Chapter 6. The performance of digital grammar checkers for native and second language learners of French

6.1 Introduction

In the last decades, many studies aimed to evaluate the effectiveness and the usefulness of digital grammar checkers in computer-assisted language learning (e.g. Heift & Schulze 2007; Biesemans 2005; Burston 1998). Heift & Schulze (2007) for instance, showed that grammar checkers are very effective in form-focused writing exercises. In agreement with Ellis (2001), form-focused instruction is a pedagogical practice undertaken by second language teachers by which the students' attention is drawn to language form. In this particular case, form-focused exercises would explicitly target overt gender marking on adjectives and past participles by confronting students with the relevant morphosyntactic contexts.

The usefulness of grammar checkers in the language learning process of French gender marking by second language learners may be defined in terms of the number of gender marking errors that the grammar checkers is able to detect and correct. In language education,

the use of grammar checkers is assumed to enrich the pedagogy in text writing in the sense that language learners will reflect, discuss and edit their written productions (Vernon 2000).

This chapter is concerned with the correction of grammatical inflection errors, as a particular category of agreement errors that are commonly found in written L1 and L2 French. Interestingly, the correction rates of French digital grammar checkers are found to differ in particular types of agreement constructions. More specifically, Biesemans (2005) observed that only 21.4% of inflection errors in past participle agreement were corrected by the Antidote grammar checker, in contrast to a correction rate of 62.6% for inflection errors in subject-verb agreement. This suggests that in French, the syntactic configuration exhibiting gender agreement has an effect on the performance of Antidote. More precisely, agreement types hosted in a relatively local syntactic configuration, such as subject-verb

1. This chapter has been prepared as a stand-alone journal paper that has been submitted to CALICO. Therefore, sections 6.5 and 6.6.2 summarize the main findings discussed in the previous chapters of this dissertation.
agreement, show an enhanced correction performance compared to those in a larger syntactic configuration, such as past participle agreement.

For both native and second language learners of French, the size of the syntactic configuration has also been shown to have an effect on the performance of written inflections. In previous research I have shown that predicative constructions trigger more writing errors in gender marking than attributive constructions (chapter 5). Here also, the difference in writing errors may be related to a difference in the size of the domain: predicative constructions are hosted in a relatively large syntactic domain, whereas attributive constructions exhibit a very local configuration pattern.

6.2 Main aim and research question

The main aim of the present chapter is to compare the performance of digital grammar checkers and language learners when it comes to gender inflection in written French. I report the results of an experiment investigating the implementation of gender marking in different types of gender agreement configurations by three digital grammar checkers (Antidote, Scribens and BonPatron) and by two populations of French language users (native speakers and second language learners).

I will address the following research question:

*Do digital grammar checkers reach a higher performance at implementing gender marking in French than native and second language learners?*

To answer this question, I will compare their performance on gender marking in three well-defined gender agreement configurations that differ from each other with respect to the size of their syntactic domain.

The chapter is organized as follows: in the third section, the main features of the three digital grammar and spell checkers under investigation will be described. In the fourth section, I will provide a state-of-the-art on the performance of grammar and spell checkers, followed by a brief overview of the gender system in French in the fifth section. In the sixth section I will present the results of the study. Finally, in the last section, I will draw conclusions based on these results.
6.3 French grammar checkers: Antidote, BonPatron and Scribens

Antidote, BonPatron and Scribens are well-known digital grammar checkers for French. More specifically, Antidote is a grammar and spell checker which has been developed by the Canadian software enterprise Druide Informatique for both native speakers and second language learners of French. It can be used as a supplementary grammar and spell corrector and can be applied to all levels of French language acquisition. For texts submitted to Antidote, the user can select the types of errors for which the program needs to correct. As such, the selected types of errors are marked by using underlines in colors marking particular categories of writing errors (e.g. the red colored underline marks writing errors in the morphosyntactic domain). The user can accept or ignore the correction proposed by the Antidote program. Antidote also provides corrective feedback to the user explaining why the underlined item has been written incorrectly. More specifically, the appropriate grammatical rule along with an example pops up when the user clicks on the underlined item.

