Individual language experience determinants of morphosyntactic variation in heritage and attriting speakers of Bosnian and Serbian

A causal inference approach

Aleksandra Tomić[1], Yulia Rodina[1], Fatih Bayram[1], and Cecile De Cat[1,2]
[1] UiT The Arctic University of Norway | [2] University of Leeds

Using a causal inference approach, we explored the relationships among the language experience determinants of morphosyntactic sensitivity, to identify the factors that indirectly and directly cause its acquisition or maintenance in immigration contexts. We probed the sensitivity to Serbian/Bosnian clitic placement violations with a self-paced listening task, in a diverse group of bilinguals in Norway ($n = 71$), born to immigrant parents, or having emigrated in childhood or adulthood. The outcomes included a metalinguistic violation detection score and a listening/processing time difference between licit and illicit structures.

Structural Equation Models revealed that literacy (as reading practices) was among the most influential determinants of the ability to detect violations, while Bosnian/Serbian use across contexts and age of bilingualism onset determined violation sensitivity in processing. We identified a significant threshold of societal language (SL) exposure at age 8. Rather than SL exposure before this age precluding bilinguals from developing and maintaining morphosyntactic sensitivity, this threshold seems to reflect a protective effect against attrition which intensifies the later after age 8 SL exposure starts. The length of residence in Norway did not determine attrition, suggesting that heritage and attrited speakers should be considered on a continuum rather than as distinct bilingualism profiles.

Keywords: heritage bilingualism, clitics, language attrition, causal inference, individual differences

https://doi.org/10.1075/lab.24016.tom | Published online: 18 November 2024
Linguistic Approaches to Bilingualism ISSN 1879-9264 | E-ISSN 1879-9272

1. Introduction

Heritage speakers (HSs) show more varied language acquisition outcomes than speakers acquiring the same language as their societal language (SL), due to more variable experience. Key factors include the age of onset of SL exposure, quality and quantity of heritage language (HL) exposure and use, and language attitudes. Nevertheless, very few studies assess a wide range of individual variables and their interrelationships in a single sample (Paradis, 2023). Herein, we investigate how a network of individual factors causally affects the acquisition and maintenance of the HL in immigrant communities. Additionally, we explore whether a categorical distinction exists in HL outcomes between bilinguals exposed to the SL before vs. after a specific age, and what that age might be, while controlling for other language experience variables. As the acquisition outcome, we examine the sensitivity to the placement of pronominal object clitics amongst Bosnian and Serbian heritage and late bilinguals in Norway.

In immigration contexts, native language acquisition and loss are defined as three distinct speaker profiles, each with unique language development and maintenance paths. HSs can be

> (a) simultaneous bilinguals, those exposed to the heritage and the majority language before the age of 3–4; (b) sequential bilinguals or child L2 learners, those exposed to the HL at home until age 4–5 and to the majority language once they start preschool; and (c) late child L2 learners, children monolingual in the HL, who received some elementary schooling in their home country and immigrated around ages 7–8. (Montrul, 2011, p. 157)

The defining feature is that their SL exposure onset should start before full HL acquisition. Despite being native speakers, HSs can manifest a broad spectrum of competence outcomes in their HL, ranging from receptive/passive bilinguals to fully balanced bilinguals that are indistinguishable from speakers in the home country. The onset of exposure to the SL can vary. Attrited speakers (AS) are those who undergo a "non-pathological decrease in a language [they] had previously acquired" (Köpke & Schmid, 2004). Changes in the native language may occur across linguistic domains, affecting both receptive and productive language skills (Bylund, 2009). The level of attrition or divergence from the monolingual baseline depends on the length of residence (LoR) in the country of immigration and the amount of usage of the native language (Kasparian & Steinhauer, 2016). If attrition starts in childhood, it can exhibit greater severity and speed in the loss of structural aspects within the native language (Pallier, 2007; Karayayla & Schmid, 2019). The assumption is that a significant erosion of the L1 system is more probable when attrition onset takes place before the end of adolescence (Bylund, 2009).

HSs can therefore be susceptible to attrition, potentially losing aspects of the HL they had previously acquired.

The third speaker profile includes recent immigrants, typically still fully proficient in their L1. As bilinguals, they provide a more reliable baseline for HS and AS than home-country monolinguals (Rothman et al., 2023). Researchers generally assume a 10-year LoR threshold after which some attrition is expected (Karayayla & Schmid, 2019). However, there is no consensus regarding the cut-off age of SL exposure onset or LoR distinguishing these presumably fully proficient speakers from HSs and ASs. Moreover, exclusively using language outcomes as a profile diagnostic would pose a risk of circularity in their identification. We argue that the relationship between the age of SL exposure onset, LoR, and native/HL outcomes must be considered alongside other sociolinguistic variables to ascertain which variables shape HL acquisition and/or attrition.

This study adopts an individual difference approach to investigate language acquisition and maintenance in immigration contexts as a continuum. We aim to determine empirically whether and how distinct profiles of speakers can be identified, based on how individual difference factors determine their language outcomes. HSs of Bosnian and Serbian in Norway provided us with a unique natural lab environment to investigate the same morphosyntactic property–pronominal clitic placement identified as a vulnerable feature in previous studies.

2. Background

2.1 The individual difference approach to language acquisition and loss

Studies on language experience effects provide valuable insights into how individual factors influence the variability in HL outcomes (Paradis, 2023, for an overview). However, this research only partially explains which factors directly cause specific HL outcomes. Predictors may be significant due to correlations with variables they influence or are influenced by. For example, early SL exposure could be significant because it causes many patterns of HL experience, in terms of quality and quantity. The amount of literacy training could be significant for HL outcomes because of the factors associated with formal schooling in that language. A theory of HL acquisition and maintenance requires a systematic approach to disentangle these multi-correlated factors.

The complex web of causal relationships among the variables of interest is a major challenge for the individual differences approach to bilingual acquisition outcomes (De Cat & Unsworth, 2023). To identify the true causes for an outcome rather than just the set of variables which can best predict that outcome, it is essential to consider the nature of relationships among variables causing the outcome. The causal inference approach (Hernán et al., 2019; Westreich & Greenland, 2013) makes this possible. The main tool in causal inference is an exhaustive theorized causal network of how an outcome variable is generated, representing all plausible direct and indirect causes of the outcome variable. This network is visually represented by a directed acyclic graph (DAG). As illustrated in Figure 1, a DAG identifies the outcome of interest (I), the *exposure* variable, i.e., the variable whose causal effect on the outcome we want to investigate (as a research hypothesis), as well as all other possible variables which could be directly or indirectly causing the outcome. The direction of causal influence is represented by arrows connecting the variables (i.e., *nodes*). In causal graphs, the arrows or paths must be one-directional (i.e., directed) and a causal path from one variable cannot cycle back to it, given the principle that a variable cannot simultaneously cause another variable *and* be caused by it, making the graphs acyclic.

DAGs are informed by the research literature and by discussions with experts on the likely causes of the outcome. The variables included in the DAG should represent the concepts (e.g., working memory) rather than how they were measured (e.g., backward digit recall). If the concept cannot be measured directly, the relevant variable is included as unobservable, i.e., latent. The direction of causal effects between variables in the DAG, i.e., what causes what, depends on the chronological order of when each variable crystalized (i.e., when its value was set), among other considerations.

The way in which variables are assumed to be causally related to each other defines their status in the causal path from a variable of interest, i.e., exposure, to the outcome variable. As shown in Figure 1, a *mediator* is a variable through which the variable of interest (i.e., the *exposure*), causes an outcome (I) indirectly, potentially in addition to its direct effect on the outcome. A *confounder* is a variable that influences both the exposure and outcome.

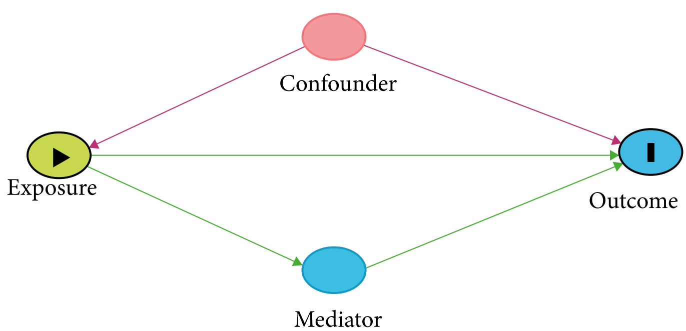


Figure 1. Directed acyclic graph (DAG)

The *direct effect* of the exposure refers to the causal effect it exerts on the outcome that does not operate through any other mediator variables, represented by the direct arrow from exposure to outcome. For example, the direct causal effect of the age of SL exposure onset on the HL outcome variable would be its effect to the exclusion of the indirect causal effects on the outcome which it exerts through causing mediator variables, such as HL education quantity, cumulative HL experience, etc. The *total effect* of an exposure is its complete causal effect on an outcome variable through all possible causal pathways/mechanisms, capturing both the direct effect of exposure on outcome and any indirect effects that pass through mediator variables. For example, the total effect of age of first exposure to the SL on an HL language outcome would include the sum of its direct effect and the indirect effects it exerts through its causal impact on mediators, such as HL education, cumulative HL experience, etc. Importantly, drawing a causal path from a variable to an outcome merely indicates the *plausibility* of a direct causal effect of this variable on the outcome. Estimating this effect using a formal statistical procedure can approximate whether this direct effect is reliable or strong.

The causal network model represented by a DAG guides the statistical analysis by specifying which variables should be included in the formal model (i.e., controlled for) and which variables must be excluded to obtain total, direct, or indirect causal effect estimates for a specific exposure. Including a variable in the model (other than the exposure) is akin to blocking the causal path that runs through that variable to the outcome. Confounding variables should be included in the model (controlled for) when calculating either direct or total effects of an exposure, to block all paths that could carry a spurious correlation between exposure and outcome, as they would otherwise confuse a causal effect with one that is only associative (i.e., non-causal). However, mediator variables should not be included in the model when estimating the total effect of exposure, as this would block the indirect causal influence of exposure on the outcome through mediators. Since the direct effect captures only the impact of exposure on outcome through pathways that do not involve the mediators, estimating the direct effect requires blocking all mediated/indirect paths by conditioning on the mediators (i.e., including them in the model).

