respectively, in what were otherwise identical grammars. The authors concluded that models of L3/L*n* transfer focusing on structural similarity as the main driver of transfer selection (e.g., the TPM, the LPM and the Scalpel model) were in a better position to explain these results, even though what the data showed was, at best, the precursors for transfer.

In a similar vein, Pereira Soares et al. (2022) aimed to test models of transfer at the initial stages of L3/L*n* acquisition through behavioral and EEG/ERP measures. Improving on the methods adopted in González Alonso et al. (2020), they introduced several modifications. First, they examined two language domains (as opposed to the single domain in González Alonso et al.) within a mini language paradigm based on Latin (mini-Latin), thereby enhancing the ecological validity of the findings. Second, although the study was not a longitudinal one and, therefore, incremental exposure to the L3 input was not possible, a consolidation period was implemented between two sessions of lexical learning. Third, two different bilingual groups were compared: heritage speakers—with Italian as a heritage language and German as the societal language—and L2 learners—German native speakers who were L2 learners of English. This comparison aimed to disentangle possible effects of multilingual transfer/acquisition and age of acquisition, since heritage speakers have two native, naturally acquired languages (Rothman & Treffers-Daller, 2014), whereas the L2ers have a marked distinction in age of acquisition, proficiency, language status and learning context between their languages. The grammatical focus in Pereira Soares et al. (2022) was on case morphology and adjective-noun word order. Like Latin, German has case morphology, while adjective position is predominantly postnominal in Latin and Italian but prenominal in German and English. In the initial learning session, participants were first exposed to a set of mini-Latin words that

[3] Incidentally, a similar limitation may have applied in another (behavioral) L3 study employing artificial language learning, Mitrofanova et al. (2023). In this study, participants were exposed to just 10 exemplars of the artificial language before moving on to the testing phase.

would be used in the experimentation phase. After a consolidation period of two to nine days, they were tested again on their mini-Latin lexical knowledge. They were then implicitly exposed to the grammatical rules through sentence reading. Finally, participants performed a grammaticality judgment task while continuous EEG was recorded.

The authors predicted two scenarios: either a replication of the P300 component found in González Alonso et al. (2020), indicating that the (still) short exposure period only allowed the data to capture an early attentional focus on grammatical elements, or a true detection of transfer effects through the usual components, as discussed in Rothman et al. (2015) and above. Results showed a negative N200/N400 deflection for heritage speakers when processing case morphology, as well as a P600-like effect for the German L2 group when processing adjective-noun word order violations. Pereira Soares et al. (2022) interpreted the early negativity to indicate increased detection among heritage speakers of novelty/mismatch patterns in mini-Latin, likely due to their extensive bilingual experience. The P600 effect in the German L2 group, where transfer was not possible given that neither German nor English have this pattern, might indicate that participants had learned the mini-language patterns and instantiated repair mechanisms, as evidenced by the late positivity. As in González Alonso et al. (2020), none of the current L3/L*n* models predicted these results.

The unexpected findings from these two studies offer some insights for the future of this experimental paradigm. The first is that, in order to achieve its original aim of informing models of transfer in L3/L*n* morphosyntax, exposure to the language might have to be substantially increased, implementing the training longitudinally and incrementally. The second is that what this method is able to capture only moments after first exposure might actually shed light on the mechanisms whereby transfer/CLI takes place, how a source is selected, how previous linguistic knowledge is actually repurposed, etc. How is similarity between the L3 and the L1/L2 assessed? What underlies parsing of the incoming L3 input? What basic mechanisms guide the learner in facing the task of processing grammar in a novel language? How do basic cognitive processes, such as attention, working memory or inhibition, become involved in these cross-evaluations of similarity and adequacy of mental linguistic representations across different sets (different grammars)? All of these questions, which are essential to a comprehensive, fully operationalized theory of transfer (in L3/L*n*, but also in L2), can be addressed with data stemming from some version of the current paradigm.

Besides shortcomings in the design of the language training sequence, what studies like González Alonso et al. (2020) and Pereira Soares et al. (2022) are lacking, especially considering what we have just discussed, is systematic attention to interindividual variability within their learner groups. In an age where individual differences have taken center stage in essentially all subareas of bilingualism research (e.g., Dörnyei, 2014; DeLuca et al., 2019; Lauro et al., 2020; Luk & Rothman, 2022; Paradis, 2023; Skehan, 1991), the disregard for individual variation in L3 studies is increasingly glaring. Individual differences—in aptitude, in cognitive function, in linguistic background, in exposure to the target language, etc.—are nearly impossible to control for, much less factor in, without careful tracking of both the developmental sequence of learners and their relative exposure to the language. However, there are few (if any) studies that follow L3 learners longitudinally with consideration of these individual differences. As we have seen, while some studies in recent years have sought to increase control over L3/L*n* exposure by employing learning paradigms based on short exposure to (semi)artificial languages (e.g., González Alonso et al., 2020; Pereira Soares et al., 2022; see also Mitrofanova et al., 2023; Sanz et al., 2015, for studies on L3 learning using mini-languages with behavioral testing paradigms), they have rarely followed the progress of these learners beyond a single session or two, and they have never modelled the role of individual differences in both linguistic background (language history) and domain-general cognitive functioning. In the remainder of this article, we present a specific methodology designed to deal with the aformentioned shortcomings. In doing so, the goal is to highlight the multifarious advantages whereby studies of similar scope can improve the empirical lanscape of L3/L*n* acquisition/processing while building bridges between related fields of study for which controlled L3/L*n* studies can serve as resource for theory building and testing (e.g. neurocognition of multilingualism, the psychology of learning and memory, and more).

METHODS

In an effort to address the gaps in methodological approaches to L3 development, we present the full methodology of a large, multisite study currently in progress. While building on previous work, it constitutes an ambitious attempt at simulating a significant portion of L3 development in the lab, using multiple grammatical properties and factoring in individual differences in linguistic background, language dominance, L2 proficiency and cognitive function (attention, memory, inhibitory control), as well as investigating relative consolidation of the acquired knowledge over time. All materials, scripts, instructions, and any other necessary elements for replication of this study are freely available at https://osf.io/974k8/.

Overall structure

The study is carried out in two sites: Madrid (Spain) and Tromsø (Norway). In each of these locations we find one group, Spanish-English bilinguals in Madrid and Norwegian-English bilinguals in Tromsø. A majority of our participants are late L2 learners (i.e., non-child L2), around upper intermediate proficiency in English, although we allow for some variation in these respects given that individual differences in linguistic background are also modelled (see below).