In a similar way, BonPatron is mainly used to correct errors in texts written by second language learners of French. This digital grammar checker has been developed at the Canadian university of Alberta and McMaster University for English learners of French and is available online (bonpatron.com). Here also, particular categories of writing errors are distinctively color-marked. In BonPatron, errors are surrounded by a colored text box over which the user can pass the cursor activating feedback concerning the grammatical rule that has been violated by the language learner. In contrast to Antidote, BonPatron has a certain degree of self-learning capacity: it collects all text data submitted to the checker to improve its own correction algorithms. An additional feature that may be of interest to second language users is the summary of signaled errors provided at the end of the text correction. This may contribute to the users’ awareness of particular writing errors.

For the digital grammar and spell checker Scribens no peculiarities need to be mentioned. Scribens corrects writing errors and provides corrective feedback on grammatical errors in roughly the same way as Antidote and BonPatron. The Scribens program is available online (scribens.fr) and has been developed for both native speakers and second language learners of French by a private software developer in France.

6.4 The performance of grammar checkers for French

In section 1 I have stated that Antidote is not able to correct all morphosyntactic writing errors (Biesemans 2005). In a similar vein, BonPatron corrects 88% of the...
morphosyntactic errors found in the written productions of L2 French language learners (Burston 2008).

Nadasdi & Sinclair (2007) provided a more detailed analysis of morphosyntactic errors that are corrected by BonPatron. They submitted 30 texts to the grammar checker that were written by English learners of French. These learners were enrolled in the first year of a French language course in a Canadian university and were asked to write a text of roughly 250 words. The writing errors that were presented to BonPatron were divided in four stylistic categories: grammar, punctuation, spelling and elision. In parallel, the same writing errors were presented to human correctors. The correction performance of BonPatron was computed in terms of the percentage of errors identified by human correctors. These results revealed that in each of the stylistic categories, human correction was more accurate than digital correction. Based on these findings, the developers made adjustments to the rule database of BonPatron. A dataset of written texts was then submitted to the new version of the grammar checker. With respect to grammar correction, an in-depth analysis shows that the correction performance was lowest in adjective-noun agreement constructions (i.e. only 61% of the errors identified by human correctors was corrected in version 1 and 87% in version 2), followed by subject-verb agreement constructions (85% in version 1 and 98% in version 2), and by determiner-noun agreement constructions (87% in version 1 and 97% in version 2). Based on these data, the authors concluded that the performance level of the enhanced version of BonPatron is equal to that of human correctors.

To the best of my knowledge, for the other two French grammar checkers (Scribens and Antidote) in-depth analysis on digital correction performance seems to be lacking altogether.

6.5 Gender marking in French

In the present chapter I will use French gender marking as test bed to compare the performance level of BonPatron, Scribens and Antidote to that of L1 and L2 learners of French. For that purpose, I will first provide a brief overview of the gender system in the French language. Firstly, the language exhibits a two-gender system with masculine and feminine gender marking. Gender marking is present on definite and indefinite articles (see (1a) and (1b) for masculine and feminine respectively), on adjectives and on past participles.

(1a) Le / Un grand cadeau
    The-M.S. / A-M.S. big-M.S. present-M.S.
    'The / A big present'
Masculine gender is not overtly expressed on regularly inflected non-derived adjectives\(^2\) (see (1a)). The feminine gender, however, exhibits overt gender marking (i.e. + e on the adjective), as in (1b).

The past participle agrees with the direct object taking the form of an object clitic (see (3)) or a noun (see (4)). To trigger overt gender (and number) agreement, the direct object must precede the past participle. Here also, masculine gender is not overtly expressed ((3a) and (4a)), whereas feminine gender marking is overtly expressed by an –e ending on the past participle ((3b) and (4b)).

\[
(3a) \quad \text{Je l’ ai arrosé}, \quad \text{(l’arbre)} \\
\text{I it-M.S. have watered-M.S. (the tree-M.S.)} \\
\text{‘I watered the tree’}
\]

\[
(3b) \quad \text{Je l’ ai arrosée}, \quad \text{(la plante)} \\
\text{I it-F.S. have watered-F.S. (the plant-F.S.)} \\
\text{‘I watered the plant’}
\]

\[
(4a) \quad \text{Le ballon qu’ on a trouvé} \\
\text{The-M.S. ball-M.S. that we have found-M.S.} \\
\text{‘The ball (that) we found’}
\]

\[
(4b) \quad \text{La boutique qu’ on a trouvée} \\
\text{The-F.S. shop-F.S. that we have found-F.S.} \\
\text{‘The shop (that) we found’}
\]

### 6.6 The experiment

The implementation rates of gender marking in native speakers and second language learners of French were taken from data that were collected for a previous study reported in chapter 4 of this dissertation.

