THE PROCESSING OF GENDER AGREEMENT DURING READING COMPREHENSION IN L2 FRENCH: THE EFFECTS OF SYNTACTIC COMPLEXITY AND WORKING MEMORY

Marco Bril and Andrea Roijen[1]

Abstract: A longstanding debate in L2 research focuses on how syntactic complexity needs to be operationalized to account for L2 performances. Whereas many studies investigated this issue in L2 production, very few studies focused on L2 sentence processing. The present study aimed at investigating the effect of syntactic complexity on gender agreement processing in L2 French, while controlling for the learners' working memory capacity. We tested 37 Dutch learners of French by means of a self-paced reading technique. The results showed decreased sensitivity to gender agreement in embedded structures, but increased sensitivity to gender agreement in non-embedded structures. We concluded that the number of clauses in gender agreement constructions accounts for the effect of syntactic complexity on gender agreement processing in L2 French and that this measure is negatively correlated to sensitivity to gender agreement. We furthermore concluded that (non-verbal) working memory does not affect L2 gender agreement processing.

Keywords: gender agreement, L2 processing, reading, syntactic complexity, working memory

1. Introduction

Defining the construct of complexity has been shown to be very challenging in many studies on second language (L2) acquisition over the past decades since the construct of complexity in L2 research can be interpreted in different ways. In L2 acquisition research the notion of complexity can be used both as an independent and a dependent variable (e.g. Pallotti 2015). Complexity as an independent variable refers to communicative task characteristics which make the task more or less complex, while complexity as a dependent variable refers to the description of L2 performances, often combined with accuracy and fluency, as an indicator of L2 proficiency. In relation to both types of complexity, Pallotti (2015) pointed out that complexity can roughly be interpreted in three ways: inherent complexity, difficulty or cognitive complexity and processing difficulty. Inherent complexity refers to objective structural complexity based on formal properties of the linguistic system (e.g. as an independent variable in communicative tasks) (cf. Pallotti 2009). Difficulty or cognitive complexity deals with the processing costs related to the difficulty of processing linguistic items by L2 learners (e.g. as a dependent variable during sentence processing) (cf. Bulté & Housen 2012). Processing difficulty is related to the developmental order in which grammatical structures are acquired by L2 learners such as the idea that complex structures are acquired late (e.g. as a dependent variable in L2 production) (cf. Ellis 2009). Whereas inherent complexity is concerned with the structural complexity of the linguistic system and builds on the idea that language does not become more complex over time, cognitive complexity and processing difficulty are taken to be more dynamic constructs related to

* Utrecht University, m.bril@uu.nl.

Bucharest Working Papers in Linguistics XXV, 2, 59-80, ISSN 2069-9239, ISSN-L 2069-9239
DOI: 10.31178/BWPL.25.2.4

L2 processing or production during which (the effect of) complexity changes over time. As such, inherent complexity is related to an interpretation of complexity which is different from interpretations of complexity in terms of difficulty (Bulté & Housen 2012, Pallotti 2015). This does, however, not mean that cognitive complexity could not be related to inherent complexity. The extent to which inherent and cognitive complexity are related is under debate. Indeed, some studies showed that inherent complexity does not fully mirror cognitive complexity as executive functions such as working memory or conflict resolution, have been found to better explain differences in processing costs instead of the inherent complexity of grammatical constructions during sentence processing (see e.g. Fernandez-Duque 2009 for L1 sentence processing, Öttl et al. 2015 for artificial grammar processing, Gilardone et al. 2023 for atypical (L1) sentence processing). This may indicate that inherent and cognitive complexity are indeed two different interpretations of complexity. However, some studies on human language processing showed that cognitive complexity goes hand in hand with inherent complexity in language processing in the sense that syntactically complex constructions require more processing costs than simpler ones (e.g. Ferreira et al. 1996, Traxler et al. 2014), which may indicate that these two interpretations of complexity share a common notion of complexity.

The present study deals with cognitive complexity as a dependent variable and inherent complexity as an independent variable, as it focuses on the cognitive costs involved in the L2 processing of gender agreement constructions manipulated with respect to their inherent complexity.

A subconstruct of inherent complexity is linguistic complexity (Bulté & Housen 2012). This type of complexity refers to the objective structural complexity based on lexical, morphological or syntactic properties of the linguistic system. As such, linguistic complexity has been defined and operationalized using different quantitative measures in L2 research. Bulté and Housen (2018: 149) pointed out that "typically, L2 complexity is operationalized in terms of average length (of different syntactic units), diversity (especially of lexical and/or morphological items) and amount of embedding (at the sentence level)". Of particular interest in the present study is syntactic complexity as a subconstruct of linguistic complexity referring to the objective structural complexity of sentences based on the syntactic properties of the L2. In this respect, the amount of embedding has been operationalized in terms of the number of clauses per sentence and has been most used in L2 research as an objective, quantitative measure of syntactic complexity as compared to the average length of syntactic units (Pallotti 2015). As the focus of the present paper is on syntactic complexity, we will not further discuss the diversity of lexical and/or morphological items as a measure of L2 linguistic complexity.

The aims of the present study are to investigate i.) the potential effect of syntactic complexity on the processing (i.e. the ability of parsing and interpreting morphosyntactic information in sentences) of gender inflection on adjectives in L2 French and ii.) which quantitative measure(s) of syntactic complexity (i.e. the average length of syntactic units and/or the number of clauses per sentence) may account for this potential effect in L2 French.

Previous studies showed that gender inflection on adjectives (i.e. the process by which adjectives agree with a noun and are marked for the noun's gender such as *le*

ballon vert 'the ball.M.SG green.M.SG' vs. *la jupe verte* 'the skirt.F.SG green.F.SG' in French) is prone to errors that persist to an advanced level of proficiency in adult L2 acquisition (e.g. Bril 2016, Bartning 2000, Bril 2021). To account for the difficulty of processing gender inflection in adult L2 the syntactic complexity of noun – adjective agreement structures has been related to inflectional variability in L2 performance (e.g. Bonilla 2015). Bonilla (2015) investigated the L2 learner's development regarding the production of gender inflection on adjectives in Spanish. The results showed that gender inflection in L2 Spanish is first produced in noun – attributive adjective constructions, then in noun – predicative adjective constructions and finally in relativized noun – adjective constructions, indicating that the syntactic complexity of noun – adjective agreement constructions affects the accuracy of gender inflection in the sense that gender inflection is more difficult to correctly produce in complex constructions (i.e. relativized noun – adjective constructions) than in simpler ones (i.e. noun – attributive adjective constructions). Under the view that linguistic competence is taken to have a dual nature comprising the implementation of grammatical knowledge in real-time computation and the production/comprehension of grammatical constructions in communicative contexts (Marinis et al. 2005, Hopp 2006, Foucart 2008), studies on the L2 processing of gender agreement (i.e. the processing of the structural dependency between the noun and the adjective agreeing with the noun in gender) may complement our understanding of operationalizing syntactic complexity and its role in the L2 acquisition of gender inflection. To the best of our knowledge, the potential effect of syntactic complexity on the L2 processing of gender agreement has, however, not been investigated in previous processing research.

Besides linguistic factors, general cognitive abilities such as working memory capacity, have also been related to grammatical processing in L2. Working memory capacity has been defined as 'the ability to maintain and manipulate information in active attention' (Reynolds et al. 2022: 1254, based on Schneider & McGrew 2018). However, previous studies did not show consistent results regarding the role of working memory capacity in grammatical processing in L2. While there seems to be growing evidence that working memory is related to morphosyntactic processing in L2 (e.g. Havik et al. 2009, Reichle et al. 2016, Dracos & Henry 2021; Gabriele et al. 2021), other studies showed that working memory is not associated with morphosyntactic processing during comprehension, i.e. the real-time interpretation of morphosyntactic information (e.g. Rodríguez 2008, Foote 2011, Baek 2012). Therefore, we will also investigate whether the learners' working memory capacity can be taken as a factor for individual variation during gender agreement processing in L2 French.

