

Plus d'erreurs de grammaire ni de typographie avec Grammalecte

Si vous utilisez un traitement de texte avec des élèves, vous avez sûrement déjà entendu cette phrase « Il n'y a pas d'erreurs car ce n'est pas souligné. » En effet, trop souvent, seul le correcteur orthographique est utilisé. Et comme son nom l'indique, il ne corrige que l'orthographe. Si vous voulez que vos élèves (et même les plus grands) questionnent leurs productions, une petite, que dis-je, une grande extension deviendra vite indispensable : Grammalecte. Laissons Olivier nous en dire un peu plus.



Bonjour Olivier, j'ai l'habitude de dire que Grammalecte est une extension qui permet d'apprendre de ses erreurs. Peux-tu nous la présenter ?

Grammalecte est un correcteur grammatical dédié à la langue française. Pour l'instant, il n'existe que pour LibreOffice et OpenOffice, mais j'ai lancé une campagne de financement pour porter l'application dans Firefox et Thunderbird.

Le but du programme, c'est bien sûr de signaler les erreurs grammaticales, mais selon le principe suivant : le moins de fausses alertes possible, car les faux positifs irritent et distraient inutilement les utilisateurs. Ce n'est pas facile à faire, car dans la langue française il y a beaucoup d'incertitudes et les confusions possibles sont innombrables. Songez par exemple que l'adjectif « évident » est aussi une forme verbale du verbe « évider » et vous aurez une idée du genre de

difficultés auxquelles il faut faire face. J'en ai parlé dans un long billet sur LinuxFR, alors je préfère ne pas me répéter ici.



Exemples d'erreurs reconnues par Grammalecte 1



Exemples d'erreurs reconnues par Grammalecte 2

Entendre que Grammalecte permet d'apprendre de ses erreurs me fait plaisir, car il n'est pas toujours facile de faire en sorte que le message d'erreur soit instructif. La place est limitée, et parfois l'imagination fait défaut pour écrire un message à la fois simple et instructif. Par ailleurs, les explications ne sont pas toujours comprises (tout le monde ne sait pas ce qu'est un COD ou un participe passé). Les exemples, ce n'est pas toujours clair. Les messages trop longs ne sont probablement pas toujours lus. Et, pour des raisons techniques, il n'est pas toujours possible d'être explicite. Il y a encore du progrès à faire sur ce point. Si je le peux, je place un hyperlien vers une page web plus complète, mais les pages web sont parfois longues et les explications ne concernent pas toujours spécifiquement l'erreur concernée. Mais il est vrai que, contrairement à Word (qui ne fournit qu'une correction sans indication), Grammalecte tente souvent d'expliquer. Car le meilleur moyen d'éviter les erreurs grammaticales, c'est d'enseigner petit à petit à l'utilisateur à ne plus en faire. Le meilleur service que puisse rendre un correcteur grammatical, c'est de devenir de moins en moins utile. Mais il le sera toujours à cause des erreurs d'inattention que même les plus doués font.

Pour aider l'utilisateur à s'y retrouver dans la langue française, il y a deux outils :

- le « lexicographe », qui, avec un clic droit sur n'importe quel mot, renseigne sur sa nature grammaticale : un nom, un adjectif, participe passé, un verbe, un article, etc.
- le conjugueur, qui est, lui aussi, accessible avec un simple clic droit sur

n'importe quel verbe.

The screenshot shows the 'Grammalecte · Conjugueur' window. At the top, there is a search bar and a 'Conjuguer' button. The main area is divided into several sections for different verb forms:

- Infinitif:** ne pas s'en être allée
- Participe présent:** ne s'en étant pas allée
- Participe passé:** allé, allés, allée, allées
- Impératif:** **Passé**
ne t'en sois pas allée
ne nous en soyons pas allées
ne vous en soyez pas allées
- Indicatif:**
 - Passé composé:** je ne m'en suis pas allée, tu ne t'en es pas allée, elle ne s'en est pas allée, nous ne nous en sommes pas allées, vous ne vous en êtes pas allées, elles ne s'en sont pas allées
 - Plus-que-parfait:** je ne m'en étais pas allée, tu ne t'en étais pas allée, elle ne s'en était pas allée, nous ne nous en étions pas allées, vous ne vous en étiez pas allées, elles ne s'en étaient pas allées
 - Passé antérieur:** je ne m'en fus pas allée, tu ne t'en fus pas allée, elle ne s'en fut pas allée, nous ne nous en fûmes pas allées, vous ne vous en fûtes pas allées, elles ne s'en furent pas allées
 - Futur antérieur:** je ne m'en serai pas allée, tu ne t'en seras pas allée, elle ne s'en sera pas allée, nous ne nous en serons pas allées, vous ne vous en serez pas allées, elles ne s'en seront pas allées
- Subjonctif:**
 - Passé:** je ne m'en sois pas allée, tu ne t'en sois pas allée, elle ne s'en soit pas allée, nous ne nous en soyons pas allées, vous ne vous en soyez pas allées, elles ne s'en soient pas allées
 - Plus-que-parfait:** je ne m'en fusse pas allée, tu ne t'en fusses pas allée, elle ne s'en fût pas allée, nous ne nous en fussions pas allées, vous ne vous en fussiez pas allées, elles ne s'en fussent pas allées
- Conditionnel:**
 - Passé (1^{re} forme):** je ne m'en serais pas allée, tu ne t'en serais pas allée, elle ne s'en serait pas allée, nous ne nous en serions pas allées, vous ne vous en seriez pas allées, elles ne s'en seraient pas allées
 - Passé (2^e forme):** je ne m'en fusse pas allée, tu ne t'en fusses pas allée, elle ne s'en fût pas allée, nous ne nous en fussions pas allées, vous ne vous en fussiez pas allées, elles ne s'en fussent pas allées

At the bottom left, there is a note: "De nombreux verbes n'ont pas encore été étiquetés et contrôlés. Pour ceux-ci, les options "usage pronominal" et "temps composés" sont désactivées."

Ce n'est pas beaucoup par rapport à ce que font des logiciels comme Cordial et Antidote, mais c'est bien mieux que ce que fait Word.

Les correcteurs du Framablog et de Framabook me soufflent également que c'est un allié particulièrement efficace pour les « typo nazis »...

Oui, j'espère qu'il l'est, attendu que c'est avant tout pour des questions de typographie que j'ai commencé ce logiciel. Grammalecte est en effet assez strict

sur ce chapitre. Au tout début, la décision de signaler les apostrophes droites avait beaucoup surpris certaines personnes, mais ça me semblait parfaitement normal. J'ai finalement mis cette règle en option pour ceux que ça gênait le plus. Grammalecte peut paraître pointilleux pour beaucoup de gens.

Mais pour soulager l'utilisateur des fastidieuses corrections typographiques, le logiciel possède un outil, appelé « formateur de texte », capable d'automatiser en quelques clics la correction des multitudes d'erreurs typographiques. Il peut par exemple :

- supprimer les espaces surnuméraires en fin de ligne, entre les mots, avant les virgules, etc.