Importantly, causal inference is a conceptual, theoretical analysis tool rather than a formal analysis tool. The causal network model helps determine which variables need to be controlled (i.e., adjusted) to estimate different causal effects. The estimation itself is done using appropriate formal analysis tools, such as linear or logistic regression, structural equation modeling (SEM), growth curve analysis, etc. SEM is a powerful multivariate analysis which can be particularly useful for estimating direct causal effects in a causal inference analysis if based on a valid causal model representable by a DAG. SEM involves modeling relationships between variables in the form of a cluster of (linear) regression equations. In a SEM based on a DAG, the direct causal effects on each endogenous variable in the causal network are estimated by a separate linear regression. We use the SEM procedure for the formal analysis of direct effects in this paper, yet it is the causal inference approach shaping and essentially limiting the analysis that enables the causal rather than associative interpretation of SEM estimates (see Section 4.3.2 for details on the implementation).

2.2 Pronominal system in Bosnian/Serbian and Norwegian

Bosnian and Serbian belong to the Bosnian–Croatian–Montenegrin–Serbian (BCMS) language continuum. While the pronominal system of Bosnian and Serbian is largely the same, Croatian and Montenegrin exhibit certain differences in clitic morphosyntax (cf. Kolaković et al. 2022). Bosnian and Serbian are morphologically rich, synthetic languages. Due to overt and specified subject–verb agreement, they allow subject pro-drop and a relatively free word order (although the canonical order is SVO). Bosnian/Serbian NPs, including pronominal clitics, are marked for gender (masculine, feminine, and neuter), number (singular and plural), and case (dative, accusative, and genitive). Table 1 illustrates the pronominal clitic paradigm by gender, number, and case.

Table 1. Pronominal clitic paradigm in standard Bosnian and Serbian

	SG				PL		
	1	2	3		1	2	3
			M/N	F			
DAT	mi	ti	mu	joj	nam	vam	im
ACC/GEN	me	te	ga	je/ju	nas	vas	ih

In Bosnian/Serbian, direct objects can be expressed as NPs, clitics, or full pronouns, illustrated in the examples in (1a–c). The choice of object realization depends on pragmatic conditions, such that NPs are typically used at first mention of a referent (1a), whereas the second mention of the referent tends to be realized as a clitic (1b). Full, stressed pronouns are reserved for special pragmatic contexts, such as contrastive focus (1c).

(1) a. Ana voli **Marka**.
Ana loves Marko.PROPN.M.ACC
'Ana loves Marko.'

b. Ana **ga** voli.
Ana him.CL.3SG.M.ACC loves
'Ana loves him.'

c. Ana voli **njega**, ne mene.
Ana loves him.PRN.3SG.M.ACC not me
'Ana loves him, not me.'

Importantly, the placement of clitics in Bosnian/Serbian is subject to typologically rare constraints. Whereas direct objects in (1a) and (1c) can change position relatively freely in the clause, clitic objects as in (1b) must stay in the second position (P2), after the first element in the clause. The first element can be the first prosodic word (a word that can carry stress; preferred in standard Bosnian) or the first syntactic/prosodic phrase (preferred in standard Serbian; cf. Kolaković et al. 2022; Pešikan, 1958). The clitic in the first position as in (2a), or in the third position (P3), as in (2b), is considered ungrammatical in standard Bosnian/Serbian.

(2) a. *__Ga__ voli.
him.CL.3SG.M.ACC loves
'S/he loves him.'

b. *Ana voli **ga**.
Ana loves him.CL.3SG.M.ACC
'Ana loves him.'

Clitic placement in Bosnian/Serbian is ruled by several constraints: clitics must appear as close to the left edge of the clause as possible, yet they cannot be the first element in a clause. They must join a syntactic constituent (stressed word or phrase) to their left forming a single prosodic phrase (see Franks & King, 2000; Radanović-Kocić, 1988; Progovac, 1996; Bošković 2020).

Pronominal clitics in Bosnian/Serbian can attach to a variety of different hosts: a subject NP (3a), an adverbial phrase in (3b), a verb in (3c), or a complementizer (3d). In most clause structures, pronominal clitics are pre-verbal, except when no other element is fronted and the pronominal subject is omitted, in which case the verb occupies the first position as in (3c). Nevertheless, verb-initial clauses with post-verbal clitics are frequent in discourse, due to frequent subject pro-drop.

(3) a. Ana **ga** je upoznala.
Ana him.CL.3SG.M.ACC be.AUX.V.CL.3SG met
'Ana met him.'

b. Jučer **ga** je upoznala.
yesterday him.CL.3SG.M.ACC be.AUX.V.CL.3SG met
'(She) met him yesterday.

c. Upoznala **ga** je jučer.
met him.CL.3SG.M.ACC be.AUX.V.CL.3SG yesterday
'She met him yesterday.'

d. Ana želi da **ga** upozna.
Ana wants that him.CL.3SG.M.ACC meet
'Ana wants to meet him.'

Unlike Bosnian/Serbian, Norwegian pronominal objects are full pronouns (not clitics) which obligatorily follow the verb (*Anna elsker* ***ham****.* 'Anna loves him.'). Nevertheless, Norwegian pronominal object realization is complex and includes apparent movement: in sentences with negation, unstressed pronouns must precede the negative particle *ikke* (*Anna elsker* ***ham*** *ikke.* 'Anna doesn't love him'; Object Shift, Anderssen et al., 2012).

In sum, Bosnian/Serbian clitics are a complex phenomenon marked for a bundle of grammatical features. Pronominal clitics occupy an early P2 position in a clause involving movement originating from the prosody–syntax interface. The pronominal clitic placement differs from pronominal object position in the SL Norwegian, potentially introducing cross-linguistic influence (CLI) in internal grammar or during processing. While pinpointing the exact linguistic origins of clitic variability is beyond the scope of this paper, the complex form and distribution of Bosnian/Serbian clitics make them a likely vulnerable or variable linguistic domain in the context of first- or second-generation immigration.

2.3 Pronominal clitics as a vulnerable aspect of language acquisition and loss

Previous research on language acquisition and bilingualism has shown that pronominal clitics can be vulnerable linguistic elements due to language internal as well as external factors (Pérez-Leroux et al., 2011; Rinke & Flores, 2014; Ivanova-Sullivan et al., 2022; Shin et al., 2023; Smith et al., 2023). Cross-linguistically, clitic pronouns are phonologically weak forms (typically unstressed, monosyllabic) that do not carry semantic weight and lack sufficient structure to stand alone, thus relying on the presence of a host element. They are complex interface phenomena that encode morphosyntactic features and referentiality. Hence, they are sensitive to reduced exposure in language contact situations.

Divergent patterns have been extensively studied in HSs of the Romance languages, Spanish (European Portuguese, and Italian). Despite mastering the finiteness constraints in Spanish (proclitics before finite verbs and enclitics after non-finite verbs), Spanish–English 3–8-year-olds produced clitic misplacement errors and showed preference towards the enclitics which was attributed to syntactic transfer and the length and timing of exposure to SL English. A combination of heritage Spanish experience, Spanish lexical proficiency, and CLI from English were also reported as predictors of child bilinguals' clitic use in Shin et al. (2023).

Adult HSs of European Portuguese where clitic placement is construction-specific and variable (enclitic/post-verbal in main clauses and proclitic/pre-verbal in subordinate clauses and with sentence negation) are reported to exhibit an enclitic bias in contexts that require proclitics (a developmental pattern typical for early stages of L1 acquisition) (Rinke & Flores, 2014). However, in contrast to the L1 speakers, HSs frequently accepted the ungrammatical proclisis order with non-finite verbs. This systematic and qualitatively different behavior was attributed to HSs' reduced experience with formal registers, rather than CLI from the SL German.

Interesting patterns of intergenerational attrition were found in adult immigrants and HSs of Italian in the UK (Smith et al., 2023). Both groups showed a weaker preference for the clitic and clitic cluster compared to the baseline (monolingual homeland population) and had unique clitic usage patterns which were argued to occur because of a creative (systematic and predictable) usage of language (Kupisch & Polinsky, 2022).

The grammar of clitics in mono- and bilingual speakers of Bosnian/Serbian is understudied. Serbian-speaking children in Varlokosta et al. (2016) are shown to have target-like morphosyntactic and pragmatic knowledge of clitic pronouns, including their distribution/realization and placement at the age of 5. Some evidence of pronominal clitic omission and misplacement is reported in the spontaneous data sample of 20 first- and second-generation Serbian–English bilingual adults in Australia (Dimitrijević-Savić, 2008). The placement of object clitic pronouns after the verb instead of the second position is argued to be a contact-induced change reflecting the word order of English.

To our knowledge, there is only one other investigation of pronominal clitic placement in heritage grammars where the placement is constrained by prosody similar to Bosnian and Serbian. Ivanova-Sullivan et al. (2022) used a self-paced listening (SPL) task and aural Acceptability Judgment task (AJT) with 13 English-dominant HSs of Bulgarian and 22 Bulgarian monolinguals. Bulgarian pronominal clitics are verb-adjacent syntactic proclitics, with encliticization to the verb possible and obligatory only when there are no stressed elements to the left of the clitic. The tasks included items with clitics and NPs in the second position (grammatical for clitics, non-felicitous for NPs), and the third position (ungrammatical for clitics, typical for NPs). In both tasks, the behavior of HSs was uniform in that they did not overtly discriminate between the grammatical (preverbal) and ungrammatical (post-verbal) position of the clitics, but they could distinguish between felicitous and infelicitous NP positions in the AJT. The uniform treatment of clitic placement was attributed by the authors to language-internal optionality caused by HSs' reduced sensitivity to prosodic domains and boundaries as well as the absence of the target structure in the SL English. It is worth noting that the implemented SPL included each word as a separate fragment, including clitics. A clitic as a lone element should be perceived as infelicitous at best. Also, any "prosodic" break before a clitic, even an artificial one before the button press, will mark the start of a new prosodic unit and the clitic would be parsed as the first element, rendering clitics in either tested position infelicitous. These task details were likely why neither speaker group showed differences for the clitic position in the SPL task, and potentially a contributing factor to why HSs did not discriminate between the clitic positions in the AJT.

In sum, previous investigations of clitic realization and clitic placement in heritage grammars across different languages show that HSs are generally sensitive to the language-internal constraints on clitic placement. It is hypothesized that patterns divergent from the baseline are caused by language-internal restructuring/reanalysis and/or CLI from a SL.

3. Research questions and hypotheses

A speaker who has internalized the P2 requirement should detect clitics in the ungrammatical P3 position, resulting in lower acceptability and/or processing interference, compared to the grammatical P2 position. On the assumption that clitics are vulnerable in HLs, we expect that P3 sensitivity in Bosnian/Serbian as HL will be affected by language experience. Our research questions are as follows:

RQ 1: Is P3 sensitivity detectable through both metalinguistic (acceptability judgements) and behavioral (self-paced listening) measures in bilingual speakers of Bosnian/Serbian living in Norway?

