Original article

Corresponding author:

Arthur Stepanov, Center for Cognitive Science of Language, University of Nova Gorica, Vipavska 13, SI-5261 Slovenia

Email: arthur.stepanov@ung.si

Anomaly detection in processing of complex syntax by early L2 learners

Arthur Stepanov

University of Nova Gorica, Nova Gorica, Slovenia

Sara Andreetta

SISSA International School for Advanced Studies, Trieste, Italy

Penka Stateva

University of Nova Gorica, Nova Gorica, Slovenia

Adam Zawiszewski

University of the Basque Country (UPV/EHU), Vitoria-Gasteiz, Spain

Itziar Laka

University of the Basque Country (UPV/EHU), Vitoria-Gasteiz, Spain

Abstract:

This study investigates the processing of long-distance syntactic dependencies by native speakers of L1 Slovenian who are advanced learners of L2 Italian, compared with monolingual Italian speakers. Using a self-paced reading task, we compare sensitivity of the early-acquired L2 learners to syntactic anomalies in their L2 in two empirical domains: 1) syntactic islands, for which the learners' L1 and L2 grammars provide a converging characterization, and 2) verb-clitic constructions, for which the respective L1 and L2 grammatical descriptions diverge. We find that although our L2 learners show native-like processing patterns in the former, converging, grammatical domain, they may nevertheless perform non-native-like with respect to syntactic phenomena in which the L1 and L2 grammars do not align, despite the early age of L2 acquisition. Implications for theories of L2 acquisition and endstate are discussed.

Keywords:

sentence processing, clitic, syntactic island, bilingualism, Slovenian, Italian

1. Introduction

One of the central issues in contemporary non-native language research is whether there exists a principled difference between how a given language is processed by native vs. non-native speakers, as well as the extent to which native-like performance in L2 is in principle possible, also taking into account the nature of biological and environmental factors that affect it. This issue is part of a larger area of study that concerns itself with the endstate of L2 acquisition, or *ultimate attainment*, which has traditionally been evaluated in the context of factors such as age of acquisition (AoA), proficiency, level of exposure, influence of L1 and working memory span (Johnson and Newport 1989; Perani and Abutalebi 2005; Juffs, 2005). Some authors argued that full success in native-like L2 attainment is possible among at least some proportion (around 15% percent; cf. Birdsong, 1999: 14-15) of learners who have been immersed in a language acquisition environment for a substantial length of time. Nevertheless, the picture that emerges when comparing L1 speakers and highly proficient L2 bilinguals is not very clear. Some authors report no differences between both populations (Kotz, Holcomb and Osterhout, 2008; Rossi et al., 2006) while others do (Chen et al. 2007; Hahne, 2001; Hahne and Friederici; 2001; Mueller et al., 2005, 2007; Weber-Fox and Neville, 1996). In particular, a number of studies found that speakers who reliably performed at the native level in global judgment tasks, can still show non-native performance in online processing tasks in their L2 (e.g. Marinis et al., 2005; Papadopoulou and Clahsen, 2003). These findings are in line, in particular, with theories such as Representational Deficit Hypothesis that maintain that global native-like attainment is in principle not possible, and that even alleged native-like knowledge is only apparent, most likely arrived at by fundamentally different means from those used by native speakers (Bley-Vroman, 1989, 2009; Hawkins and Hattori, 2006; Newmeyer, 1998). The question is then which parts of language could be acquired native-like, and which could not (Lardiere, 2013).

In this regard, it has been conjectured that grammatical phenomena in L2 whose grammatical characterization substantially differs from that in L1 may present substantial obstacles in achieving a native-like performance. For instance, in the domain of syntax, non-native L2 performance in the form of age and/or proficiency effects in processing is often observed with syntactic phenomena that had no correlate in the participants' L1 (e.g. Weber-Fox and Neville, 1996, Mueller et al., 2005, Ojima et al., 2005, Chen et al., 2007).

In a different, though closely related, line of L2 research, a question is asked whether language learners acquire not only the grammar of their target languages, but also language-particular processing strategies, routines, and preferences or, rather, there is something like a "universal parser" which will respond to any language with a given set of structural properties in predictable ways. Theories supporting the latter alternative include, in particular, the "full transfer/full access/full parse" model (Dekydtspotter et al., 2006; Schwartz and Sprouse, 1996; White, 2000), and the Fundamental Identity Hypothesis (Hopp, 2007). These models share the prediction that with increased proficiency, L2 processing signatures will approximate, if not coincide with, the processing routines of L1, as both are ultimately based on the fundamentally same psychological and/or neurocognitive mechanisms. Given the considerations of explanatory parsimony, this alternative can even be viewed as the experimental null hypothesis (Schwartz, 1987). In contrast, theories favoring the former alternative may include the Shallow Structure Hypothesis (Clahsen and

Felser, 2006), and the procedural/declarative model (Ullman, 2001 and later works). These theories argue that the L1 and L2 processing mechanisms must be fundamentally different, the latter being in some way impoverished and/or based on different cognitive structures from the former. The different, often diverging, views on the possibility and extent of the ultimate L2 attainment as well as the nature of the task involved highlight the need to better understand the connection between speakers' internal grammars, on the one hand, and processing mechanisms that come into play in L2 acquisition and its relation to L1, on the other.

In this article we report the results of a behavioral study of processing patterns of L1 Slovenian-L2 Italian speakers in comparison with monolingual L1 Italian speakers in a specific area of complex syntax which involves syntactic long-distance dependencies, namely, syntactic islands and verb-clitic constructions. Here we ask whether advanced L2 learners with an early AoA are sensitive to syntactic anomalies in complex L2 sentences if these sentences receive a converging characterization in the learners' L1 and L2 grammars, as compared to those in which the two respective grammatical descriptions diverge yielding different surface representations. The principal purpose of focusing on anomaly detection in these constructions during sentence processing is to try to obtain a finer grained measure of interaction between grammar and processing routines in L2, as well as the kind(s) of information the L2 speakers use to identify the anomalies.

The choice of long-distance dependencies for the present study is particularly interesting because these dependencies cannot be learned simply from positive evidence available in the input, and therefore are only learnable, if at all, either through L1 transference or indirect negative evidence (on the common assumption that there is no direct negative evidence). In other words, the main rationale for our inquiry comes from considerations of learnability: a L2 learner faces a challenging learning task in syntactic domains where positive evidence, in the form of direct surface cues or otherwise, is generally unavailable. These are domains of knowledge typically described in linguistic theory by syntactic constraints and discussed shortly below. The role of potential substitutes for such negative data, such as postulating an innate mechanism of Universal Grammar (UG), the role of L1 language transfer, and other factors have been a long-standing point of debate in the L2 research (e.g. Schwartz and Sprouse, 1996; Bley-Vroman, 1989; Belikova and White, 2009; White, 1989, 2003). Without taking a stand on this debate, we believe that comparing L2 learners' knowledge in the domains where the L1 and L2 grammars converge to those in which both grammars diverge is warranted, as it has a potential of contributing to comparisons between different theories of L2 learning, as well as better understanding of mechanisms involved in the L2 endstate.

2. Theoretical preliminaries

2.1. Long-distance dependencies: Subject islands

Long-distance dependencies are a hallmark of natural language. A given lexical item, such as a *wh*-phrase, can appear potentially indefinitely far away from a position to which it is thematically related, such as a verb. For instance, in (1) the *wh*-phrase *which senator* is separated from its original position of a complement of *arrest* by two clausal boundaries:

(1) Which senator$_i$ do you believe [$_{CP}$ they said in the news [$_{CP}$ the police would arrest e_i tomorrow]]?

Since the early stages of modern syntactic theory it has been known that certain types of structural boundaries are opaque for long-distance displacement (Ross, 1967). In particular, an "island" effect obtains when extracting a wh-phrase out of the subject constituent, in contrast to extraction of the object:

(2) a. *Who$_i$ did [pictures of e_i] attract everybody's attention?
b. cf. Who did everybody see [pictures of e_i]?

The nature of the constraints precluding such extraction in (2a), while allowing cases like (2b) is a topic of lively debate in L1 as well as L2 research (cf. Chomsky, 1986; White, 2003; Omaki and Schulz, 2011). Relevant to our concerns here, Sprouse et al. (2015) reports a robust subject island effect in wh-questions like (3), though not in relative clauses, in Italian (see also Rizzi, 1982).

(3) *Di chi$_i$ pensi [$_{CP}$ che [$_{DP}$ il quadro e_i] raffiguri la nascita di Venere]?
of whom think.2SG that the painting depicts the birth of Venus
'*Who do you think the painting of depicts the birth of Venus?'
Sprouse et al., 2015; (23d)

For Slovenian, Golden (1995) reports examples of subject island violations (cf. (2)) judged as grammatical by native speakers. She argues that Slovenian does not observe a subject island constraint in wh-constructions, essentially implying that Slovenian might behave like Italian with respect to selective islandhood. However, more recent studies using fine-grained experimental designs similar to that in Sprouse et al. (2015) found a stable subject island effect in Slovenian *wh*-questions, in contrast with the initial reports (Stepanov et al., 2016, 2018). For the purposes of the present study, we assume, with the latter authors, that the subject island effect, exemplified in (4), is fully operative in Slovenian.

(4) *Od koga$_i$ je Rok mislil, [$_{CP}$ da je [$_{DP}$ velika slika e_i] visela v dvorani]?
Of whom is Rok thought that is large picture hang in hall
"*Of whom did Rok think that a large picture _hangs in the hall?"
Stepanov et al., (2016): (7d)

Thus the grammars of Italian and Slovenian converge in disallowing wh-extraction out of subject constituents.