---

\(^2\) In derived adjectives (e.g. *menteur*-M.S. vs. *menteuse*-F.S. ‘lying’) the masculine gender is overtly expressed by the suffix bearing the +M gender feature
6.6.1 Method

6.6.1.1 Digital grammar checkers and participants
The digital grammar checkers tested in this experiment, were Antidote, BonPatron and Scribens. With respect to the French language users, the participants tested in the experiments presented in chapter 4, were monolingual native speakers of French (n = 28; age range = 14 – 15 years) from the west of France. Each participant had to confirm that the home language was French and that he/she had no known language disorder, such as dyslexia. The Dutch learners of French (n = 26; age range = 17 – 18 years) attended courses in French language for 5 years at the highest level of the Dutch secondary school system (i.e. voorbereidend wetenschappelijk onderwijs ‘academic university preparatory education’). Participants did not have any known language disorder, such as dyslexia.

6.6.1.2 Materials
To allow for comparison, the text stimuli that were presented to the digital grammar checkers were identical to those used with human language learners in the experiments presented in chapter 4. More precisely, the test stimuli were presented as a fill-in-the-gap task requiring the learner or the digital grammar checker to fill in overt feminine gender marking on adjectives or past participles when deemed necessary. The test items were controlled for three types of gender agreement constructions representing different sizes of syntactic domains: with attributive adjectives, with past participles preceded by an object clitic and with past participles precede by an object noun.

Native speakers’ task. Each test condition contained 40 test items which were all in a feminine singular context. To avoid mistakes with respect to the lexical gender of the noun or the clitic, the gender was overtly expressed in all conditions. Furthermore, 30 filler items were added in order to verify whether the participants were able to comply with the test requirements. Only the results of the test items were included in the analysis.

Second language learners’ task. A subset of gender marking contexts was further used for in-depth analysis. In this stimulus set only contexts with feminine singular marking were further taken up for analysis (n = 46). Test items targeting plural and/or masculine agreement (n = 104), were considered as filler items and were not taken into consideration any further.

From a total of 46 items, 5 test items which represented the noun-adjective agreement condition (5), 7 test items exhibited the clitic-past participle agreement condition (6) and 34 test items represented the noun-past participle agreement condition (7).
(5) **Attributive adjective**

La meilleure joueuse a été sélectionnée pour la compétition.

*The best player has been selected for the competition’*

---

(6) **Past participle with object clitic**

La plante n’avait plus d’eau. Je l’ai arrosé.

*The plant did not have water. I watered it.*

---

(7) **Past participle with fronted noun**

C’est la fleur exotique que j’ai vu dans le jardin.

*It’s the exotic flower (that) I saw in the garden’*

---

### 6.6.1.3 Procedure

The participants were asked to write down on paper the correct gender morpheme in each test item if needed. The experimental setting of the native speakers and second language learners was in a classroom at a secondary school in the west of France and in the west of The Netherlands respectively.

The task was done individually and under supervision of the teacher. Furthermore, the presentation order of the test items was counter-balanced in three versions. All participants finished the task well within the 50 minutes.

The outcomes from the digital grammar checkers were obtained by feeding the native speakers’ test items described above to Antidote and Scribens for correction. Similarly, the test items of the second language learners’ test were submitted to BonPatron and Antidote. Here, the filler items were removed as there was no need to control for the task in a digital experiment environment.
6.6.2 Results

For the digital grammar checkers, I computed the implementation rate of gender marking per test condition. This rate was expressed in terms of the percentage of correctly adjusted test items per condition. An overview of the implementation rates is presented in table 1.