RQ 2: What individual difference factors significantly cause modulations in P3 sensitivity, directly or through causing other factors? We will consider:

1. SL exposure onset and caregiver HL input (i.e., only one or both caregivers speaking HL)
2. LoR in the SL country
3. Amount of experience in Bosnian/Serbian (exposure and use; current and cumulative)
4. Richness of the Bosnian/Serbian environment (diversity of interlocutors, formal education, visits to HL country, literacy practices)
5. Current code-switching experience (exposure or use)

RQ 3: Is there a significant cut-off point in SL exposure onset, distinguishing HS from other speakers?

In relation to RQ 1, we predict that AJ will reveal sensitivity to P2 violations (at least in some participants) and that longer listening times (LT) will be observed at the critical region in P3 trials vs. P2 trials (at least for some participants).

Addressing RQ 2 requires (i) clarifying the causal relationships we assume between the variables of interest (based on existing literature) in the form of a DAG, then (ii) modeling the direct and total effects of these variables on the outcome measures and evaluating the fit of the formal models. We expect the individual variables capturing chronologically earlier experiences to influence the ones which relate to later and current language experience. For example, the onset of SL exposure should affect most of the variables measured later, particularly the cumulative exposure to HL, the amount of HL literacy training, the LoR in the SL country, and the amount of current HL experience. The LoR in the SL country is in turn hypothesized to have an impact on the frequency of speaking and reading in the SL, the connection to the HL country proxied by the number of visits to the HL country, etc. The causal relationships among the variables relevant to RQ 2 are represented in a simplified DAG in Figure 2 (excluding additional control variables specified in the final analysis model for ease of visual inspection).

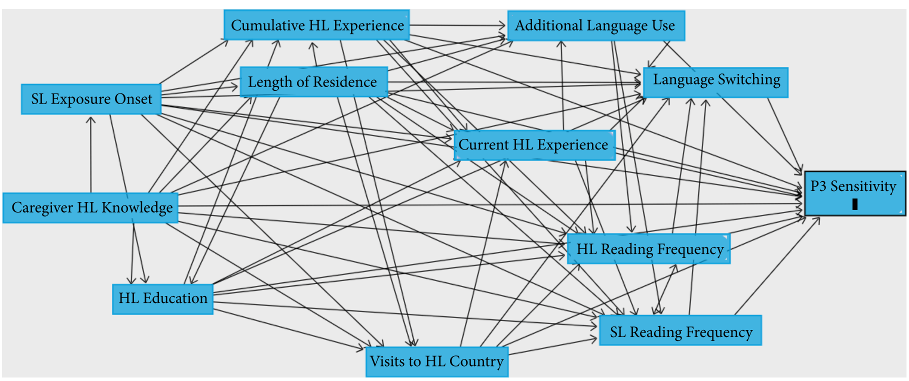


Figure 2. DAG representing the hypothesized process of how different language experience aspects cause P3 sensitivity outcomes in immigrant populations (dagitty R package, Textor et al., 2016)

We hypothesize that several variables would increase P3 sensitivity in both metalinguistic and behavioral measures. These variables include the age of onset of SL exposure, caregiver HL input, HL literacy training (in formal education or informal language training), cumulative HL experience, the size of the HL speaking community, current HL experience, HL input quality (e.g., number of proficient HL interlocutors), visits to the HL country, frequency of reading in the HL, HL dominance (in proficiency or experience).

Conversely, we expect that the LoR in the SL country, frequency of code-switching, frequency of reading in SL, and use of additional languages would have a negative direct effect on P3 Sensitivity. This could be attributed to a combination of reduced opportunities to use the HL and potential sources of CLI. Given that the P2 clitic rule is specific to Bosnian/Serbian, the use of any other language, particularly Norwegian and English (the most common L2 in Norway), could lead to greater acceptance of the ungrammatical clitic order and/or reduce the P3 interference during listening.

In terms of the decomposition into total and direct effects, we hypothesize that earlier variables, such as SL exposure onset, might not have a direct effect on the HL outcome once the direct effect of all other variables they cause is explained. The effect of the SL exposure onset on the outcome measures would therefore be indirect, through shaping the cumulative and current language experience.

Finally, in relation to RQ 3, as late child L2ers are assumed to have acquired the core properties of their L1 prior to becoming bilingual (Montrul, 2011), we expect that SL onset prior to age 7 or 8 will be associated with a reduction in P3 sensitivity. This threshold may however turn out to be non-significant once the impact of the other language experience variables is taken into account.

4. Method

4.1 Experiment design

We developed a self-paced listening (SPL) task combined with acceptability judgments (AJ) to test participants' sensitivity to the clitic position in Bosnian/Serbian. Each target sentence featured a pronominal clitic either in the grammatical P2 position or in the ungrammatical P3 position. As longer LTs are thought to reflect interference or difficulties in processing (Papadopoulou et al., 2013), sentences with clitics in P3 should yield longer processing times than sentences with clitics in P2 position, indicating P3 violation sensitivity. We expect that reduced sensitivity to P2 violations, i.e., the P3 Sensitivity, will be manifested as a smaller difference in listening time between P2 and P3 conditions, and a reduced ability to reject P3 in the AJT.

Participants listened to 40 trials with clitics in P2 (20) and P3 position (20) from one of four different lists. We included two practice trials and 20 fillers similar to experimental trials, all grammatical. The larger ratio of grammatical vs. ungrammatical trials (2:1) was meant to decrease potential shallow processing of items or desensitization to the ungrammatical items (Ferreira et al., 2009). All trials started with a preamble sentence introducing the referent to be cliticized in the following target sentence. The target sentence was divided into three fragments: the target fragment, followed by two post-target fragments. In the P3 condition, the clitic was in the third position, with either a verb or adverb occupying the second position, illustrated in (4).

The target fragment always began with a nominative-case noun or pronoun to which the clitic attached in the target-like P2 conditions. The first element was thus both the first syntactic phrase and the first prosodic word in the clause, satisfying both Serbian and Bosnian preferences for the first element type. Three additional elements followed the noun/pronoun in the target fragment, in various orders depending on the condition: the verb in present simple tense (to avoid the additional complexity of auxiliary verb and pronominal clitic clusters), pronominal clitic, and adverb. The target fragment for an item in different conditions included the same lexical items, only differently ordered, to enable closer comparison across different conditions, as the LT were likely to be affected by the frequency and length of words. We included both verb and adverb as intervening elements to introduce variation in the design, and potentially test linguistic sources of clitic variability in a future study.

Each participant heard the stimulus as pronounced by a native speaker of their HL variety which differed only in terms of pronunciation and some vocabulary. The items in the two versions were created by the first author and native speakers born and residing in Bosnia and Serbia. Task instructions and trials were recorded by two female native Bosnian-/Serbian-speakers, using the Zoom H1n MP3/Wave Handy Recorder. The assistants were instructed to pronounce all fragments as naturally as possible, regardless of the grammaticality, to avoid the ungrammatical fragments being longer or additionally marked by fillers. All cliticization rules other than the position were respected in the P3 conditions, i.e., clitics were kept unstressed and prosodically cliticized to the element on the left. Assistants were also instructed not to introduce a prosodic utterance break before the intervening element in the P3 conditions, which would have made the P3 clitic position more felicitous (in that case, the prosodic utterance boundary would have become the boundary relative to which the clitic must appear second). Also, they were instructed to make brief pauses between fragments, while maintaining the sentence intonation contour, to avoid coarticulation between fragments and ensure smooth fragmentation. Recordings were subsequently denoised, equalized in terms of volume, fragmented, and padded by 5 ms silences in Praat (Boersma & Weenink, 2022). There were no significant effects of different conditions on the duration of target fragments. The experiment was implemented in Gorilla Experiment Builder 1. LT for the target fragment constituted the time elapsed from the end of the target fragment audio to the participant's button press/signal to start the following fragment. The AJ question followed each item (presented aurally in the first two practice trials): "What did the previous sentence sound like to you, from 'very bad' on the left, to 'very good' on the right?", eliciting answers on a 7–point Likert scale, with a sad face at the left end and a smiley face at the right. We avoided the binary acceptability scale so as not to draw attention to the grammaticality of specific structures and to acknowledge the fact that heritage and attriting grammars are likely to be more variable.

The appropriateness of using this paradigm to test clitic position sensitivity was validated by a pilot study with speakers in Bosnia ($n=46$): P2 (grammatical) clitic order resulted in reliably faster LT in ms compared to the ungrammatical P3 clitic order (in the target fragment only; linear mixed effects model effect estimate: $b=-76.913$, $SE=25.098$, $t=-3.064$) and increased AJs ($b=0.876$, $SE=0.103$, $t=8.515$; Tomić et al., 2022).

(4) Example of an experimental item in four target conditions. If a word in the item had a separate version for Bosnian and Serbian, both are included separated by a slash. The underscore marks the same referent for the noun in the preamble and the clitic in the target region.

a. Preamble: Sejo/Stefan i Mersiha/Milena su uzeli novog cuku/psa.
Sejo/Stefan and Mersiha/Milena got a new dog$_i$.

b. Target: Mersiha/Milena ga dresira danima. P2 clitic pre-V
Mersiha/Milena <u>him</u>$_i$ trains for_days
*Mersiha/Milena dresira ga danima. *P3 clitic post-V
Mersiha/Milena trains <u>him</u>$_i$ for_days
Mersiha/Milena ga danima dresira. P2 clitic pre-ADV
Mersiha/Milena <u>him</u>$_i$ for_days trains.
*Mersiha/Milena danima ga dresira. *P3 clitic post-ADV
Mersiha/Milena for_days <u>him</u>$_i$ trains

c. Post — target 1: ali cuko/pas i dalje
but the dog still

d. Post — target 2: uništava sve po kući.
is destroying everything around the house.