In L1 processing, speakers' sensitivity to subject islandhood was investigated in a number of self-paced reading, speeded grammaticality and eye-tracking studies in English (Stowe, 1986; Pickering et al., 1994; Traxler and Pickering, 1996; Kurtzman and Crawford, 1991; Clifton and Frazier, 1989; Phillips, 2006, among others) and Italian (De Vincenzi, 1991). Regarding L2 processing of filler-gap dependencies, in general it was shown to be similar to that in L1 (Cunnings et al., 2010; Williams et al., 2001). Syntactic islands were also tested in L2 speakers, using both off-line (e.g. acceptability judgments; see Belikova and White, 2009) and online (e.g. plausibility evaluation) testing paradigms. In particular, Johnson and

Newport (1991) examined wh-islands in L1 Chinese-L2 English participants who arrived in the USA after the age of 17 and revealed that they performed at chance level when judging the grammaticality of the wh-island sentences, contrary to the native speakers of English who rated correctly more than 90% of the sentences. This result was expected given that, unlike English, Chinese lacks wh-movement. Schachter (1990) with materials using similar constraints showed that native Chinese, Indonesian and Korean speakers of L2 English were significantly less accurate when judging island violation sentences than their native English peers. Hawkins and Chan (1997) investigated the L2 acquisition of syntactic island constraints among Chinese and French speakers of English. The results revealed that unlike the French advanced learners who correctly rejected 85% of wh-island violation sentences, the Chinese advanced learners performed poorly on rejecting Subjacency violation sentences (41%). Another piece of evidence comes from the study by Li (1998) who also tested wh-island constraints in adult Chinese speakers who were studying English in China. Again, the results revealed that they performed at chance level rejecting only 52% of ungrammatical wh-island sentences.

More recently, Kim et al. (2015) tested how proficiency-matched Spanish and Korean learners of English processed island constraints in English. The results of this study revealed that the Spanish participants were able to apply the island constraint while the Korean learners of English showed a delayed sensitivity to it. The authors interpret these findings as "suggesting that the L1 influenced L2 learners' processing patterns" (p. 409). At the same time, based on both offline (acceptability judgment) and online (reading time) measures, Omaki and Schulz (2011) argued that Spanish learners of L2 English are sensitive to the island constraints just as much as L1 speakers of English, an expected outcome if one assumes the shared syntax hypothesis (e.g. Hartsuiker et al., 2004). Boxell and Felser (2017), using an off-line judgment task as well as eye-movement measures, found that German speaking late L2 learners of English demonstrated native-like awareness of subject islands and related structural constraints known as parasitic gaps in an off-line judgment task, despite their unavailability in the participants' L1 and rareness in the L2 input. But they also showed somewhat delayed sensitivity to island-inducing cues during online processing, as measured in the eye-movement task. Note that the L2 speakers' success in the latter two studies, at least in their respective off-line portions, correlates with similar grammatical characterization of these speakers' L1 and L2 in each case, namely, overt wh-movement. In contrast, the diverging sensitivity observed in the off-line tasks involving syntactic islands above correlates with different grammatical characterizations of L1 and L2, namely, overt wh-movement (e.g. English) vs. wh-in-situ (Chinese, Korean). In other words, the experimental evidence reviewed above suggests that sentence acceptability evaluation, and possibly also processing strategies used by the L2 speakers, may rely, among other things, on the grammatical specifications of their L1: when the grammatical trait at play is present in both L1 and L2, no native vs. non-native differences are observed; when it is absent, differences emerge.[1]

2.2. Long-distance dependencies: The clitic-second phenomenon

Clitics are grammatical morphemes that do not bear phonological stress. They are roughly distinguished as verbal clitics and pronominal clitics by their grammatical function. The placement of clitics within a sentence is language-specific. In Italian, a reflexive verbal clitic *si*

is generally adjacent to the inflected verb (5a). At the same time, it can appear indefinitely long from the non-inflected verb (5b), in particular, with restructuring verbs such as *want*, where a clitic undergoes a process known as "clitic climbing" (Burzio, 1986):

(5) a. Maria ***si*** ***alza*** sempre alle sette [Italian]
Maria refl. gets up always at seven
"Maria always gets up at seven"

b. Maria ***si*** vuole sempre ***alzare*** alle dieci
Maria refl. wants always get up at ten
"Maria always wants to get up at ten"

Slovenian, too, features a reflexive clitic, *se*. Unlike in Italian, this clitic obeys a general rule of following the first syntactic constituent in their clause (Golden and Sheppard, 2000). A long-distance dependency may obtain between the clitic and the respective verb, extending potentially indefinitely far.

(6) a. Ta knjiga ***se*** lahko ***bere*** [Slovenian]
This book refl. easily reads
"This book reads easily"
b. Filmski kritik ***se*** [čeprav je zunaj trideset stopinj] ***ovija*** v debelo odejo.
Movie critic refl. although is outside thirty degrees cover in thick blanket
"Although it is thirty degrees outside, the movie critic covers himself in a thick blanket"

Thus, with a reflexive clitic, Italian allows long-distance dependencies only in the presence of a restructuring verb (cf. also (9b) below), whereas in Slovenian there is no requirement on the verb class: rather, such dependency is simply a manifestation of the clitic-second requirement.

Thus the grammars of Slovenian and Italian diverge with respect to regulating the distribution of the reflexive clitic. A task for a L1 Slovenian-L2 Italian speaker would then be to somehow suppress the clitic-second requirement, while still acknowledging the possibility of establishing a long-distance dependency with the restructuring verb. In other words, for L1 Slovenian-L2 Italian learners, while there is positive evidence from Italian supporting long-distance clitic-verb dependencies, in the form of restructuring constructions as in (5b), there is no negative evidence effectively *restricting* these dependencies to the restructuring contexts in this language. In a way, this gives rise to a superset-subset relation between L1 and L2 whereby the set of admissible constructions in L1 Slovenian (including both adjacent verb-clitic combinations as well as remote verb-clitic dependencies with any verb supporting a reflexive clitic) is a superset of the admissible constructions in L2 Italian (which include adjacent verb-clitic combinations as well as remote verb-clitic dependencies with restructuring verbs only). This may present challenges to L2 learners (cf. also White 1989: Ch.6; see also Section 4) with respect to combinations that involve a *non*-restructuring verb and are thus ungrammatical in this language, but are grammatical in Slovenian. This challenge could then also be expected to show up in the processing routines (see below).

Although, to our knowledge, no studies to date have looked at L1-driven overgeneralization of clitic placement in L2, previous work shows that the use of clitic constraints is variable and/or error-prone in L2, and likely affected by L1 knowledge. Specifically, clitics were argued to be rather difficult to acquire by L2 learners (e.g. Adiv, 1984; Montrul, 2011; Herschensohn, 2004), and even by children acquiring their first language (e.g. Guasti, 1994). They are also likely to be easily affected in aphasic speakers. For instance, Rossi (2015) using the syntactic priming paradigm to study placement of object clitics in sentences with finite and restructuring verbs in Italian (cf. (5b)) reports only a limited performance with object clitics in agrammatic speakers who produced less clitics than controls during spontaneous speech production (Experiment 1) as well as in the context of a priming paradigm (Experiments 2 and 3). With respect to L2 research, Rossi et al. (2014), in an ERP study, report that a subset of their highly proficient English-Spanish late bilingual speakers performed comparably on the detection of wrong uses of clitics in terms of their featural specification for gender and number, while less proficient late bilinguals showed a more limited performance than monolingual controls. Clitics in L2 were also argued to be subject to positive L1 transfer, whenever clitic placement rules in L1 are similar to those in L2. Grüter and Crago (2010), for instance, used an elicited production and truth-value judgment task in order to test how school-aged children whose parents spoke Chinese (a language without clitics) and Spanish (a language with clitics) produce and understand object clitics and null objects in French. The results revealed that children from the Spanish group produced significantly more clitics than those in the Chinese group and omitted clitics significantly less often than their Chinese peers. An important role of working memory in clitic processing has also been explored. In one study that investigated comprehension and production of object clitic pronouns in the context of a specific type of classroom teaching technique, termed “processing instructions”, low-span L2 learners of French whose L1 was English manifested greater difficulties producing clitic pronouns associated with right-dislocated sentential objects in spoken French, compared to high-span L2 learners (cf. Santamaria and Sunderman, 2015; VanPatten, 2005). These studies suggest that clitics offer an interesting and relevant testing ground regarding different views on the success in L2 learning including ultimate attainment, including the domains of grammar that involve long-distance dependencies.

2.3. Predictions for an experimental study

In the present study we were interested in performance of L1 Slovenian-L2 Italian speakers as regards detecting syntactic anomalies online in subject island and clitic-second constructions outlined above. In particular, if the theories of L2 predicting some non-native performance even in the advanced learners due to L1 transfer, are on the right track, we may then expect that Slovenian learners will perform native-like on the converging, subject island phenomenon in Italian, but not necessarily native-like on the diverging, clitic-second phenomenon. These predictions crucially rely on the assumption that the L2 learners’ endstate ensures a native-like performance at least in some grammatical domains.

3. The experiment

We conducted a self-paced reading experiment whose main purpose was to investigate the processing of long-distance dependencies by advanced Slovenian speakers of L2 Italian. The experiment included two manipulations testing participants' performance on the converging and diverging syntactic phenomena, namely, a subject island manipulation and a verb-clitic manipulation, as outlined below.

3.1. The subject island manipulation

A typical syntactic long-distance dependency contains a "filler", or head of the dependency, and a corresponding "gap". In the case of *wh*-dependencies as in (1), for instance, the filler is the fronted *wh*-phrase, and the gap is its thematically determined position (e.g. as an object of the verb). In the case of clitics, a (fronted) clitic and its host, e.g. a verb, form a dependency that can be described in similar terms (cf. (5b) and (6b)). In processing terms, postulation of an appropriate gap of the dependency implies successful integration of the filler/head into the syntactic structure according to grammatical rules of the language.

For testing sensitivity of Slovenian L2 learners to subject islands in Italian, we compared the reading times of sentences involving extraction of the wh-phrase *di qui* "of whom" out of sentential objects (the "object extraction" condition, cf. (7b)) and out of sentential subjects (the "subject island" condition, cf. (7a)). For the object extraction condition, we used sentences with postverbal subjects that speakers find more natural in normal speech than those involving preverbal subjects (on the grammar of postverbal subjects in Italian see, e.g. Longobardi, 2000; Cardinaletti, 2004).

(7)a. *Di chi$_i$ pensi [che [il pianto e_i] abbia interrotto *la recita del bambino* giovedì]?
of-who think-2sg that the crying has interrupted the recital of-the child Thursday?
"Whose crying do you think _ interrupted the child's recital on Thursday?"

b. Di chi$_i$ pensi [che abbia interrotto [la recita e_i] *il pianto del bambino* giovedì]?
of-who think-2sg that have interrupted the recital the crying of-the child Thursday
"Whose recital do you think that the crying of the baby interrupted _ on Thursday?"