Participants in each of these learner groups are randomly assigned to one of two semiartificial languages (hence ALs, for "artificial language"), and will complete all sessions of learning and testing in that language. The three languages developed for the study are called mini-English (common to both sites), mini-Spanish and mini-Norwegian. Participants in Madrid are assigned to mini-English or mini-Spanish; participants in Norway, to mini-English or mini-Norwegian. Each of these languages is lexically based on English, Spanish or Norwegian, respectively, and instantiates through novel morphology three properties that are subject to testing across the study's multiple sessions: noun-adjective gender (and number) agreement, differential object marking (hence DOM), and object-verb number agreement.

This combination of mini-languages and participant groups is particularly interesting for several reasons, which relate to how these properties align across the languages and what, therefore, they can tell us about early parsing and transfer of grammatical representations (see details in *Artificial language training and testing*, below). In short, gender agreement is present in at least one previous language in all groups, although instantiated in different ways; DOM is only available to those participants who speak Spanish, and verb-object number agreement is a novel form (novel context) of agreement for all groups involved, although the abstract property itself is present in all three background languages.

In a first session, conducted online, learners complete a language history questionnaire (the LHQ 3.0) as well as a battery of computerized behavioral tests to measure individual differences in cognition. After that, participants come to the lab and begin their training in the AL they have been assigned to. The first lab-based session, Session 2, implicitly trains learners on noun-adjective gender and number agreement. After training, participants complete a behavioral sentence-picture matching task to demonstrate command of the property. If they show above-chance performance, they move on to an event-related potential (ERP) experiment where they are exposed to violations of this grammatical property, as well as correct sentences. Resting-state EEG measures are also acquired at the beginning of the session. This procedure is repeated in two more sessions, where new grammatical properties are progressively introduced. Session 3 introduces DOM. Session 4, object-verb agreement. In these sessions, the fillers from Session 2 are progressively replaced by instances of previously encountered grammatical properties (gender and DOM in Session 4, for example), so that participants continue to receive exposure to these properties. Overall, this provides three instances of first exposure to a grammatical property (in each session), a long(er) longitudinal sequence (for gender, across all three sessions), and a shorter one (for DOM, across two sessions).

After Session 4, a fifth online session retests participants with the same battery conducted in Session 1, which gives an estimate of the potential change in cognitive function as a result of intensive language training. Four months after session 4, participants return to the lab for a final session (Session 6). The procedure in this session is the same as in Session 4, containing all three properties, but the learners receive no refreshment exposure before taking the ERP experiment. This attempt at measuring the relative consolidation of the learned properties is based on Morgan-Short, Finger, Grey and Ullman (2012), as well as suggestions in González Alonso et al. (2020) and Pereira Soares et al. (2022), and seeks to determine the stability of the knowledge the learners have developed during the training sessions.

Participants

All participants are neurotypical adults with normal or corrected-to-normal vision. On each site, two main groups are established according to which AL participants are randomly assigned to: Mini-English or Mini-Spanish. To address the problematic issue of experimental mortality that ails all longitudinal studies, participants are paid for each session, and receive a bonus payment if they finish all sessions. Assuming medium effect sizes, and considering the possibility of participant attrition inherent to any longitudinal study, target sample size per group in the first experimental session is n=50.

Materials

Language history questionnaire

We employ the Language History Questionnaire (LHQ3; Li, Zhang, Yu & Zhao, 2020), a computerized, multilingual, comprehensive linguistic background questionnaire that gathers information about the languages of a multilingual, their proficiency, their relative use, and the contexts in which these are used, among other variables. The questionnaire includes preprogrammed factor analyses that estimate four aggregated scores on different dimensions for each language: proficiency, dominance, immersion, and multilingual language diversity.

These aggregated scores are used as predictors of the different dependent variables collected throughout the study.

Cognitive tasks and resting-state EEG

Cognitive factors have received substantial attention in L2 acquisition research (e.g., Chen, Li, Wang & Wang, 2022), but much less in L3/L*n* acquisition (e.g., Huang, Loerts & Steinkrauss, 2022). While several factors have been studied, working memory has been highlighted as the most relevant contributor to L2 learning (e.g., Chen et al., 2022; Reichle, Tremblay & Coughlin, 2016). Additionally, inhibitory control has been linked to language-switching ability and resistance to interference from non-target languages (e.g., Prior & Gollan, 2011), which becomes increasingly relevant in multilingual scenarios. Finally, an aptitude for implicit learning has been found to correlate to the acquisition of morphosyntactic features in a second language (e.g., Frost, Siegelman, Narkiss & Afek, 2013; Godfroid & Kim, 2021; Williams & Rebuschat, 2016).

In the current study, we evaluate these three cognitive factors at the onset and offset of L3/L*n* acquisition using an artificial language paradigm, contributing to cross-validation of the role of executive functions in multilingualism (e.g., Jylkkä, Laine & Lehtonen, 2021). We also aim to evaluate the longitudinal stability of each of these cognitive factors, as well as the participants' resting-state EEG patterns, by associating the predictive power of each cognitive factor in the performance of the participant in the learning sessions to the pre-post change of each cognitive factor between Sessions 1 and 5. To measure resting-stage EEG, electrical activity is recorded before the start of the learning session, in the absence of any task, for five minutes with the participant's eyes closed, and five minutes with their eyes open. Activity across these two epochs is then averaged.

The cognitive tasks employed are the Stroop task (Stroop, 1935), the (forward) Digit Span Task (DGS), and the Alternating Serial Reaction Time (ASRT) task (e.g., Sævland & Norman, 2016). The Stroop task, where participants are asked to name the color a word is written in while inhibiting production of the actual word—which may be congruent or incongruent with the name of the color—has been identified as a good measure of executive control, cognitive flexibility, and conflict resolution (e.g., Bialystok, 2009), because it involves a conflict between automatic and controlled processes of language processing. Furthermore, several authors have pointed out that the task engages selective attention (e.g., Wright, 2017), an important skill in the early stages of language processing. To measure working memory capacity, participants take a classic version of the DGS, where they are asked to repeat increasingly long, random sequences of digits. Simplifying slightly, the digit span of a person is set at the digit sequence length at which they repeat a minimum number of sets without error.

Finally, the ASRT measures implicit learning by asking the participant to press on a button corresponding to one of four positions where a stimulus appears. The movement of the stimulus across the possible spatial positions alternates between random and pattern sequences. Differences in reaction time (RT) on blocks of random trials vs blocks following the rule-governed pattern are typically used to infer implicit learning when the participant exhibits an associated lack of conscious knowledge of the learned sequence (Sævland & Norman, 2016). Specifically, if the RT consistently decreases as participants progress through the blocks, it indicates that they are improving their recognition and processing of the pattern. Conversely, if the RT remains consistent or increases throughout the blocks, it suggests difficulty in learning or internalizing the patterns.