4.2 Participants

Speakers of Bosnian and Serbian constitute large immigrant groups in Norway, with 21,599 people reporting Bosnian background and 11,888 Serbian background (Statistics Norway, January 2023). The peak of immigration for the Bosnian group occurred in early 1990s during the dissolution of Yugoslavia, the Balkan wars and Serbian aggression, when as many as 14,000 people fled Bosnia and Herzegovina to Norway, and most came as whole families (Dzamarija, 2016). Today there are three generations of immigrant or immigrant-origin Bosnian speakers in Norway (including passive Bosnian bilinguals). As for the Serbian group, a large wave of immigration from Serbia took place later, in early 2000's (Hadžibulić & Manić, 2016), mainly as guest and migrant workers. Today there are mainly first and second-generation speakers of Serbian in Norway. We recruited participants from both communities which differ in terms of social variables such as the time and reason for immigration, the size and connectedness of HL communities, etc. This increased the likelihood of obtaining a variable sample in terms of HL experience while keeping the SL influence and dominant-society variables constant. We recruited a sample with a wide range of language experiences and proficiency, with specific campaigns targeting and incentivizing passive HL speakers to participate (e.g., emphasizing that the task requires no writing). This aimed to avoid self-selection bias (i.e., only high-proficiency HL speakers participating), which could have distorted the causal effect estimates.

Seventy-five participants ($n=75$; 27 male) completed either the Serbian ($n=31$) or Bosnian version ($n=44$) of the SPL task and the Heritage Language Experience questionnaire (HeLEx) designed to document proficiency and use of up to five languages (Tomić et al., 2023). Four participants were excluded: two who were born and raised in Serbia but performed the task in Bosnian, and two with outlier responses to HeLEx questions (a balanced trilingual and a participant reporting only using their HL in Norway). Seventy-one participants (mean age 29.69, age range 13–61) were included in the final analysis: 29 Serbian and 42 Bosnian (25 male).

Figure 3 shows participant distribution based on age of SL onset and LoR in the SL country. As per the literature reviewed above, participants who arrived in Norway before the age of 6 are likely to be HSs (labeled *Heritage?*) and participants who have lived in Norway for 10 years or more are potentially AS (*Attrited?*). The remaining portion (*Baseline?*) is presumed to represent baseline speakers, individuals whose L1 has been minimally affected by residing in the SL country. We will not use this categorical distinction as a factor in the analyses. Additional figures S1–S3 showing the distribution of language experience variables for the hypothetical speaker profiles and groups are available in the OSF repository https://osf.io/4x5e3.

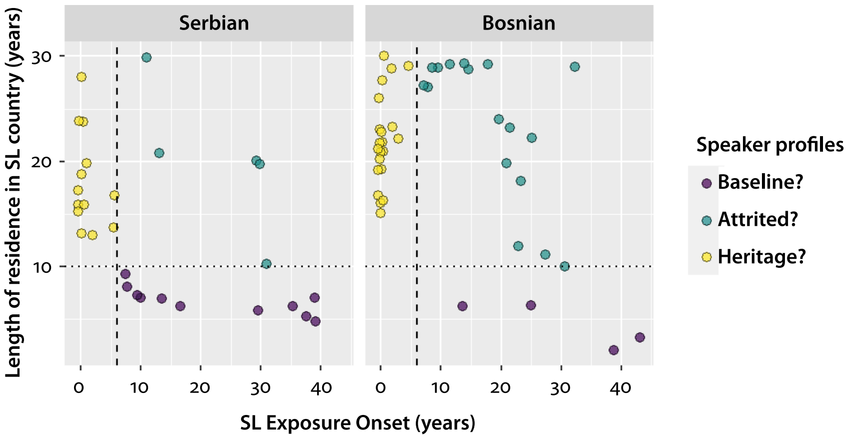


Figure 3. Scatterplot of LoR and SL exposure onset in years for each participant, with the SL exposure onset by the age 6 inclusive (vertical line) delineating the presumed threshold for HSs (in yellow) vs. others, and the LoR of 10 or more years (horizontal line) delineating presumed AS (in green) from baseline speakers (in purple)

4.3 Analysis methods

4.3.1 *Outcome variables*

In our experiment, P3 Sensitivity was measured in two ways: (i) based on the LT of the critical segment, and (ii) based on the AJ after each sentence. For each measure, we derived sensitivity scores for each participant as explained below. Before the calculation, we removed the data points for three items which were overall judged as particularly unacceptable (mean AJ below 2 standard deviations of the distribution; 213 data points, i.e., 7.46% data loss). We also removed trials with outlier LT of 3,000 ms and above (further 3.03% data loss), since the processing of these items was likely interfered with by outside factors, e.g., noisy conditions or distractions.

P3 Sensitivity as listening times difference
To obtain individual measures for the difference in LT from P2 to P3 clitic order, we fitted a linear regression model predicting LT for the target fragments, including a P3–by–Participant interaction as a main effect[1] and pre-target-fragment reading time as a covariate.[2] The goodness of fit measure, percent deviance

1. We treated Participant as a main effect rather than a random effect because we are primarily interested in individual variation and wanted to maximize the variation preserved in coefficients.

2. The average LT for each item (in the P2 and P3 conditions) by participant can be obtained by summing the participant's P2 or P3 coefficient with the individual item residuals (representing 'random error').

explained, was 75.5% for the model, indicating that most of the variance in the data was captured. We obtained a single difference measure from the two coefficients by using the percent change formula to control for individual differences in the baseline: ((P3–P2)/P2)*100), referred to as LT Difference. Larger values represent a larger estimated listening interference by P3 conditions, i.e., the sensitivity to P3 conditions. Figure S4 in the OSF repository (https://osf.io/4x5e3) shows the individual coefficients for P2 vs. P3 trials for each participant.

P3 Sensitivity as detection score

To analyze the AJ data, we adopted signal detection theory to derive P3 detection scores, i.e., D-prime (Huang & Ferreira, 2020; Swets & Green, 1978). Our goal was to determine how reliably each participant detected the ungrammatical P3 order, which would manifest as consistently judging P3 items lower than P2 items, relative to their individual judgment distribution, i.e., judgment mean.

Considering P3 order as the signal to detect (identifying it as "bad"), AJ scores were categorized either as *hits* (the correct rejection of P3), if a P3 trial was judged lower than the individual's mean acceptability score, *false alarms* (the incorrect rejection of P2), if a P2 trial was judged lower than the individual's mean score, *misses* (the incorrect acceptance of P3), if a P3 trial was judged higher than the mean score, or *correct non-detection* (the correct acceptance of P2), if a P2 trial was judged higher than the mean score.[3] We did not use the true scale midpoint to define hits or false alarms, as the metalinguistic judgment question was nonspecific and the scale was only visually labeled, leading to individual patterns of use.

The signal detection score, or D-prime, represents the difference between the sample-normalized participant's hit rate and false alarm rate, with the hit rate being the proportion of P3 hits out of all P3 trials, and the false alarm rate being the proportion of P2 trials where P3 was erroneously detected (lower than mean score). The D-prime score, hereinafter referred to as the P3 Detection Score, is thus higher in participants with a high hit rate and a low false alarm rate.

4.3.2 *Directed acyclic graph for variables of interest and structural equation modeling formal analysis*

We used SEM to estimate the direct causal effects of variables on each other and on the latent HL outcome of interest (i.e., P3 Sensitivity). An expanded causal DAG from Figure 2, which included covariate relationships and added control variables Gender and Group (Bosnian/Serbian; Figure 4), was directly trans-

3. Out of the 2563 data points, there were only two (from the same participant) where the judgment aligned with the mean score for that participant, which were counted as neither a hit nor a false alarm).

lated to SEM syntax. The DAG-based SEM model code used for the analyses, where each equation represents the theorized causes of the variable on the left of the tilde (~), is available in the associated online repository https://osf.io/x9zku. We detail how each independent variable from the DAG was operationalized in Section 4.3.3.

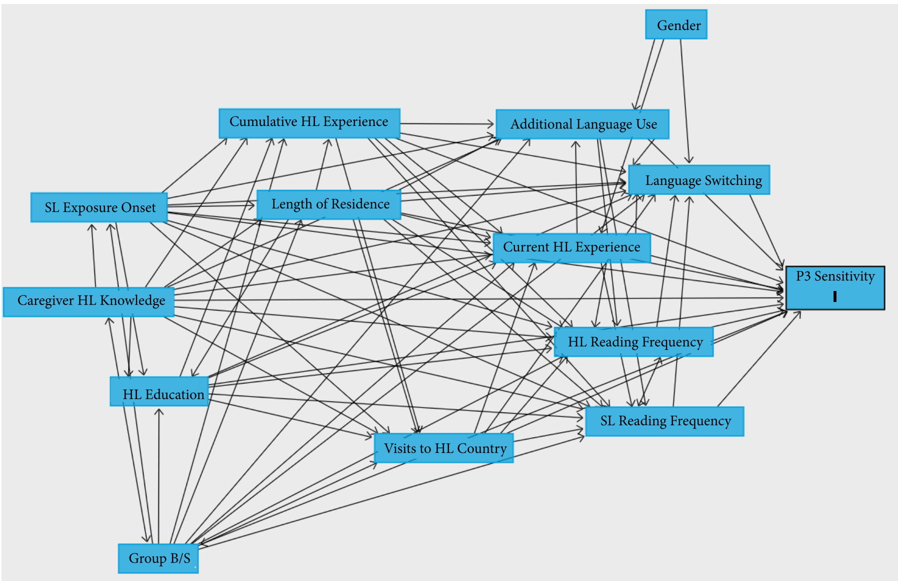


Figure 4. DAG used in the SEM analysis, including control variables of group and gender, and covariate relationships used to control for unknown covariance (bidirectional arrows in the graph)

Chronologically early determined variables, such as caregiver HL knowledge (determined at birth), SL exposure onset (age of birth or of immigration), and the LoR in the SL-dominant country (determined at SL exposure onset) were presumed to cause most of the other language experience variables.

Gender and Group (Bosnian or Serbian) were allowed to either cause or covary with many language experience variables, to control for unknown gender-based differences in socialization and unknown differences between Bosnian and Serbian communities or language practices, as well as to control for any sample imbalances in collected data (e.g., Serbian HL speakers being on average younger and more recent immigrants than Bosnian speakers).

We carried out all analyses using the same structural model, mapping the causal relationships between all variables of interest, including the individual language experience variables listed in Table 2 and the latent outcome variable of clitic position sensitivity, i.e., P3 Sensitivity. We varied the indicator variables (dependent variables) used to measure the latent P3 Sensitivity, resulting in three different measurement models. We first modeled P3 Detection Score (D-prime) and LT Difference as jointly measuring the latent P3 Sensitivity, since including several measurement variables as manifestations of a single latent outcome provides more robust estimates of direct causal effects on the outcome. We also modeled the two measures separately to determine whether they were directly caused by the same set of language experience variables.