Each word constituted its own processing region in our study. The by-region distribution of the two tested conditions is illustrated in (8):

(8)

Region	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Obj.Ext.	Di	chi	pensi	che			abbia	interrotto	la	recita	il	pianto	del	bambino	giovedi.
Subj.Ext.	Di	chi	pensi	che	il	pianto	abbia	interrotto			la	recita	del	bambino	giovedi.
Gloss	of	whom	think-2sg.	that	the	crying	has	interrupted	the	recital	the	crying/ recital	of-the	child	Thursday

We assume, for the present purposes, that participants attempt to integrate the wh-phrase in accord with the Active Filler processing strategy which favors the integration of the filler at the earliest point allowed by the grammar (Fodor, 1978; Frazier and Clifton, 1989; cf. also Stowe 1986; Frazier, 1978; Frazier and Fodor, 1978). This is also in line with the common trend in the literature that recognizes that the parser deals with processing costs of two different kinds: those related to temporary storage of the filler in the working memory, and those pertaining to its integration at the grammatically licensed position in the input (see, e.g. Gibson, 1998, 2000). The Active Filler strategy is often seen as the result of the parser's effort to reduce these costs, and is thus a function of available computational resources and computational efficiency.[2]

Consider first *wh*-extraction out of the object position as in (7b). If the Active Filler strategy is in place, participants will attempt to integrate the stored filler *di chi* at the first opportunity licensed by the grammar. In this particular case, this opportunity arrives at the NP *la recita* (Regions 9-10). Presumably, there is no longer a processing storage cost for keeping this *wh*-phrase in the working memory at the later regions. The situation is different in subject island sentences as in (7a). All else equal, the first available opportunity for integrating *di chi* in this case arrives at the NP *il pianto* (Regions 5-6). However, unlike in the case of *wh*-extraction from the object, this potential integration slot is not grammatically licensed. As noted in the previous section, there are good reasons to believe that in L1 and L2 processing, participants are sensitive to island configurations, especially if they are similarly constrained in both. The first grammatically licensed integration possibility arises at the NP *la recita.* Hence, given that its complement is another prepositional phrase (cf. *del bambino* in (8a)), structural principles are likely to prevent successful integration of the wh-phrase *di chi* (cf. Stowe, 1986). Thus the complement of the second NP and the following sentence-final adverb followed by a period (Regions 13-15) are critical regions where the difference in reading times with the object extraction and the subject island conditions may potentially show up. Specifically, reading in the critical area is expected to trigger a subject island effect in the respective condition, causing greater reading latencies. The adverb in the final region can be used as the locus of a potential spill-over effect. It is also important to note that in Italian, right-peripheral adverbs scoping over the entire sentence typically indicate the end of the sentence. Being sentence-final, the adverb may indicate to the parser that there are no further integration possibilities, thus offering a convenient tool to mark a potential integration failure.

3.2. The clitic placement manipulation

We aimed to test whether L2 learners are sensitive to the distinction between sentences with grammatical placement of *si* as in (9a), compared with sentences with ungrammatical placement of *si* in constituent-second position, as in (9b). The by-region of the two tested conditions is shown in (10).

(9) a. Il postino tutte le mattine si veste con la divisa delle poste
The postman all the mornings refl. dresses with the uniform of-the post
"Every morning the postman puts on the uniform of the postal service"

b. *Il postino si tutte le mattine veste con la divisa delle poste
The postman refl. all the mornings dresses with the uniform of-the post

"Every morning the postman puts on the uniform of the postal service"

(10)

Region	1	2	3	4	5	6	7	8	9	10	11	12	13
Correct	Il	postino		tutte	Le	mattine	si	veste	con	la	divisa	delle	poste
Incorrect	Il	postino	si	tutte	Le	mattine		veste	con	la	divisa	delle	Poste
Gloss	the	postman	refl.	All	The	mornings	refl.	dress-3sg.	with	the	uniform	of-the	post

From the online perspective, the clitic *si* in the grammatically correct condition *triggers* an expectation for a verb in order to be successfully integrated. This can be either a finite or a restructuring verb (cf. (5)). We compare this grammatical condition with one where *si* and the following verb are separated by the intervening material. An anomaly in the ungrammatical condition can be expected both pre-verbally, that is, at the intervening regions themselves (Regions 4-6 in the Incorrect condition in (10)), as well as at the verb region (Region 8). Postverbally, we can expect a spill-over effect at two or three successive regions, in our case Region 9 and possibly Region 10 and 11, as is typical for morphosyntactic violations (e.g. De Vincenzi et al., 2003; Ditman et.al., 2007). A priori, we do not expect reading differences across the remaining regions.

3.3. Participants

Sixty-one adult volunteers were recruited from the University of Nova Gorica and the surrounding communities in the Primorska region of Slovenia bordering Italy, as well as from the University of Udine and the surrounding communities in the Friuli-Venezia-Giulia region of Italy, including several high schools and school centers. Some participants received no material compensation for their participation, while others received a gift worth €1.20. The participant pool consisted of 35 self-reported monolingual native speakers of Italian (10 males; mean age= 22.89, *SD*=3.70) and 26 self-reported native speakers of Slovenian who were also non-native speakers of Italian (6 males; mean age= 23.73, *SD*=6.95). The Slovenian participants were born and raised in Slovenian families, went to elementary schools in Slovenia (all except one participant who went to an Italian elementary school), but were exposed to Italian early in life, either at home, at kindergarten or school, via communicating with Italian-speaking relatives and/or peers, as well as via other kinds of active (e.g. cross-border interactions) and passive (e.g. TV) exposure. The average AoA of Italian for the non-native group was 4.39 years (*SD*=3.13). The majority of the tested non-native participants (65%) also passed the state secondary school matriculation exam ("matura") in Italian, and some of the participants (35%) were also past or current university students of Italian philology. The participants indicated that they use Italian on a regular basis in various social contexts, not in the least because of their geographical proximity to Italy. All participants had normal or corrected to normal vision and reported no history of neurological disorders.

To evaluate the overall proficiency in L2 Italian in the non-native group, a c-test was administered to both native and non-native groups. A c-test is a version of a cloze test based

on a short text in which the second half of every second word is truncated (the first two and the last two sentences are left intact). The participants are asked to fill in the missing word parts. The ratio of correctly reconstructed words to the total words is taken to be a measure of general language proficiency. C-tests are commonly regarded in the literature as an integrative measure of linguistic knowledge. They can be used as a viable and efficient tool for evaluating this knowledge using the written modality, especially in the domain of morphology, syntax and the lexicon (Eckes and Grotjahn, 2006; Karimi, 2011). In our study, we constructed a c-test based on two short original texts taken from Italian online newspapers. The texts were balanced by size, each being about 100 words long, and by the word type/token ratio (about 73%). Each text was presented separately on a computer screen in its entirety. The deleted word parts were substituted with mini-windows that were all of the same size, where the participants could type in their responses. There was no time limit on this task.

In terms of results, the percentage of correct reconstructions for the native group was 92.40 (*SE*=1.66). The ratio for the non-native group was 74.55 (*SE*= 2.58). In other words, Slovenian non-native speakers performed in the task at 80% compared to the native Italian group. Following common practice in the literature (cf. Omaki and Schulz, 2011), we consider this ratio, together with the early AoA, as indicative of the rather advanced level of L2 knowledge in our participants.

3.4. Materials

3.4.1. The subject island materials

Sixteen pairs of sentences for testing the subject island phenomenon (cf. (7)) were constructed. They involved extraction of the prepositional wh-phrase *di qui* "of whom" out of (preverbal) subject NP, compared with similar extraction out of the object NP. The sentences were matched by syntactic complexity and length, and were constructed following the general template illustrated in (8). In particular, they involved extraction out of a finite clause introduced by the verb *pensare "to* think"*.* The preverbal subject NP in the subject sub-extraction condition (7a), and the postverbal object NP in the object sub-extraction condition (7b) were always of the type [Det N]. The critical area included the rightmost postverbal NP, part of the direct object in the subject extraction condition, and part of the postverbal subject in the object extraction condition (Regions 11 and 12). Liaised articles, e.g. *l'*, were avoided. Region 13 included the sentence-final adverb as a potential spill-over area.

We first conducted a pre-test intended to ensure that the baseline contrast between target sentences involving wh-extraction out of object and out of subject in Italian holds at the level of acceptability judgments. The sixteen target sentence pairs were mixed with 40 filler sentences of similar length and complexity, half of which was fully grammatical, and the other half ungrammatical. The resulting 56 items were split into two lists, in which the target items were distributed across conditions in a Latin square design, so that no two similar lexicalizations of target conditions appeared in the same list. The filler items remained the same in both lists.

We invited 35 adult native speakers of Italian (9 males; mean age= 32.76, *SD*=4.33) who did not participate in the main experiment and were not aware of its aims/goals to

evaluate the acceptability of the target sentences on a Likert scale from 1 to 5. There was no time limit on the task, which took about 20 minutes to complete. In order to exclude scale biases between participants, the obtained Likert scores were transformed to z-scores which represent the number of standard deviations the actual score was away from each participant's mean score over both target and filler sentences. Scores that were above or below 3 standard deviations from the average were excluded from the count.

We found a contrast in the speakers' judgments regarding wh-extraction out of subjects and wh-extraction out of objects, the former being judged significantly worse than the latter (Welch $t(190)=5.92$, $p < 0.0001$). The results are illustrated in Figure 1. We attribute a somewhat lower than average score on the sentences involving extraction out of objects to a "complex structure" effect involving an embedded clause with a complex subject (see Sprouse et al., 2012 for discussion). Sprouse et al (2016:328), using a similar design, report extraction out of object in embedded sentences in Italian receiving the z-score of (-0.30)-(-0.40), whereas sentences involving extraction out of subject received a z-score around (-0.70)-(-0.80). Our present results corroborate those findings.