The battery of cognitive tasks employed at the Norwegian site is openly available at https://app.gorilla.sc/openmaterials/782880, whereas the battery for the Spanish site is available at https://app.gorilla.sc/openmaterials/782877.

Artificial language training and testing

For consistency, the written instructions in the lab-based sessions are provided in the natural language consistent with the artificial language of the session—namely, English for Mini-English sessions, Norwegian for Mini-Norwegian sessions, and Spanish for Mini-Spanish sessions. The linguistic and the graphic stimuli for the three artificial languages were created using *R* (scripts available at the OSF repository, see link above). As in González Alonso et al. (2020), each artificial language is based on the lexicon of the original natural language, whereas the novel morphology is shared across all artificial languages. In Mini-Norwegian, gender is based on the two-gender system of some current Norwegian varieties, whereby etymological masculine and feminine are merged into a 'common' gender that retains the original masculine morphology, such as the definite article *-en* (see Rodina & Westergaard, 2015, 2021). Table 1 shows the definite articles classified by the gender and the number in each site.

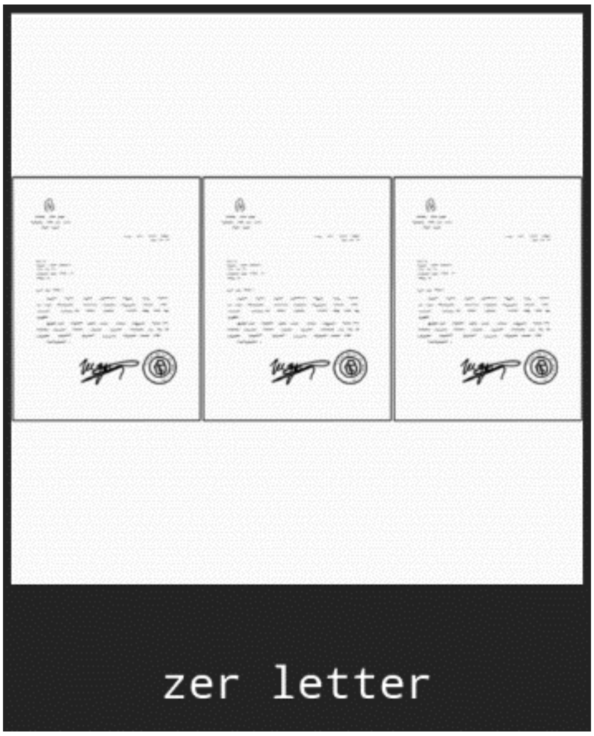
Table 1. Definite articles classified by gender and number in each site.

	Feminine (Spain) Common (Norway)	Masculine (Spain) Neuter (Norway)
Singular	*je*	*ze*

Plural	*jer*	*zer*

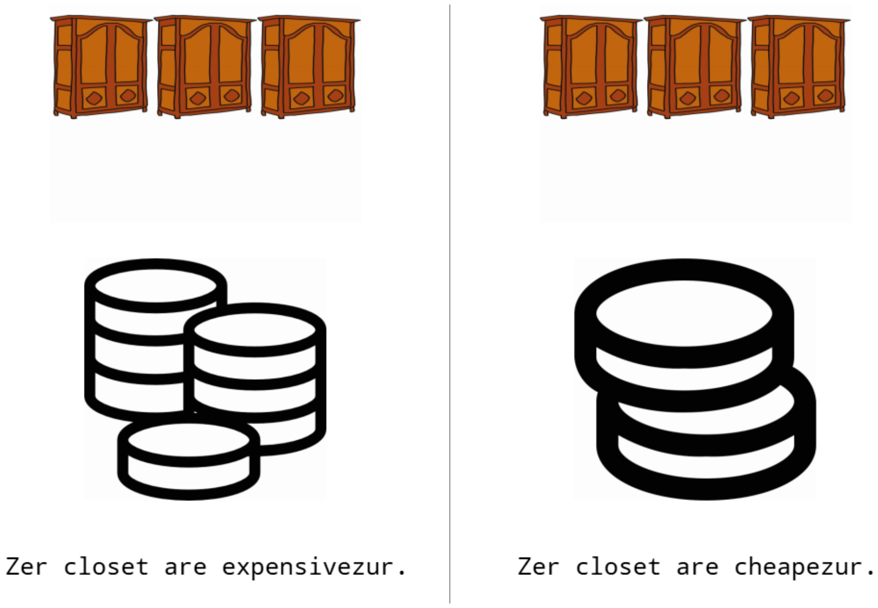
All lab-based sessions (2 through 4) begin with a pre-training that familiarizes the participant with the relevant nouns and their gender. Participants passively loop through two rounds of the target nouns, with pictures, once in singular form and once in plural form—thus being exposed to the plural marker as well. Figure 1 provides one such example.

Figure 1. An example of the pre-training in Mini-English (noun: *the letter*).



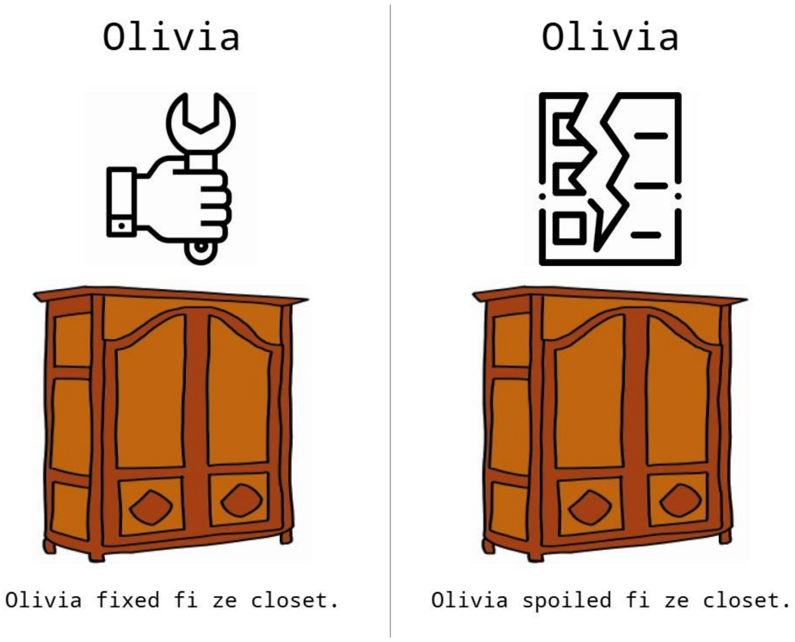
After this phase, the training for each session begins. Session 2 introduces the language for the first time, with gender and number agreement between nouns and adjectives. Gender markers take the form of CV suffixes with the vowel "u" and a consonant matching that of the corresponding article ("z" for feminine/common, "j" for masculine/neuter). As in González Alonso et al. (2020), implicit training in nominal agreement consists of passive exposure to picture pairs showing some semantic contrast but keeping the noun, gender and number morphology constant (see Figure 2).