2. Defining the syntactic complexity of gender agreement constructions in French

As mentioned by Pallotti (2015), previous L2 research has typically taken the average length of syntactic units or the amount of embedding as an objective, quantitative measure of syntactic complexity at the sentence level.

Regarding the average length of syntactic units as a measure of syntactic complexity, this length can be operationalized in terms of the number of words per

constituent or the number of (syntactic) phrases per clause (Pallotti 2015). Based on the operationalizations described in Pallotti (2015), complex structures in L2 are taken to comprise more words per constituent, more phrases per clause or more clauses than simpler structures. With respect to the complexity measure based on the number of words per constituent, noun – attributive adjective gender agreement constructions in French (e.g. *Louise achète une jupe violett**e** dans une boutique* 'Louise buys a purple.F.SG skirt.F.SG in the shop') for instance, consist of 3 words (i.e. *une jupe* (2 words) and *violette* (1 word)), noun – predicative adjective gender agreement constructions (e.g. *La beauté est important**e** pour Claude et moi* 'Beauty.F.SG is important.F.SG for Claude and me') consist of 4 words (i.e. *la beauté* (2 words), *est* (1 word) and *importante* (1 word)) and relativized noun – adjective gender agreement constructions (e.g. *Momo déteste la popularité qui est important**e** à l'école* 'Momo hates the popularity.F.SG which is important.F.SG at school') consist of 5 words (i.e. *la popularité* (2 words), *qui* (1 word), *est* (1 word) and *importante* (1 word)). As such, noun – attributive adjective constructions are considered as less complex than noun – predicative adjective constructions and relativized noun – adjective constructions, and noun – predicative adjective constructions as less complex than relativized noun – adjective constructions (see Bartning 2000 for the same operationalization of syntactic complexity in French gender agreement constructions).

As regards the measure of syntactic complexity based on the number of clauses, both noun – attributive adjective and noun – predicative adjective gender agreement constructions exhibit the same level of syntactic complexity since both constructions yield 1 clause. However, relativized noun – adjective gender agreement constructions yield 2 clauses due to embedding. As such, the latter can be taken as more complex than noun – attributive adjective and noun – predicative adjective gender agreement constructions (see Alarcón 2021 for the same operationalization of syntactic complexity in Spanish adjectival agreement constructions).

The complexity measure based on the number of (syntactic) phrases per clause defines noun – attributive adjective gender agreement constructions as the least complex gender agreement construction in French since they consist of 3 phrases (i.e. DP, NP and AP), whereas noun – predicative adjective gender agreement constructions consist of 4 phrases (i.e. DP, NP, VP and AP) and relativized noun – adjective gender agreement constructions consist of 5 phrases (i.e. DP and NP in the main clause and CP, VP and AP in the embedded clause). The latter taken to be the most complex (see Bril 2021 for the same operationalization of syntactic complexity in French gender agreement constructions). Table 1 gives an overview of the hierarchy of gender agreement constructions in French based on the operationalizations of syntactic complexity described. 1 represents the gender agreement construction exhibiting the lowest level of syntactic complexity, whereas 3 represents the gender agreement construction exhibiting the highest level of syntactic complexity.

Table 1. Hierarchy of gender agreement constructions per measure of syntactic complexity

Type of gender agreement construction	Example	Measure 1 (number of words per constituent)	Measure 2 (number of clauses)	Measure 3 (number of phrases per clause)
Noun – attributive adjective	*Une jupe violett**e***	1	1	1
Noun – predicative adjective	*La beauté est important**e***	2	1	2
Relativized noun – adjective	*La population qui est important**e***	3	2	3

3. Syntactic complexity and working memory in the L2 processing of gender agreement

As mentioned in the introduction, the syntactic complexity of gender agreement constructions has been shown to affect the performance at producing gender inflection in L2. Studies on the L2 processing of gender agreement (e.g. de Jong 2005, Foucart & Frenck-Mestre 2012, Gabriele et al. 2013, Alemán Bañón et al. 2018) reported effects of adjective position on the L2 processing of this type of agreement, which *may* suggest that syntactic complexity also affects the L2 processing of gender agreement. De Jong (2005) for instance, investigated the comprehension and production of noun-adjective gender agreement in L2 Spanish. Dutch learners of Spanish received implicit instruction on a miniature linguistic system targeting gender agreement, followed by a receptive or a receptive-productive training session (or no training in case of the control group). They were tested by means of a self-paced listening test, a match-mismatch test and a grammaticality judgment test to test receptive knowledge, and a picture description task to test productive knowledge. The results showed that both types of training lead to knowledge of gender agreement in comprehension, but much less in production. Regarding the self-paced listening test, the L2 processing of noun - adjective gender agreement revealed to be affected by the position of the adjective. More precisely, longer listening times were found for attributive adjectives as compared to predicative ones. This seems to suggest that the L2 processing of gender agreement is influenced by syntactic complexity in the sense that in more complex constructions (e.g. noun – predicative adjective constructions) gender agreement processing is reflected by shorter reaction times than in simpler constructions (e.g. noun – attributive adjective constructions).

Similar effects of adjective position on noun – adjective gender agreement processing in L2 were reported in studies in which event-related potentials (ERPs) were measured (e.g. Foucart & Frenck-Mestre 2012; Gabriele et al. 2013; Alemán Bañón et al. 2018). ERPs are brain responses which are elicited by linguistic elements of stimuli such as inflections. These responses are a measure of the real-time interpretation of linguistic elements and can be captured by a P600 or a N400 effect. Typically, the P600 is a

positive waveform that shows up when there is a syntactic violation making the sentence ungrammatical, while the N400 is a negative waveform which is elicited by lexical mismatches (e.g. Friederici et al. 1996) or by morphosyntactic violations in early stages of L2 acquisition (McLaughlin et al. 2010). Both ERP effects are taken to reflect high sensitivity to syntactic or lexical violations, while the lack of ERP effects indicates no sensitivity to syntactic or lexical violations (e.g. Gabriele et al. 2013). Foucart and Frenck-Mestre (2012) investigated the processing of gender agreement in native and L2 French by means of an ERP experiment. Participants were presented with grammatical and ungrammatical noun - adjective agreement constructions in which the adjective was in an (prenominal or postnominal) attributive or a predicative position. In the L2 learners group gender agreement violations on postnominal adjectives triggered P600 effect, while violations on prenominal adjectives triggered a N400 effect. In contrast, gender agreement violations on predicative adjectives did not trigger any ERP effect. These results show that L2 learners of French are less sensitive to gender agreement violations on adjectives in a predicative position than in an attributive position, indicating that less attention is paid to gender agreement in predicative adjectives than in attributive ones. In a similar vein, Gabriele et al. (2013) investigated the processing of number and gender agreement in native and L2 Spanish by means of an ERP experiment, while controlling for proficiency. With respect to gender agreement, participants were presented with grammatical and ungrammatical noun – adjective constructions in which the adjective was in an attributive or a predicative position. For the native, intermediate and advanced L2 learners group the ERP results showed that noun – adjective constructions in which the adjective was in an attributive position yielded more P600 effects than noun – adjective constructions in which the adjective was in a predicative position, regardless of the sentences' grammaticality. This means that the predicative position of the adjective reduces the sensitivity to gender agreement in both grammatical and ungrammatical agreement constructions. For low proficient participants, however, this effect of adjective position on sensitivity to gender agreement did not show up. Alemán Bañón et al. (2018) tested native speakers and L2 learners of Spanish with different levels of proficiency in the same type of experiment as in Gabriele et al. (2013). By focusing on number and gender agreement they investigated how morphosyntactic development is modulated by typological similarities and the adjective's position in agreement constructions. In line with Gabriele et al. (2013), the ERP results showed that advanced learners elicited more P600 effects in noun – adjective constructions in which the adjective was in an attributive position as compared to noun – adjective constructions in which the adjective was in a predicative position. As in Gabriele et al. (2013), this effect did not show up in the low proficient learners group. Overall, these findings indicate that the L2 processing of gender agreement is affected by the adjective's position in the sense that noun – adjective constructions in which the adjective is in an attributive position trigger longer reaction times or more P600 effects on the adjective than noun – adjective constructions in which the adjective is in a predicative position. The reason for these findings can be sought in the fact that L2 learning is cognitively more taxing as compared to native language learning. Consequently, they lack the ability to retain linguistic information of the noun in their working memory during gender processing in constructions in which the adjective is separated from the noun, such as by a verb in noun – predicative adjective constructions

(Clahsen & Felser 2006). This means that less attention is paid to gender agreement on adjectives in a predicative position than in an attributive position. While this effect of adjective position on L2 gender agreement processing seems to suggest that the syntactic complexity of gender agreement constructions influences the sensitivity to gender agreement (violations) in L2, the effect of syntactic complexity has, however, not been investigated in previous studies on gender agreement processing in L2.