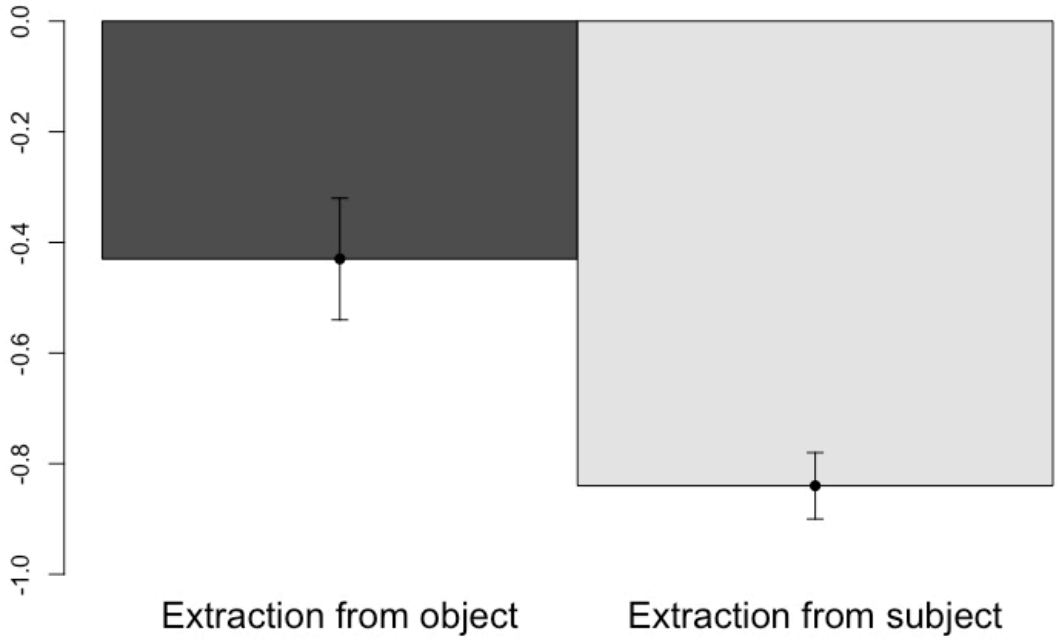

Figure 1. Acceptability ratings of sentences involving extraction out of object and out of subject in Italian in the pre-test study

3.4.2. Clitic placement materials

Sixteen pairs of sentences involving clitic-second constructions (cf. (9)) were constructed. Similarly to the first set, these sentences were all built on the same template (cf. (10)), were matched for syntactic complexity and length. In the grammatical condition, the reflexive clitic *si* immediately precedes the finite verb. In the ungrammatical condition, *si* is separated

from the verb by a temporal phrase such as “every Sunday evening”. The length of the intervening phrase was always three words.

All 32 sets of target sentences were counterbalanced across two sets of stimuli so that each participant saw only one lexical version of a particular item per condition. Consequently, each subject saw eight different lexical versions per each of the four conditions in total. In addition, 64 filler sentences of similar length and complexity were constructed and added to each stimulus set, making it the total of 96 items per each set.

3.5. Procedure

The experiment was administered to both groups of participants in the following order: a) language background questionnaire; b) two c-tests; c) self-paced reading task. In the self-paced reading task, stimulus sentences were presented word-by-word in the moving-window presentation mode (Just et al., 1982). In the present experiment, the participants were instructed to read the sentences at a natural pace and to be sure they understand what they read. Each sentence (including fillers) was followed by a yes-no comprehension question about the event or state of affairs described in the target sentence (for instance, a question following item (9) was: *Does the postman wear a uniform?*). The proportion of the correct “yes” and “no” answers to comprehension questions was equally distributed within each of the two stimulus sets. Feedback was provided when an incorrect response to a comprehension question was given. No feedback was given in cases of correct answer. Failure to respond within 4 sec counted as an incorrect response. Before beginning the task, subjects read five practice sentences. The order of stimulus presentation was pseudo-randomized for each participant by the experimental software and it was ensured that at least one filler sentence intervenes between any two target items. The entire experimental session lasted 20-25 minutes for the native group and 25-30 minutes for the non-native group. A somewhat longer session for the non-native speakers was due to a more comprehensive version of the language questionnaire for this group. All parts of the experiment were programmed on the Ibex web-based platform (http://spellout.net/ibexfarm). Participants performed the task at remote locations of their choice without having to come to the lab. They were specifically instructed to make sure external disturbance is reduced to the minimum while performing the tasks.

3.6. Data analysis

Reading times faster than 100ms or longer than 3000ms were omitted as (presumably) unlikely to be generated by relevant linguistic processes. This affected about 2% of the data. Reading times for the target sentences in both manipulations were examined for each successive region. To control for differences in word length across conditions as well as overall differences in participants' reading speed, a regression equation predicting reading time from word length was constructed for each participant, on the basis of all filler items (cf. Ferreira and Clifton, 1986). Specifically, log-transformed raw reading times were input into a linear regression model to calculate *residual reading times,* or RRTs (a logarithmic transformation was chosen as a result of applying the Box-Cox procedure, cf. Box and Cox 1964; Enochson and Culbertson, 2015). Thus words read faster will have negative RRTs, whereas words read slower will have positive RRTs. Outliers were then identified and

excluded from further analyses, using 3 standard deviations from the mean RRT for a given condition and region as criterion, for each participant.

For comprehension question accuracy we used logistic mixed effects models for binary data (Jaeger, 2008). Reading time data were analyzed using linear mixed-effects models run under *lme4* package in R 3.5.0 environment (Baayen 2008, Bates et al., 2014, R Core Team, 2014). For each manipulation, we fitted models crossing fixed factors GROUP (native, non-native) and GRAMMATICALITY (yes, no), as well as their potential interactions. For each fixed factor, sum coding was used. In both manipulations, the ungrammatical condition and the native (Italian) group received the coding of -0.5, whereas the grammatical condition and the non-native (Slovenian) group were coded with the value of 0.5. Participants and stimuli were entered as random factors. In each analysis, the best fitting model was chosen by comparing models with different degree of complexity using the likelihood ratio test. The simplest model considered included only by-subject and by-item random intercepts. More complex models included by-subject and by-item random intercepts, a random slope for GRAMMATICALITY varying across subjects and a random GRAMMATICALITY*GROUP slope varying across items. Starting with the maximal model and using the backward-selection heuristic, the results of the likelihood ratio test were used to compare pairs of models to evaluate whether inclusion of additional random-effects parameters provided a better fit to the data, so that more complex models were excluded only if the *p*-value for the significance of the difference between the two models was above 0.2 (cf. Matuschek et al., 2017). In the analysis of main effects and interactions, we report *p*-values based on Satterthwaite's approximations for degrees of freedom using the *lmerTest* package (Kuznetsova et al., 2014). Posthoc planned comparisons were performed as Tukey's estimations via the *multcomp* package (Hothorn et al., 2008).

3.7. Results

3.7.1. Comprehension accuracy

For the native speaker group, the mean comprehension question accuracy was 96.25 (*SD*=3.11). For the non-native group, the mean accuracy was 89.93 (*SD*=10.46). This indicates that participants were generally attentive during the reading. The data from one participant in the non-native group were excluded from the analyses because of the low comprehension question accuracy (below 50%). Logistic mixed effect models constructed as explained above showed no main effects of either GRAMMATICALITY or GROUP ($p > 0.10$ for each), indicating that none of these factors affected comprehension accuracy in target sentences.

3.7.2. Reading times

The data from one participant in the native group were removed because of the overall slow reading time (3.22 *SD* from the participants' average). This left the data from the total of 34 participants in this group for the analysis.

3.7.2.1. Subject island results

The preverbal areas 1-4 where the conditions did not differ lexically or structurally were not part of our statistical analysis. Because of different preceding contexts in respective sentences, response data in Regions 7 and 8 would be difficult to interpret in light of the present research hypothesis; we therefore do not report quantitative analyses in these

regions as well. The time course of reading the postverbal regions for both native and non-native groups is represented in Figure 2. Table 1 shows estimates, standard errors, test statistics and the *p*-values for main effects and interaction in the critical regions, derived from our models.

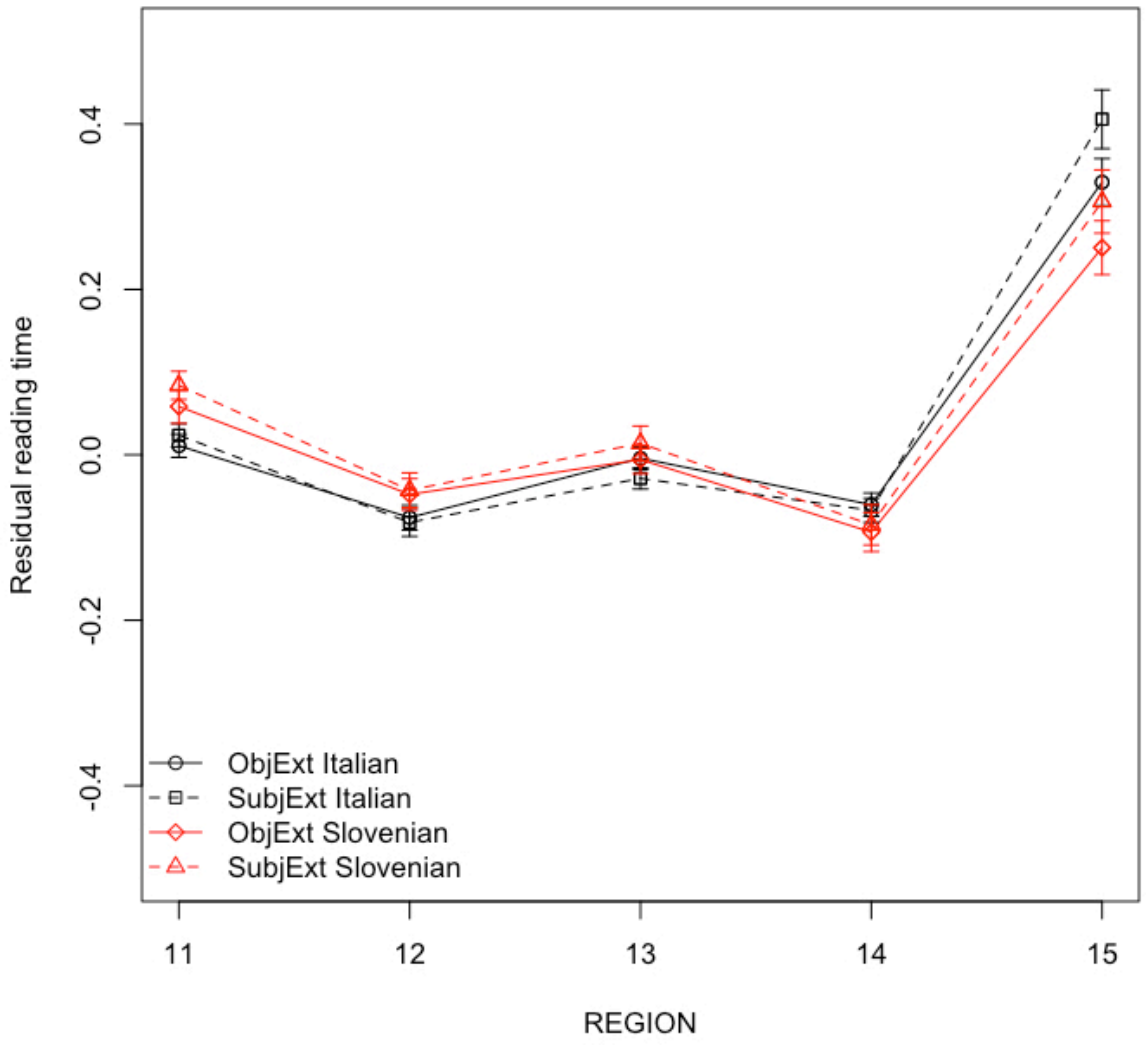

Figure 2. Plot of mean (standard error) residual RTs per word by region in the subject island sub-experiment, Regions 11-15 (see the regioning schema in (8))