To explore which language experience variables affect outcomes for both heritage and potentially attrited bilinguals, we selected a subset of participants excluding the participants from the "Baseline?" group, as the latter are likely to have an intact HL system. The subset included the speakers whose exposure to SL started before or at the age of 6 ($n=34$) and the speakers who had resided in the SL country for 10 or more years ($n=22$), amounting to a total of 56 participants. The same SEM analyses as for the full sample were performed.

The SEM analyses and visualizations of direct effects were carried out in R (v4.2.2, R Core Team, 2022), using the following packages for data analysis and visualization: lavaan (Rosseel, 2012), semptools (Cheung & Lai, 2023), semPlot (Epskamp, 2015). The SEM specifications, the analysis script, and the model outputs are available in the OSF repository https://osf.io/x9zku. The total causal effects of the variables hypothesized to have direct causal effects on P3 Sensitivity were estimated using linear regression and adjustmentSets function from the dagitty package (Textor et al., 2016). The function lists the variables which need to be adjusted for (i.e., included in the statistical model) to obtain the total effect of a variable of interest on the outcome, given the full DAG-represented causal model. The resulting linear regression formulas for determining total effects of each variable of interest can be found in the analysis script in the OSF repository.

4.3.3 *Individual variable choice and operationalization*

The choice and operationalization of variables was informed by causal inference principles and the best practices for formal analysis with SEM (Lei & Wu, 2007; West et al., 2012). To avoid multicollinearity, we chose a representative variable for each of the relevant dimensions of language experience, by either focusing on one modality (use or exposure) when the two modalities were highly correlated or by averaging or summing across related variables (e.g., averaging HL use proportion across conversational contexts). The use of composite variables is generally not recommended in SEM, but it was a necessary practical compromise to ensure reliable SEM results with a relatively small sample.

Proficiency estimates, initially assumed to cause HL outcomes in the original DAG, were removed from the reported SEM since they led to poor model fit or model nonconvergence. This is likely due to proficiency estimates and morphosyntactic ability, measured by P3 sensitivity, being two instantiations of the same latent Language Proficiency variable, rather than one causing the other.

Upon inspecting the variables across different groups (heritage vs. later bilinguals, Bosnian vs. Serbian origin), we discarded variables showing measurement variance, i.e., differences in how a measure is conceptualized across the sample. For example, we had indications that the HL attitudes, e.g., how much one identifies with being a national of the HL country, were conceptualized differently by different age-groups and Bosnians vs. Serbians, so we discarded this variable. Another reason to discard HL attitudes was because the younger speakers were likely to be still developing their HL attitudes under the influence of language experience, whereas the HL attitudes were likely fully formed in the older population at the time we recorded this data.

Table 2 presents the calculation breakdown of language experience variables included in the final models. The frequency of use/exposure questions had a verbal response scale which was converted to numerical values (Tomić et al., 2023, for scales and conversion). All variables were subsequently scaled to optimize model fitting.

Table 2. Variable names and operationalization summary

Variable name	Variable description
SL Exposure Onset	Age of arrival in Norway (continuous variable)
Length of Residence	The difference between current age and age of arrival in Norway
Current HL Experience	Proportion of HL used or spoken (out of the total amount of HL and SL), averaged across the 5 contexts
Cumulative HL Experience	Number of full-month equivalent use of the HL, over the lifetime
Language Switching	Summed frequencies of (intra-sentential) language switching use (scale: 0, .25, .75, 1) across 5 contexts.
HL Education	0: no HL literacy training 1: no formal HL education, but some HL literacy training (e.g., Sunday school) 2: formal HL education
Visits to HL Country	Number of visits to the HL country in a typical year.
HL Reading Frequency SL Reading Frequency	Frequencies of reading in HL and SL. 1–7 scale.
Additional Language Use	Summed frequencies of speaking languages other than HL and SL. 1–7 scale
Caregiver HL Knowledge	One caregiver speaking HL–1, Two caregivers speaking HL–2

4.3.4 *Threshold regression modeling*

To assess potential direct threshold effects of the SL Exposure Onset on P3 Sensitivity, we fit threshold regression models for the two outcome variables separately, while controlling for the direct effects of all other variables. This analysis enables both categorical and continuous treatment of the SL Exposure Onset variable, potentially capturing different aspects of HL bilingualism development (Kremin & Byers-Heinlein, 2021).

Threshold regression or two-phase regression models assume that the effect of a variable on the outcome differs before and after a certain threshold value of the variable (Fong et al., 2017; Hinkley, 1971), whereas the effects of other variables remain constant. We used a segmented model (Pastor & Guallar, 1998) to estimate two different slopes and the threshold value before and after the threshold. The hypothesis is that an earlier second language onset reduces morphosyntactic sensitivity. This would be confirmed in a segmented threshold model by a significant effect of SL exposure onset up to age 6–8, where it would stabilize, i.e., approach 0.

For this analysis, we chose a subsample of bilinguals exposed to SL before or at the age of 20 and who lived in the SL country for at least 10 years. The reasoning was to look for a developmental threshold in a tight sample with the range of SL Exposure including the proposed developmental threshold for language (6–8), with a relatively equal span of values on both sides of it. The sample excludes very late bilinguals, who are likely to differ markedly in language experience variables. This resulted in 45 participants, reduced to 42 when three participants with NAs in some responses had been excluded.

5. Results

5.1 HL outcome measures: Descriptive results

Figure 5A shows the distribution of the P3 Sensitivity scores as LT Difference and Figure 5B shows the distribution of the P3 Detection Score (D-prime) for the three initially proposed groups (Baseline?, Attrited?, Heritage?).[4] The distributions of

4. We include (a) the distribution of raw acceptability judgments by condition and purported profile, (b) the scatterplot of individual acceptability means for grammatical vs. ungrammatical items, and (c) the by-profile distribution of D-prime scores calculated by using the scale midpoint of 4 as the acceptability threshold instead of participants' individual acceptability means, in the associated Open Science Framework repository at https://osf.io/3g6ey. Individual patterns of scale use represented by the scatterplot justify the use of individual acceptability means as acceptability thresholds. Nevertheless, the distribution of the D-prime score calculated with the true midpoint as the cutoff for hits/false alarms closely resembles the reported one.

the two outcome measures showed trends, but also a great variability. The potential heritage group expectedly had the lowest average scores on both measures. Potential AS showed greater P3 Sensitivity in LT Difference, whereas the potential baseline speakers had the highest average score for the P3 Detection Score. Nevertheless, the distributions largely overlapped among the three groups, suggesting that SL Exposure Onset and the LoR in the SL country did not fully account for the score variability.

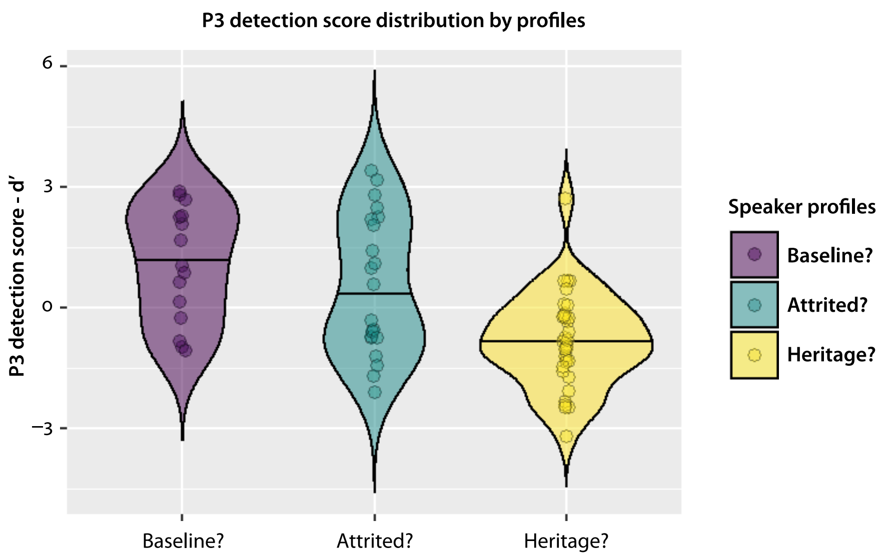
5.2 Causes of variability in outcomes: SEM results

Despite the relatively modest sample sizes, all SEM models showed marginal to excellent fit on all available model fit indices available, including the parsimony-adjusted index titled the Root Mean Square Error of Approximation (RMSEA) (the models' RMSEA was at or below .05; see Lei & Wu, 2007, West et al., 2012, for cut-offs), as shown in Tables S1 to S3 in the project OSF repository https://osf.io/qfnjt. This indicates that the theorized causal network can reliably reproduce the observed data and is likely to be a good approximation of the true causal system of producing HL outcomes.

The latent factor loadings of LT Difference and P3 Detection Score as the measures for the latent variable P3 Sensitivity for the whole participant sample ($n = 71$) suggested that they are both reliable measures for latent P3 Sensitivity (P3 Detection Score was fixed to $\lambda = 1.0$, symbolized by dashed line in Figure 6, and assumed to be significant; LT Difference: $\lambda = 0.42$, $SE = 0.1$, $z = 4.1$, $p < .001$, symbolized by asterisks on the arrow leading to in Figure 6). The latent factor loading estimate for the P3 Detection Score measure in the full sample was higher in relation to the factor loading of LT Difference, meaning that the P3 Detection Score relatively more strongly responded to the changes in the variables which we hypothesized to cause P3 Sensitivity, i.e., the variables which latent P3 Sensitivity was regressed on. In the participant subset (n = 56), the LT Difference factor loading estimate ($\lambda = 0.64$) had a closer estimate to the factor loading of P3 Detection Score (also set to $\lambda = 1.0$), suggesting that P3 Detection Score and LT Difference were even closer in estimate in this subset.

Table S1 (https://osf.io/qfnjt) presents estimates for all direct effects on all variables, including the latent P3 Sensitivity measured by the two outcome variables, for both sample sets (Model 1 for the full sample, Model 2 for the subset). The variable SL Exposure Onset had a significant direct positive effect on P3 Sensitivity in both samples (full sample $b = 0.87$, $SE = 0.4$, $z = 2.2$, $p = .028$; subset $b = 0.97$, $SE = 0.41$, $z = 2.36$, $p = .018$), as well as HL Reading Frequency (full sample $b = 0.67$, $SE = 0.24$, $z = 2.85$, $p = .004$; subset $b = 0.51$, $SE = 0.25$, $z = 2.07$, $p = .038$). Language Switching across contexts had a significant negative effect on P3 Sen-

A. The distribution of LT difference measure in P2 to P3 listening time percent change, by speaker profiles. Full lines represent the means for each profile and the dotted line marks the 0% change, i.e., no change



B. The distribution of P3 detection score measure by speaker profiles. Full lines represent the means for each profile

Figure 5.

sitivity in the entire sample ($b = -0.36$, $SE = 0.13$, $z = -2.77$, $p = .006$) and was a trending effect when the presumed baseline speakers were excluded ($b = -0.26$, $SE = 0.14$, $z = -1.89$, $p = .059$).