General cognitive abilities have also been related to the L2 processing of grammatical inflection. Of particular interest in studies investigating morphosyntactic processing in L2 is the learners' working memory capacity (e.g. Sagarra 2007, Havik et al. 2009, Sagarra & Herschensohn 2010). In this respect, working memory concerns the temporary storage and simultaneous manipulation of linguistic information during tasks demanding much cognitive effort such as the processing of linguistic input and is considered as a multicomponent model consisting of the central executive component, the phonological loop, the visuo-spatial sketchpad and the episodic buffer (Baddeley 2003). The central executive component principally deals with attentional control while coordinating and planning linguistic input. The phonological loop and the visuo-spatial sketchpad are considered as temporary store systems for phonological information and visuo-spatial information respectively. The episodic buffer component is principally concerned with the processing and temporary storage of linguistic information from different dimensions. Regarding processing, it assembles linguistic information from different dimensions (e.g. from phonological, visual or semantic representations, and from long-term memory) into a unitary multidimensional representation. Regarding storage, it temporarily stores unitary multidimensional representations of linguistic information during language processing (Rönnberg et al. 2009). As the processing of grammatical inflection involves the temporary storage of the first word of the agreement configuration until the second word is encountered, one may predict that working memory capacity affects the processing of gender inflection in L2. As mentioned in the introduction, an ongoing debate in L2 literature concerns the question of whether working memory constrains morphosyntactic processing in L2. Caplan & Waters (1999) argued that working memory constrains linguistic performances if the meaning of a sentence is used to execute these performances. More precisely, the accuracy of interpreting sentences decreases if working memory capacity decreases as well. However, processes during which syntactic structures need to be assigned to sentences are not influenced by working memory capacity (see also Rodríguez 2008, Foote 2011, Baek 2012). Based on different online (i.e. real-time) measures, recent studies showed contrastive results in the sense that working memory capacity affects the processing of L2 morphosyntax (self-paced reading: Havik et al. 2009, Sagarra & Herschensohn 2010, Dracos & Henry 2021; ERP: Reichle et al. 2016, Gabriele et al. 2021; Eye-tracking: Sagarra 2021). These studies demonstrated that the L2 learners' difficulty to process agreement constructions is due to working memory limitations (Clahsen & Felser 2006). Specifically, learners with decreased working memory capacity can store less linguistic information which leads to more cognitive efforts to process this information. As a result, these increased processing demands affect the implementation of linguistic cues of the first word of agreement dependencies (Hopp 2007). Sagarra (2007) found similar results in English learners of Spanish when focusing on the sensitivity to gender agreement violations. Reading times

collected by means of a self-paced reading experiment showed that learners with higher working memory capacity were more sensitive to gender agreement violations than those with lower working memory capacity (see also Sagarra & Herschensohn 2010 based on the same experimental design as Sagarra 2007). This study also showed an interaction between the effect of working memory capacity and the syntactic complexity of gender agreement constructions. Specifically, the sensitivity to gender agreement violations within the noun phrase (e.g. in attributive constructions) was found to be moderated by working memory, whereas the sensitivity to gender agreement violations across clauses (e.g. in relative constructions) was not. Since syntactic constructions in which gender agreement takes place within the noun phrase can be considered as less complex than those in which gender agreement takes place across clauses, the effect of working memory on L2 gender agreement processing might be mediated by the syntactic complexity of gender agreement constructions.

4. Research questions and hypotheses

The present study deals with the L2 processing of gender agreement in agreement constructions exhibiting different levels of syntactic complexity, while controlling for the learner's working memory capacity. Specifically, we will focus on gender inflection in noun – attributive adjective, noun – predicative adjective and relativized noun – adjective agreement constructions in French. We addressed the following research questions:

RQ1: Does syntactic complexity affect the processing of gender agreement in Dutch intermediate L2 learners of French during reading comprehension?
RQ2: If yes, which quantitative measure(s) of syntactic complexity account(s) for this effect in L2 French?
RQ3: Does working memory affect the processing of gender agreement in Dutch intermediate L2 learners of French during reading comprehension?

Based on previous studies on syntactic complexity, gender agreement processing and the role of working memory in gender agreement processing in L2, we hypothesized that

(i) the L2 processing of gender agreement reveals shorter reaction times in relativized noun – adjective agreement constructions than in noun – attributive adjective and noun – predicative adjective agreement constructions (De Jong 2005, Clahsen & Felser 2006, Pallotti 2015),
(ii) the L2 processing of gender agreement reveals shorter reaction times in noun – predicative adjective agreement constructions than in noun – attributive adjective agreement constructions, based on the complexity measures in terms of the number of words per constituent and the number of (syntactic) phrases per clause (De Jong 2005, Clahsen & Felser 2006, Foucart & Frenck-Mestre 2012, Gabriele et al. 2013, Pallotti 2015, Alemán Bañón et al. 2018),
(iii) the L2 processing of gender agreement reveals equal duration of reaction times in noun – predicative adjective agreement and noun – attributive adjective agreement constructions, based on the complexity measure in terms of the number of clauses per sentence (Pallotti 2015), and that

(iv) working memory affects the L2 processing of gender agreement in French during reading comprehension in noun - attributive and noun – predicative adjective agreement constructions and does not in relativized noun – adjective agreement constructions (Sagarra 2007).

5. Method

5.1 Participants

Thirty-seven Dutch learners of French (M_{age} = 17.7 years; SD_{age} = 4.3) took part in this experiment. Bilingual learners were excluded to avoid transfer effects from other languages than Dutch. By bilingual learners we mean language learners who acquired more than one language during early childhood and use both languages in their everyday lives (Grosjean 2008). They were all enrolled in French language classes for 5 to 7 years and were at an intermediate (B1 or B2) level of proficiency, which was tested by means of a standardized reading test. As proficient L2 learners have been shown to be sensitive to the effect of adjective position on L2 gender agreement processing (Gabriele et al. 2013' Alemán Bañón et al. 2018), this population can be taken as appropriate for this experiment. All participants confirmed that they had no language disorder such as dyslexia, and gave written informed consent for the experiment. They did not receive any information about the purpose of this study.

5.2 Materials and procedures

The materials consisted of a self-paced reading task to measure the speed of processing gender inflection and a forward-backward digit span task to measure the learner's working memory capacity. All data and experimental stimuli are available in the OSF repository at https://osf.io/m7xfh/?view_only=29d90e1025a04ca5909aabba4ab9dee4.

5.2.1 Self-paced reading task

The self-paced reading task was based on a non-cumulative moving window technique (Just et al. 1982) which has been shown to be appropriate to measure the L2 processing of inflectional morphology in previous studies (e.g. Hopp 2006, Jackson 2008; Pliatsikas & Marinis 2013). In the present experiment this technique provides the reaction time on the adjective which is assumed to reflect the sensitivity to gender agreement during the processing of the adjective combined with the structural dependency between the noun and the adjective agreeing with the noun in gender. Each segment comprises one word (determiners are integrated in the same segment as the noun (see Pan & Felser 2011 for instance)). Participants are presented with the test stimuli word-by-word on a computer screen and need to push on a pacing button once they have read the segment's word.