Figure 2. Example of the implicit training for Session 2 (gender agreement) in Mini-English.



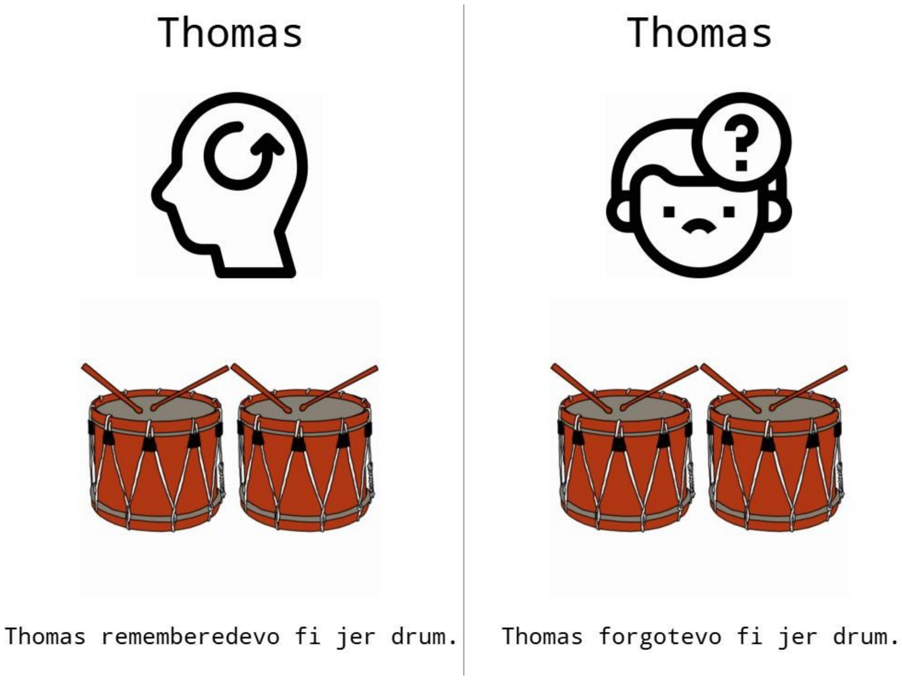
The grammatical properties of interest are incorporated incrementally across sessions. Session 3 introduces DOM. Unlike gender agreement, which is present in both Spanish and Norwegian but not in English (outside of some residual agreement related to natural gender in the pronominal system), DOM is only instantiated in Spanish. In our ALs, the DOM marker is a free morpheme (*fi*) appearing between the verb and the determiner phrase (DP). Figure 3 shows an example of the training phase for DOM.

Figure 3. Example of the implicit training for Session 3 (DOM) in Mini-English.



Finally, Session 4 introduces object-verb number agreement (OVA). This property, present in languages like Basque, is absent in all three base languages in this study, but unlike DOM (which is truly novel in the Norwegian site, because it is not present in the grammar of either Norwegian or English), constitutes more of a novel context for an otherwise known property in all three languages (number agreement). Mirroring Basque morphological expression of this property—to a certain extent—, the agreement marker takes the form of a suffix (*-(e)vo*) that is attached to the verb only when its direct object is plural. Figure 4 provides an example.

Figure 4. Example of the implicit training for Session 4 (VOA) in Mini-English.



Each of these training sessions is followed by a test. Participants must score above 80% to continue to the ERP experiment, and they have two attempts at the test (following González Alonso et al., 2020). If the second attempt is unsuccessful, the learner is thanked for their participation and removed from further sessions of the study. If the participant scores above threshold in either attempt, they move on to the experiment. The test consists of a series of multiple-choice picture-sentence matching questions, where the correct sentence describing the picture is one among five options. Figure 5 provides an example from the test in Session 2 (gender agreement).

Figure 5. Example of a test question in Session 2 (gender agremeent) for Mini-English. The options include the correct sentence as well as four distractors: a gender violation, a number violation, a double violation, and a semantic violation (correct morphology).

