

University of Groningen

## Computer programming skills: A cognitive perspective

Graafsma, Irene

DOI:  
[10.33612/diss.168003240](https://doi.org/10.33612/diss.168003240)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2021

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Graafsma, I. (2021). *Computer programming skills: A cognitive perspective*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.168003240>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

## Chapter 5

Processing of violations in human and computer languages:

An EEG study

## 5.1 INTRODUCTION

The role that natural human language plays when learning to computer program is still to be determined. Some researchers have argued that programming is, at least partly, a language skill (e.g., Fedorenko et al., 2019; O'Regan, 2012; Paulson, 2007). It has also been suggested that, from a theoretical point of view, there are many similarities between programming and natural language writing. Specifically, it has been argued that both skills rely on specific structure and style rules (Hermans & Aldewereld, 2017) and that these rules might be similar to some extent, especially because modern programming languages were intentionally designed to resemble natural languages (Fedorenko et al., 2019; Paulson, 2007). This natural language-like design of modern programming languages may facilitate transfer between natural language and programming skills, leading to the possibility that natural language skills may be a predictor of programming success. However, other researchers have argued that programming is essentially a problem-solving skill that should be classed as a Science Technology Engineering and Mathematics (STEM) subject (Fedorenko et al., 2019). Understanding how both STEM and language skills relate to programming is important for optimising programming education (Fedorenko et al., 2019).

Experimentally, only a handful of studies have looked into the role of language when learning to program. In an earlier study, we found that vocabulary learning skills predicted programming performance, but grammar learning skills did not (see Chapter 3). A second study on this topic, by Prat et al. (2020), showed that scores on language aptitude tests explained some of the variance in learned programming skill. In addition, three neuro-imaging studies, using functional Magnetic Resonance Imaging (fMRI), have investigated the relationship between programming languages and natural languages. The first, by Siegmund et al. (2014), showed that reading computer code for comprehension activated similar brain areas to silent reading of a natural language. In contrast, Floyd et al. (2017) showed differences in activation patterns between reading of natural and programming languages. However, this difference was only found in beginning programmers. Thus, it is possible that a programming language is processed in a similar way as natural language once a programmer has become sufficiently proficient. In a recent fMRI study Ivanova et al. (2020) found that reading a programming language depended much less on language systems in the brain than reading sentence problems. They argued that reading and

understanding programming languages was more reliant on the multiple demand system, which is the brain network most associated with mathematics. However, their results suggested that although the language system was not activated in the same way for programming languages as for natural languages, parts of the language system were still involved when reading code. They tested two programming languages: Python, which is text-based, and Scratch Junior, which is a more graphical programming language, and therefore is structured less like a natural language. Although they did not directly compare these programming languages, their results suggest that the language system was more involved for the text-based programming language Python, compared to the graphical language Scratch Junior. Therefore, there still seems to be a role for language in programming, and this role appears to be related to the exact nature of the programming language.

In sum, the previous studies suggest that natural language processing is involved when reading programming languages. What remains unclear is which aspects of language relate to programming skills. One can argue that the structural rules in a programming language, referred to as 'programming syntax', have a similar role to grammar in a natural language, as both prescribe the rules that a writer has to follow to create correct and interpretable text or code (Hermans & Aldewereld, 2017). However, in the study reported in Chapter 3, we saw that programming skills were not predicted by artificial grammar learning. This raises the question to what extent learning the syntax of a programming language resembles learning grammar in a natural language. To answer this question, further research into the linguistic properties of programming languages is necessary. Specifically, it is relevant to investigate whether there are fundamental differences between processing programming syntax and natural language grammar in the brain.

### **5.1.1 Event-Related Potentials (ERP) and language processing**

One way to study language processing in natural languages is by analysing brain responses to different kinds of violations (e.g., semantic or morphosyntactic) in written or spoken language (Osterhout & Holcomb, 1995). For natural languages, a considerable amount of research has been performed using such violations with Event-Related Potentials (ERPs; Carreiras & Clifton, 2004). In these studies, the electrical brain responses evoked by these

violations are recorded with electroencephalography (EEG) and compared to those evoked by stimuli without violations (Friederici et al., 2002). Previous studies have shown that different types of violations elicit different types of responses (e.g., Carreiras & Clifton, 2004; Friederici et al., 2002).

For grammatical violations in natural language, most studies have shown two types of brain responses: a left-anterior negativity (LAN), which occurs approximately between 300 and 500ms (Carreiras & Clifton, 2004), and a positive deflection in the signal at approximately 500ms post-onset of the stimulus, which is called the P600 (Carreiras & Clifton, 2004; Friederici et al., 2002; Hagoort et al., 1993; Osterhout & Holcomb, 1992). The LAN effect (when present) is believed to signal recognition of a violation, while the P600 response is usually considered to reflect repair and reanalysis (Friederici et al., 2002). The P600 indicates that when this type of violation occurs, the reader has to exert additional mental effort to repair and/or reanalyse the sentence. Molinaro et al. (2008) suggest that in the first stage of the P600, around 500 to 700ms, the incongruence of the sentence is registered. In the second stage, from 700 to 900ms, the reanalysis of the stimulus occurs. Studies have shown clear P600 effects for, for example, subject-verb violations, where the form of the verb does not match the number or person of the subject in the sentence, for example: \**“He are a good swimmer.”* (e.g., Gouvea et al., 2010; Kaan, 2002; Nevins et al., 2007).

Previous studies have also shown that violations that are not grammatical in nature elicit different responses. Specifically, semantic violations (where the grammar of the sentence is correct, but a word is presented that does not fit in the conceptual context of the sentence; e.g., *“The cloud was buried.”*) have been shown to elicit a negative deflection at approximately 400ms, called the N400 (Friederici, 1995; 2002; Holcomb, 1993; Kutas & Hillyard, 1980). The N400 has also been elicited in non-language contexts, for example, for errors in mathematics (e.g., Jost et al., 2004; Szucs & Csépe, 2004; Wang et al., 2000).

Several studies have shown earlier positive deflections at around 300ms. The interpretation of this effect compared to the P600 effect is still contested. While some studies interpret this effect as an early P600, arguing that it is still a language specific effect (Vissers et al., 2006), others argue that this should be considered as a separate component

(P300; Osterhout, 1999; Osterhout et al., 1996). The P300 is found in response to a wide variety of unexpected, task relevant stimuli, both within and outside language contexts, and is typically seen in the frontocentral and centroparietal regions (Donchin, 1981; Osterhout et al., 1996). The amplitude and the duration of the P300 effect are thought to reflect the amount of attentional resources allocated to a stimulus, with stimuli that demand more attentional resources eliciting longer P300s with larger amplitudes (Suárez-Pellicioni et al., 2013). Coulson et al. (1998) and Sassenhagen and Fiebach (2019) argue that the P600 is part of the same family as the P300. They argue that both the P300 and the P600 reflect a domain-general brain response to salience and are not language specific, supported by Sassenhagen et al.'s (2014) finding that the P600, like the P300, is reaction-time aligned. The P300 has also been found in language paradigms, for example by Osterhout et al. (1996), who found a positive deflection at around 300ms, which they interpreted as a P300 effect, when they presented a word in uppercase while the rest of the sentence was in lowercase. Vissers et al. (2006) found similar effects for orthographic violations (spelling errors) in expected words, which could also have been interpreted as a P300, but they interpreted the effect as an early, language-related P600. Overall, regardless of the name assigned to this early positive deflection, the literature suggests that the early positive response (either a P300 or an early P600) in a language context is related to unexpected formats or spellings, rather than grammatical or semantic violations.

### **5.1.2 The current study**

In the current study, we aimed to test whether syntax errors in a programming language are processed similarly to grammatical violations in a natural language. If they are, then they would be predicted to elicit a P600 effect. We used the programming language Java, as this is a text-based programming language that is widely used and has strict syntax rules, and, thus, allows for unambiguous violations (Gosling et al., 2000). We chose to present a typical violation that is programming specific, and grammatical in nature: a violation in bracket-use in “if” and “while” statements. For example, presenting “while {x==1}” which uses the incorrect type of brackets, while the correct structure would be “while (x==1)”.