Region	Grammaticality				Group				Grammaticality*Group			
	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*P*	Estimate	*SE*	*t*	*p*
11	0.018	0.16	1.173	0.241	0.053	0.02	2.786	0.007	0.012	0.03	0.397	0.691
12	-0.002	0.02	-0.110	0.912	0.032	0.02	1.095	0.286	0.004	0.04	0.108	0.914
13	0.005	0.02	0.265	0.793	0.042	0.03	1.312	0.204	0.061	0.03	1.626	0.120
14	0.003	0.01	0.203	0.838	0.019	0.06	0.339	0.737	0.021	0.03	0.657	0.511
15	0.069	0.03	2.189	0.034	-0.048	0.08	-0.705	0.484	-0.002	0.07	-0.042	0.966

Table 1. Summary statistics for main effects and interactions on the basis of linear mixed models predicting reading times in the critical regions in the subject island manipulation.

In Region 11 corresponding to a definite article pertaining to the postverbal subject in the object extraction condition, and the direct object in the subject extraction condition, we find a main effect of GROUP, with non-natives reading this article slower than the natives. This effect is likely to be due to a spill-over effect from the preceding region, given the (partially) differing structural contexts. There was no main effect of GRAMMATICALITY, and no significant interaction between the two factors in this region. In Region 12, the first postverbal noun, we find no main effect of either factor, as well as no significant interaction ($p > 0.10$).

In the lexically matched Regions 13 and 14, the preposition+article combination starting the PP complement of the noun in the direct object position, we find no significant differences concerning either factor and no interaction among them. Finally, in Region 15, our models revealed a main effect of GRAMMATICALITY: the sentence-final adverb was read slower in the ungrammatical subject island condition than in the grammatical object extraction condition. We interpret this effect as reflecting a failure to integrate the wh-phrase after the list of all potential candidates for the integration site has been exhausted. There was, however, no effect of GROUP at that region. Interaction between GRAMMATICALITY and GROUP did not reach significance either. Therefore, there was no evidence that the strength of the grammaticality effect differed between the two groups.[3]

3.7.2.2. Clitic placement results

The initial material (Regions 1-3) in which the conditions do not differ was not included in our statistical analysis. As noted in Section 3.2, from the perspective of the GRAMMATICALITY factor, we are mainly concerned with the pre-verbal regions intervening between the "quasi-clitic-second" position and the main verb (Regions 4-6), as well as with the post-verbal regions that represent a possible spill-over (Regions 8-11). The time course of reading the respective conditions in these regions is represented for the native (black) and non-native (red) groups in Figure 3.

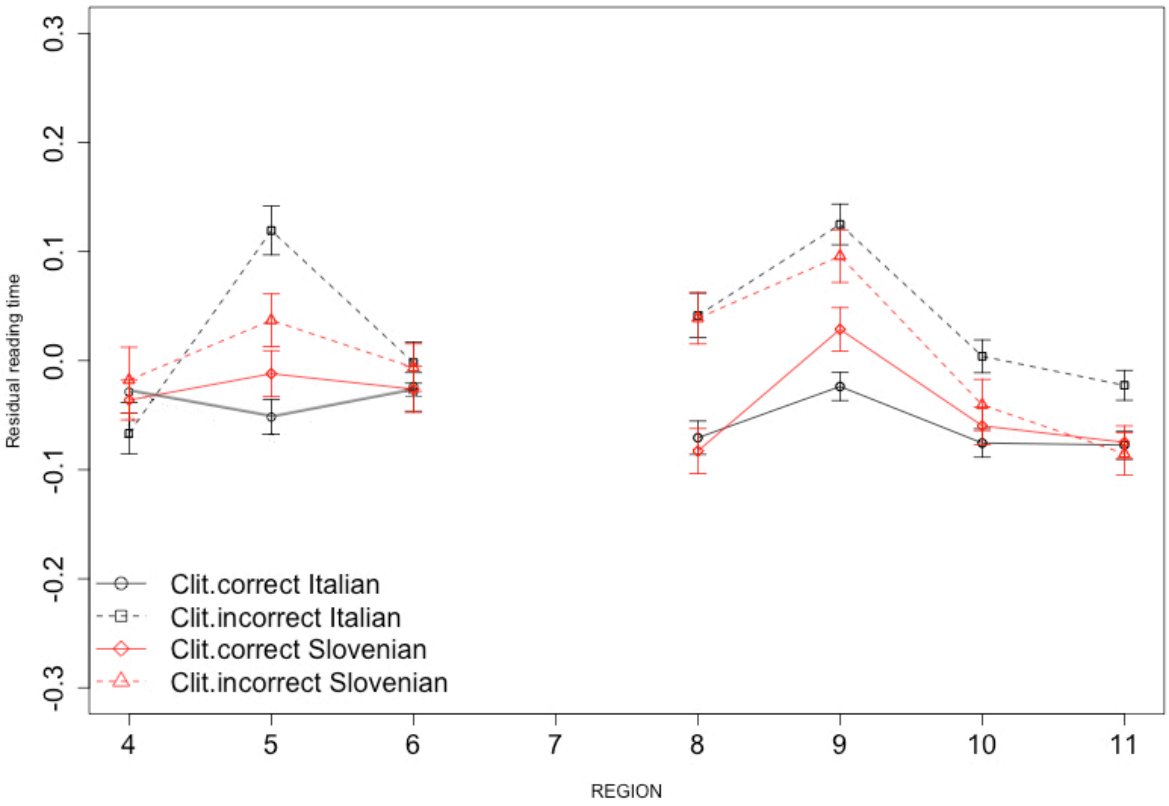

Figure 3. Plot of mean (standard error) residual RTs per word by region in the clitic placement manipulation, Regions 4-6 and 8-11 (see the regioning schema in (10))

Region	Grammaticality				Group				Grammaticality*Group			
	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*p*
4	0.017	0.03	0.580	0.566	0.054	0.03	1.473	0.150	0.059	0.060	0.987	0.330
5	0.112	0.02	4.841	<0.001	-0.016	0.03	-0.622	0.536	-0.118	0.04	-2.547	0.012
6	0.024	0.02	1.060	0.294	0.001	0.02	0.058	0.954	-0.002	0.04	-0.050	0.960
8	0.117	0.02	5.511	<0.001	-0.007	0.02	-0.346	0.730	0.010	0.04	0.247	0.805
9	0.107	0.02	5.782	<0.001	0.012	0.02	0.636	0.528	-0.081	0.03	-2.188	0.029
10	0.045	0.01	2.812	0.005	0.004	0.03	0.138	0.895	-0.067	0.03	-2.053	0.040
11	0.024	0.01	1.621	0.103	-0.028	0.01	-1.450	0.152	-0.063	0.02	-2.140	0.032

Table 2. Summary statistics for main effects and interactions on the basis of linear mixed models predicting reading times in critical regions in the clitic manipulation.

Table 2 shows estimates, standard errors, test statistics and the *p*-values for main effects and interaction in the critical regions, derived from our models. In the preverbal regions 4 and 6, there was no main effect of either factor, and no significant interaction between them ($p > 0.10$ for all comparisons). In Region 5, we find a main effect of GRAMMATICALITY: it took participants longer to read this item in the clitic-incorrect condition than in the clitic-correct condition. There was also a significant interaction between GROUP and GRAMMATICALITY suggesting that nativeness of the speakers affects the pattern of reading in this intervening region. Sensitivity to the wrong placement of the clitic was different in each group, as revealed by pairwise comparisons on the interaction term. Within the L1 group, participants read the respective item slower in the clitic-incorrect condition than in the clitic-correct condition (Estimate=0.172, *SE*=0.03, z=5.880, $p < 0.001$). In contrast, within the L2 group, there was no significant difference in the reading times between the two conditions (Estimate=0.053, *SE*=0.03, z=1.475, $p = 0.446$).

We now move on to the verb and the postverbal regions. At Region 8 (the main verb), there was again a main effect of GRAMMATICALITY, with the clitic-incorrect condition being read slower than the clitic-correct condition. Factor GROUP was not significant, and there was no interaction among the two factors. In other words, both L1 and L2 groups detected an anomaly here independently.

At Region 9, the first postverbal region, GRAMMATICALITY is again significant, and it interacts with GROUP. Pairwise comparisons on the interaction term revealed that the L1 group detect an anomaly by slowing down on the clitic-incorrect condition compared to the clitic-correct condition (Estimate=0.148, *SE*=0.02, z=4.207,$p<0.001$), whereas the L2 group did not show a significant reading time difference (Estimate=0.089, *SE*=0.04, z=2.075, $p = 0.154$).

At Region 10, the second postverbal region, there was a main effect of GRAMMATICALITY, a main effect of GROUP and their interaction, with the L1 group again sensing an anomaly in the ungrammatical condition by slowing down (Estimate=0.792, *SE*=0.02, z=3.842, $p < 0.001$), while the L2 group did not show a slowdown in the ungrammatical condition (Estimate=0.012, *SE*=0.02, z=0.494, $p = 0.957$). Similarly, at Region 11, GRAMMATICALITY significantly interact with GROUP, with the contrast in grammaticality mainly due to the L1 group (Estimate=0.055, *SE*=0.02, z=2.984, $p = 0.014$), but not the L2 group (Estimate=-0.007, *SE*=0.02, z=-0.328, $p = 0.987$).