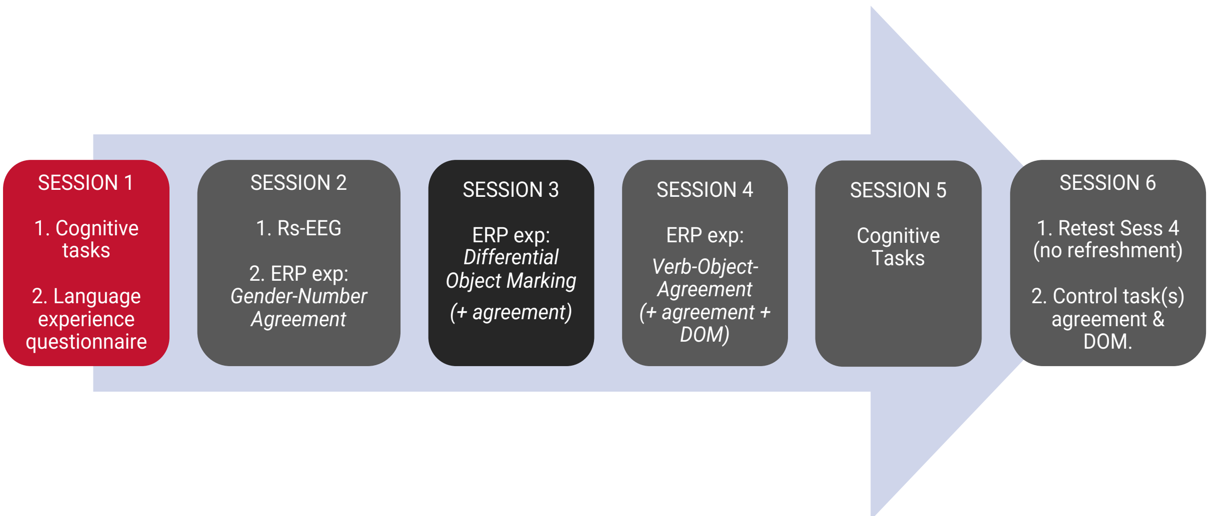
ERP experiments

The ERP experiment, taking place after the implicit training (and the test) in each session 2 through 4, consists of a standard grammaticality judgment task at the end of word-by-word rapid serial visual presentation (RSVP) of sentences, with EEG recording. Besides filler sentences, each session contains a grammatical condition, critical condition—a violation of the property of interest—, and a control condition—a violation of an ancillary property that the implicit training has not focused on (number agreement for Session 2, article location for Sessions 3 and 4). Behavioral and electrophysiological responses to these conditions are collected and statistically modelled to investigate overt and covert sensitivity to violations of these grammatical properties—as well as development and maintenance over time of such sensitivity. At the end of the study, participants are given a gender assignment task and a DOM task (only Spain site) to ascertain that these properties are indeed part of their Spanish and Norwegian grammars (see González Alonso et al., 2020, for a similar control).

Summary and predictions

The study we have described is complex in structure and rich in data types, sources and timepoints (see Figure 6 for a diagram of the overall design). For this reason, and given that we cannot rest on precedents for many of the innovations introduced with respect to previous studies, providing an outline of predictions and theoretical implications is not straightforward. In light of both González Alonso et al. (2020) and Pereira Soares et al. (2022), we expect the processing of gender agreement in Session 2 to show similar results, with high behavioral accuracy in all groups and ALs but ERP indices of attentional resource allocation (P300, N2) showing up *only* for those groups where the AL is lexically based on the previous language instantiating gender agreement: participants exposed to mini-Spanish in the Spanish-English and English-Spanish groups, and participants exposed to mini-Norwegian in the Norwegian-English group. In terms of individual differences, we expect the measured aspects of memory and attention to be predictors of overall language learning performance, as found in previous studies. More interestingly, however, we also expect them to function as modulators of the processing effects that obtain differentially for groups exposed to different mini-languages—potentially correlating to aspects such as the amplitude of any ERP effects indexing allocation of attentional resources.

Figure 6. Overall structure of the study.



DISCUSSION

This paper has presented the design and materials of a large-scale longitudinal study of L3 development employing (semi)artificial language learning. The methodology described here is designed to investigate L3/L*n* development following three guiding principles that are epistemologically motivated. First, it aims to do so starting from the very first exposure to the language. We believe that acquiring a solid understanding of the initial conditions (the initial state/stages) is the soundest, most principled way to investigate the nature of the L3 grammar at later stages, because the nature of the trajectory leading up to them is part of the actual data, and does not need to be inferred. Second, the study is designed to track progress longitudinally within the same learners, instead of relying on cross-sectional comparisons between groups of learners at different stages of development. Third, our study employs semiartificial languages to maximize control over input exposure, as well as over L1/L2/L3 grammatical mismatches. The fact that these languages have a lexical base known to participants allows us to quickly focus on morphosyntax overcoming the lexical bottleneck at first exposure, and helps retain ecological validity in comparison to fully artificial languages. Besides what these methodological choices can contribute to our understanding of L3/L*n* development, however, we believe that there is much value in the way our study gauges, through both behavioral and electrophysiological measures, the mechanisms at play upon first exposure to a third language. It is indeed this aspect that connects most meaningfully to larger questions about human cognition. For that reason, we will dedicate part of this discussion to highlighting those connections.

Almost 10 years after the publication of Rothman et al. (2015), we stand at a different place as a field. We have learned that cross-linguistic influence at the level of representation (i.e., transfer), at least at the neural signature level, may not be detectable as early in the L3/L*n* acquisition process as originally predicted. Nevertheless, in doing the initial empirical work we have uncovered some consistent electrophysiological indices related to attentional processing, which we have good reason to link to the eventual manifestation of transfer in (slightly) later stages. This work, thus, brings to the forefront a question which had not yet been addressed, but is essential to the operationalization of all current theoretical models: *how* does CLI/transfer come to be, in precise cognitive terms?

This might be one of the most fundamental questions we can pose about our object of study, because it is impossible to answer it without realizing that linguistic transfer is, essentially, a reflex of a much larger principle operating in human cognition: economy. Powerful as they are, human cognitive resources are limited. Those limitations impose constraints on the interaction between our mind/brain and the world, and establish guiding principles of information processing and acquisition that seek a maximal benefit-to-effort ratio (e.g., Finton, 2012). The *cognitive economy* approach has received empirical support on several fronts, such as the "predictive brain" hypothesis, according to which human brains maximize resources by attempting to predict incoming input on the basis of experience (e.g., Clark, 2013). However, predictive processing is not the only way in which the mind/brain implements cognitive economy. In the context of learning (in the most general sense), reliance on prior knowledge to avoid redundancy is a hallmark of human cognition, and language acquisition is no exception (Rothman, 2013; 2015). In this context, it should come as no surprise that the influence of previous linguistic knowledge on the learning of subsequent languages has become a central theme in nonnative language acquisition research.

Using prior knowledge to ease and shorten the learning task requires, at a minimum, three things: (1) that the input is parsed to an extent that allows for extraction and identification of at least some of its properties; (2) that this input is compared with existing representations (of the same kind, e.g., linguistic) in long-term memory, sometimes belonging to competing sets that may not be simultaneously active; and (3) that these comparisons yield some outcome based on an evaluation of (degrees of) identity between the input and existing representations, resulting in the selection—i.e., activation—of the most similar available representations for reuse, repurpose, or copy. Given the operations involved in (1) through (3), it is likely that (selective) attention, working memory and inhibition, or probably different combinations of all three, are recruited to make knowledge reuse possible. The fact that this basic description should largely be true of all kinds of knowledge across all kinds of human learning makes linguistic transfer in L3/L*n* acquisition—where, crucially, prior knowledge is organized in multiple, potentially mutually exclusive sets—a privileged window into one of the most important questions in cognitive science. The study we outline here, as well as the research program that emerges from it, are driven by this insight.

Beyond what it can contribute to a better understanding of the cognitive mechanisms underlying CLI, the research presented in this paper is designed to inform several other inquiries in L3/L*n* acquisition. Here, we have sought to provide a detailed, transparent description of our methods to more easily discuss the (clusters of) questions this study hopes to address. In what follows, we offer a summary of the data types that will be collected and analyzed, and the theoretical and epistemological issues we will be focusing on through such analyses, as well as the larger bridges we hope to build toward other fields of enquiry.