“If” and “while” statements are very common structures that occur in almost all programming languages. Both structures start a conditional instruction for the execution of

the code that follows. For example, “if (x==1)” means “if variable x is currently equal to 1, execute the following code”. Hence, the code that follows is only executed if variable x is equal to 1. Similarly, “while (x<10)” instructs the computer to execute the following code repeatedly until variable x is no longer smaller than 10. In this structure, variable x will typically change with every execution of the following code, causing it to eventually reach or exceed 10; then the computer will stop running that section of the code.

Although “if” and “while” statements occur in almost every programming language, the exact syntax rules to present such a statement vary across languages. For example, in Java and C++ the correct formulation is “if (x==1)”, while in Python the brackets can be omitted: “if x==1”. The absence or presence of the brackets, therefore, does not change the meaning of the statement. Instead, they are a part of a syntax rule that is specific to each programming language, in which they indicate where the conditions of the “if” or “while” statements start. Although the use of brackets is somewhat flexible across programming languages, within Java the use of round brackets indicates a strict syntax rule. If the use of brackets is incorrect, the code will not run. Consequently, we propose that a bracket in Java has a function akin to functional elements (e.g., auxiliaries or inflectional morphology) in natural language. Furthermore, it is used in a certain form in conjunction with the “if” or “while” statement to create correct syntax. Therefore, we suggest that this kind of violation is of a grammatical nature and hence might be predicted to elicit a similar P600 response to grammatical violations in natural languages.

In order to closely match the type of violation across the languages, we used subject-verb disagreement as the grammatical violation in Dutch (e.g., *\*Ik is een geweldige leraar.*: “I is a great teacher.”). This type of violation, like the Java bracket violations, is also detected upon presentation of the second word, and involves a similar type of agreement between a function word (the subject in Dutch, and the “while” or “if” functions in Java) and an incorrect inflection of the following word (the verb in Dutch and the bracket in Java).

When considering the language aspects of programming languages, it is important to take into account that a programming language is never a native language. Therefore, Pandža (2016) suggests that, to further understand the role of language in programming, more research should take a second language learning perspective. Consequently, we

added a third condition: subject-verb disagreements in a second language, in this case, English, which our participant population acquired after the critical age of language acquisition (approximately 10 years old). Second language ERP studies have shown that ERP responses to syntactic violations in a second language vary according to proficiency and age of acquisition (Kotz, 2009). Late learners with low proficiency may even show a complete absence of both the LAN and the P600 effect, while more proficient learners may show no LAN effect but still a P600 effect (Weber-Fox & Neville, 1996). Both the LAN and P600 effects can also be delayed and decreased in amplitude, with typically shorter latencies and higher amplitudes for more proficient L2 speakers (Rossi et al., 2006). If the violation that is tested in the second language also exists in the native language, ERP responses for second language speakers resemble those of native speakers more closely in latency (Tokowicz & MacWhinney, 2005). In general, ERP components for second language speakers tend to have a similar scalp distribution as for native speakers, although some studies have found responses to be more bilateral for second language speakers compared to the more left-lateralised effects for native speakers (Van Hell & Tokowicz, 2010).

In sum, our research question was: Is a programming language processed similarly to a natural language? And, as noted above, to address this question we compared bracket violations in Java with subject-verb violations in Dutch (L1) and English (L2).

For Dutch, based on previous studies on subject-verb violations in natural languages, we predicted a P600 type response to the violations (Chen et al., 2007). On the basis of literature on second language processing and our participants' high proficiency in English, we also predicted a P600 of similar scalp distribution, but that this may be delayed (Kotz, 2009; Rossi et al., 2006; Weber-Fox & Neville, 1996).

This is the first study to study ERP responses to syntactic violations in a programming language. Hence, we had no previous research to guide predictions regarding the effect of violations in Java. However, based on the selection of our stimuli, we expected that the Java bracket violations are of a language-like grammatical nature, and thus, would elicit a P600, that is, an effect comparable in latency and distribution to that for Dutch and English. We predicted that this response would be delayed compared to Dutch (L1), and more similar to English (L2), as a programming language is always a non-native language. However, if the



effect for Java differed in its characteristics from those for Dutch and English, we could not rule out that we might observe a different effect. For example, it is possible that the P300, that can be elicited by a spelling error (Vissers et al. 2006) or a deviant capitalisation (Osterhout et al., 1996), could be evoked from the bracket errors. This P300 effect should then be characterised by an earlier latency, possibly with a shorter duration and a more frontal and bilateral distribution than we would expect from a P600. In order to establish the exact nature of the violations, we also examined the topography of the ERP effects. Specifically, we included anteriority (i.e., how frontal the effect is) and hemisphere as variables and studied the scalp distribution of responses over time.

## 5.2 METHODS

### 5.2.1 Ethics statement

The protocol for the current study received ethical approval from the Research Ethics Committee at Groningen University (Reference number: 67378826). We followed the approved protocol where all participants received an information sheet at the start of the experiment and gave their informed consent prior to participation. Participants received 15 Euros for their participation.

### 5.2.2 Participants

We recruited 12 male participants<sup>1</sup> (mean age 26.58, range 18-39) with at least 2 years of programming experience, and at least basic knowledge of programming in the Java programming language specifically. All participants spoke Dutch as their native language. They had all completed Dutch high school English education and thus had at least B1 level of English (Europees Referentiekader Talen, n.d.). The group consisted of six university students and six participants who worked as programmers professionally. All were right-handed as confirmed with the Edinburgh Handedness Questionnaire (Oldfield, 1971).

---

<sup>1</sup> The current sample size is smaller than intended because testing was cut short due to COVID-19.

### 5.2.3 Materials

#### Demographic questionnaire

A questionnaire was administered in which the participants were asked about general demographic details such as age and gender, as well as about their language and programming backgrounds to ensure that participants did not learn English before the age of 10, and thus could not be considered native English speakers, and to ensure that they all had at least basic knowledge of Java programming.

#### Stimuli

The experiment consisted of three language blocks (Dutch, English and Java), each containing 80 items. The order of language blocks was counterbalanced across participants. The items were equally divided between grammatical (40 stimuli) and ungrammatical (40 stimuli). In order to prevent participants from seeing one and the same item in both its grammatical and ungrammatical form, we created two lists using the Latin Square design. This means that each participant saw 20 grammatical and 20 ungrammatical sentences for each language. This adds up to 120 experimental items per list. Each participant only saw one list, meaning that half of the participants saw the grammatical version of an item, while the other half saw the ungrammatical version of that item. In addition, there were 120 filler items per list, bringing the total number of items to 240 per participant. The fillers were included to prevent participants from developing a strategy where they would stop reading the whole stimulus, and instead would only focus on one specific violation. Examples of target and filler stimuli are provided in Tables 5.1 and 5.2 respectively, and all stimuli can be found in Appendix B.

#### Programming language stimuli in Java

For the Java stimuli, the grammatical condition for each participant consisted of 20 “if” and “while” statements with the corresponding round brackets (e.g., *if (x>10)*), while the ungrammatical condition consisted of 20 “if” and “while” statements with incorrect curly brackets (e.g., *\*if {x>10}*) (see Table 5.1). In our pilot questionnaires, proficient Java programmers indicated that they found these types of Java bracket errors easy to

recognise. Each participant was presented with 20 grammatical and 20 ungrammatical fillers. We used two types of fillers. The first type were “else” statements with grammatical or ungrammatical bracket use. In this case, the grammatical brackets were the curly brackets (e.g., *else{x=5}*) while the ungrammatical strings contained round brackets (e.g., *\*else(x=5)*). This type of filler ensured that participants could not simply conclude that curly brackets were always ungrammatical. The second type of fillers were “if”, “while” and “else” statements with grammatical (e.g., *while(x==50); else{x=5}*) or ungrammatical (e.g., *\*while(x=50), else{x==5}*) expressions inside the brackets (see Table 5.2). These fillers prevented the participants from only paying attention to the brackets and made sure that they also read the content of the brackets.