In this study the experimental items consisted of 75 stimuli which were categorized in 3 types of gender agreement constructions (i.e. noun – attributive adjective, noun – predicative adjective and relativized noun – adjective agreement constructions), resulting

in 25 stimuli per category. This categorization was based on the level of syntactic complexity of these gender agreement constructions as shown in section 2. Furthermore, 15 filler items targeting other grammatical elements than gender agreement were added. All items were created in accordance with the criteria described in Jegerski (2014). As such, each sentence contained 6 à 8 segments of which 1 contained a noun and 1 contained an adjective (e.g. for noun – attributive adjective constructions : *Louise / achète / #une jupe / #violette / dans / une boutique*, for noun – predicative adjective constructions: *La beauté / #est / #importante / pour / Claude / et / moi* and for relativized noun – adjective constructions: *Momo / déteste / la popularité / qui / #est / #importante / à / l'école*). Pronoun – adjective constructions (e.g. *elle est importante* 'she.F.SG is important.F.SG') were not used in our experiment. The # indication shows the regions of measurement. In each test item processing speed was measured in both the adjective segment and the segment preceding the adjective (i.e. a noun in the noun – attributive adjective construction and *est* 'is' in the noun – predicative adjective and the relativized noun – adjective constructions). As the segment preceding the adjective is not comparable across conditions (*est* vs. a noun), we only focused on the processing of *est* in the noun – predicative adjective and the relativized noun – adjective construction in order to ensure that potential differences in reaction times on the adjective between experimental conditions are exclusively related to effects of syntactic complexity on the processing of gender agreement. As such, the segment preceding the adjective (i.e. *est*) is a control segment by which latencies on the adjective may exclusively be related to effects of syntactic complexity on gender agreement (see Sagarra & Herschensohn 2010 for the same procedure). Furthermore, we did not focus on the potential effect of working memory on the processing of the segment preceding the adjective. With respect to the noun, all nouns were inanimate and in a feminine singular context. Masculine contexts were avoided because these contexts lack overt gender inflection on the adjective in French (e.g. *le ballon vert-ø* 'the ball.M.SG green.M.SG' vs. *la jupe violett-e* 'the skirt.F.SG purple.F.SG'). As correct gender assignment (i.e. categorizing a noun in a gender class such as masculine or feminine in French) has been shown to be a prerequisite for accurate gender agreement (Hopp 2016), gender was expressed by an overt determiner so that gender agreement processing was not affected by incorrect gender assignment. The noun's segment did not contain determiners in which elision (i.e. the suppression of unstressed vowels in French such as *l'* 'the') has taken place. Regarding the adjective, all adjectives contained 1 to 4 syllables and exhibited regular overt gender inflection ending in *-e* (e.g. *la jupe violette* 'the skirt.F.SG purple.F.SG'). Adjectives exhibiting stem alternations such as *beau* 'nice.M.SG' vs. *belle* 'nice.F.SG', or those derived from other lexical categories such as *travailleur* 'diligent.M.SG' vs. *travailleuse* 'diligent.F.SG', were not included in the experimental items. The adjectives included did not differ with respect to the number of syllables across experimental conditions ($X^2(2) = 1.38$; $p = .50$). This means that the length of adjectives varies to the same extent across conditions. Furthermore, only phonologically expressed gender inflection on adjectives was used to avoid effects of phonological cues on processing speed (as shown in Carrasco-Ortiz & Frenck-Mestre 2014). To minimize the effect of word frequency on L2 processing during reading comprehension (as shown in Kim & Kim 2012) all adjectives included were based on the Brunet (2014)'s corpus for word frequency in French and did not show

differences in word frequency across experimental conditions ($F(2,26) = 2.21$; $p = .13$). This means that the frequency of adjectives varies to the same extent across conditions. With respect to the noun – attributive adjective condition, all adjectives were in a postnominal position since previous research demonstrated that the L2 processing of gender inflection is affected by the (postnominal vs. prenominal) position of the adjective (Foucart & Frenck-Mestre 2012). Since the adjective segment was controlled for length, frequency and phonological expression of gender inflection, differences in reaction times on the adjective segment may be related to the processing of gender agreement.

To collect the reaction times in the regions of measurement self-paced reading software (ZEP) was installed on a Linux computer. In a quiet classroom, participants were seated in front of a computer screen on which the test stimuli were presented segment-by-segment. They proceeded to the next segment by pressing on a button box. During short instruction participants were told to press the button at their own pace. Before the task started 4 practice stimuli were presented so that participants were familiar with the experiment. After these practice stimuli each participant was presented with all experimental items (i.e. items from all experimental conditions) and these items were presented using one and the same self-paced reading task in which the presentation order was counterbalanced for each participant. A yes-no comprehension question showed up after a random set of stimuli to verify whether participants paid attention to these stimuli. Answers were used to check whether all participants understood the sentences. All participants finished the task well within 20 minutes.

5.2.2 Forward-backward digit span task

The participants' working memory capacity was measured by means of a forward-backward digit span task. This task was part of the WISC intelligence test (Wechsler 2003) and has been shown to highly predict L2 reading comprehension (Kormos & Sáfár 2008). Since other working memory tasks such as non-word repetition tasks and reading span tasks (e.g. Daneman & Carpenter 1980), seem to measure proficiency and experience rather than the learners' working memory (MacDonald & Christiansen 2002), a forward-backward digit span task has been used as an appropriate task to measure working memory capacity in previous studies (e.g. Kormos & Sáfár 2008, Juffs & Harrington 2011). In this task, a series of digits (i.e. numerals) was orally presented to the participants and they were asked to reproduce the digits. Each digit was 1 second in duration and was presented once. In case of the forward version, participants needed to reproduce the digits in the same order, while in case of the backward version, participants needed to reproduce the digits in reverse order. The first series comprised 2 digits and the last one comprised 9 digits for the forward version and 8 digits for the backward version (per 2 trials the series increased with 1 digit). The measure of working memory capacity was defined as the highest number of correctly reproduced series. All participants were in a quiet classroom and the task was administered by one of the researchers.

5.3 Data analysis

As mentioned in section 5.2, reaction times were collected on both the adjective segment and the segment preceding the adjective. These reaction times (in milliseconds) were defined as the time lapse between the segment's onset and the button press, which reflects the speed of processing the segment under investigation. Based on predetermined cut-off points for non-native speakers, reaction times lower than 200 milliseconds (as in Luce 1986) and higher than 4000 milliseconds (as in Havik et al. 2009) were removed. For each participant the mean of reaction times were calculated per segment of measurement for each experimental condition. Per condition and segment of measurement the data were inspected for outliers. Outliers were removed by winsorizing the data. This means that outliers were replaced with the lowest or highest value that is not an outlier (see Field 2013). The data were analyzed by means of a mixed ANCOVA to detect differences in reaction times between the types of agreement constructions on both the adjective segment and the segment preceding the adjective (see Bril et al. 2021 for the same statistical analysis used in a self-paced listening study on syntactic complexity and reaction times in L2 listening). The dependent variable was the reaction time in milliseconds per segment, the independent within-subjects variable was the Segment of Measurement yielding 2 levels: the segment preceding the adjective and the adjective segment. The independent between-subjects variable was Sentence Type (i.e. the type of agreement construction) containing 3 levels: noun – attributive adjective, noun – predicative adjective and relativized noun – adjective constructions. To account for potential mediating effects of working memory, the scores of the forward-backward digit span task were included as a covariate. For an ANCOVA the covariate and the between-subjects variable need to be independent to avoid reduction of the experimental effect (Wildt & Ahtola 1978). Since the working memory scores were equal for all sentence types (i.e. each participant performed all sentence types), these scores did not affect the potential effect of Sentence Type across conditions and can therefore be considered as independent of the Sentence Type variable. Following the procedures described in Field (2013), the assumption of homogeneity of regression slopes was tested and did not reveal to be violated ($F(2,102) = .43$; $p = .65$). Except the reaction times on the segment preceding the adjective in the noun – attributive adjective condition, all segments of measurement were non-parametric ($p < .05$) and violated the assumption of normality. Since ANCOVA is robust for non-normality if group sizes are equal (Donaldson 1968), this statistical analysis can be taken as appropriate for data analysis. Post-hoc contrast analyses with Bonferroni correction were run to detect differences in reaction times between types of agreement constructions and segments of measurement. For all statistical analyses the α level of significance was .05.