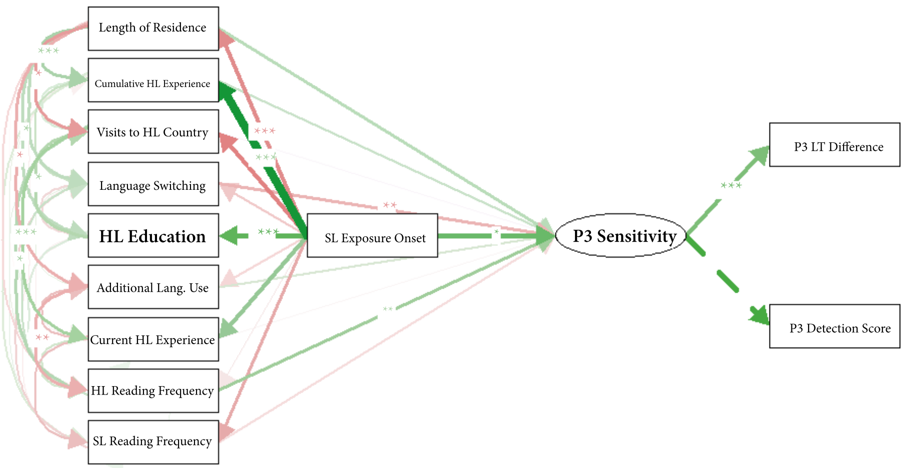


Figure 6. Standardized estimate visualizations of direct effects from the SE model for the latent outcome P3 Sensitivity, as measured by both P3 Detection Score and LT Difference, for all analyzed participants ($n = 71$). The arrow thickness and opacity represent the standardized estimates (z-values) of the effects, while color denotes the direction (green for positive, red for negative). Asterisks indicate statistical significance (based on p-values). Control variables such as Group (Bosnian/Serbian) and Gender, along with Caregiver HL Input (due to minimal variation), were excluded from this and subsequent plots

In terms of direct effects on language experience variables, Cumulative HL Experience was significantly increased, as expected, by both later SL Exposure Onset and, to a lesser extent, by LoR for both the full and subset samples (Table S1, both models). The HL Education degree was significantly increased by the later SL Exposure Onset for both sample sets (Model 1, full sample: $b = 0.77$, $SE = 0.13$, $z = 5.81$, $p < .001$; Model 2: $b = 0.82$, $SE = 0.16$, $z = 5$, $p < .001$). The number of Visits to the HL Country significantly increased the proportion of Current HL Experience for both samples (Table S1, Model 1: $b = 0.35$, $SE = 0.1$, $z = 3.39$, $p = .001$; Model 2: $b = 0.39$, $SE = 0.11$, $z = 3.47$, $p = .001$). The number of Visits to the HL Country also significantly increased HL Reading Frequency in the full sample (Table S1, Model 1: $b = 0.2$, $SE = 0.08$, $z = 2.57$, $p = .01$), while it was the Current HL Experience proportion that increased HL Reading Frequency in the subset (Model 2: $b = 0.18$, $SE = 0.08$, $z = 2.18$, $p = .029$). The Additional Languages Use reduced the HL Reading Frequency for both the full sample (Model 1: $b = -0.21$, $SE = 0.07$, $z = -3.13$, $p = .002$) and the subset (Model 2: $b = -0.19$, $SE = 0.08$, $z = -2.36$, $p = .018$). As for predictors of Language Switching, no single predictor was significant for the full sample, yet the SL Reading Frequency decreased Language Switching for the subset (Model 2: $b=-1.04$, $SE=0.42$, $z=-2.48$, $p=.013$). Longer Length of SL Residence reduced the frequency of Visits to the HL Country in both samples ($b=-0.73$, $SE=0.29$, $z=-2.53$, $p=.011$; $b=-1.05$, $SE=0.39$, $z=-2.72$, $p=.006$), whereas more HL education increased the number of Visits to the HL Country for the full sample ($b=0.3$, $SE=0.13$, $z=2.33$, $p=.02$).

5.3 Comparing causes of the P3 Detection Score vs. LT Difference

5.3.1 *Direct effects on P3 Detection Score*

The direct effect estimates from the models predicting P3 Detection Score for both the full sample and the subset (Tables S2 and S3, Model 1, https://osf.io/qfnjt) suggest that the metalinguistic judgment accuracy was increased by the HL Reading Frequency (full sample: $b=0.72$, $SE=0.25$, $z=2.83$, $p=.005$), and negatively affected by Language Switching (full sample: $b=-0.43$, $SE=0.14$, $z=-3.14$, $p=.002$). The SL Exposure Onset was not a significant direct cause of latent P3 Sensitivity when measured by P3 Detection Scores in either sample set.

5.3.2 *Total effects on P3 Detection Score*

The SL Exposure Onset had a significant total effect on P3 Detection Score both in the full sample ($b=1.107$, $SE=0.180$, $t=6.147$, $p<.001$, Table S4, Model 1, https://osf.io/qfnjt) and the subset ($b=1.145$, $SE=0.265$, $t=4.328$, $p<.001$, Table S5, Model 1). HL Reading Frequency total effect is similar to its direct effect, so it was significant in the full sample ($b=0.596$, $SE=0.291$, $t=2.046$, $p=.046$) and trending in the subset. Since Language Switching effect amounts to its direct effect only, i.e., it does not indirectly affect P3 Sensitivity through any other variables, its total effect was significant as well using the adjustment sets and the linear regression procedure, in the full sample and the subset. No other variables had a significant total effect on P3 Detection Score.

5.3.3 *Direct effects on LT Difference*

The latent P3 Sensitivity as measured by LT Difference was increased by later SL Exposure Onset for both the full sample (Table S2, Model 2; $b=0.7$, $SE=0.34$, $z=2.08$, $p=.038$) and the subset (Table S3, Model 2; $b=1$, $SE=0.37$, $z=2.69$, $p=.007$). In the subset model (Table S3, Model 2), Current HL Experience, i.e., higher proportion of HL over SL speaking and hearing across conversational contexts, was also a significant positive cause of P3 Sensitivity ($b=0.28$, $SE=0.13$, $z=2.24$, $p=.025$).

5.3.4 *Total effects on LT Difference*

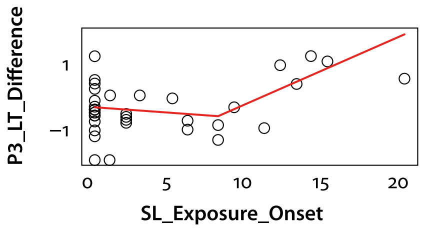
As for total effects on LT Difference, the SL Exposure Onset was a positive significant cause in both samples, full ($b = 0.618$, $SE = 0.124$, $t = 4.966$, $p < .001$, Table S4, Model 2) and the subset ($b = 0.826$, $SE = 0.167$, $t = 4.945$, $p < .001$, Table S5, Model 2). No other effects had a significant total causal effect on LT Difference in the full sample. As its direct effect is nearly identical to its total effect, Current HL Experience also had a significant total positive effect on LT Difference in the subset of participants ($b = 0.349$, $SE = 0.137$, $t = 2.540$, $p = .015$) using the linear regression procedure.

5.4 Exploring possible threshold age for SL exposure onset for HL outcomes

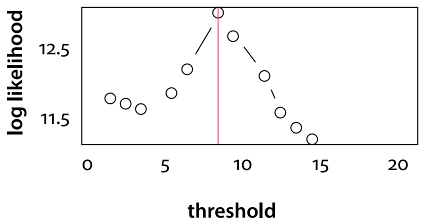
We used data-driven threshold regression to explore whether there is a developmental-like threshold of age of SL Exposure Onset before and after which bilinguals (SL exposure < 21, LoR > 10, $n = 42$) would show different trends of P3 sensitivity, while controlling for the direct effects of other language experience variables and confounders. The threshold analysis included observed SL Exposure Onset values as candidates for the threshold point.

As for P3 Detection Score, the segmented threshold model identified the threshold age of 5 inclusive. Nevertheless, the SL Exposure Onset effect was not significant before or after this age, and the threshold model did not perform better than the standard linear model. The full model fit is in Table S6, Model 1, https://osf.io/qfnjt.

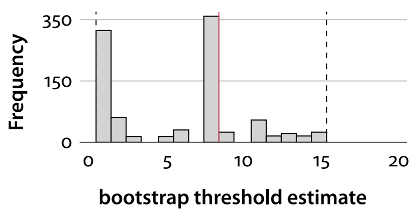
The results of the segmented threshold model for LT Difference indicated a significant threshold point of 8 years of age for SL Exposure Onset (std. error for the threshold point estimate ± 3.827), with two different regression slopes up to and after this threshold. The slope for the direct effect of SL Exposure Onset up to the threshold age of 8 inclusive was slightly negative and not significant ($b = -0.032$, $SE = 0.054$, $t = -0.588$), whereas it was positive and significant after the threshold ($b = 0.231$, $SE = 0.108$, $t = 2.145$, $p = .041$). The variable Current HL Experience was also a significant positive predictor of LT Difference in the threshold model ($b = 0.335$, $SE = 0.162$, $t = 2.064$, $p = .049$). The threshold model results suggest that bilinguals exposed to SL up to the age of 8 do not show a unified variation in clitic position sensitivity driven by the age of SL Onset. However, the P3 Sensitivity significantly increases for those exposed to SL after age 8, with each additional year of exposure. Figure 7 shows the threshold variable fit, log-likelihoods for different thresholds, and bootstrap threshold confidence interval for the threshold estimate. The threshold model output for LT Difference is in Table S6, Model 2, https://osf.io/qfnjt.



A. Estimated SL Exposure Onset as a threshold variable (red line) vs. observed data (points)



B. Log-likelihood estimates for models with different threshold points, with the highest log-likelihood marked at threshold age 8 (in red)



C. Distribution of the threshold point estimates from a bootstrap process with 1000 iterations, with the age 8 marked in red as the most frequent outcome for the threshold point

Figure 7.