### **First language stimuli in Dutch**

For the Dutch stimuli, we manipulated subject-verb agreement. The grammatical condition consisted of 20 sentences beginning with a subject pronoun, followed by a verb agreeing in person and number (e.g., *Hij is*: “He is”); the 20 sentences in the ungrammatical condition started with a subject pronoun with which the verb did not agree (e.g., *\*Hij ben*: “He am”) (see Table 5.1). These first two words of the sentences were critical for the experiment and aligned with the experimental stimuli for Java. After those first two words the experimental sentences were completed with no further violations. As fillers, we used 20 grammatical sentences in which the noun agreed with the preceding adjective and article in gender and number. The 20 ungrammatical sentences contained either gender or number disagreement between the adjective and the noun or number disagreement between the article and the noun (see Table 5.2). The violations for the fillers can be recognised in various positions in the sentences, thereby ensuring that participants would read the whole sentence, and not only focus on the second word, the position in which the violations in the experimental stimuli occurred.

### **Second language stimuli in English**

For English, we also manipulated subject-verb agreement. The grammatical condition consisted of 20 sentences beginning with a subject pronoun and verb agreeing in person and number (e.g., *He is*), while the 20 sentences in the ungrammatical condition started

with a subject pronoun and verb disagreeing in person or number (e.g., *\*He am*) (see Table 5.1). As fillers for the English condition we used 20 grammatical sentences and 20 ungrammatical sentences with incorrect affixation (pluralised uncountable nouns) or the indefinite article with uncountable (mass) nouns (see Table 5.2). Violations for the fillers can be identified in different positions in the sentences, thereby making sure that participants would read the whole sentence, and not only focus on the second word, in which position the violations in the experimental stimuli occurred.

**Table 5.1.** Examples of the experimental stimuli.

(1) Java						
Grammatical	While	(	x	>	10	)
Ungrammatical	While	*{	x	>	10	}
(2) Dutch						
Grammatical	Ik	ben	een	geweldige	docent.	
	I am a great teacher					
Ungrammatical	Ik	*is	een	geweldige	docent.	
	I is a great teacher.					
(3) English						
Grammatical	We	have	the	nicest	dog.	
Ungrammatical	We	*has	the	nicest	dog.	

*Note:* Examples of experimental stimuli for Java, Dutch and English. For the full list of stimuli see Appendix B. The words or symbols are spaced out to show how they were presented in fragments. \* indicates where the violation occurs.

**Table 5.2.** Examples of the fillers.

(1) Java						
Grammatical	Else	{	x	=	5	}
Ungrammatical	Else	*(	x	=	5	}
Grammatical	While	(	x	==	50	)
Ungrammatical	While	(	x	*=	50	)
(2) Dutch						
Grammatical	De	lange	touwen	hangen	daar.	
		The long ropes hang there.				
Ungrammatical	De	*lang	touwen	hangen	daar.	
		The <sub>ART,DEF</sub> long <sub>ADJ,INDEF</sub> ropes <sub>SN,PL</sub> hang there.				
(3) English						
Grammatical	They	perform	maintenance	on	roads.	
Ungrammatical	They	perform	*maintenances	on	roads.	

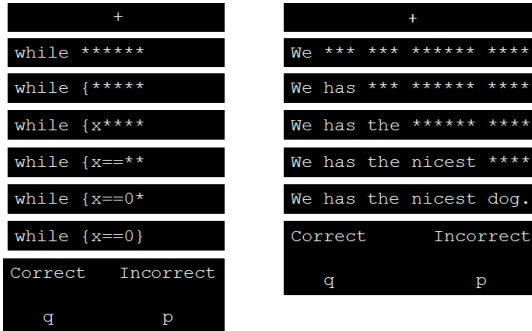
*Note:* Examples of the different types of fillers for Java, Dutch and English. For all three languages the position of the violation in the sentence/code varied between stimuli. For the full list of fillers see the Appendix B. The words or symbols are spaced out to show how they were presented. \* indicated where the violation first occurs.

**5.2.4 Procedure**

After participants gave informed consent and completed the handedness questionnaire, they were seated at a comfortable distance from the screen (approximately 80 cm) and the EEG cap was fitted. E-Prime 2.0 (Psychology Software Tools, Inc.) was used as the presentation software. Once the EEG system was fully set up, the experimenter explained that the experiment consisted of three blocks: Dutch, English and Java. The task was to judge a series of sentences or snippets of code for grammaticality using the “q” or “p” key to indicate whether each sentence or snippet had been correct or incorrect. Participants were not told what kind of violations they would encounter but there were five practice items at the start of each block to give them an impression of the experimental stimuli. Instructions were also given on the screen and participants could ask for clarification after the practice items.

Due to the spoken nature of natural languages, people are used to processing sentences serially. Although this is not a common presentation for written language, previous studies have shown that it is possible to present written sentences word by word without readers losing track of the meaning of the sentence (e.g., Carreiras & Clifton, 2004; Friederici et al., 2002). However, programming languages are always presented in written format in such a way that the complete code is visible simultaneously. Programming languages have no spoken form where programmers learn to process code one word or symbol at a time. Therefore, the traditional way of showing stimuli for a reading ERP experiment, with only one word at a time shown in the middle of the screen, would be unnatural and confusing in this condition. To present the programming stimuli in a natural way, while still allowing for accurate timing, for all three languages, we first presented a series of asterisks representing the full length of the stimulus on the screen, and then filled in the words or symbols by replacing the asterisks segment by segment (word by word, or per semantic entity for code; see Figure 5.1 for examples of this method of presentation). Each stimulus started with a fixation cross for 1000ms, followed by a blank screen for 500ms. Then the first segment (word or symbol) appeared with asterisks filling in the space of the following words or symbols for 600ms. Every 600ms another word/symbol was filled in, replacing its asterisk place holder, until the whole sentence or code snippet was on the screen. The sentence was then followed by a blank screen for 500ms and then participants had 3000ms to indicate whether the stimulus was correct or incorrect using the “q” or “p” key. Once a response was given or 3000ms had elapsed, a blank screen appeared for 500ms before the next trial began, starting again with the fixation cross. Stimuli were written in white on a black background (Courier New, 25pt). The assignment of the Correct and Incorrect responses to left or right hands was counterbalanced across participants. The three blocks for the different languages were counterbalanced in order across participants. The task took 30 min on average. After the ERP experiment was completed, participants filled out the demographic questionnaire.

**Figure 5.1.** Examples of stimulus presentation.



*Note:* These examples of an ungrammatical stimulus in Java and in English show how each stimulus started with a fixation cross, followed by the sentence or code with the first word written out and the rest of the stimulus length filled in by asterisks. Words or semantic entities then replaced the asterisks one by one, at 600ms intervals. After the full presentation of the stimulus the participant had three seconds to judge whether the previously presented statement had been correct (grammatical) or incorrect (ungrammatical) by pressing the “p” or “q” keys on the keyboard.

### 5.2.5. EEG Recording and Data processing

The continuous EEG was recorded from 32 Ag/AgCl scalp electrodes (WaveGuard) using the eego system (ANT Neuro Inc, Enschede, The Netherlands). An additional electrode was used (EOG, above the left eye) in order to detect, and allow correction for eye movement artifacts. Impedances were kept below 10 kΩ. Data were acquired at a 500 Hz sampling rate using the common average reference.