6. Results

Table 2 displays the descriptive data of reaction times per segment of measurement and sentence type, and the working memory scores. As the data were found to be non-parametric, 5 parameter statistics were reported. These data are set out in Figure 1 and 2.

Table 2. Descriptive data of reaction times (in milliseconds) per segment and sentence type, and working memory scores

	Segment	Sentence type	*Min*	*Lower quartile*	*Median*	*Upper quartile*	*Max*
Reaction times	Segment preceding adjective	noun-attributive adjective	313.00	481.78	674.80	1304.82	2194.00
		noun-predicative adjective	270.00	421.12	524.44	722.78	955.32
		relativized noun-adjective	253.96	389.56	484.04	687.68	1036.16
	Adjective	noun-attributive adjective	339.64	477.26	654.28	954.78	1461.76
		noun-predicative adjective	273.68	485.82	637.56	908.32	1492.88
		relativized noun-adjective	284.20	423.14	562.08	873.12	1477.44
Working memory			12.00	16.25	20.00	23.00	27.00

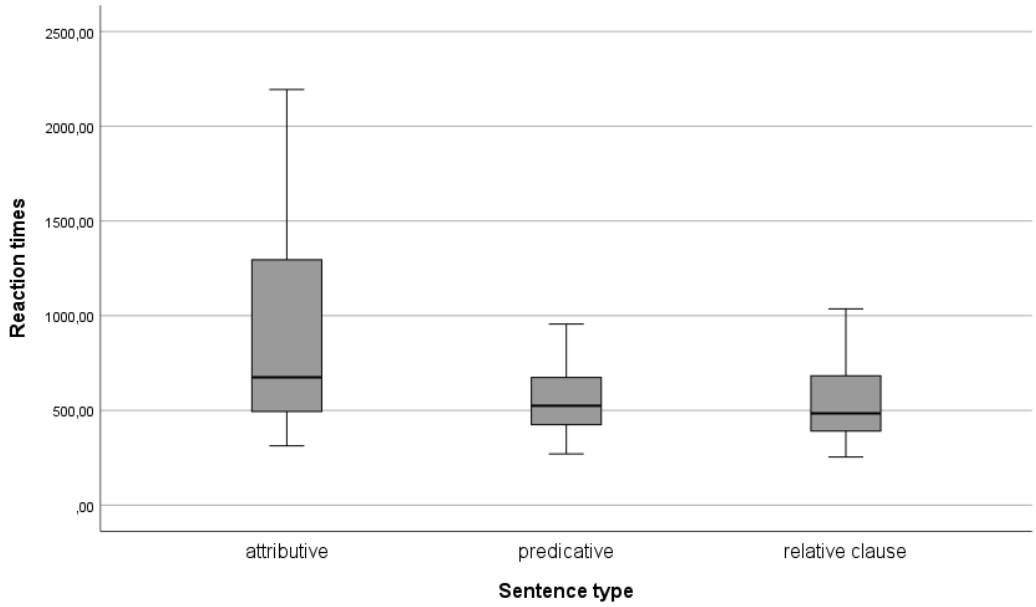

Figure 1. Mean reaction times (in milliseconds) on preceding segment per sentence type

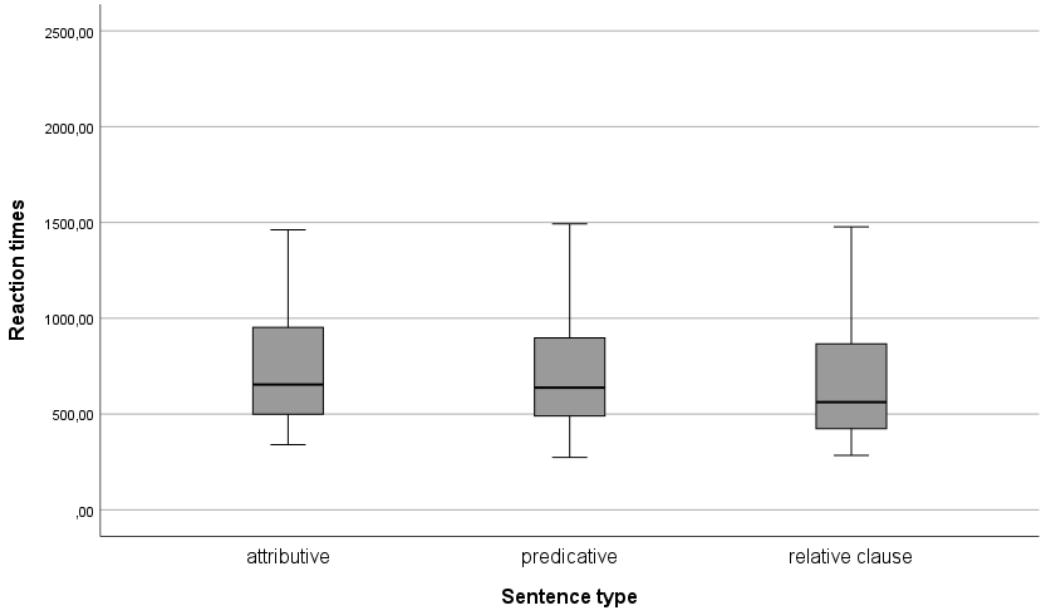

Figure 2. Mean reaction times (in milliseconds) on adjective per sentence type

The ANCOVA revealed a main effect for Sentence Type ($F(2,104) = 4.85$; $p = .01$; $\eta^2 = .09$). Bonferroni post-hoc analyses showed that reaction times in noun – attributive adjective constructions (e.g. *une jupe violette* 'a purple.F.SG skirt.F.SG') were longer than in noun – predicative adjective (e.g. *la beauté est importante* 'beauty.F.SG is important.F.SG') ($t(104) = 2.41$; $p = .05$; $r = .23$) and relativized noun – adjective constructions (e.g. *la popularité qui est importante* 'the popularity.F.SG which is important.F.SG') ($t(104) = 2.91$; $p = .01$; $r = .27$). However, reaction times in noun – predicative adjective constructions vs. relativized noun – adjective constructions did not differ ($t(104) = .51$; $p = 1.00$; $r = .05$). No main effect was found for Segment of Measurement ($F(1,104) = .001$; $p = .98$; $\eta^2 = .00$), indicating that reaction times did not differ between the adjective segment and the segment preceding the adjective. Furthermore, working memory capacity was not found to modulate the reaction times ($F(1,104) = .101$; $p = .75$; $\eta^2 = .001$). Interaction between Segment of Measurement and Sentence Type turned out to be significant ($F(2,104) = 25.32$; $p < .001$; $\eta^2 = .33$). This means that differences in reaction times between sentence types differed between segments of measurement. More precisely, contrast analyses showed that reaction times on the adjective segment significantly differed between noun – predicative adjective and relativized noun – adjective constructions ($t(104) = 3.89$; $p < .001$; $r = .36$), and noun – attributive adjective and relativized noun – adjective constructions ($t(104) = 4.53$; $p < .001$; $r = .41$). The reaction times measured on the adjective in predicative and attributive constructions were longer than those measured on the adjective in relative constructions. In contrast, no significant difference was observed in reaction times on the adjective segment between noun – attributive adjective and noun – predicative adjective constructions ($t(104) = .84$; $p = .39$; $r = .08$). Regarding the segment preceding the

adjective, we only focused on noun – predicative adjective and relativized noun – adjective constructions (as mentioned in section 5.2). A contrast analysis showed that the reaction times on the segment preceding the adjective did not significantly differ between these constructions ($T = 233.00$; $p = .07$; $r = .21$).