The results of the Maximum Likelihood Ratio Test testing the null hypothesis (SL Exposure Onset as a continuous linear effect) vs. the alternative hypothesis (SL Exp. Onset as two different slopes up to and after age 8) confirmed that there is a significant, model-robust threshold effect at the age of 8 ($p = .032$).

6. Discussion

We investigated sensitivity to pronominal clitic placement in adult bilingual speakers of Bosnian and Serbian in Norway. Participants were of three hypothetical profiles: heritage, attrited, and fully proficient speakers. Employing an online self-paced listening task complemented by AJ questions, we aimed to investigate the effects of individual experience factors on the sensitivity to clitic placement violations, and identify the aspects of language experience exerting the strongest impact on this language outcome. Crucially, we aimed to unveil causal relationships between the variables of interest (Paradis, 2023; De Cat & Unsworth, 2023). The HeLEx questionnaire (Tomić et al., 2023) was used to gather data on individual variables, and statistical modeling was led by causal inference principled, informed by a DAG (a directed acyclic graph) representing the hypothesized causal relationships between variables (Hernán et al., 2019; Westreich & Greenland, 2013).

Our results align with previous research showing variable performance with pronominal clitic placement in adult bilinguals (e.g., Ivanova-Sullivan et al., 2022; Rinke & Flores, 2014). The impact of individual differences in language experience on P3 Sensitivity was reliably detected both through self-paced LT *and* AJ (RQ 1), unlike previous research on heritage Bulgarian (a language with a similar clitic system) showing no differences in self-paced listening time (cf., Ivanova-Sullivan et al., 2022). This is likely since we increased sample size and diversity, maximized signal-to-noise ratio in dependent variables, and considered more individual variables. Differences in the Bosnian/Serbian vs. Bulgarian clitic systems or the particular HL/SL pairs could also have played a role.

The varied, inter-generational bilingual sample presented a unique testing ground for RQ 2. Based on assumed thresholds in the traditional variables of bilingual language experience, SL Exposure Onset and LoR, we hypothesized three speaker profiles: HSs (onset of SL exposure up to age 6), AS (LoR of 10 years or more) or fully proficient speakers (i.e., the baseline, with SL Exposure Onset > 6 and LoR < 10). The presumed HSs scored lower as a group on both measures. Nevertheless, the substantial overlap between the groups suggested from the outset that traditional variables such as age of exposure to SL and LoR in the SL country would be insufficient to account for the variability in outcome measures. Our analyses aimed to uncover which individual variables explained P3 Sensitivity, and how these variables predicted each other. To answer RQ 2, we considered both the full sample of participants and a sub-sample excluding the presumed baseline speakers.

A Structural Equation Model (SEM) for the latent variable P3 Sensitivity using both outcome measures revealed that the P3 Detection Scores (based on AJ) were better predicted by the network of individual background variables, compared to differential listening times (LT). This could be due to LT being more confounded with other factors not considered in our study, such as word frequency, participants' vocabulary size or ease of access, executive function abilities, external listening conditions, etc. Nevertheless, the relatively lower latent factor loading for LT Difference could also be due to it being caused by a few selected variables (e.g., SL Exposure Onset), rather than the entire causal network.

We proceeded to test the two outcome measures for P3 Sensitivity separately to determine their respective significant causes. Using causal inference, we decomposed the effects of language experience variables on the clitic position sensitivity in terms of direct and total effects. We hypothesized that the SL Exposure Onset effect would not be a direct cause of P3 Sensitivity, but that its total effect would be significant due to its indirect effects. This was confirmed in the P3 Detection Score model, where the SL Exposure Onset did not have a significant direct effect on the measure, but it did have a significant total effect through all causal paths, as SL Exposure Onset significantly caused most of the other variables.

For differential LT, on the other hand, there was a significant total and direct effect of SL Exposure Onset, suggesting that later exposure to SL may be necessary to solidify P3 Sensitivity, or limit CLI from SL. Nevertheless, this direct effect may have been amplified by unmeasured causes or confounds of P3 Sensitivity correlated with SL Exposure Onset, such as age. For example, older age could have globally increased LT, amplifying the differential LT, or it could have increased the time to recover from hearing the violation. We explored the effect of SL Exposure Onset further in the threshold analysis.

In addition to SL Exposure Onset, the current HL use across contexts was also a significant direct positive cause of differential LT if the presumed baseline speakers (with SL Exposure Onset > 6 and LoR < 10) were excluded. This suggests that modulations in current HL experience might not significantly affect clitic sensitivity in presumed baseline speakers, as their HL is stable, whereas a higher proportion of current HL experience might be critical for maintaining clitic sensitivity in presumed attriters and HSs.

For the metalinguistic P3 Detection Score, the only significant direct effects were Code Switching (with a negative effect) and HL Reading Frequency (with a positive effect). HL Reading Frequency was itself affected by various aspects of the HL experience.

Interestingly, LoR was not a significant direct or total causal effect for any of the models or subsamples. Indeed, many of our participants who have been in Norway the longest maintained their HL well (e.g., through reading frequently in the HL). This shows that attrition (at least in relation to P3 Sensitivity) is not automatic for those who immigrated long ago. How much and how the native language continues to be used is more impactful.

It is important to note that causal interpretations from observational data have limitations (Pearl, 2012). These limitations arise because results depend on assuming the correct causal network and including all relevant variables. There is a risk of missing important variables or including biasing ones. Additionally, ensuring the unidirectionality of variable effects is challenging. When measured, a caused variable A contains the variance of the variable which caused it, B. Nevertheless, the variable A also likely affected the variable B which had originally *caused* it by the time that they were measured. To disentangle the directionality of effects measured at a single point, each variable would need to be measured at many different points. Accordingly, the direct effects of HL Reading Frequency (positive effect) and Language Switching (negative effect) on P3 Sensitivity can be interpreted as effects in their own right, but they can also contain some residual variance of unobserved variables they are caused by, and/or the 'backflow' of the variables that they cause. HL Reading Frequency appears to be a true direct cause of P3 Sensitivity. Unlike spoken input, written texts more rarely contain non-standard examples of clitic positioning, allowing for more stable acquisition of target-like forms. Additionally, reading requires more attention and cognitive effort, further solidifying P2 clitic order representation. Nevertheless, the HL Reading Frequency is likely also significant because it is caused by the extent of HL literacy training and education, a variable which was difficult to capture invariantly in a sample of bilinguals with diametrically different HL training opportunities (formal schooling for potential attriters vs. weekend school for HSs). On the other hand, Language Switching had a negative direct effect on P3 Sensitivity. Frequent code-switching may obscure the correct positioning of HL clitics or introduce variability in clitic positioning due to CLI. Nevertheless, it is also likely that more code-switching use is a potential HL avoidance strategy, i.e., a symptom of lower proficiency or dominance in HL.

RQ 3 asked whether there is a significant cut-off point in SL Exposure Onset, distinguishing HS from other bilingual speakers in terms of clitic position sensitivity. We investigated this in a restricted sample ($n = 42$), excluding presumed baseline speakers and those first exposed to the SL after age 20. We found a threshold for the outcome measure of LT difference only, yet such that later SL Exposure Onset increased the measure only for speakers exposed to SL after the age of 8. Previous research on the acquisition of clitics suggests that P2 is mastered by monolingual children by the age of 5 (Varlokosta et al., 2016). If a child's experience in their L1 is reduced before that age (as a result of exposure to another language), mastery would be delayed and/or take a different development path. Indeed, the onset of exposure to the societal language has been shown to predict proficiency in the HL (Flores and Barbosa, 2014). The lack of a significant effect of SL Exposure Onset up to age 8 is surprising on that account. The significant positive impact of SL Exposure Onset after age 8, likely after P2 syntax was acquired, suggests a protective effect against attrition for those exposed to SL after age 8, rather than marking a clear threshold for HSs. This coincides approximately with the onset of formal schooling in Norway (the SL country), resulting in a steep increase of SL exposure and use (mirrored by a decrease in HL exposure and use), and the development of literacy in the SL rather than the HL. Our findings suggest that every year of schooling in HL (prior to SL onset) appears to confer a protective effect against attrition (after 10 years or more of SL experience).

The lack of significant threshold on P3 Detection scores and the variable distribution of LT Difference up to the age of 8 lead us to conclude that the onset of SL exposure is better interpreted as a proxy for the truly causal language experience variables rather than a determining factor in itself. Instead of discrete groups, there appears to be a continuum between HSs and AS, with some speakers affected both in their acquisition of the HL and their loss of it. Further research is needed to ascertain whether and how distinct profiles of speakers in immigrant contexts can be identified empirically based on how their individual variables shape their language outcomes.

Conclusion

This study has demonstrated that P2 clitic placement in Bosnian and Serbian is vulnerable in bilingual speakers in immigrant contexts. Using an individual differences approach, we studied a diverse group of adult speakers who either emigrated to Norway in childhood or adulthood or were born there to immigrant parents. A key methodological advancement of our study is the use of causal inference modeling, which enables us to untangle the complex interactions between language experience variables and uncover their direct and indirect effects on morphosyntactic sensitivity. This demonstrates that age-related effects are proxies for the quantity and quality of language exposure. Our findings underscore the importance of literacy practices, particularly reading, as one of the most significant factors promoting sensitivity to P2 violations as a metalinguistic measure. In contrast, current HL use across various contexts was the key driver of clitic position sensitivity in the processing measure, along with later SL Exposure Onset (if later than the age of 8). Notably, the LoR in the SL country did not predict attrition, suggesting that long-term residence alone is insufficient to explain HL outcomes. This study challenges the binary distinction between heritage and AS, proposing instead that language outcomes exist along a continuum influenced by a web of individual language experiences.

Funding

This work received funding from the AcqVA Aurora Center at UiT The Arctic University of Norway. Open Access publication of this article was funded through a Transformative Agreement with UiT The Arctic University of Norway.
Open Access publication of this article was funded through a Transformative Agreement with UiT The Arctic University of Norway.

CRediT statement

A.T.: Conceptualization, Methodology, Exp. Programming, Investigation, Validation, Formal analysis & Visualization, Data Curation, Writing — original draft, Writing — editing. Y.R.: Conceptualization, Methodology, Project administration, Funding acquisition, Writing — original draft, Writing — editing. F.B.: Conceptualization, Project administration, Funding acquisition, Writing — original draft. C.D.C.: Conceptualization, Methodology, Supervision of formal analysis, Mentorship, Writing — original draft, Writing — editing.