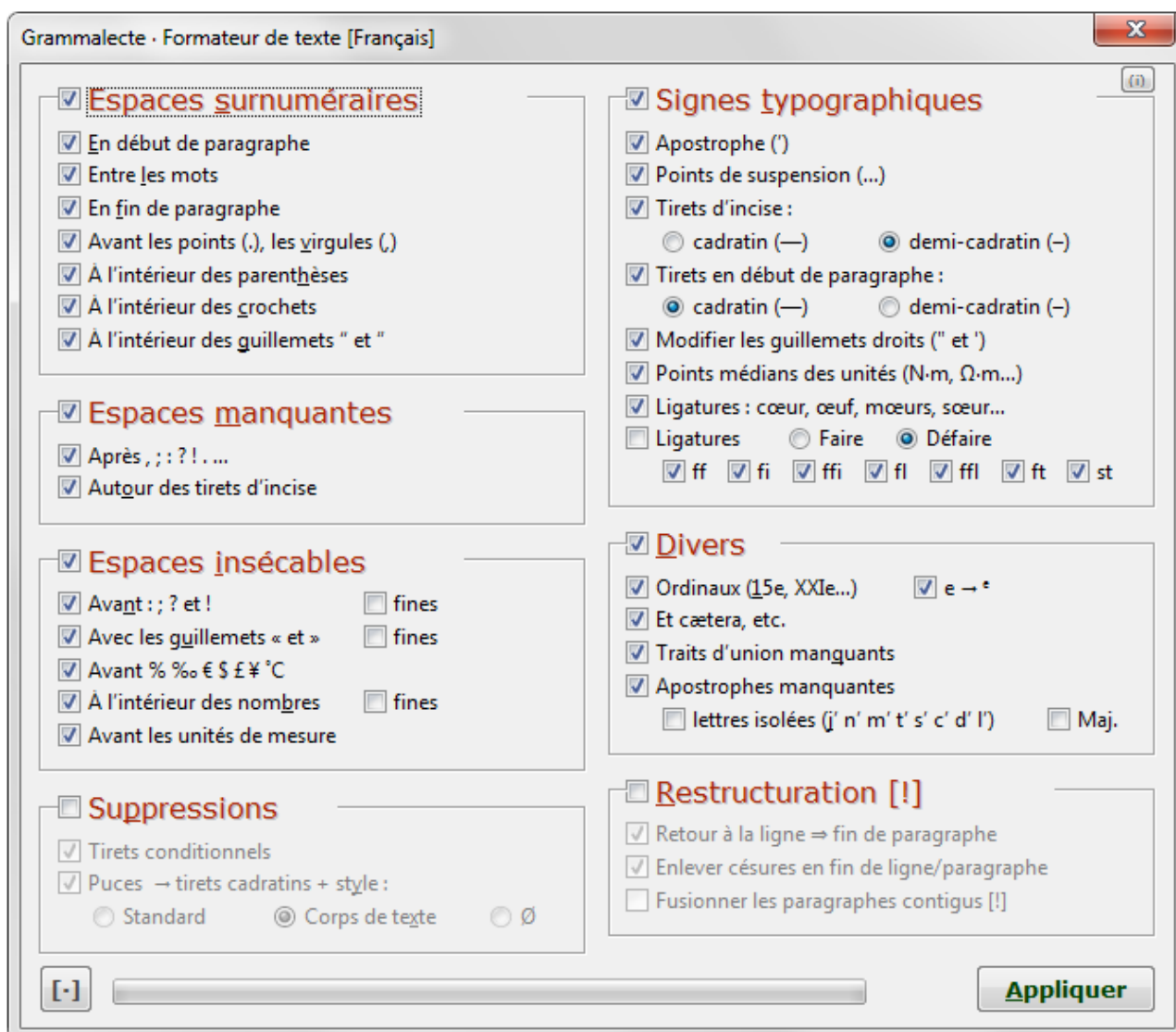
- ajouter les espaces insécables là où elles sont requises,

- transformer les apostrophes droites en apostrophes typographiques,

- placer des tirets cadratins pour les dialogues,

et toutes sortes d'autres choses pénibles à faire manuellement, une par une.

C'est très utile quand on récupère des textes mal formatés sur le Net, car ça fait économiser un temps considérable de mise en forme.