7. Discussion

As mentioned in Section 5.2, we focused on the segment preceding the adjective (*est* 'is') to ensure that differences in reaction times on the adjective segment are exclusively related to the processing of gender agreement. Since the significant main effect of Sentence Type is based on an overall effect (i.e. including both the segment preceding the adjective and the adjective segment) of types of agreement constructions on reaction times, this effect is not exclusively related to the processing of gender agreement. To focus on the processing of gender agreement we need to look at the Segment of Measurement x Sentence Type interaction and focus on the contrast analyses run on the adjective segment. This interaction revealed a medium-sized effect, indicating that this effect was robust (Cohen 1988). These results showed that reaction times in processing (noun – adjective) gender agreement in L2 French were shorter in relativized noun – adjective constructions as compared to noun – predicative adjective and noun – attributive adjective constructions, which confirms our first hypothesis. The effect sizes for both contrast analyses revealed to be medium. However, no difference in reaction times was observed between noun – predicative adjective and noun – attributive adjective gender agreement constructions. The effect size for this contrast analysis was small, which indicates that this difference was not robust. This does not confirm our second hypothesis predicting shorter reaction times in noun – predicative adjective constructions as compared to noun – attributive adjective constructions. Instead, this finding confirms our third hypothesis predicting equal duration of reaction times in both constructions. Regarding working memory, no effect of the learners' working memory capacity was found on the processing of (noun – adjective) gender agreement. Although this observation revealed a small-sized effect, this means that the fourth hypothesis predicting that working memory modulates the L2 processing of gender agreement during reading comprehension in noun – attributive and noun – predicative adjective agreement constructions cannot be confirmed by the results of this study.

We also analyzed the segment preceding the adjective in noun – predicative adjective and relativized noun – adjective constructions as a control segment (i.e. *est* 'is'). Whereas the reaction times measured on the adjective segment differed between both types of constructions, those measured on the segment preceding the adjective did not significantly. Although the effect size for this contrast analysis was small, this indicates that differences in reaction times measured on the adjective segment are related to noun – adjective gender agreement (since we controlled for the adjective's frequency, position, length in terms of syllables and the phonological expression of gender inflection across experimental conditions). As defined in section 2, relativized noun – adjective gender agreement constructions can be considered as more complex than noun – predicative adjective and noun – attributive adjective gender agreement constructions in French.

Whereas the measure of syntactic complexity based on the number of clauses per sentence predicted the same level of syntactic complexity in noun – predicative adjective and noun – attributive adjective gender agreement constructions in French, the measures of syntactic complexity based on the number of words per constituent or the number of (syntactic) phrases per clause predicted a higher level of syntactic complexity for noun – predicative adjective gender agreement constructions as compared to noun – attributive adjective ones in French. Based on the results found in our experiment, the number of clauses per sentence may be taken as an appropriate quantitative operationalization of syntactic complexity in relation to the processing of noun – adjective agreement constructions in L2 French. This means that syntactic complexity affects the L2 processing of (noun – adjective) gender agreement in the sense that the processing of gender agreement in non-embedded constructions triggers longer reaction times, indicating more sensitivity to gender agreement, than in embedded constructions. For L2 French we, thus, argue that embeddedness may account for the effect of syntactic complexity on the processing of (noun – adjective) gender agreement and that the level of syntactic complexity based on the number of clauses in gender agreement constructions is negatively correlated to the learners' sensitivity to noun - adjective gender agreement. It is important to mention that these results may not be found in low proficient learners of French, as this population has been shown to be less sensitive to gender agreement in L2 than intermediate learners (Gabriele et al. 2013, Alemán Bañón et al. 2018). As our results show longer reaction times on the adjective in noun – attributive adjective and noun – predicative adjective gender agreement constructions in French (as compared to relativized noun – adjective constructions), the L2 population under investigation may be assumed to be more sensitive to noun – adjective gender agreement in these constructions than to noun – adjective gender agreement in relativized noun – adjective constructions. This contrasts with previous studies on the L2 learners' sensitivity to noun – adjective gender agreement (e.g. de Jong 2005, Foucart & Frenck-Mestre 2012, Gabriele et al. 2013; Alemán Bañón et al. 2018) which showed that L2 learners are more sensitive to gender agreement on adjectives in an attributive position than in a predicative position. Although there was a small-sized effect, our experiment showed an equal level of sensitivity to gender agreement in these constructions in L2 French, but a reduced level of sensitivity to gender agreement in embedded constructions. This contrast may be explained by the research method used in Foucart and Frenck-Mestre (2012), Gabriele et al. (2013) and Alemán Bañón et al. (2018). Whereas these studies used the ERP technique to measure the L2 processing of noun – adjective gender agreement, we used the self-paced reading technique. Although both methods have been shown to be appropriate to measure the L2 processing of gender agreement, ERP is principally used to provide insights into the ongoing neurocognitive processes involved in processing L2 gender agreement and reflects immediate neural responses. Self-paced reading, however, is used to measure the L2 learner's processing efficiency during reading comprehension and provides the processing costs involved in processing L2 gender agreement. As such, the cognitive processes indexed by these two online measures are different and tap into different types of processing (e.g. Ditman et al. 2007). In self-paced reading experiments the processing of gender agreement spills over to the next segment in the sentence due to the participant's fixed button-press pace (Mitchell 2004). Participants may therefore press

the button before gender agreement has completely been processed, which affects the reaction times measured. As ERP reflects immediate responses and does not detect spill-over, ERP measures of gender agreement processing may differ from self-paced reading measures (cf. Bicknell et al. 2010 in which different results from self-paced reading vs. ERP were reported for noun-verb constructions), resulting in the contrast between the results found in Foucart & Frenck-Mestre (2012), Gabriele et al. (2013) and Alemán Bañón et al. (2018), and our results when it comes to the L2 processing of gender agreement on adjectives in attributive vs. predicative positions. Further research needs to focus on these potential differences between ERP and self-paced reading in measuring the L2 processing of (noun – adjective) gender agreement. In addition to these ERP studies, our results also contrast with De Jong (2005) in which longer reaction times on adjectives were found in attributive constructions as compared to predicative ones. Since De Jong (2005) focused on the processing of (noun – adjective) gender agreement in L2 Spanish, the perceptual salience of gender inflection in Spanish may have affected the learners' sensitivity to this type of agreement. Renaud (2014) for instance, investigated the processing of gender agreement in intermediate to advanced L2 learners of Spanish and French by means of a self-paced reading experiment and an acceptability task. The experiment focused on anaphoric adjective agreement in superlative constructions such as *el más sucio* 'the.M.SG most dirty.M.SG' vs. *la más sucia* 'the.F.SG most dirty.F.SG' for Spanish and *le moins lourd* 'the.M.SG least heavy.M.SG' vs. *la moins lourde* 'the.F.SG least heavy.F.SG' for French. Regarding the comparison between L2 learners of Spanish and French, the results showed more robust grammatical knowledge of gender agreement in the L2 Spanish group than in the L2 French group. The author concluded that this contrast may be explained by differences in the perceptual salience of gender agreement on adjectives, since Spanish exhibits more salient gender inflection on adjectives than French. This may also be an explanation for the contrast found between De Jong (2005) and the results found in the present study. A question for further research could be to investigate to what extent perceptual salience mediates the effect of syntactic complexity on the processing of (noun – adjective) gender agreement in L2 French.