In sum, our L2 speakers were generally slower in reading the critical regions of the clitic construction, compared to the native Italian speakers. At the same time, both groups reacted differently to the grammatical anomaly. Italian L1 speakers notice an anomaly already preverbally (Region 5), and later show a broad slowdown from the verb (Region 8) through several postverbal regions up to Region 11. In contrast, the Slovenian L2 speakers do not seem to detect an anomaly preverbally, but detect some anomaly at the verb, and show no substantial slowdown through the postverbal regions. The two groups thus differ in their preverbal responses, and in how broad the slowdown effect is at and after the verb. Put differently, the L2 subjects' performance on the clitic-incorrect sentences seems to display no sensitivity preverbally, and an attenuated (mild) sensitivity to the core of the violation.

4. General discussion

The main focus of this study was to test whether the L2 knowledge as reflected in online sentence processing is affected by converging vs. diverging grammatical characterizations of complex syntactic phenomena in L1 and L2, in the domain of long-distance dependencies where direct positive evidence is generally unavailable to a L2 learner. Based on the earlier literature, we hypothesized that in this area, advanced Slovenian learners of L2 Italian will perform like native speakers as concerns the phenomena whose grammatical descriptions converge, and non-native-like in the areas whose grammatical characteristics diverge. In the domain of subject islandhood, the converging domain, L2 speakers could not simply rely on the L2 input. As for the domain of clitic placement, characterized by diverging grammatical descriptions, the challenge was basically in the subset-superset character of the available data, whereby, even though long-distance placement of the clitic was generally available in the L2, it is restricted to a subset of contexts available in the speakers' L1.

Our results show that in the area of subject islands, advanced Slovenian learners of L2 Italian performed on a par with Italian native speakers, as predicted. They reliably manifest sensitivity to the deviant character of the island sentences at the same point in processing as the native speakers do. This replicates the earlier findings (Section 1), but also adds a new dimension to this line of inquiry, as it shows that native-like performance is also possible in complex syntactic domains that cannot be learned only on the basis of surface morphosyntactic cues, unlike, e.g. subject-verb agreement. This result is also in line with the findings reported in Omaki and Schulz (2011) who show that advanced L2 learners are capable of building complex syntactic representations when processing island structures (see also White 1989, 2003; Belikova and White 2003, among others). It is also compatible with the Shared Syntax hypothesis (Hartsuiker et al., 2004), according to which non-native speakers do not "learn" any new specification in the island condition, but apply the same knowledge as they do in L1.

In terms of the diverging grammatical characterizations in the domain of clitic placement, our non-native participants generally showed a good performance. However, unlike native speakers, they showed reduced sensitivity to the ungrammatical structure at Regions 5 and 9-11. These differences point to deviations from native performance in non-native speakers regarding the clitic-placement phenomenon.

What is the nature of the particular processing strategy used by our L2 speakers? We conjectured that the process of retaining the clitic trace in working memory in languages like Slovenian with subsequent integration with the verb is akin to processes that take place in processing filler-gap dependencies, e.g. in wh-movement, where the filler and the gap can be separated indefinitely far from each other (Section 3.1). Grammatically speaking, the overarching clitic-second requirement in their L1 prepares the Slovenian speakers for the possibility of encountering material of potentially unlimited complexity (including any number and combination of restructuring verbs, modifying adverbials, parentheticals etc.) intervening between the reflexive clitic and the verb in the input. In a way, encountering the reflexive clitic signals that the verb is coming, but, as discussed in Section 2.2, there is no adjacency proviso that the verb is coming *next* in the Slovenian grammar. This proviso,

however, is part of Italian grammar. Italian native speakers allow for a similar grammatical possibility, but only in a subset of the Slovenian-compatible contexts, namely, those restricted to restructuring verbs. Thus, when processing the native input, not finding a verb in the vicinity of the clitic would signal an anomaly in Italian, but not necessarily in Slovenian. This is indeed what we find in the present study. Therefore, the observed anomaly at Region 5 may have to do with violating the adjacency requirement on clitics in Italian (Section 2.2). In other words, then, the processing difficulty in the L2 Italian learners stems from the constraints imposed by the grammatical knowledge of their L1, expected under the view that the grammar informs parsing routines and may therefore constrain them as well (see also Grillo and Costa, 2014).

It could be argued that the (partially) attenuated performance on the incorrect clitic-placement condition manifested by the non-native speakers indicates a latency effect typical for non-native parsing: the latter is generally slower and less automatic than native processing (e.g. Hopp, 2007). However, in our subject island sub-experiment, non-native speakers clearly followed a time course similar to that of the native speakers up to the final region area. As a post-test, we also compared the reading times on the filler sentences grouped by sentence length and found no GROUP effect (all *p*s> 0.10). This speaks against general latency as a possible explanation. Instead, we suggest that, in processing Region 5 of the clitic-second configurations, L1 Slovenian-L2 Italian speakers use a processing strategy transferred from their L1, in contrast to native Italians. This transferred strategy is in some sense more "permissive" in that it may allow non-adjacent clitic-verb combinations not present in the target L2. Under the assumption that the parser or online processing routines (whether universal or L2-specific, see Section 1 and below) consult the grammatical knowledge during processing L2 material (see Hopp, 2007 for a review), this processing transfer, in turn, may be guided by the erroneous "clitic-second" grammatical representation activated on the basis of the speakers' L1. In this case, more "permissive" grammatical representation(s), as compared to the target ones, lead to non-native performance, as expected.

It is still an open question whether and to what extent L2 processing strategies can be affected by grammatical representations in L1. A cursory overview of the current literature offers a somewhat mixed picture. For instance, Marinis et al. (2005) report that speakers of languages with no overt wh-movement such as Chinese, as well as languages with overt wh-movement such as German, process complex long-distance wh-questions in English in a similar way (not identical with native speakers). On the other hand, since it is not always clear what those allegedly language-dependent processing routines are in each case, this claim may be hard to maintain. In contrast, this difficulty may not arise in light of the existing substantial evidence that the parser or the set of processing mechanisms are universal, hence employed in both L1 and L2 on an equal basis. The present study contributes to this debate by highlighting the case of parsing routines guided by grammatical representations transferred from L1 in the domain of a grammatical phenomenon the learning of which cannot be triggered simply by positive evidence available in the input (see Section 1).

Previous studies found that speakers who performed at the native level in global judgment tasks, can still show non-native performance in online procedures (e.g. Marinis et al., 2005; Papadopoulou and Clahsen, 2003). Zawiszewski et al. (2011) suggest a possible

transfer of processing routines from L1 Spanish in processing Basque structures as an explanation for the differing performance of non-native participants on the grammatically diverging phenomena such as word order (see also Erdocia and Laka, 2018).[4] Our results corroborate this line of literature, by showing a non-native behavior in Region 5 and postverbal regions (9-11) in the clitic manipulation, although our non-native participants did detect the respective anomaly concerning clitic placement at one point, namely, the verb, in Italian.

We suggest that the (partial) processing deviance observed by the non-native speakers in the clitic placement sentences may be due to difficulty in retrieving the respective parts of L2 grammatical knowledge that bear on the lack of positive evidence, during online processing of L2 structures. In other words, the observed processing deviance may not necessarily be predicated on the respective grammatical deviance, but on the difficulty in consulting the relevant parts of the L2 grammar in order to assign a structural representation to the input sentence online. In particular, what might be difficult for the L2 learners to do in our case is not activate the clitic placement rule of Italian (roughly, referring to adjacency to the verb), but pick this particular rule in the presence of the competing L1 rule that relates to the former in a superset-subset fashion (see Section 1). The non-trivial character of the relationship between the respective grammatical rules in L1 and L2 may affect actual linguistic performance in the latter (see also Lardiere, 2013).

This scenario entails that L2 processing may not be qualitatively different from L1 processing, consistently with our results from the subject island sub-experiment. Indeed, with respect to the diverging grammatical phenomenon (clitic-placement), we do find a (partially) divergent processing pattern between the native and non-native groups. However, the fact that non-native speakers process certain syntactic configurations differently from native speakers does not imply, in and of itself, different processing mechanisms for L1 and L2 (cf. Dekydtspotter et al., 2006). Rather, what may not be native-like, is "the quality of access" to the relevant parts of L2 grammatical knowledge that stand in a complicated, in our case superset-subset, relation with the respective parts of the L1 grammar. In a sense, then, the parsing deviance observed by the non-native speakers in our study boils down to a version of a grammatical L1 interference problem reflected in parsing, but not in global acceptability evaluations.[5]

If this latter line of analysis is correct, it supports the class of theories that maintain that the L1 and L2 processing routines and strategies are fundamentally the same, despite possible differences in age, proficiency and other factors such as the "full transfer/full access/full parse" model (Dekydtspotter et al., 2006; Schwartz and Sprouse, 1996; White, 2000), and the Fundamental Identity Hypothesis (Hopp, 2007). These models share the prediction that with increased proficiency, L2 processing signatures will approximate, if not coincide with, the processing routines of L1, as both are ultimately based on the fundamentally same psychological and/or neurocognitive mechanisms. Under the latter hypothesis, processing differences can be attributed to L1 transfer effects as well other performance factors, e.g. computational limitations.

At the same time, our results do not favor the models that capitalize on various kinds of overt cues in the input, in order to explain that difference between L1 and L2 processing. For instance, the Shallow Structure Hypothesis (Clahsen and Felser, 2006) postulates that

the L1 and L2 processing mechanisms must be fundamentally different, L2 learners employing "shallow" processing strategies that lack syntactic detail compared to the native ones but, instead, rely more or lexical-thematic or pragmatic cues in lieu of computing abstract phrase structure. In this sense, the SSH may not sit well with the option of syntactic L1 transfer effects in processing L2 because such transfer effects presuppose the availability of a deeper grammatical structure in L2. Similarly, the Competition model (MacWinney, 1997) posit that speakers rely on different competing cues such as word order, animacy or case, during direct "form-to-function" mapping. In cases when L1 and L2 differ, the respective processing cues may compete, resulting in a negative transfer from L1 which hampers the learning of L2.[6] In contrast, in our clitic-verb manipulation we observed non-native performance in the absence of obvious overt cues in the input that would differentiate between the two types of structures. In particular, the adjacent clitic-verb combination in the speakers' L2 Italian cannot serve as a reliable cue since, as we saw above, it is also present in their L1 Slovenian. The only "cue" available to the learners in this case is an abstract one accessible by a reference to grammatical representations.