The ALs used in this study are learned incrementally: gender agreement, DOM and verb-object agreement, in that order, are introduced in different sessions, and add to previous properties to increase the complexity of the cumulative language material the learner is exposed to. This means that the first property, gender agreement, has been the subject of substantial incidental exposure by the end of the third training session. This is less so for DOM, although it does feature in two different sessions, both directly—in sentences designed to expose the learner to this property—and indirectly—as part of sentences designed to familiarize the learner with object-verb agreement. This last property is only present in the third session. What this incremental design achieves is two-fold. First, it provides a large data set on first exposure to three different properties of different complexity and typological nature, and with different presence or alignment across the L1, the L2 and the L3 (ALs) of the participants. Second, it offers the first longitudinal data sets of this kind, with sessions two and three weeks apart, respectively, to explore the evolution of both behavioral and electrophysiological sensitivity to newly learned grammatical properties. Furthermore, the 4-month delay testing in Session 6 will allow us to investigate the relative consolidation of linguistic knowledge on the basis of the patterns observed on first exposure and longitudinally across sessions.

Collecting data on individual differences in cognitive functions will allow us to evaluate their role in the general learning patterns of our participants. This is consistent with a growing body of literature focusing on the role of executive functions, inclusive of individual differences in them, in L2 acquisition (e.g., Chen et al., 2022; Frost et al., 2013; Godfroid & Kim, 2021; Prior & Gollan, 2011; Reichle et al., 2016; Williams & Rebuschat, 2016, among many others). However, what makes these measures particularly interesting in the context of this study is their potential to show whether attention, working memory, inhibition and other such processes are involved in the detection—and exploitation—of cross-linguistic similarity. While the correlation of cognitive function scores to the (purported) differences between groups of learners assigned to different ALs would be an indirect measure of their involvement, we rely on the qualitative differences between electrophysiological responses (ERP components) to further explore the real-time deployment of these resources during language processing.

The fact that participants take the cognitive testing battery both before and after the block of language training sessions will allow us to explore the bidirectional relationship between core cognitive processes and multilingual language learning and processing. If our hypotheses about the involvement of executive functions in CLI/transfer processes is on the right track, we might see (potentially short-lived) changes in these measures of cognitive functions between Sessions 1 and 5. For instance, the learning process in and of itself may stress and thus train certain aspects of these functions, for example working memory, which may manifest as changes in performance on relevant tasks after the longitudinal period. Conversely, the mechanisms behind adaptations in other domains (e.g., attentional control) may not manifest in the same way given that the learning process does not necessarily entail active use of the languages after learning, and thus the cognitive adaptations required may not manifest at the level of performance in inhibition tasks (e.g., Bialystok & Craik, 2022; Rothman, 2024). Moreover, resting-state EEG measures, also collected before and after the language training phase of the study, could provide further data to gauge potential changes to a speaker's base-line cognitive functioning as a result of acquiring and processing multiple languages (see, e.g., Bice, Yamasaki & Prat, 2020; Pereira Soares, Kubota, Rossi & Rothman, 2021). The point we wish to make here is that a study of this particular type is in a privaledged position to serve as a bridge between the multilingual language learning/processing and the neurocognition of multilingualism literatures, whereby the extent of and links between their bidirectional near and far transfer effects are better documented and understood.

Finally, by collecting detailed information through the LHQ3, we also seek to capture another dimension of interindividual variability: that which relates to language experience. This is important because theories and models are expected to not only cover explanatorily the trends observed in aggregated data sets, but indeed in the individual data points that contribute to them. While there has been a shift to individual differences approaches in much of the bi-/multilingualism literatures over the past decade, this trend has not manifested in the landscape of L3/L*n* theorizing. And yet, there is no reason to believe that cross-linguistic transfer in L3/L*n* (or for that matter in L2) learning would be any different from other cognitive processes that are clearly subject to individual differences. In fact, to take one example, insofar as differences in cognitive priors—e.g., an individual's own baseline for working memory, attentional control, implicit learning—factor into cognitive and linguistic task performances and perhaps are subject to changes as a function of the AL intervention, it stands to reason that these individual priors could have some explanatory coverage of how CLI manifests differentially for some individuals as compared to others. If this is shown to be the case and/or if quantities and qualities of language experience (e.g., contextualized usage patterns of the previous languages) prove to be non-trivial predictors, then it would require the existing theoretical models to go beyond the expectations of group tendencies, to also address how they can (or not) accommodate individual patterns. Surprisingly, to date, no L3/L*n* study has sought to examine this, an important gap the present study will be positioned to address.

CONCLUSION

The study of L3/L*n* development has become a burgeoning subfield of third language acquisition studies in recent years. This focus on intermediate stages constitutes the next step in our theoretical development as a field and brings us closer to more comprehensive models of L3/L*n* acquisition, models that may take scope over the whole developmental sequence and consider variables well beyond the interaction and multidirectional influence between the languages of a multilingual speaker. Taking these further steps, however, requires a sound epistemological approach that ensures our knowledge is built incrementally, slowly accumulating over the foundations of our research into progressively more advanced stages of development. The study we presented in this paper takes this approach, investigating L3/L*n* acquisition longitudinally from first exposure to the language. We do so while considering individual differences in executive functions and history of bilingual exposure/engagement, in order to estimate their role both in learning trajectories as a whole and in modulating patterns of linguistic transfer/CLI. We believe that this type of studies has the potential to contribute to a more detailed understanding of L3/L*n* development under controlled conditions, and stimulate research into L3/L*n* acquisition and processing that is grounded in, and connected to, inquiries into human cognition more broadly.

REFERENCES

Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2014). Morphosyntactic processing in advanced second language (L2) learners: An event-related potential investigation of the effects of L1-L2 similarity and structural distance. *Second Language Research, 30*, 275-306.

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research*, 23(4), 459-484.

Bayram, F., González Alonso, J., Kush, D., Lohndal, T., & Rothman, J. (2021). *Individual Differences in Bilingualism*. Special issue of *Applied Psycholinguistics*, 42(2).

Bialystok, E., & Craik, F. I. (2022). How does bilingualism modify cognitive function? Attention to the mechanism. *Psychonomic Bulletin & Review*, *29*(4), 1246-1269.

Bice, K., Yamasaki, B. L., & Prat, C. S. (2020). Bilingual language experience shapes resting-state brain rhythms. *Neurobiology of Language*, *1*(3), 288-318.