As mentioned in the introduction, a central question in L2 studies is to define and operationalize the construct of complexity. Based on our results from L2 French, we suggest that, to some extent, syntactic complexity being a subconstruct of inherent complexity may be related to the cognitive complexity of (noun – adjective) gender agreement constructions in L2 French. More specifically, the cognitive complexity of this type of agreement construction by L2 learners of French may be related to (non-) embeddedness in (noun – adjective) gender agreement constructions. This is in line with the fact that the number of (embedded) clauses is the quantitative operationalization of syntactic complexity that is most used in L2 research (Pallotti 2015) as compared to other operationalizations of syntactic complexity in L2 research such as the number of words per constituent or the number of syntactic phrases per clause. As no difference in reaction times was observed between noun – predicative adjective and noun – attributive adjective gender agreement constructions, syntactic complexity was not found to fully mirror cognitive complexity in our experiment, which may indicate that syntactic complexity can be related to a slightly different interpretation of complexity in L2 acquisition, as compared to cognitive complexity.

Regarding working memory capacity, the results showed no effect of the learners' working memory capacity on the processing of gender agreement in noun – attributive adjective and noun – predicative adjective gender agreement constructions. In contrast to previous studies (e.g. Havik et al. 2009, Sagarra & Herschensohn 2010, Dracos & Henry 2021) in which working memory was demonstrated to affect L2 morphosyntactic processing, our findings are in line with studies (e.g. Caplan & Waters 1999, Foote 2011, Baek 2012) in which working memory was shown to not affect structural processing in L2. As described in section 3, these studies showed that the L2 learners' working memory capacity only affects sentence processing when meaning is necessary to interpret these sentences such as animacy processing. Although our results did not show an effect of working memory on the L2 processing of gender agreement, these results need to be interpreted carefully as only one measure of working memory capacity was included in the present study. Within the context of language-related cognitive abilities, research on working memory often distinguishes between non-verbal and verbal working memory (e.g. White 2021). Whereas non-verbal working memory involves the temporal storage of digits while processing a non-verbal task such as arithmetic, verbal working memory involves the temporal storage of verbal units such as words or letters. In research on language interpretation an ongoing debate concerns the question of whether working memory needs to be considered as a cognitive system separate from other cognitive processes involved in language processing (e.g. Just & Carpenter 1992) or as a cognitive system equating to experience in language processing (e.g. Macdonald & Christiansen 2002). Based on the results of the present study, working memory considered as a separate cognitive system such as non-verbal working memory, has been found to not affect the processing of (noun – adjective) gender agreement in L2 French. As a non-verbal working memory task was used in this study, adding a verbal working memory task such as a reading span task, to future studies may complement our understanding of the role of working memory in the L2 processing of gender agreement.

A limitation of the present study is that few fillers have been used in the self-paced reading task. As such, participants may have known the purpose of the task. Although no research has been done on the potential effect of experimental item – filler ratio on reaction times in self-paced reading, Marsden et al. (2018) recommend a 1:1 experimental item – filler ratio in this particular task. Another limitation is that there were no stimuli containing noun – adjective gender agreement violations included in the experiment. In addition to the inclusion of the segment preceding the adjective and control for the adjective's frequency, the number of syllables, the phonological expression of gender inflection and the adjective's position in the sentence, the inclusion of this type of stimuli would have been an additional control condition to ensure that differences in reaction times are exclusively related to gender agreement processing.

8. Conclusions

In this study we investigated the role of syntactic complexity in the processing of (noun – adjective) gender agreement in L2 French, while controlling for the learners' working memory capacity. By means of a self-paced reading technique we tested

intermediate Dutch learners of French with respect to their performances at processing gender agreement in sentences exhibiting different levels of syntactic complexity. The results showed decreased sensitivity to gender agreement in embedded structures, but increased sensitivity to gender agreement in non-embedded structures. Based on these results, we may conclude that the quantitative measure of syntactic complexity based on the number of clauses in gender agreement constructions is an appropriate operationalization to account for effects of syntactic complexity on the L2 processing of (noun – adjective) gender agreement and that this measure is negatively correlated to sensitivity to gender agreement. Furthermore, we may conclude that (non-verbal) working memory does not affect the processing of gender agreement in L2 French.

Abbreviations

F	Feminine
M	Masculine
SG	Singular

Ethics and consent

This study has received approval from the Ethical Assessment Committee Linguistics – reference number: 'BRIL'0002-07-2019. Participants of this study gave written informed consent.

References

Alarcón, I. 2021. Adjectival and verbal agreement in the oral production of early and late bilinguals: Fluency, complexity, and integrated knowledge. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics* 34 (2): 371-401.

Alemán Bañón, J., Fiorentino, R., & Gabriele, A. 2018. Using event-related potentials to track morphosyntactic development in second language learners: The processing of number and gender agreement in Spanish. *PloS One* 13 (7): e0200791.

Baddeley, A. 2003. Working memory and language: An overview. *Journal of Communication Disorders* 36 (3): 189-208.

Baek, S. 2012. Processing of English Relative Clauses by Adult L2 Learners. PhD dissertation, University of Illinois.

Bartning, I. 2000. Gender agreement in L2 French: pre-advanced vs. advanced learners. *Studia Linguistica* 54 (2): 225-237.

Bicknell, K., Elman, J. L., Hare, M., McRae, K. & Kutas, M. 2010. Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language* 63 (4): 489-505.

Bonilla, C. L. 2015. From number agreement to the subjunctive: Evidence for Processability Theory in L2 Spanish. *Second Language Research* 31 (1): 53-74.

Bril, M. 2016. Syntactic complexity and inflections in the written production of L1 and L2 French. *Bucharest Working Papers in Linguistics* XVIII (2): 99-114.

Bril, M. 2021. Gender marking in written L1 and L2 French: Syntactic complexity and phonological expression. In L. Avram, A. Sevcenco, & V. Tomescu (eds.), *L1 Acquisition and L2 Learning: The View from Romance,* 289-314. Amsterdam/Philadelphia: John Benjamins.

Bril, M., Gerrits, A.C.A. & Visser, A.M. 2021. The effects of linguistic and cognitive factors on the L2 processing of oral input: a self-paced listening experiment. *International Journal of Listening* 36 (3): 268-282: 1-15.

Brunet, E. 2014. Liste de mots classés par fréquence décroissante. http://eduscol.education.fr/cid47916/liste-des-mots-classee-parfrequence-decroissante.html

Bulté, B. & Housen, A. 2012. Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (eds.), *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*, 23-46. Amsterdam/Philadelphia: John Benjamins.

Bulté, B. & Housen, A. 2018. Syntactic complexity in L2 writing: Individual pathways and emerging group trends. *International Journal of Applied Linguistics* 28 (1): 147-164.

Caplan, D. & Waters, G. S. 1999. Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences* 22 (1): 77-94.

Carrasco-Ortiz, H. & Frenck-Mestre, C. 2014. Phonological and orthographic cues enhance the processing of inflectional morphology. ERP evidence from L1 and L2 French. *Frontiers in Psychology* 5: 888.

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

Cohen, J. 1988. *Statistical Power Analysis for the Behavioural Sciences*. Mahwah, NJ: Erlbaum.

Daneman, M. & Carpenter, P. A. 1980. Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior* 19 (4): 450-466.

Ditman, T., Holcomb, P. J. & Kuperberg, G. R. 2007. An investigation of concurrent ERP and self-paced reading methodologies. *Psychophysiology* 44 (6): 927-935.

Donaldson, T. S. 1968. Robustness of the F-test to errors of both kinds and the correlation between the numerator and denominator of the F-ratio. *Journal of the American Statistical Association* 63 (322): 660-676.

Dracos, M. & Henry, N. 2021. The role of task-essential training and working memory in offline and online morphological processing. *Languages* 6 (1): 24.

Ellis, R. 2009. The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics* 30 (4): 474-509.