5. Conclusion

The present study showed that complex grammatical characterizations may interact in non-trivial ways at the processing level in L2 speakers. In particular, our L2 learners manifested native-like processing patterns in the domain of syntactic islands, for which the learners' L1 and L2 grammars provide a converging characterization. At the same time, despite the early age of L2 acquisition, they performed non-native-like with respect to clitic placement phenomenon, for which the L1 and L2 grammars do not align. We argued that each case of a non-native parsing should be seen in the context of competing grammatical characterizations in L1 and L2 and their possible interactions, as reflected in the online performance. In this regard, the grammar, in terms of pinpointing precise points of grammatical variation, may be a good predictor of the seemingly non-native patterns.

Acknowledgments

The authors thank Renata Bonet, Ilyana Krapova, Chiara Melloni and Manca Mušič for help with recruiting, and three anonymous SLR reviewers for valuable comments on the previous versions of this article.

Funding:

This work was supported by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613465. Adam Zawiszewski and Itziar Laka acknowledge funding from the Basque Government (IT665-13).

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplementary material

Supplementary material for this article is available online

Notes

[1] Under this generalization, the difference in L2 online processing between speakers of two overt wh-movement languages such as Spanish and German, discussed in the text, could possibly be attributed to further grammatical factors which affect processing, for instance, canonical word order (SVO in Spanish, SOV in German). See also Zawiszewski et al. (2011) and Erdocia and Laka (2018).

[2]A related type of cost-reducing efficiency principle suggested in the literature is Minimality, which states that "in the absence of explicit information to the contrary, the human language comprehension system assigns minimal structures" (Bornkessel and Schlesewsky 2009:1542). This principle entails that only required dependencies and relations are created.

[3] Earlier views on structural integration in behavioral sentence processing maintained that sentence-final regions may be subject to wrap-up effects that may potentially obscure investigation of the relevant integration processes (Just and Carpenter, 1980). More recently, Stowe et al. (2018) argue that the phenomena that were originally thought to constitute the set of alleged wrap-up effects (search for referents, construction of inter-clause relations or lexical ambiguity resolution) may occur also at the earlier points in the sentence (see also Kuperberg et al., 2006), and that the current set of the electrophysiological (ERP) studies do not support the wrap-up view on the sentence-final regions. This is not to deny that there exist clause- and sentence-final occurrences, for instance, punctuation effects operating over and above the structural integration and affecting reading times (Hill and Murray, 2000). Stowe et al. (2018) emphasize that sentence-final regions are legitimate sites for analyzing on-line measures, as long as the latter are not mixed with post-test decision-making processes such as an acceptability judgment task. That was indeed the case in our subject island manipulation.

[4] Another processing domain involving different grammatically-regulated options concerns relative clause attachment in sentences like *Someone shot the servant of the actress who was at the balcony,* where speakers of English and some other languages prefer a "low" attachment of the relative clause to the second NP (Frazier and Clifton, 1996), whereas speakers of Spanish, Greek or French were reported to prefer the "high" attachment, that to the first NP (e.g. Cuetos and Mitchell, 1988). This domain was also investigated in the L2 studies (Fernandez, 2003; Frenck-Mestre, 2002; Felser et al., 2003). However, Grillo and Costa (2014) argued that there is a confound in the previous studies regarding attachment preferences, with the availability of so-called pseudo-relative constructions in the high-attachment preference languages. Their findings thus potentially undermine the claim that different parsing strategies are involved in this empirical domain.

[5] This account also predicts that speakers whose L1 happens to entertain clitic placement possibilities that resemble those in Italian, in particular, avoiding the subset-superset trap, would show more native-like performance on Italian verb-clitic constructions than that of our Slovenian speakers in this study. This seems to us a promising venue for future research. We thank a SLR reviewer for highlighting this aspect.
[6] It should be noted that even though the Competition model focuses on learning and does not address L2 processing as such, we consider it relevant for discussing the online L2 strategies as well, in line with many other works on the subject.

References

Adiv E (1984) Language learning strategies: The relationship between L1 operating principles and language transfer in L2 development. In Anderson R (ed.), *Second languages: A cross-linguistic perspective*. Rowley: Newbury House, pp125–142.

Baayen RH (2008) *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge, UK ; New York: Cambridge University Press.

Bates D, Maechler M, Bolker B and Walker S (2014) Lme4: Linear mixed-effects models using Eigen and S4_. R package version 1.1-7.

Belikova A and White L (2009) Evidence for the fundamental difference hypothesis or not? Island Constraints Revisited. *Studies in Second Language Acquisition* 31: 199–223.

Birdsong D (ed.) (1999) *Second language acquisition and the Critical Period hypothesis*. Mahwah, NJ: Lawrence Erlbaum.

Bley-Vroman R (1989) What is the logical problem of foreign language learning? In: Gass SM and Schachter J (eds.) *Linguistic Perspectives on Second Language Acquisition*. Cambridge: Cambridge University Press, pp. 41–68..

Bley-Vroman R (2009) The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition* 31: 175–98.

Bornkessel-Schlesewsky I and Schlesewsky M (2009) Minimality as vacuous distinctness: Evidence from cross-linguistic sentence comprehension. *Lingua* 119(10): 1541–1559.

Box G and Cox D (1964) An Analysis of Transformations. *Journal of the Royal Statistical Society, Series B (Methodological)* 26: 211-252.

Boxell O and Felser C (2017) Sensitivity to parasitic gaps inside subject islands in native vs. non-native processing. *Bilingualism, Language and Cognition* 20: 494–511.

Burzio L (1986) *Italian syntax*. Dordrecht: Reidel.

Cardinaletti A (2004) Towards a cartography of subject positions. In: Rizzi L. (ed.), *The structure of CP and IP: The cartography of syntactic structures,* 2: Oxford University Press, New York, pp. 115-165.

Cunnings I, Batterham C, Felser C and Clahsen H (2010) Constraints on L2 learners' processing of wh-dependencies. Evidence from eye movements. In: VanPatten B and Jegerski J (eds.) *Research in Second Language Processing and Parsing*. John Benjamins Publishing Company, Amsterdam / Philadelphia, pp. 87-110.

Chen L, Shu H, Liu Y, Zhao J, and Li P (2007) ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition*, 10 (2): 161–174.

Chomsky N (1986) *Barriers*. Cambridge, MA: MIT Press.

Clahsen H and Felser C (2006) Grammatical processing in language learners. *Applied Psycholinguistics* 27(1): 3–42.

Clifton C andFrazier L (1989) Comprehending sentences with long distance dependencies. In Carlson GN and Tannenhaus MK (eds) *Linguistic structure in language processing*. Dordrecht: Kluwer, pp. 272–317.

Cuetos F and Mitchell DC (1988) Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition* 30(1): 73–105.

De Vincenzi M (1991) *Syntactic parsing strategies in Italian: The minimal chain principle*. Springer Science & Business Media.

De Vincenzi M, Job R, Di Matteo R, Angrilli A, Penolazzi B, Ciccarelli L and Vespignani F (2003) Differences in the perception and time course of syntactic and semantic violations. *Brain and Language* 85(2). 280–296.

Dekydtspotter L, Schwartz BD and Sprouse RA (2006) The Comparative Fallacy in L2 Processing Research. In: O'Brien MG, Shea C and Archibald J (eds) *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)*. Somerville, MA: Cascadilla Proceedings Project, pp. 33–40..

Ditman T, Holcomb PJ and Kuperberg GR (2007) An investigation of concurrent ERP and self-paced reading methodologies. *Psychophysiology* 44(6): 927–935.

Eckes T and Grotjahn R (2006) A closer look at the construct validity of C-tests. *Language Testing* 23(3): 290–325. Felser C, Leah R, Marinis T and Gross R (2003) The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics* 24(3): 453–489.

Enochson K and Culbertson J (2015) Collecting psycholinguistic response time data using Amazon Mechanical Turk. *PLoS ONE* 10(3): e0116946. https://doi.org/10.1371/journal.pone.0116946

Erdocia K and Laka I (2018) Negative transfer effects on L2 word order processing. *Frontiers in Psychology* 9:337. doi: 10.3389/fpsyg.2018.00337

Fernández EM (2003) *Bilingual Sentence Processing: Relative clause attachment in English and Spanish*. Amsterdam: John Benjamins Publishing Company.

Ferreira F and Clifton C (1986) The independence of syntactic processing. *Journal of memory and language* 25(3): 348–368.

Fodor JD (1978) Parsing strategies and constraints on transformations. *Linguistic Inquiry* 9: 427–473.

Frazier L (1978) *On Comprehending Sentences: Syntactic Parsing Strategies*. Doctoral dissertation, University of Connecticut, Storrs, CT.

Frazier L and Fodor JD (1978) The Sausage Machine: A New Two-Stage Parsing Model. *Cognition* 6: 291-325.

Frazier L and Clifton C (1989) Successive cyclicity in the grammar and the parser. *Language and cognitive processes* 4(2): 93–126.

Frazier L and Clifton C (1996) *Construal*. Cambridge, MA: MIT Press.

Frenck-Mestre C (2002) An on-line look at sentence processing in the second language. In: Heredia RR and Altarriba J (eds) *Bilingual Sentence Processing*. Amsterdam: Elsevier, pp. 217–236..

Gibson E (1998) Linguistic complexity: locality of syntactic dependencies. *Cognition* 68: 1–76.

Gibson E (2000) The dependency locality theory: A distance-based theory of linguistic complexity. In: Miyashita Y, Marantz A and O'Neil W (eds), *Image, language, brain*. Cambridge, MA: MIT Press , pp. 95–126. .

Golden M (1995) Interrogative wh-movement in Slovene and English. *Acta Analytica* 14: 145–187.

Golden M and Sheppard M (2000) Slovene Pronominal Clitics. In: Beukema and den Dikken (eds) *The Clitic Phenomena in European Languages*. Amsterdam: Benjamins, pp. 191–207.

Grillo N and Costa J (2014) A novel argument for the Universality of Parsing principles. *Cognition* 133(1): 156–187.