Data Availability

All experiment materials, anonymized data, and analysis scripts can be found in the Open Science Framework repository at https://osf.io/x9zku.

Acknowledgements

We would like to thank the AcqVA Aurora Center community and affiliates for valuable feedback, and our assistants Aida Hajduković Kristiansen, Bedrana Alić, Azra Talam-Habib, and Andrijana Tomić, for their valuable assistance on task development. We thank our participants and the Bosnian and Serbian communities in Norway.

References

doi Anderssen, M., Bentzen, K., & Rodina, Y. (2012). Topicality and complexity in the acquisition of Norwegian object shift. *Language Acquisition*, *19*(1), 39–72.

Boersma, P., & Weenink, D. (2022). *Praat: doing phonetics by computer [Computer program].* Version 6.2.06.

doi Bošković, Ž. (2020). On the syntax and prosody of verb second and clitic second. In R. Woods & S. Wolfe (Eds.), *Rethinking Verb Second* (pp. 503–535). Oxford University Press.

Bylund, E. (2009). Maturational constraints and first language attrition. *Language Learning*, *59*(3), 687–715.

Cheung, S., & Lai, M. (2023). *semptools: Customizing Structural Equation Modeling Plots*. R package version 0.2.10.1 [computer software], https://sfcheung.github.io/semptools/

De Cat, C., & Unsworth, S. (2023). So many variables, but what causes what? *Journal of Child Language*, *50*(4), 832–836.

Dimitrijević-Savić, J. (2008). Convergence and attrition: Serbian in contact with English in Australia. *Journal of Slavic Linguistics*, *16*(1), 57–90.

Dzamarija, M. T. (2016). Bosniere — integreringsvinnerne? [Bosnians — intergration winners?] https://www.ssb.no/befolkning/artikler-og-publikasjoner/bosniere-integreringsvinnerne

Epskamp, S. (2015). semPlot: Unified visualizations of structural equation models. *Structural Equation Modeling*, *22*(3), 474–483.

Ferreira, F., Engelhardt, P. E., & Jones, M. W. (2009). Good enough language processing: A satisficing approach. In *Proceedings of the 31st Annual conference of the Cognitive Science Society* (Vol. *1*, pp. 413–418). Cognitive Science Society.

Flores, C., & Barbosa, P. (2014). When reduced input leads to delayed acquisition: A study on the acquisition of clitic placement by Portuguese heritage speakers. *International Journal of Bilingualism*, *18*(3), 304–325.

Fong, Y., Huang, Y., & Gilbert, P. B. (2017). chngpt: threshold regression model estimation and inference. *BMJC Bioinformatics*, *18*, 454.

Franks, S., & King, T. H. (2000). *A handbook of Slavic clitics*. Oxford University Press.

Hadžibulić, S., & Manić, Z. (2016). My life abroad: The nostalgia of Serbian immigrants in the Nordic countries. *Crossings: Journal of Migration and Culture*, *7*(1), 97–114.

Hernán, M. A., Hsu, J., & Healy, B. (2019). A second chance to get causal inference right: A classification of data science tasks. *CHANCE*, *32*(1), 42–49.

Hinkley, D. V. (1971). Inference in Two-Phase Regression. *Journal of the American Statistical Association*, *66*(336), 736–743.

Huang, Y., & Ferreira, F. (2020). The application of signal detection theory to acceptability judgments. *Frontiers in Psychology*, *11*, 498268.

Ivanova-Sullivan, T., Sekerina, A. I., Tofighi, D., & Polinsky, M. (2022). Language-internal reanalysis of clitic placement in heritage grammars reduces the cost of computation: evidence from Bulgarian. *Languages*, *7*, 24.

Karayayla, T., & Schmid, M. S. (2019). First language attrition as a function of age at onset of bilingualism: First language attainment of Turkish-English bilinguals in the United Kingdom. *Language Learning*, *69*(1), 106–142.

Kasparian, K., & Steinhauer, K. (2016). Confusing similar words: ERP correlates of lexical-semantic processing in first language attrition and late second language acquisition. *Neuropsychologia*, *93*(A), 200–217.

Kolaković, Z., Jurkiewicz-Rohrbacher, E., Hansen, B., Filipović Đurđević, D., & Fritz, N. (2022). *Clitics in the wild: Empirical studies on the microvariation of the pronominal, reflexive and verbal clitics in Bosnian, Croatian and Serbian* (OpenSlavicLinguistics7). Language Science Press.

Köpke, B., & Schmid, M. S. (2004). First language attrition: The next phase. In M. S. Schmid, B. Köpke, M. Keijzer, & L. Weilemar (Eds.), *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues* (pp. 1–43). John Benjamins.

Kremin, L. V., & Byers-Heinlein, K. (2021). Why not both? Rethinking categorical and continuous approaches to bilingualism. *International Journal of Bilingualism*, *25*(6), 1560–1575.

Kupisch, T., & Polinsky, M. (2022). Language history on fast forward: Innovations in heritage languages and diachronic change. *Bilingualism: Language and Cognition*, *25*(1), 1–12.

Lei, P. W., & Wu, Q. (2007). Introduction to structural equation modeling: Issues and practical considerations. *Educational Measurement: Issues and Practice*, *26*(3), 33–43.

Montrul, S. (2011). Introduction: The Linguistic Competence of Heritage Speakers. *Studies in Second Language Acquisition*, *33*(2), 155–161.

Pallier, C. (2007). Critical periods in language acquisition and language attrition. In B. Köpke, M. S. Schmid, M. Keijzer, & S. Dostert (Eds.), *Language attrition: Theoretical perspectives* (pp. 155–168). John Benjamins.

Papadopoulou, D., Tsimpli, & Amvrazis, N. (2013). Self-paced listening. In J. Jegerski, & B. VanPatten (Eds.), *Research Methods in Second Language Psycholinguistics*. Routledge.

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

Pastor, R., & Guallar, E. (1998). Use of two-segmented logistic regression to estimate change-points in epidemiologic studies. *American Journal of Epidemiology*, *148*(7), 631–642.

Pearl, J. (2012). The causal foundations of structural equation modeling. In R. H. Hoyle (Eds.), *Handbook of structural equation modeling* (pp. 68–91). The Guilford Press.

Pérez-Leroux, A. T., Cuza, A., & Thomas, D. (2011). Clitic placement in Spanish-English bilingual children. *Bilingualism: Language and Cognition*, *14*(2), 221–232.

Pešikan, M. (1958). O mestu enklitike u rečenici [About the place of the enclitic in the sentence]. *Naš jezik*, *9*(710), 305–311.

Progovac, L. (1996). Clitics in Serbian/Croatian: Comp as the second position. In A. Halpern & A. Zwicky (Eds.), *Approaching Second: Second Position Clitics and Related Phenomena* (pp. 411–428). CSLI Publications.

Radanović-Kocić, V. (1988). The Grammar of Serbo-Croatian Clitics: A Synchronic and Diachronic Perspective. Ph.D. dissertation, University of Illinois, Urbana.

R Core Team (2022). *R: A language and environment for statistical computing, v. 4.2.2*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rinke, E., & Flores, C. (2014). Morphosyntactic knowledge of clitics by Portuguese heritage bilinguals. *Bilingualism: Language and Cognition*, *17*(4), 681–699.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Rothman, J., Bayram, F., DeLuca, V., Di Pisa, G., Dunabeitia, J. A., Gharibi, K., … & Wulff, S. (2023). Monolingual comparative normativity in bilingualism research is out of "control": Arguments and alternatives. *Applied Psycholinguistics*, *44*(3), 316–329.

Shin, N., Cuza, A., & Sánchez, L. (2023). Structured variation, language experience, and crosslinguistic influence shape child heritage speakers' Spanish direct objects. *Bilingualism: Language and Cognition*, *26*(2), 317–329.

Smith, G., Spelorzi, R., Sorace, A., & Garraffa, M. (2023). Inter-generational attrition Language transmission between long-term UK residents and heritage speakers of Italian on production of clitic pronouns. *Linguistic Approaches to Bilingualism.*

Swets, J.A., & Green, D.M. (1978). Applications of signal detection theory. In H.L. Pick, H.W. Leibowitz, J.E. Singer, A. Steinschneider, & H.W. Stevenson (Eds.), *Psychology: From Research to Practice* (pp. 311–331). Springer.

Tomić, A., Rodina, Y., Bayram, F., & De Cat, C. (2023). Documenting heritage language experience using questionnaires. *Frontiers in Psychology*, *14*, 1131374.

Tomić, A., Rodina, Y., Bayram, F., & De Cat, C. (2022). *Individual variation in heritage language morphosyntactic processing: Bosnian Clitics in Norway*. Poster presented at the 13th International Symposium on Bilingual and L2 Processing in Adults and Children, Tromsø, Norway. Available at https://osf.io/63u48

Textor, J., van der Zander, B., Gilthorpe, M., Liskiewicz, M., & Ellison, G. (2016). Robust causal inference using directed acyclic graphs: The R package 'dagitty'. *International Journal of Epidemiology*, *45*(6), 1887–1894.

Varlokosta, S., Belletti, A., Costa, J., Friedmann, N., Gavarró, A., Grohmann, K.K., Guasti, M.T., Tuller, L., Lobo, M., Anđelković, D., & et al. (2016). A cross-linguistic study of the acquisition of clitic and pronoun production. *Language Acquisition*, *23*(1), 1–26.

West, S.G., Taylor, A.B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R.H. Hoyle (Eds.), *Handbook of Structural Equation Modeling* (pp. 209–231). The Guilford Press.

Westreich, D., & Greenland, S. (2013). The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients. *American Journal of Epidemiology*, *177*(4), 292–298.

Address for correspondence

Aleksandra Tomić
Department of Language and Culture
UiT The Arctic University of Norway
Hansine Hansens veg 18
N-9037 Tromsø
Norway

aleksandra.tomic@uit.no
https://orcid.org/0000-0002-2561-4602

Co-author information

Yulia Rodina
Department of Language and Culture
UiT The Arctic University of Norway
yulia.rodina@uit.no

Fatih Bayram
Department of Language and Culture
UiT The Arctic University of Norway
fatih.bayram@uit.no

Cecile De Cat
School of Languages, Cultures and Societies
University of Leeds
C.DeCat@leeds.ac.uk

Publication history

Date received: 27 February 2024
Date accepted: 20 October 2024
Published online: 18 November 2024