The offline processing was carried out in MATLAB using the EEGLAB package. Data were re-referenced to the average of the mastoids. Offline filtering was performed using a band-pass filter (0.1–30 Hz). The data were then segmented into epochs starting 200ms before presentation of the critical word or symbol, until 1200ms post onset. Data were baseline-corrected, with the baseline period of -200-0ms. We then used Independent Component Analysis (ICA) to remove artifacts due to eye movement, muscle activity, heart rate and electrical interference. Only components that were labelled “brain” or “other” by ICA remained. Finally, we used automatic epoch rejection with a threshold of  $\pm 100 \mu\text{V}$  to reject epochs that still contained too much noise. In total 6.4% of all trials were excluded

(5.8% for Dutch grammatical, 5.8% for Dutch ungrammatical, 4.2% for English grammatical, 6.7% for English ungrammatical, 9.2% for Java grammatical and 6.7% for Java ungrammatical). There was no difference in the number of retained epochs between conditions ( $F(5, 55) = .938, p = .464$ ). Finally, the data were averaged per subject and per condition. Grammaticality judgement accuracy was sufficient for all participants in all conditions (70% correct or higher), so no participants were excluded from any of the analyses. Grammaticality judgements in the current study were used to maintain the participants' attention and to give an indication of the overall recognition of violations, they were not used to exclude data on individual trials.

### 5.2.6 Analysis

Averaged values (in  $\mu\text{V}$ ) were extracted per participant, per condition, and per region of interest. Scalp electrodes were divided into 9 regions of interest, each containing either 2 or 3 electrodes: left anterior (F7, F3, FC5), midline anterior (Fz, FC1, FC2), right anterior (F4, F8, FC6), left central (C3, CP5), midline central (Cz, CP1, CP2), right central (C4, CP6), left posterior (P7, P3), midline posterior (Pz, POz), and right posterior (P4, P8). The grand average was computed for each region of interest over the full time-span from 200ms before until 1200ms after presentation of the violation. Topographic maps and ANOVAs were carried out for four time-windows, from 200 to 400ms, 400 to 600ms, 600 to 800ms and 800 to 1000ms.

For the statistical analysis, which was carried out using R, two repeated measures ANOVAs (using the "ez" package in R) were performed for each language and for each of the four time-windows. The first included the lateral regions of interest (Left Anterior, Left Central, Left Posterior, Right Anterior, Right Central and Right Posterior) and had grammaticality (2 levels: grammatical and ungrammatical), anteriority (3 levels: anterior, central, and posterior), and hemisphere (2 levels: left and right hemisphere) as within subject factors. The second repeated measures ANOVA only looked at the regions of interest in the midline (Midline Anterior, Midline Central and Midline Posterior) and had grammaticality (2 levels: grammatical and ungrammatical), and anteriority (3 levels: anterior, central, and posterior) as within subject factors.



Additionally, to analyse whether there were differences in the grammaticality effects across the three languages, we performed similar ANOVAs on the difference waves between the grammatical and ungrammatical levels of each condition. For the lateral regions, the ANOVAs had language (3 levels: Dutch, English and Java), anteriority (3 levels: anterior, central, and posterior), and hemisphere (2 levels: left and right hemisphere), as within subject factors. For the midline regions, the ANOVAs had language (3 levels: Dutch, English and Java), and anteriority (3 levels: anterior, central, and posterior) as within subject factors. The significance level was set to  $p < .05$ . Post-hoc paired  $t$ -tests for the individual languages were carried out with those interactions that were significant ( $p < .05$ ) or marginally significant ( $.1 > p > .05$ ), and that included the factor grammaticality or language. As an effect size statistic for the ANOVA analyses, we reported generalised eta squared ( $\eta^2G$ ). For the pairwise comparisons,  $p$ -values were corrected using Bonferroni corrections. If the assumption of sphericity for the ANOVAs was violated at  $p < .05$ , the Greenhouse and Geisser (1959) correction was applied.

## 5.3 RESULTS

### 5.3.1 Accuracy data

Overall accuracy in the grammaticality judgment task was 97.5%. Of the 12 participants, 3 achieved 100% accuracy; the number of errors for the other participants ranged from 1 to 7. Accuracy of grammaticality judgements across languages ranged from 99.58% correct for both grammatical and ungrammatical items in Dutch to 94.58% correct for the ungrammatical items in Java. However,  $t$ -tests showed that none of these differences in grammaticality judgements across languages reached significance (all  $p$ -values  $> .08$ ).

### 5.3.2 ERP results

For the analysis we compared grammatical and ungrammatical sentences for each of the three languages (Dutch, English and Java). A visual inspection of the waveforms showed a positive effect elicited by ungrammatical stimuli in all three languages. For Dutch (see Figure 5.2), visually, the main deflection of this effect appeared to start at approximately

500ms, lasted until approximately 800ms and occurred mainly in the central and frontal regions. This effect was followed by a smaller ongoing positivity in the posterior regions which was still present at 1000ms.

For the English language (see Figure 5.3), the effect started at approximately 600ms and was broadly distributed, with the strongest initial effect in the frontal left and central regions, lasting until approximately 800ms, and an extended effect in the posterior regions that was still present at 1000ms.

For the programming language Java (see Figure 5.4), the effect started at around 400ms, mainly in the frontal and central areas and lasted until approximately 800ms. There was no extended positive effect after 800ms.

We now discuss the ANOVA results for each language in each of the four time windows. In the text we only report the ANOVA results that were significant or marginally significant. All results, including non-significant results can also be found in Appendix C.

## Dutch

In the Dutch language condition, in the time window from **200 to 400ms** there was a significant main effect of grammaticality in the lateral regions ( $F(11) = 10.591, p = .008, \eta^2G = .038$ ), with ungrammatical stimuli eliciting a more positive response than grammatical stimuli. There were no significant interactions between grammaticality and anteriority or between grammaticality and hemisphere.

In the analysis of the midline regions there was a marginally significant main effect of grammaticality ( $F(11) = 4.648, p = .054, \eta^2G = .040$ ), with ungrammatical stimuli eliciting a more positive response than grammatical stimuli. There was no significant interaction between grammaticality and anteriority.

In the time window from **400 to 600ms** there was a marginally significant main effect of grammaticality in the lateral regions ( $F(11) = 3.641, p = .083, \eta^2G = .039$ ) with ungrammatical stimuli eliciting a more positive response than grammatical stimuli. There was also a significant interaction between grammaticality and anteriority ( $F(22) = 12.076, p < .001, \eta^2G = .048$ ). Post hoc *t*-tests for the interaction indicated that the effect of grammaticality was present in the left anterior region ( $t(11) = -3.400, p = .006$ ), but not in

the other regions. In this ROI ungrammatical stimuli elicited a more positive response than grammatical stimuli. There was also a marginally significant interaction between grammaticality and hemisphere ( $F(11) = 3.631, p = .083, \eta^2G = .014$ ). Post hoc  $t$ -tests for this interaction indicated that the effect of grammaticality was significant in the left hemisphere ( $t(11) = -2.239, p = .047$ ) but not in the right hemisphere.

In the analysis of the midline regions there was no significant main effect of grammaticality. However, there was a significant interaction between grammaticality and anteriority ( $F(22) = 8.079, p = .013, \eta^2G = .040$ ). Post hoc  $t$ -tests for this interaction indicated that the effect of grammaticality was present in the midline anterior region ( $t(11) = -2.851, p = .016$ ) but not in the other regions. In this ROI ungrammatical stimuli elicited a more positive response than grammatical stimuli.

In the time window from **600 to 800ms** there was a significant main effect of grammaticality in the lateral regions ( $F(11) = 15.198, p = .002, \eta^2G = .208$ ), with ungrammatical stimuli eliciting a more positive response than grammatical stimuli. There was also a significant interaction between grammaticality and anteriority ( $F(22) = 10.883, p < .001, \eta^2G = .026$ ). Post hoc  $t$ -tests for this interaction indicated that the effect of grammaticality was present in the left anterior region ( $t(11) = -5.287, p < .001$ ), left central region ( $t(11) = -3.842, p = .003$ ), right anterior region ( $t(11) = -3.833, p = .003$ ) and right central region ( $t(11) = -3.175, p = .009$ ), but not in the posterior regions. In these ROIs ungrammatical stimuli elicited a more positive response than grammatical stimuli. There was no significant interaction between grammaticality and hemisphere.