Cabrelli, J., Chaouch-Orozco, A., González Alonso, J., Pereira Soares, S. M., Puig-Mayenco, E., & Rothman, J. (2023). Introduction – Multilingualism: Language, Brain and Cognition. In J. Cabrelli, A. Chaouch-Orozco, J. González Alonso, S. M. Pereira Soares, E. Puig-Mayenco, & J. Rothman (Eds.), *The Cambridge Handbook of Third Language Acquisition* (pp. 1-20). Cambridge University Press.

Cabrelli, J., & Iverson, M. (2023). Why do learners overcome non-facilitative transfer faster from an L2 than an L1? The cumulative input threshold hypothesis. *International Journal of Multilingualism*. Epub ahead of print, DOI: 10.1080/14790718.2023.2200252

Cabrelli, J., Iverson, M., Giancaspro, D., & González, B. H. (2020). The roles of L1 Spanish versus L2 Spanish in L3 Portuguese morphosyntactic development. *Linguistic approaches to Portuguese as an additional language*, *24*, 11-33.

Chen, Y., Li, L., Wang, M., & Wang, R. (2022). Which cognitive factors predict L2 grammar learning: Cognitive control, statistical learning, working memory, or attention? *Frontiers in Psychology, 13,* 943988.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36,* 181-204.

Datta, A., Cusack, R., Hawkins, K., Heutink, J., Rorden, C., Robertson, I. H., & Manly, T. (2007). The P300 as a marker of waning attention and error propensity. *Computational intelligence and neuroscience*, *2007*(1), 093968.

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences*, *116*(15), 7565-7574.

Dörnyei, Z. (2014). *The psychology of the language learner: Individual differences in second language acquisition*. Routledge.

Finton, D. J. (2012). Cognitive-Economy Assumptions for Learning. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning*. Boston, MA: Springer.

Flynn, S., Foley, C., & Vinnitskaya, I. (2004). The cumulative-enhancement model for language acquisition: Comparing adults' and children's patterns of development in first, second and third language acquisition. *International Journal of Multilingualism, 1*(1), 3-16.

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science, 24*(7), 1243–1252.

Gabriele, A., Fiorentino, R., & Alemán Bañón, J. (2013). Examining second language development using event-related potentials: A cross-sectional study on the processing of gender and number agreement. *Linguistic Approaches to Bilingualism, 3*, 213-232.

Gass, S. M. (1984). A review of interlanguage syntax: Language transfer and language universals. *Language Learning*, 34(2), 115-132.

Gass, S. M. (1988). Second language acquisition and linguistic theory: The role of language transfer. In *Linguistic theory in second language acquisition* (pp. 384-403). Dordrecht: Springer Netherlands.

Godfroid, A., & Kim, K. M. (2021). The contributions of implicit-statistical learning aptitude to implicit second-language knowledge. *Studies in Second Language Acquisition*, *43*(3), 606-634.

González Alonso, J. (2023). Generative approaches. In J. Cabrelli, A. Chaouch-Orozco, J. González Alonso, S. M. Pereira Soares, E. Puig-Mayenco, & J. Rothman (Eds.), *The Cambridge Handbook of Third Language Acquisition* (pp. 18-34). Cambridge University Press.

González Alonso, J., Alemán Bañón, J., DeLuca, V., Miller, D., Pereira Soares, S. M., Puig-Mayenco, E., Slaats, S., & Rothman, J. (2020). Event related potentials at initial exposure in third language acquisition: Implications from an artificial mini-grammar study. *Journal of Neurolinguistics, 56*, 100939.

Grey, S. (2023). Using Artificial Languages to Study Third Language Learning and Processing. In J. Cabrelli, A. Chaouch-Orozco, J. González Alonso, S. M. Pereira Soares, E. Puig-Mayenco, & J. Rothman (Eds.), *The Cambridge Handbook of Third Language Acquisition* (pp. 724-743). Cambridge University Press.

Grey, S., Williams, J. N., & Rebuschat, P. (2014). Incidental exposure and L3 learning of morphosyntax. *Studies in Second Language Acquisition*, *36*(4), 611-645.

Herrmann, C. S., & Knight, R. T. (2001). Mechanisms of human attention: event-related potentials and oscillations. *Neuroscience & Biobehavioral Reviews*, *25*(6), 465-47

Hopp, H. (2014). Working memory effects in the L2 processing of ambiguous relative clauses. *Language Acquisition*, *21*(3), 250-278.

Huang, T., Loerts, H., & Steinkrauss, R. (2022). The impact of second-and third-language learning on language aptitude and working memory. *International Journal of Bilingual Education and Bilingualism, 25*(2), 522–538.

Jylkkä, J., Laine, M., & Lehtonen, M. (2021). Does language switching behavior rely on general executive functions? *Bilingualism: Language and Cognition, 24*(3), 583–595.

Kellerman, E. (1977). Toward a characterization of the strategy of transfer in second language learning. *Interlanguage Studies Bulletin*, 2(1), 58-145.

Klein, E. C. (1995). Second versus third language acquisition: Is there a difference? *Language Learning*, 45(3), 419-465.

Lauro, J., Core, C., & Hoff, E. (2020). Explaining individual differences in trajectories of simultaneous bilingual development: Contributions of child and environmental factors. *Child Development*, *91*(6), 2063-2082.

Leung, Y.-K. I. (1998). The acquisition of French as a third language in Hong Kong: Interlanguage and typology (Master's thesis). University of Hong Kong.

Leung, Y.-K. I. (2001). The initial state of L3A: Full transfer and failed features? *In Proceedings of theTh 3rd Generative Approaches to Second Language Acquisition Conference (GASLA 2000)* (pp. 55-62). Cascadilla Press.

Leung, Y.-K. I. (2002). Functional categories in third language acquisition: A cross-linguistic study of Cantonese-English-French learners. *The 6th Generative Approaches to Second Language Acquisition Conference (GASLA 2002)*, 199-207.

Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, 23(5), 938-944.

Luk, G., & Rothman, J. (2022). Experience-based individual differences modulate language, mind and brain outcomes in multilinguals.

McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., & Osterhout, L. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning, 60*(S2), 123-150.

Mägiste, E. (1979). The competing language systems of the multilingual: A developmental study of decoding and encoding processes. *Journal of Verbal Learning and Verbal Behavior*, 18(1), 79-89.

Mägiste, E. (1986). Selected issues in second and third language learning. *Acta Universitatis Upsaliensis: Studia Psychologica et Paedagogica*, 3, 34-45.

Mitrofanova, N., Leivada, E., & Westergaard, M. (2023). Crosslinguistic influence in L3 acquisition: Evidence from artificial language learning. *Linguistic Approaches to Bilingualism, 13*(5), 717-742.