Table 1. Percentages of corrected test items per condition

<table>
<thead>
<tr>
<th>Type of agreement</th>
<th>Antidote L1 test items</th>
<th>Scribens L1 test items</th>
<th>BonPatron L2 test items</th>
<th>Antidote L2 test items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjective</td>
<td>100</td>
<td>100</td>
<td>75</td>
<td>100</td>
</tr>
<tr>
<td>Object clitic</td>
<td>40</td>
<td>0</td>
<td>57</td>
<td>43</td>
</tr>
<tr>
<td>Fronted noun</td>
<td>98</td>
<td>95</td>
<td>53</td>
<td>100</td>
</tr>
</tbody>
</table>

For both groups of French language users, I will present the implementation rates of gender marking, taken from the experiments reported in chapter 4. These rates were expressed in terms of the percentage of correct gender inflections per condition. The data set of implementation rates found in the native speakers’ group revealed to be non-parametric and were therefore reported by means of 5 parameter statistics (table 2). In the second language learners’ group the data set revealed to be parametric. Therefore, the means and standard deviations were reported (table 3).

Table 2. Percentages of correct gender inflections in native speakers of French

<table>
<thead>
<tr>
<th>Type of agreement</th>
<th>minimum</th>
<th>lower quartile</th>
<th>median</th>
<th>upper quartile</th>
<th>maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjective</td>
<td>75</td>
<td>85</td>
<td>95</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Object clitic</td>
<td>0</td>
<td>42.50</td>
<td>95</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Fronted noun</td>
<td>3</td>
<td>45</td>
<td>76.50</td>
<td>85.75</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 3. Percentages of correct gender inflections in second language learners of French

<table>
<thead>
<tr>
<th>Type of agreement</th>
<th>mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjective</td>
<td>81</td>
<td>22</td>
</tr>
<tr>
<td>Object clitic</td>
<td>54</td>
<td>30</td>
</tr>
<tr>
<td>Fronted noun</td>
<td>45</td>
<td>34</td>
</tr>
</tbody>
</table>

To compare the implementation rates of the grammar checkers to those of the French language users, I conducted a one-sample Wilcoxon test for the L1 grammar checkers and a one-sample T-test for the L2 grammar checkers on each of the agreement constructions. More precisely, the implementation rates of the grammar checkers were taken as the test values to which those of the native speakers
and second language learners were contrasted. For all statistical analyses the α level of significance was set at .05.

With respect to the L1 grammar checkers (i.e. Antidote and Scribens), significant contrasts for all types of agreement constructions are observed between these grammar checkers and native speakers of French. More specifically, Antidote performs better in implementing gender marking than native speakers both in noun-adjective agreement constructions (T (28) = -3.54; p = .000) and in noun-past participle agreement constructions (T (28) = -4.47; p = .000). In clitic-past participle agreement constructions, however, native speakers outperform Antidote (T (28) = 3.83; p = .000).

Similar contrasts in the correct implementation of gender marking are found between Scribens and French language users. More precisely, Scribens performs better than native speakers both in noun-adjective agreement constructions (T (28) = -3.54; p = .000) and in noun-past participle agreement constructions (T (28) = -4.44; p = .000). In clitic-past participle agreement constructions, however, native speakers outperform Scribens (T (28) = 4.57; p = .000).

With respect to the L2 grammar checkers (i.e. BonPatron and Antidote), no significant contrasts are observed between BonPatron and second language learners of French (i.e. noun-adjective agreement constructions: T (25) = 1.40; p = .172; clitic-past participle agreement constructions: T (25) = -.44; p = .664; noun-past participle agreement constructions: T (25) = -1.19; p = .244). Concerning Antidote, significant contrasts between this grammar checker and second language learners of French reveal in noun-adjective agreement constructions and noun-past participle agreement constructions. More specifically, Antidote reaches a higher performance at implementing gender marking than second language learners in both noun-adjective agreement constructions (T (25) = -4.46; p = .000) and noun-past participle agreement constructions (T (25) = -8.18; p = .000). The implementation of gender marking in clitic-past participle agreement constructions, however, did not reach a significant contrast between Antidote and second language learners (T (25) = 1.93; p = .066).