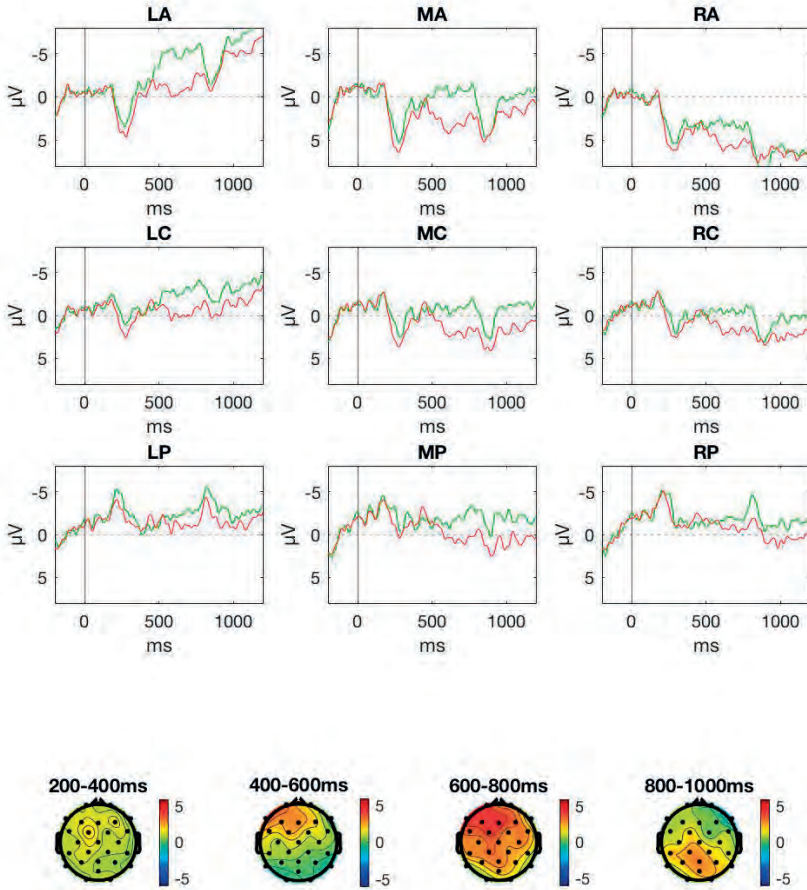
In the analysis of the midline regions there was a significant main effect of grammaticality ( $F(11) = 9.972, p = .009, \eta^2G = .281$ ), with ungrammatical stimuli eliciting a more positive response than grammatical stimuli. There was no significant interaction between grammaticality and anteriority.

In the time window from **800 to 1000ms** there was a significant main effect of grammaticality in the lateral regions ( $F(11) = 6.304, p = .029, \eta^2G = .051$ ), with ungrammatical stimuli eliciting a more positive response than grammatical stimuli. There was also a significant interaction between grammaticality and anteriority ( $F(22) = 6.452, p = .006, \eta^2G = .018$ ). Post hoc  $t$ -tests for this interaction indicated that the effect of

grammaticality was present in the left central region ( $t(11) = -2.489, p = .030$ ), the left posterior region ( $t(11) = -3.112, p = .010$ ), and the right posterior region ( $t(11) = -4.099, p = .002$ ), but not in the other regions. In these ROIs ungrammatical stimuli elicited a more positive response than grammatical stimuli. There was no significant interaction between grammaticality and hemisphere.

In the analysis of the midline regions there was a significant main effect of grammaticality ( $F(11) = 8.228, p = .015, \eta^2G = .137$ ), with ungrammatical stimuli eliciting a more positive response than grammatical stimuli. There was also a significant interaction between grammaticality and anteriority ( $F(22) = 8.656, p = .009, \eta^2G = .025$ ). Post hoc  $t$ -tests for this interaction indicated that the effect of grammaticality was present in the midline central region ( $t(11) = -3.022, p = .012$ ) and the midline posterior region ( $t(11) = -4.779, p < .001$ ), but not in the midline anterior region. In these ROIs ungrammatical stimuli elicited a more positive response than grammatical stimuli.

**Figure 5.2.** Line plots per region of interest and head plots over time for Dutch stimuli.



*Note:* The green line shows the response to grammatical stimuli, and the red line shows the response to ungrammatical stimuli. L=left, M=midline, R=right, A=anterior, C=Central, P= posterior.

## English

In the English language condition, in the time window from **200 to 400ms** there were no significant main effects of grammaticality in either the lateral or midline analysis nor any significant interactions between grammaticality and anteriority, nor between grammaticality and hemisphere (lateral analysis only).

In the time window from **400 to 600ms** there was no significant main effect of grammaticality in the lateral regions. There was also no significant interaction between grammaticality and anteriority. There was a marginally significant interaction between grammaticality and hemisphere ( $F(11) = 3.974$ ,  $p = .072$ ,  $\eta^2G = .011$ ). Post hoc  $t$ -tests for this interaction indicated that the effect of grammaticality was significant in the left hemisphere ( $t(11) = -2.507$ ,  $p = .029$ ) but not in the right hemisphere.

In the analysis of the midline regions there was no significant main effect of grammaticality. There was also no significant interaction between grammaticality and anteriority.

In the time window from **600 to 800ms** there was a significant main effect of grammaticality in the lateral regions ( $F(11) = 11.341$ ,  $p = .006$ ,  $\eta^2G = .188$ ), with ungrammatical stimuli eliciting a more positive response than grammatical stimuli. There were no significant interactions between grammaticality and anteriority or between grammaticality and hemisphere.

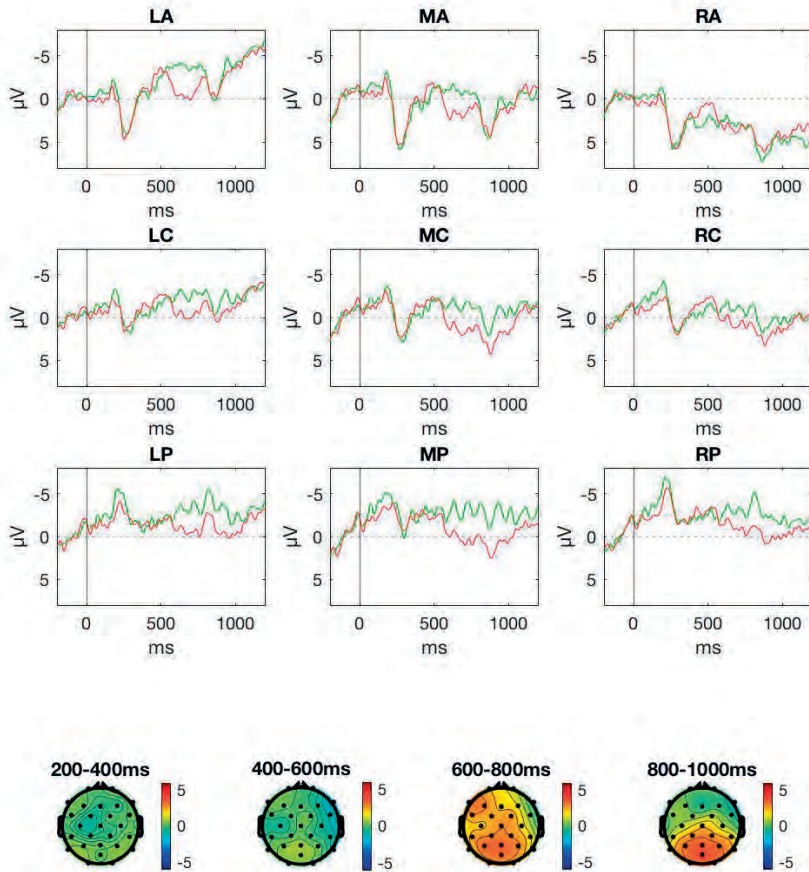
In the analysis of the midline regions there was a significant main effect of grammaticality ( $F(11) = 13.342$ ,  $p = .004$ ,  $\eta^2G = .296$ ), with ungrammatical stimuli eliciting a more positive response than grammatical stimuli. There was no significant interaction between grammaticality and anteriority.