Odlin, T. (1989). *Language transfer* (Vol. 27). Cambridge: Cambridge University Press.

Paradis, J. (2023). Sources of individual differences in the dual language development of heritage bilinguals. *Journal of Child Language*, *50*(4), 793-817.

Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: selected studies of the event-related potential. *International Journal of Medical Sciences*, *2*(4), 147.

Pereira Soares, S. M., Kubota, M., Rossi, E., & Rothman, J. (2021). Determinants of bilingualism predict dynamic changes in resting state EEG oscillations. *Brain and Language*, *223*, 105030.

Pereira Soares, S. M., Kupisch, T., & Rothman, J. (2022). Testing Potential Transfer Effects in Heritage and Adult L2 Bilinguals Acquiring a Mini Grammar as an Additional Language: An ERP Approach. *Brain Sciences*, *12*(5), 669.

Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: Evidence from Spanish–English and Mandarin–English bilinguals. *Journal of the International Neuropsychological Society, 17*(4), 682–691.

Puig-Mayenco, E. (2023). The sandwich effect in L3 acquisition: A look at the Catalan and Spanish grammars of highly advanced speakers of L3 English. *International Journal of Bilingualism*, *27*(6), 996-1015.

Puig-Mayenco, E., González Alonso, J., & Rothman, J. (2020). A systematic review of transfer studies in third language acquisition. *Second Language Research*, *36*(1), 31-64.

Puig-Mayenco, E., Rothman, J., & Tubau, S. (2022). Language dominance in the previously acquired languages modulates rate of third language (L3) development over time: a longitudinal study. *International Journal of Bilingual Education and Bilingualism, 25*(5), 1641-1664.

Reichle, R. V., Tremblay, A., & Coughlin, C. (2016). Working memory capacity in L2 processing. *Probus*, *28*(1), 29-55.

Rodina, Y., & Westergaard, M. (2015). Grammatical gender in Norwegian: Language acquisition and language change. *Journal of Germanic Linguistics, 27*(2), 145–187.

Rodina, Y., & Westergaard, M. (2021). Grammatical gender and declension class in language change: A study of the loss of feminine gender in Norwegian. *Journal of Germanic Linguistics, 33*(3), 235–263.

Rothman, J. (2010). On the typological economy of syntactic transfer: Word order and relative clause high/low attachments preference in L3 Brazilian Portuguese. *International Review of Applied Linguistics in Language Teaching*, 48(3-4), 245-273.

Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. *Second language research*, *27*(1), 107-127.

Rothman, J. (2013). Cognitive Economy, Non-Redundancy and Typological Primacy in L3 Acquisition: Evidence from initial stages of L3 Romance. In Baauw, S., Dirjkoningen, F., Meroni, L. & Pinto, M. (eds.), *Romance Languages and Linguistic Theory 2011* (pp. 217-247). Amsterdam/Philadelphia: John Benjamins.

Rothman, J. (2015). Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered. *Bilingualism: language and cognition*, *18*(2), 179-190.

Rothman, J. (2024). Harnessing the bilingual descent down the mountain of life: Charting novel paths for Cognitive and Brain Reserves research. *Bilingualism: Language and Cognition*. Epub ahead of print, 13 March 2024. DOI: 10.1017/S1366728924000026.

Rothman, J., Alemán Bañón, J., & González Alonso, J. (2015). Neurolinguistic measures of typological effects in multilingual transfer: Introducing an ERP methodology. *Frontiers in psychology*, *6*, 1087.

Rothman, J., & Cabrelli Amaro, J. (2010). What variables condition syntactic transfer? A look at the L3 initial state. *Second Language Research*, 26(2), 189-218.

Rothman, J., González Alonso, J., & Puig-Mayenco, E. (2019). *Third language acquisition and linguistic transfer*. Cambridge University Press.

Rothman, J., & Treffers-Daller, J. (2014). A prolegomenon to the construct of the native speaker: Heritage speaker bilinguals are natives too!. *Applied linguistics*, *35*(1), 93-98.

Sanz, C., Park, H. I., & Lado, B. (2015). A functional approach to cross-linguistic influence in ab initio L3 acquisition. *Bilingualism: Language and Cognition*, *18*(2), 236-251.

Sævland, W., & Norman, E. (2016). Studying different tasks of implicit learning across multiple test sessions conducted on the web. *Frontiers in Psychology*, *7*, 808.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access hypothesis. *Second Language Research*, 12, 40-72.

Skehan, P. (1991). Individual differences in second language learning. *Studies in second language acquisition*, *13*(2), 275-298.

Slabakova, R. (2017). The scalpel model of third language acquisition. *International Journal of Bilingualism*, 21(6), 651-665

Sprouse, R. A., & Schwartz, B. D. (2023). L3 Cognitive States and the Abbreviated Grappling Period Model. In V. Sheu, A. Zhou & J. D. Weirick (Eds.), *Variation in Linguistics: Second Language Acquisition, Discourse Studies, Sociolinguistics, Syntax* (pp. 2-25). Cambridge Scholars Publishing.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643-662.

Treffers-Daller, J., & Sakel, J. (2012). Why transfer is a key aspect of language use and processing in bilinguals and L2-users. *International Journal of Bilingualism*, *16*(1), 3-10.

Weinreich, U. (1953). *Languages in contact: Findings and problems*. New York, NY: Linguistic Circle of New York.

Westergaard, M. (2021). Microvariation in multilingual situations: The importance of property-by-property acquisition. *Second Language Research*, 37(3), 379-407.

Westergaard, M., Mitrofanova, N., Mykhaylyk, R., & Rodina, Y. (2017). Crosslinguistic influence in the acquisition of a third language: The Linguistic Proximity Model. *International Journal of Bilingualism*, 21(6), 666-682.

Westergaard, M., Mitrofanova, N., Rodina, Y., & Slabakova, R. (2023). Full Transfer Potential in L3/Ln Acquisition: Crosslinguistic Influence as a Property-by-Property Process. In J. Cabrelli, A. Chaouch-Orozco, J. González Alonso, S.M. Pereira Soares, E. Puig-Mayenco & J. Rothman (Eds.), *The Cambridge handbook of third language acquisition* (pp. 219-242). Cambridge University Press.

Williams, J., & Rebuschat, P. (2016). *Implicit learning and second language acquisition*. Evanston, IL, USA: Routledge.

Wright, B. C. (2017). What Stroop tasks can tell us about selective attention from childhood to adulthood. *British Journal of Psychology*, *108*(3), 583-607.