Grüter T and Crago M (2010) The roles of L1 transfer and processing limitations in the L2 acquisition of French object clitic constructions: Evidence from Chinese- and Spanish-speaking learners. In: Franich K, Iserman KM and Keil LL (eds..) *Proceedings of the 34th Annual Boston University Conference on Language Development (BUCLD)*, Somerville, MA: Cascadilla Press, pp.150–161.

Guasti MT (1994) Verb syntax in Italian child grammar: Finite and nonfinite verbs. *Language Acquisition* 3: 1–40.

Hahne A (2001) What's different in second-language processing: Evidence from event-related brain potentials. *Journal of Psycholinguistic Research* 30(3): 251–266.

Hahne A andFriederici AD (2001) Processing a second language late learners' comprehension mechanisms as revealed by event-related brain potentials. *Language and Cognition* 4 (2): 123–141.

Hartsuiker RJ, Pickering MJ and Veltkamp E (2004) Is Syntax Separate or Shared Between Languages?: Cross-Linguistic Syntactic Priming in Spanish-English Bilinguals. *Psychological Science* 15(6): 409–414.

Hawkins R and Chan CY (1997) The partial availability of Universal Grammar in second language acquisition: the 'failed functional features hypothesis'. *Second Language Research*, 13 (3): 187 – 226.

Hawkins R and Hattori H (2006). Interpretation of multiple wh-questions by Japanese speakers: a missing uninterpretable feature account. *Second Language Research* 22: 269–301.

Herschensohn J (2004) Functional categories and the acquisition of object clitics in L2 French. In: Prévost P and Paradis J (eds) *The Acquisition of French in Different Contexts: Focus on functional categories*, Amsterdam: Benjamins, pp.207–242.

Hill R, and Murray W (2000) Commas and spaces: Effects of punctuation on eye movements and sentence processing. In: Kennedy A, Heller D, and Pynte J (eds) *Reading as a perceptual process,* Amsterdam: Elsevier, pp. 565–589.

Hopp H (2007) *Ultimate Attainment at the Interfaces in Second Language Acquisition: Grammar and Processing*. Doctoral dissertation, University of Groningen (NL) .

Hothorn T, Bretz F and Westfall P (2008) Simultaneous Inference in General Parametric Models. *Biometrical Journal* 50(3): 346–363.

Jaeger T F (2008) Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4): 434–446.

Johnson JS and Newport EL (1989) Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 21(1): 60–99.

Johnson J and Newport EL (1991) Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition* 39(3): 215–258.

Juffs A (2005) The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research* 21(2): 121–151.

Just MA and Carpenter PA (1980) A theory of reading: From eye fixations to comprehension. *Psychological Review* 87(4): 329–354.

Just MA, Carpenter PA and Woolley JD (1982) Paradigms and processes and n reading comprehension. *Journal of Experimental Psychology: General* 3: 228–238.

Karimi N (2011) C-test and Vocabulary Knowledge. *Language Testing in Asia* 1(4): 7-38.

Kim E, Soondo Band Tremblay A (2015) The role of island constraints in Second Language Sentence Processing. *Language Acquisition* 22: 384 – 416.

Kotz S, Holcomb PH, and Osterhout L (2008) ERPs reveal comparable syntactic sentence processing in native and non-native reader of English. *Acta Psychologica*, 128: 514–527.

Kuperberg GR, Kreher DA, Goff D, McGuire PK and David AS (2006) Building up linguistic context in schizophrenia: Evidence from self-paced reading. *Neuropsychology* 20(4): 442–452.

Kurtzman HS and Crawford LF (1991) Processing parasitic gaps. In: Sherer T (ed) *Proceedings of the 21st Annual Meeting of the North Eastern Linguistic Society*, Amherst, MA: GLSA Publications, pp. 217–231.

Kuznetsova A, Brockhoff P, and Christensen R (2014). LmerTest: Tests for random and fixed effects for linear mixed effect models. R package, version 2.0-3.

Lardiere D (2013) Native and non-nativelike attainment. In: Herschensohn J and Young-Scholten M (eds) T*he Cambridge Handbook of Second Language Acquisition,* Cambridge: Cambridge University Press, pp. 670-691.

Li X (1998) Adult L2 accessibility to UG: An issue revisited. In: Flynn S, Martohardjono G and O'Neil W (eds) *The Generative Study of Second Language Acquisition*, New York: Psychology press, pp. 89 – 110.

Longobardi G (2000) Postverbal subjects and the mapping hypothesis. *Linguistic Inquiry* 31(4): 691-702.

MacWhinney B (1997) Second language acquisition and the competition model. In: de Groot AMB and Kroll JF *Tutorials in Bilingualism: Psycholinguistic Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 113–142.

Marinis T, Leah R, Felser C and Clahsen H (2005) Gaps in second language sentence processing. *Studies in Second Language Acquisition* 27(1): 53-78.

Matuschek H, Kliegl R, Vasishth S, Baayen H and Bates D (2017) Balancing Type I error and power in linear mixed models. *Journal of Memory and Language* 94: 305–315.

Montrul S (2011) Morphological errors in Spanish second language learners and heritage speakers. *Studies in Second Language Acquisition* 33(2): 163–192.

Mueller JL, Hahne A, Fujii Y and Friederici AD (2005) Native and Nonnative Speakers' Processing of a Miniature Version of Japanese as Revealed by ERPs. *Journal of Cognitive Neuroscience* 17(8): 1229–1244.

Mueller J, Hirotani M, and Friederici AD (2007) ERP evidence for different strategies in the processing of case markers in native speakers and non-native learners. *BMC Neuroscience*, 8 (18): 1–16.

Newmeyer FJ (1998) *Language form and language function*. Cambridge, MA: MIT Press.

Ojima S, Nakata H, and Kakigi R (2005) An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience* 17(8): 1212–1228.

Omaki A and Schulz B (2011) Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition* 33(4): 563–588.

Papadopoulou D and Clahsen H (2003) Parsing strategies in l1 and l2 sentence processing. *Studies in Second Language Acquisition* 25(4): 501-528.

Perani D and Abutalebi J (2005) The neural basis of first and second language processing. *Current Opinion in Neurobiology* 15(2): 202–206.

Phillips C (2006) The Real-Time Status of Island Phenomena. *Language* 82(4): 795–823.

Pickering M, Barton S and Shillcock R (1994) Unbounded dependencies, island constraints and processing complexity. In: Clifton CJ, Frazier L and Rayner K (eds) *Perspectives on sentence processing*, London: Lawrence Erlbaum Associates, pp. 199–224..

R Core Team (2014) R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Http://www.R-project.org/.

Rizzi L (1982) Violations of the wh-island constraint and the subjacency condition. In: Rizzi L (ed) *Issues in Italian syntax*, Dordrecht: Foris, pp. 49–76.

Ross JR (1967) Constraints on variables in syntax. Doctoral dissertation, MIT, Cambridge, MA.

Rossi E (2015) Modulating the sensitivity to syntactic factors in production: Evidence from syntactic priming in agrammatism. *Applied Psycholinguistics* 36(3): 639–669.

Rossi S, Gugler MF, Friederici AD, and Hahne A (2006). The impact of proficiency on second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, 18 (2): 2030–2048.

Rossi E, Kroll JF and Dussias PE (2014) Clitic pronouns reveal the time course of processing gender and number in a second language. *Neuropsychologia* 62: 11–25.

Santamaria K. and Sunderman G (2015) Working Memory in Processing Instruction: The Acquisition of L2 French Clitics. In: Zhisheng Z, Mota M, and MacNeil (eds) *Working Memory in Second Language Acquisition and Processing; Theory, Research & Practice,* Multilingual Matters pp. 43)..

Schachter J (1990) On the issue of completeness in second language acquisition. *Second Language Research*, 6(2): 93 – 124.

Schwartz BD and Sprouse RA (1996) L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research* 12(1): 40–72.

Schwartz, BD (1987) The modular basis of second language acquisition. Doctoral dissertation, University of Souther California, Los Angeles, CA..

Sprouse J, Caponigro I, Greco C and Cecchetto C (2016) Experimental syntax and the variation of island effects in English and Italian. *Natural Language & Linguistic Theory* 34(1): 307-344.

Sprouse J, Wagers M and Phillips C (2012) A test of the relation between working-memory capacity and syntactic island effects. *Language* 88(1): 82–123.

Stepanov A, Mušič M and Stateva P (2016) Asymmetries in sub-extraction out of NP in Slovenian: a magnitude estimation study. *Linguistica* 56: 253-271

Stepanov A, Mušič M and Stateva P (2018) Two (non-) islands in Slovenian: a study in experimental syntax. *Linguistics: An interdisciplinary journal of the language sciences* 56(3): 435-476.

Stowe LA (1986) Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes* 1: 227–245.

Stowe LA, Kaan E, Sabourin L and Taylor RC (2018) The sentence wrap-up dogma. *Cognition* 176: 232–247.

Traxler, MJ and Pickering MJ (1996). Plausibility and the processing of unbounded dependencies: an eye-tracking study. *Journal of Memory and Language* 35: 454– 475.

Ullman MT (2001) The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition* 4(2): 105-122.

VanPatten B (2005) Processing instruction. In: Sanz C (ed) *Mind and Context in Adult Second Language Acquisition*, Washington, DC: Georgetown University Press, pp. 267–281.

Weber-Fox CM and Neville HJ (1996) Maturational Constraints on Functional Specializations for Language Processing: ERP and Behavioral Evidence in Bilingual Speakers. *Journal of Cognitive Neuroscience* 8(3): 231–256.

White L (1989) *Universal grammar and second language acquisition*. Amsterdam: Benjamins.

White L (2000) Second language acquisition: From initial to final state. In: Archibald J (ed), *Second language acquisition and linguistic theory*, Oxford: Blackwell, pp. 130–155.

White L (2003) *Second language acquisition and universal grammar*. Cambridge, U.K.: Cambridge University Press.

Williams JN, Möbius P and Kim C (2001) Native and non-native processing of English wh-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics*, 22(4): 509–540.

Zawiszewski A, Gutiérrez E, Fernández B and Laka I (2011) Language distance and non-native syntactic processing: Evidence from event-related potentials. *Bilingualism: Language and Cognition* 14(3): 400–411.

Words (excluding the abstract and the references): 9809