In the time window from **800 to 1000ms** there was a significant main effect of grammaticality in the lateral regions ( $F(11) = 5.535$ ,  $p = .038$ ,  $\eta^2G = .079$ ), with ungrammatical stimuli eliciting a more positive response than grammatical stimuli. There was also a significant interaction between grammaticality and anteriority ( $F(22) = 12.735$ ,  $p = .002$ ,  $\eta^2G = .063$ ). Post hoc  $t$ -tests for this interaction indicated that the effect of grammaticality was present in the left posterior region ( $t(11) = -4.409$ ,  $p = .001$ ) and the right posterior region ( $t(11) = -4.339$ ,  $p = .002$ ), but not in the other regions. In these ROIs,

ungrammatical stimuli elicited a more positive response than grammatical stimuli. There was no significant interaction between grammaticality and hemisphere.

In the analysis of the midline regions there was a significant main effect of grammaticality ( $F(11) = 10.428, p = .008, \eta^2G = .205$ ), with ungrammatical stimuli eliciting a more positive response than grammatical stimuli. There was also a significant interaction between grammaticality and anteriority ( $F(22) = 13.739, p = .002, \eta^2G = .090$ ). Post hoc *t*-tests for this interaction indicated that the effect of grammaticality was present in the midline central region ( $t(11) = -3.585, p = .004$ ) and the midline posterior region ( $t(11) = -5.249, p < .001$ ) but not in the midline anterior region. In these ROIs, ungrammatical stimuli elicited a more positive response than grammatical stimuli.

**Figure 5.3.** Line plots per region of interest and head plots over time for English stimuli.



*Note:* The green line shows the response to grammatical stimuli, and the red line shows the response to ungrammatical stimuli. L=left, M=midline, R=right, A=anterior, C=Central, P= posterior.



## Java

For the Java language condition, in the time window from **200 to 400ms** there was a significant main effect of grammaticality in the lateral regions ( $F(11) = 5.444$ ,  $p = .040$ ,  $\eta^2G = .050$ ), with ungrammatical stimuli eliciting a more positive response than grammatical stimuli. There were no significant interactions between grammaticality and anteriority or between grammaticality and hemisphere.

In the analysis of the midline regions there was a marginally significant main effect of grammaticality ( $F(11) = 3.519$ ,  $p = .087$ ,  $\eta^2G = .059$ ), with ungrammatical stimuli eliciting a more positive response than grammatical stimuli. There was no significant interaction between grammaticality and anteriority.

In the time window from **400 to 600ms** there was a significant main effect of grammaticality in the lateral regions ( $F(11) = 15.200$ ,  $p = .002$ ,  $\eta^2G = .140$ ), with ungrammatical stimuli eliciting a more positive response than grammatical stimuli. There was also a marginally significant interaction between grammaticality and anteriority in the lateral regions ( $F(22) = 3.478$ ,  $p = .078$ ,  $\eta^2G = .019$ ). Post hoc  $t$ -tests for this interaction indicated that the effect of grammaticality was present in the left anterior region ( $t(11) = -4.548$ ,  $p < .001$ ), the left central region ( $t(11) = -2.985$ ,  $p = .012$ ), the right anterior region ( $t(11) = -3.218$ ,  $p = .008$ ) and the right central region ( $t(11) = -3.842$ ,  $p = .003$ ), but not in the posterior regions. In these ROIs ungrammatical stimuli elicited a more positive response than grammatical stimuli. There was no significant interaction between grammaticality and hemisphere.

In the analysis of the midline regions there was a significant main effect of grammaticality ( $F(11) = 10.890$ ,  $p < .007$ ,  $\eta^2G = .184$ ), with ungrammatical stimuli eliciting a more positive response than grammatical stimuli. There was also a significant interaction between grammaticality and anteriority ( $F(22) = 14.160$ ,  $p = .001$ ,  $\eta^2G = .037$ ) with post hoc  $t$ -tests for this interaction indicating that the effect of grammaticality was present in all three ROIs: the midline anterior region ( $t(11) = -4.089$ ,  $p = .002$ ), the midline central region ( $t(11) = -2.636$ ,  $p = .023$ ) and the midline posterior region ( $t(11) = -2.441$ ,  $p = .033$ ). In all ROIs, ungrammatical stimuli elicited a more positive response than grammatical stimuli.

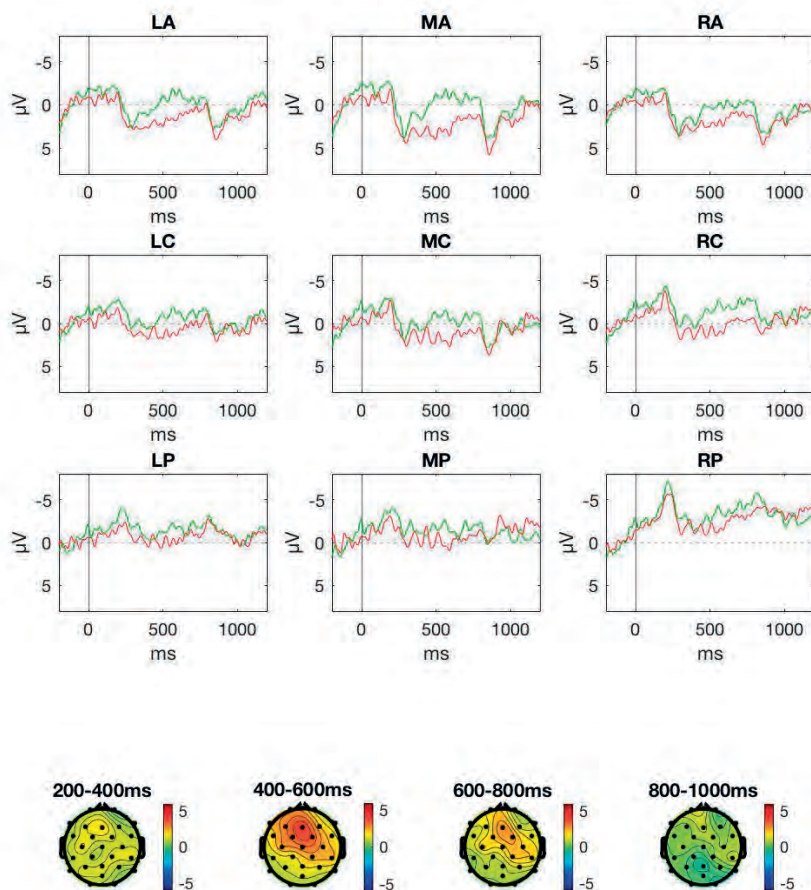
In the time window from **600 to 800ms** there was a marginally significant main effect of grammaticality in the lateral regions ( $F(11) = 4.327, p = .062, \eta^2G = .052$ ), with ungrammatical stimuli eliciting a more positive response than grammatical stimuli. There were no significant interactions between grammaticality and anteriority or between grammaticality and hemisphere.

In the analysis of the midline regions there was no significant main effect of grammaticality, but there was a significant interaction between grammaticality and anteriority ( $F(22) = 6.224, p = .007, \eta^2G = .015$ ), with post hoc *t*-tests for this interaction indicating that the effect of grammaticality was present in the midline anterior region ( $t(11) = -2.915, p = .014$ ) but not in the other regions. In this ROI ungrammatical stimuli elicited a more positive response than grammatical stimuli.