Fernandez-Duque, D. 2009. Cognitive and neural underpinnings of syntactic complexity. In M. Shibatani & T. Givón (eds.), *Syntactic Complexity: Diachrony, Acquisition, Neuro-cognition, Evolution,* 433-460. Amsterdam/Philadelphia: John Benjamins.

Ferreira, F., Henderson, J. M., Anes, M. D., Weeks, P. A. & McFarlane, D. K. 1996. Effects of lexical frequency and syntactic complexity in spoken-language comprehension: Evidence from the auditory moving-window technique. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22 (2): 324-335.

Field, A. 2013. *Discovering Statistics Using IBM SPSS Statistics*. London: Sage.

Foote, R. 2011. Integrated knowledge of agreement in early and late English–Spanish bilinguals. *Applied Psycholinguistics* 32 (1): 187-220.

Foucart, A. 2008. Syntactic Gender Processing in French as a First and a Second Language. PhD dissertation. Université Aix-Marseille and University of Edinburgh.

Foucart, A. & Frenck-Mestre, C. 2012. Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language* 66 (1): 226-248.

Friederici, A. D., Pfeifer, E. & Hahne, A. 1993. Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research* 1 (3): 183-192.

Gabriele, A., Fiorentino, R. & Bañón, J. A. 2013. Examining second language development using event-related potentials: A cross-sectional study on the processing of gender and number agreement. *Linguistic Approaches to Bilingualism* 3 (2): 213-232.

Gabriele, A., Alemán Bañón, J., Hoffman, L., Covey, L., Rossomondo, A. & Fiorentino, R. 2021. Examining variability in the processing of agreement in novice learners: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 47 (7): 1106-1140.

Gilardone, G., Viganò, M., Costantini, G., Monti, A., Corbo, M., Cecchetto, C. & Papagno, C. (2023). The role of verbal short-term memory in complex sentence comprehension: An observational study on aphasia. *International Journal of Language & Communication Disorders*.

Grosjean, F. 2008. *Studying Bilinguals*. Oxford: Oxford University Press.

Havik, E., Roberts, L., Van Hout, R., Schreuder, R. & Haverkort, M. 2009. Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. *Language Learning* 59 (1): 73-112.

Hopp, H. 2006. Syntactic features and reanalysis in near-native processing. *Second Language Research* 22: 369-397.
Hopp, H. 2007. Ultimate Attainment at the Interfaces in Second Language Acquisition: Grammar and Processing. PhD dissertation. University of Groningen.
Hopp, H. 2016. Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research* 32 (2): 277-307.
Jackson, C. 2008. Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning* 58 (4): 875-909.
Jegerski, J. 2014. Self-paced reading. In J. Jegerski & B. VanPatten (eds.), *Research Methods in Second Language Psycholinguistics*, 20-49. New York: Routledge.
De Jong, N. 2005. Can second language grammar be learned through listening?: An experimental study. *Studies in Second Language Acquisition* 27 (2): 205-234.
Juffs, A. & Harrington, M. 2011. Aspects of working memory in L2 learning. *Language Teaching* 44 (2): 137-166.
Just, M. A., Carpenter, P. A., & Woolley, J. D. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111: 228-238.
Just, M. A. & Carpenter, P. A. 1992. A capacity theory of comprehension. *Psychological Review* 99 (1): 122-149.
Kim, S. H. & Kim, J. H. 2012. Frequency effects in L2 multiword unit processing: Evidence from self-paced reading. *Tesol Quarterly* 46 (4): 831-841.
Kormos, J. & Sáfár, A. 2008. Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition* 11 (2): 261-271.
Luce, R. 1986. *Response times*. Oxford: Oxford University Press.
MacDonald, M. C. & Christiansen, M. H. 2002. Reassessing working memory: A comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review* 109 (1): 35–54.
Marinis, T., Roberts, L., Felser, C. & Clahsen, H. 2005. Gaps in second language sentence processing. *Studies in Second Language Acquisition* 27 (1): 53-78.
Marsden, E., Thompson, S., & Plonsky, L. 2018. A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics* 39 (5): 861-904.
McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., & Osterhout, L. 2010. Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning* 60: 123-150.
Mitchell, D. C. 2004. Online methods in language processing: Introduction and historical review. In M. Carreiras & C. Clifton Jr. (eds.), *The Online Study of Sentence Comprehension: Eyetracking, ERPs and Beyond*, 15-32. New York: Psychology Press.
Öttl, B., Jäger, G. & Kaup, B. (2015). Does formal complexity reflect cognitive complexity? Investigating aspects of the Chomsky hierarchy in an artificial language learning study. *PloS One*, 10 (4): e0123059.
Pallotti, G. 2009. CAF: Defining, refining and differentiating constructs. *Applied Linguistics* 30 (4): 590-601.
Pallotti, G. 2015. A simple view of linguistic complexity. *Second Language Research* 31 (1): 117-134.
Pan, H. Y. & Felser, C. 2011. Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua* 121 (2): 221-236.
Pliatsikas, C. & Marinis, T. 2013. Processing of regular and irregular past tense morphology in highly proficient second language learners of English: A self-paced reading study. *Applied Psycholinguistics* 34 (5): 943-970.
Reichle, R. V., Tremblay, A. & Coughlin, C. 2016. Working memory capacity in L2 processing. *Probus* 28 (1): 29-55.
Renaud, C. 2014. A processing investigation of the accessibility of the uninterpretable gender feature in L2 French and L2 Spanish adjective agreement. *Linguistic Approaches to Bilingualism* 4 (2): 222-255.
Reynolds, M. R., Niileksela, C. R., Gignac, G. E. & Sevillano, C. N. 2022. Working memory capacity development through childhood. *Developmental Psychology* 58 (7): 1254-1263.
Rodríguez, G. A. 2008. Second Language Sentence Processing: Is it Fundamentally Different? PhD dissertation. University of Pittsburgh.
Rönnberg J., Rudner M. & Foo C. 2009. The cognitive neuroscience of signed language: Applications to a working memory system for sign and speech. In L. Bäckman & L. Nyberg (eds.), *Memory, Aging and the Brain: A Festschrift in Honour of Lars-Göran Nilsson,* 275-296. New York: Psychology Press.
Sagarra, N. 2007. Online processing of gender agreement in low proficient English-Spanish late bilinguals. In M. J. Cabrera, J. Camacho, V. Déprez, N. Flores & L. Sánchez (eds.), *Current Issues in Linguistic Theory Series,* 240-253. Amsterdam/Philadelphia: John Benjamins.

Sagarra, N. 2021. When more is better Higher L1/L2 similarity, L2 proficiency, and working memory facilitate L2 morphosyntactic processing. In M. J. Leeser, G. D. Keating & W. Wong (eds.), *Research on Second Language Processing and Processing Instruction*, 125-150. Amsterdam/Philadelphia: John Benjamins.

Sagarra, N. & Herschensohn, J. 2010. The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua* 120 (8): 2022-2039.

Schneider W. J. & McGrew, K. S. 2018. The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (eds.), *Contemporary Intellectual Assessment: Theory, Tests, and Issues*, 73-163. New York: Guilford Press.

Traxler, M. J., Corina, D. P., Morford, J. P., Hafer, S., Hoversten, L. J. & NSF Science of Learning Center for Visual Language & Visual Learning (VL2). 2014. Deaf readers' response to syntactic complexity: Evidence from self-paced reading. *Memory & Cognition* 42: 97-111.

Wechsler, D. 2003. *Wechsler Intelligence Scale for Children: Technical and Interpretive Manual*, 4th edition. San Antonio, TX: Psychological Corporation.

White, M. J. 2021. Phonological working memory and non-verbal complex working memory as predictors of future English outcomes in young ELLs. *International Journal of Bilingualism* 25 (1): 318-337.

Wildt, A. R. & Ahtola, O. 1978. *Analysis of covariance*. London: Sage.