6.6.3 Discussion

Based on previous evaluations of grammar checkers (e.g. Nadasdi & Sinclair 2007; Biesemans 2005) and the implementation rates by native speakers and second language learners of French (Bril 2016), one would expect that language users perform better in implementing gender marking than grammar checkers. This expectation is not borne out by this study: apart from clitic-past participle agreement constructions, grammar checkers reach a higher performance at implementing gender marking than native speakers and second language learners of French. The reason for this unexpected finding can be sought in the fact that in
previous studies (Biesemans 2005; Nadasdi & Sinclair 2007) the outcomes did not differentiate between different types of adjectival agreement. In Biesemans (2005) for instance, number and gender inflection errors were analyzed within the nominal group (e.g. determiner-noun and adjective-noun agreement constructions) without further specification of the type of agreement construction. In a similar vein, no specification was made for the type of construction in which past participle agreement or adjectival agreement takes place (Nadasdi & Sinclair 2007).

In the present chapter, however, the correction rates were specified per type of agreement construction. An interesting observation is the fact that all grammar checkers tested in this study, seem to struggle with clitic-past participle agreement constructions. For this condition only, native French language users performed better than digital grammar checkers. Here, I believe that this may be due to the fact that the use of a clitic pronoun is a ‘Last Resort strategy’. Following Avram & Coene (2008), such a Last Resort strategy allows the clitic to spell out the features it has copied from the R-expression mentioned earlier in the discourse. Last Resort elements are costly operations that involve a high level of computational complexity. This may explain among others why they appear relatively late in native language acquisition. For similar reasons, the processing of anaphoric reference may pose problems for grammar checkers who need to identify the referent in a non-local domain. It has been shown that most grammar checkers are capable of identifying grammatical features of words in one and the same sentence (e.g. Teixeira Martins et al. 1998). Against this background, the effect of the size of the syntactic domain on the performance of digital grammar checkers seems to negatively affect the expression of person features on clitic pronouns.

The results may have important implications for the educational practice, especially with respect to second language teaching. The accuracy of gender marking by French grammar checkers is clearly not uniform across syntactic constructions and grammar checkers. Teachers using a grammar checker in form-focused writing instructions, may need to decide on which grammar checker to use depending on the particular grammatical condition they are teaching and the target population. For native speakers of French, Antidote and Scribens are very effective when it comes to correcting written gender marking in noun-adjective and noun-past participle agreement constructions. Within this context, native speakers receive corrective feedback to reflect and edit their written productions (cf. Vernon 2000). However, these very same grammar checkers might not be as effective when it comes to teaching gender marking in clitic-past participle agreement constructions.

Concerning grammar checkers for second language learners of French, all grammar checkers revealed to be less effective than those developed for native speakers. Specifically, BonPatron did not do any better than second language
learners themselves on gender agreement. Therefore, BonPatron does not provide the appropriate corrective feedback to improve the learners’ awareness of gender marking. In a similar vein, Antidote does not provide second language learners the appropriate corrective feedback when it comes to gender agreement in clitic-past participle agreement constructions. Yet, to improve second language learners’ awareness of written gender marking in noun-adjective and noun-past participle agreement constructions, Antidote might be effective.

6.7 Conclusion

In this study I conducted an experiment aiming to investigate whether digital grammar checkers reach a higher performance at implementing gender marking than native speakers and second language learners of French. More specifically, I focused on noun-adjective, clitic-past participle and noun-past participle agreement constructions. The results showed that grammar checkers developed for native speakers of French (i.e. Antidote and Scribens), outperformed language learners in noun-adjective and noun-past participle agreement constructions. However, in clitic-past participle agreement constructions, digital grammar checkers could not be of assistance to native speakers of French.

With respect to grammar checkers developed for second language learners of French (i.e. Antidote and BonPatron), no difference between BonPatron and second language learners was observed with respect to the correct implementation of gender marking in all three types of gender agreement constructions tested in this experiment. Antidote, however, did reach a higher performance than second language learners in noun-adjective and noun-past participle agreement constructions, but not in clitic-past participle agreement constructions.

I may thus conclude that Antidote and Scribens might be of help to native speakers and second language learners by improving their awareness of written gender marking in noun-adjective and noun-past participle agreement constructions. BonPatron is clearly less effective for this purpose.