In the time window from **800 to 1000ms** there was no significant main effect of grammaticality in the lateral regions. There were also no significant interactions between grammaticality and anteriority or between grammaticality and hemisphere.

In the analysis of the midline regions there was no significant main effect of grammaticality, but there was a marginally significant interaction between grammaticality and anteriority ( $F(22) = 3.577, p = .074, \eta^2G = .024$ ). However, post hoc *t*-tests for this interaction revealed no significant effects of grammaticality in any of the ROIs.

**Figure 5.4.** Line plots per region of interest and head plots over time for Java stimuli.



*Note:* The green line shows the response to grammatical stimuli, and the red line shows the response to ungrammatical stimuli. L=left, M=midline, R=right, A=anterior, C=Central, P= posterior.

### Comparison of the ERP patterns per language

When comparing the difference waves between grammatical and ungrammatical conditions between the three languages, in the time window from **200 to 400ms** there were no significant main effects of language in either the lateral or midline analysis nor any significant interactions between language and anteriority, nor between language and hemisphere (lateral analysis only).

In the time window from **400 to 600ms** there was a significant main effect of language in the lateral regions ( $F(22) = 5.820, p = .009, \eta^2G = .146$ ). Post hoc *t*-tests indicated that the effect in response to Java was stronger than the effects in response to Dutch ( $t(71) = -2.896, p = .015$ ) and English ( $t(71) = -5.637, p < .001$ ) while the effect in response to Dutch was also stronger than the effect in response to English ( $t(71) = 3.301, p = .005$ ). There was also a significant interaction between language and anteriority ( $F(44) = 4.716, p = .020, \eta^2G = .055$ ). Post hoc *t*-tests for this interaction indicated that there was a significant effect of language only in the left central region where the effect in response to Java was stronger than the effect in response to English ( $t(11) = -2.823, p = .050$ ), in the right anterior region where the effect in response to Java was stronger than the effect in response to English ( $t(11) = -3.316, p = .021$ ), and the effect in response to Dutch was stronger than the effect in response to English ( $t(11) = 3.274, p = .022$ ) and in the right central region where the effect in response to Java was stronger than the effect in response to Dutch ( $t(11) = -3.580, p = .013$ ) and stronger than the effect in response to English ( $t(11) = -3.259, p = .023$ ). There was no significant interaction between language and hemisphere.

In the analysis of the midline regions, there was a marginally significant main effect of language ( $F(22) = 3.098, p = .065, \eta^2G = .088$ ). Post hoc *t*-tests indicated that the effect in response to Java was stronger than the effects in response to Dutch ( $t(35) = -2.527, p < .048$ ) and English ( $t(35) = -3.232, p = .008$ ). There was also a significant interaction between language and anteriority ( $F(44) = 6.017, p = .017, \eta^2G = .054$ ). Post hoc *t*-tests for this interaction indicated that there was a significant effect of language only in the midline anterior region where the effect in response to Java was stronger than the effect in response to English ( $t(11) = -2.867, p = .046$ ).

In the time window from **600 to 800ms** there was a marginally significant main effect of language in the lateral regions ( $F(22) = 2.719, p = .088, \eta^2G = .055$ ). In contrast to the earlier time windows, post hoc  $t$ -tests indicated that the effect in response to Java was *weaker* than the effects in response to Dutch ( $t(71) = 4.002, p < .001$ ) and English ( $t(71) = 2.825, p = .018$ ). There was also a significant interaction between language and anteriority ( $F(44) = 4.420, p = .004, \eta^2G = .025$ ). Post hoc  $t$ -tests for this interaction indicated that there was a significant effect of language only in the left anterior region, where the response to Dutch was stronger than the response to Java ( $t(11) = 2.873, p = .045$ ). There was no significant interaction between language and hemisphere.

In the analysis of the midline regions, there was a marginally significant main effect of language ( $F(22) = 2.736, p = .087, \eta^2G = .076$ ). Post hoc  $t$ -tests indicated that the effect in response to Java was weaker than the effects in response to Dutch ( $t(35) = 3.660, p = .003$ ) and English ( $t(35) = 2.822, p = .023$ ). There was also a significant interaction between language and anteriority ( $F(44) = 4.816, p = .012, \eta^2G = .020$ ). Post hoc  $t$ -tests for this interaction indicated that there was a significant effect of language only in the midline posterior region where the effect in response to English was stronger than the effect in response to Java ( $t(11) = 3.092, p = .031$ ).

In the time window from **800 to 1000ms** there was no significant main effect of language in the lateral regions, but there was a significant interaction between language and anteriority ( $F(44) = 6.400, p = .005, \eta^2G = .069$ ). Post hoc  $t$ -tests for this interaction indicated that there was a significant effect of language only in the left posterior region where the effect in response to English was stronger than the effect in response to Java ( $t(11) = 2.953, p = .039$ ). There was no interaction between language and hemisphere.

In the analysis of the midline regions, there was a significant main effect of language ( $F(22) = 4.331, p = .026, \eta^2G = .162$ ). Post hoc  $t$ -tests indicated that, once again, the effect in response to Java was weaker than the effects in response to Dutch ( $t(35) = 3.379, p = .005$ ) and English ( $t(35) = 3.347, p = .006$ ). There was also a significant interaction between language and anteriority ( $F(44) = 10.180, p < .001, \eta^2G = .101$ ). Post hoc  $t$ -tests for this interaction indicated that there was a significant effect of language only in the midline

posterior region where the effect in response to Java was weaker than the effects in response to Dutch ( $t(11) = 3.082, p = .031$ ) and English ( $t(11) = 4.002, p = .006$ ).

### **5.3.3 Summary of ERP results**

The statistical ERP results are summarised in Tables 5.3 and 5.4. Overall, we find a positive deflection in responses to ungrammatical sentences for all three languages that is present between 400 and 800ms. Statistically, the effects for Java and Dutch start first, reaching significance in the time window from 200ms to 400ms, followed by English which reaches significance in the left hemisphere in the time window from 400-600ms, and reaches overall significance in the time window from 600 to 800ms. Visually, the main deflection for Java appears to start first, at around 400ms. For Dutch, although there is already an early statistically significant effect of grammaticality, the main deflection appears to start at around 500ms. The effect for English appears to start last, at around 600ms. We see that English and Dutch show prolonged positive effects in the time window from 800 to 1000ms as well, while this effect is absent for Java. The effects for Dutch and Java start more frontally, while the effect for English is more broadly distributed. The effects for English and Dutch in the time window from 400-600ms appear to occur mostly in the left hemisphere. The late effect for Dutch and English from 800 to 1000ms is mainly found in the central and posterior regions.

**Table 5.3.** Summary of ANOVA results by language.

		200-400ms	400-600ms	600-800ms	800-1000ms
Dutch	Main effect	✓		✓	✓
	Anterior		✓	✓	
	Midline		✓	✓	✓
	Posterior				✓
	Left hemisphere		✓		
	Right hemisphere				
	English	Main effect			✓
Anterior					
Midline					✓
Posterior					✓
Left hemisphere			✓		
Right hemisphere					
Java		Main effect	✓	✓	
	Anterior		✓	✓	
	Midline		✓		
	Posterior		✓		
	Left hemisphere				
	Right hemisphere				

*Note:* This table shows in which time windows there were significant main effects of grammaticality for each language, and, if there was a significant interaction with anteriority in either the lateral or the midline regions, whether effects were present in the anterior, midline or posterior regions. If there was an interaction between grammaticality and hemisphere the table shows whether the effects were present in the left or the right hemisphere.

**Table 5.4.** Summary of statistical results comparing grammaticality effects between the languages.

	200-400ms	400-600ms	600-800ms	800-1000ms
Main effects		Java > Dutch&English		Dutch&English > Java
Grammaticality * Anteriority	Anterior	Dutch&Java > English	Dutch > Java	
	Central	Java > Dutch & English		
	Posterior		English>Java	Dutch&English > Java

*Note:* This table shows in which time windows there were main effects of language, and the directions of those effects. If there was an interaction between language and anteriority in either the midline or lateral regions, the table also shows in which region an effect was found (anterior, central or posterior) and what the direction of the effect was in that region.

## 5.4 DISCUSSION

This study aimed to examine whether a programming language is processed similarly to natural languages. This was done by examining brain responses elicited by syntax errors in a programming language and comparing those to responses elicited by grammatical violations in participants' first (Dutch) and second (English) natural languages. Specifically, we compared ERP effects in response to bracket errors in "if" and "while" statements in Java to responses in reaction to subject-verb disagreements in Dutch and English. We visually and statistically inspected the responses per language over four time windows in nine different ROIs.

We found a positive deflection in response to violations for all three languages. This suggests that participants were sufficiently proficient in all three languages to process the presented violations as errors. This was confirmed by the high accuracy on the grammaticality judgment questions. Additionally, finding the current ERP responses confirms that the violations in all three languages were not semantic in nature, as semantic



violations are known to elicit a negative wave around 400ms after the violation (Friederici, 1995; 2002; Kutas & Hillyard, 1980). However, the differences in latency, offset and scalp distribution between Java, Dutch and English was unexpected. In this section we summarise these differences and discuss various possible explanations.

The early frontal effect for Dutch that reached significance between 200 and 400ms was unexpected. Visually, the main deflection of the effect appeared to start at around 500ms, with a broad scalp distribution. It is possible that the early significant effect was noise due to the small sample size of the study. However, the early positivity seems to precede the main P600 effect, and may be a separate effect related to visual processing and expectancy (Potts, 2004), perhaps brought on by the way stimuli were presented and their relative brevity. The visual and statistical effects for Dutch and English at around 500-600ms are in line with the P600 effect in response to sentences with a violation or unpreferred use of syntax (Friederici et al., 2002; Hagoort & Brown, 2000). Additionally, the prolonged late positivity effects in both Dutch and English in the posterior regions were in line with a typical P600 effect (Hagoort & Brown, 2000), and probably reflect reanalysis of the stimulus (Molinero et al., 2008).

For Java the effect occurred early. It was statistically detectable from 200ms and became visually prominent at around 300ms. It occurred mainly in the frontal and central regions. Unlike for Dutch, the effect seemed to be uniform (consisting of only one *peak* or component) throughout its entire duration (200-800ms). Visually, as well as statistically, the effect for Java had a shorter duration and ended earlier than the effects for Dutch and English. Furthermore, the effect for Java was never predominantly distributed in the posterior regions, which did happen for the effects for Dutch and English and is expected for the later part of a P600 effect (Hagoort & Brown, 2000). Therefore, the effect for Java cannot be unequivocally interpreted as an early P600 effect, as its characteristics are more in line with the P300 effect (Coulson et al., 2010; Osterhout, 1999; Osterhout et al., 1996; Sassenhagen & Fiebach, 2019; Sassenhagen et al. 2014). Based on the early latency, the short duration and the frontocentral distribution, the effect favours a P300 interpretation (Osterhout, 1999; Osterhout et al., 1996). In this case, the bracket error in Java is considered as a more superficial violation of expectations rather than a grammatical violation. This violation can still be language-related, such as incorrect spelling, or it can be a formatting

error that is not cognitively integrated in the same way as in natural languages. When comparing our findings for Java to the findings by Osterhout et al. (1996) for unexpected capitalisation and by Vissers et al. (2006) for orthographic (spelling) violations in English, we see that the ERP responses to orthographic errors resembled the responses to our Java violations most closely. The orthographic errors elicited an early response, starting around 400ms (which can also be classified as a P300). This response was also fronto-centrally distributed and, thus, matches the current response in Java in both latency and topography. Hence, interpretation of the early positive wave as a P300 effect is a plausible explanation for the differences in onset, offset and topography between Java and the natural languages in the current study.

However, we cannot completely rule out the possibility that the effect is an early P600 effect, especially given the fact that the P600 for Dutch also had an earlier component. If we were to interpret the effect as the P600, we have to address several more issues, mainly related to its latency. Since Dutch was the participants' native language, we expected the violations in Dutch to elicit the earliest ERP response, which would then be followed by a later onset for English and Java. Friederici (1995) proposed that the latency of the P600 effect reflects the complexity of processing necessary for the revision of the initially preferred reading. In the current study, there are two factors that may have influenced the ease of processing across the languages. The first is the level of fluency in each language. Based on the literature, we expect the most fluent language to elicit the earliest response (Kotz, 2009; Rossi et al., 2006; Weber-Fox & Neville, 1996). We assumed that Dutch, being the native language, was also the most fluent language. However, it is possible that although Dutch was the programmers' native language, and English was typically learned before Java, Java is currently the written language to which they are exposed the most. Professional programmers and programming students spend the majority of their worktime working with code, rather than reading texts in natural languages such as Dutch or English. Unless these programmers read regularly in their free time, it is possible that they have been primarily exposed to written Java rather than written texts in Dutch or English for multiple years. This higher exposure and proficiency in Java reading could explain why the errors were processed more quickly and, therefore, elicited an earlier P600 compared to Dutch and English.

A second factor that may have influenced the ease of processing for the Java violations is stimulus length. The violation in Java becomes clear when a single bracket appears, while for Dutch and English the violation becomes clear at the appearance of a multiple-letter word, that is decoded and then processed. It is possible that the single character length of the bracket versus the multiple character words makes processing of the Java violations quicker than processing of the violations in Dutch and English, leading to an earlier P600 effect in Java. We cannot exclude this possibility with the current results. However, future studies could investigate this by conducting a similar experiment where the Java error consists of multiple characters.

If the effect in Java does indeed reflect an (early) P600, it only elicits the first stage of registering, where the reader identifies the violation, without later reanalysis (Molinaro et al, 2008). This may be because the bracket violation in Java is more easily recognised and resolved than the violations in Dutch and English. Programmers may assume that the “if” or “while” function was chosen deliberately, and, thus, the curly bracket must simply be replaced with a round one (rather than replacing “if” or “while” with, for example, “else” which requires a curly bracket). In Dutch and English this resolution may be more complicated because the error can be resolved in several ways. For example, readers may consider either the inflection on the verb or the pronoun as incorrect. The resolution of this violation is more dependent on the context of the sentence and, therefore, may require more in-depth processing from the reader. A lack of reanalysis is also expected if Java bracket errors are indeed comparable to orthographic violations. In that case, they are quickly resolved as a spelling error with no alternative meaning. Therefore, a lack of the repair and reanalysis stage would explain the short duration of the effect, as well as its lack of a posterior distribution, which is mainly related to repair and reanalysis (Hagoort & Brown, 2000).

### **5.4.1 Conclusion**

Based on previous studies, the effect in response to Java bracket violations can be interpreted either as an early P600 or as a P300 effect. Regardless of the exact classification of the effect, the latency and topography of the Java response suggest that this type of error is similar to an orthographic violation in natural language. Based on the current

results, the most likely conclusion is, thus, that the bracket violation in Java is language related, but not of a grammatical nature. However, we cannot exclude the possibility that the Java response reflects a P300 effect that is unrelated to language, or that it reflects an early P600 that is related to grammar. This means that, based on the current experiment, we are unable to conclude whether syntax in a programming language is processed similarly to grammar in a natural language. Future studies with different stimuli will be necessary to disentangle these different possibilities.

The current study lays the foundations for these future studies by showing that processing of programming languages can be studied through ERPs. Future research could build on this by studying a wide variety of violations in different programming languages, aiming to determine the linguistic nature of different elements in these types of languages. Through multiple ERP studies, we should eventually be able to determine which elements of a programming language are of a grammatical, orthographic, semantic or entirely un-languagelike nature. This will help us understand the way in which people process and understand programming languages, and the way in which programming languages differ from or resemble natural languages.