Le formateur de texte de Grammalecte

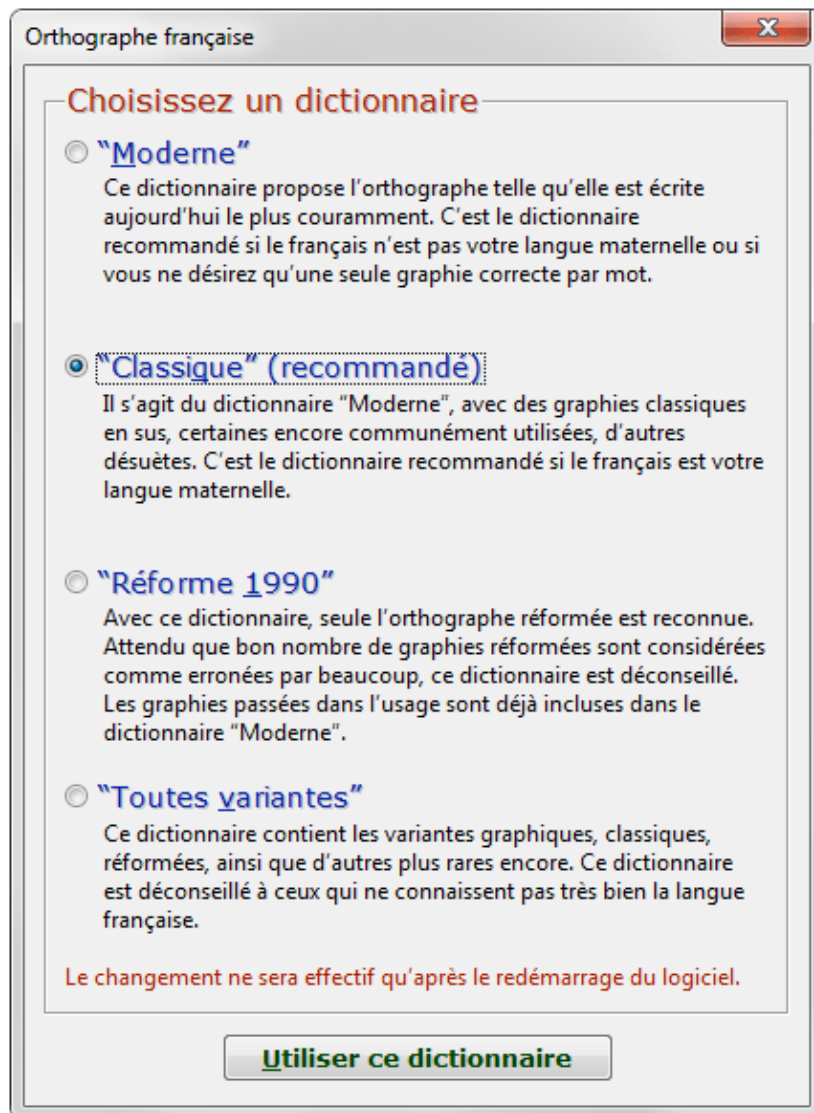
Cela dit, Grammalecte n'intégrera pas tout à fait les mêmes règles de contrôle typographique dans Firefox, attendu que, dans ce contexte, certaines seront plus une gêne qu'une aide. Ce qui sera supprimé ou modifié reste encore à déterminer. Par exemple, il est possible que je change ou supprime certaines règles sur les espaces.

De manière générale, peux-tu nous présenter les personnes derrière Grammalecte ? Tu travailles seul ?

Alors, oui, je travaille seul sur le moteur interne de Grammalecte. Mais ça ne signifie pas que je sois le seul à avoir travaillé dessus, indirectement ou directement. Le logiciel est un dérivé de Lightproof, un correcteur grammatical minimaliste (d'où son nom, qu'on pourrait traduire par « vérificateur léger ») écrit

pour LibreOffice par un Hongrois, mais qui est peu utilisé en raison du manque de ressources lexicales. Ce correcteur fait appel à Hunspell, le correcteur orthographique, mais la plupart des dictionnaires orthographiques n'étant pas grammaticalement étiquetés, son potentiel est limité et il sert surtout pour des corrections basiques ou typographiques. D'ailleurs, tel quel, il ne pouvait être d'une grande utilité pour le français, même avec un dictionnaire étiqueté, c'est pourquoi il a fallu que je triture le code pour pouvoir en faire un correcteur plus puissant (et moins léger). Mais je ne blâme pas Lightproof d'être léger et rudimentaire. Au contraire, ça m'a permis de mettre le pied à l'étrier et de constater que beaucoup de choses n'étaient pas si compliquées à faire. Ensuite, peu à peu, j'ai commencé à réfléchir à des choses plus complexes et à avoir des idées plus vastes que ce que j'avais imaginé faire en premier lieu.

Cependant il y a pas mal d'autres personnes qui ont travaillé et travaillent sur la base indispensable de Grammalecte : le dictionnaire orthographique grammaticalement étiqueté. Ça peut paraître anecdotique, mais gérer un dictionnaire c'est une tâche qui requiert un temps considérable et certains contributeurs y ont consacré une énergie qui méritent toute votre estime. Il y a quelques années, j'avais confié l'administration du dictionnaire à d'autres personnes, ce qui m'a permis d'avoir du temps pour améliorer le site web dédié à l'amélioration du dictionnaire et surtout pour concevoir les premières versions du correcteur grammatical. Gérer une base lexicale, c'est très loin d'être négligeable, c'est pourquoi Grammalecte et LanguageTool utilisent la même ressource. L'un des plus gros contributeurs au dictionnaire, c'est d'ailleurs le mainteneur de la partie française de LanguageTool. Quant aux autres, c'est une petite poignée de passionnés très investis ou de simples passants qui avaient besoin qu'on ajoute certains mots au dico. Mais je n'ai jamais demandé de comptes à quiconque, je ne sais pas qui ils sont. Un pseudo, une adresse e-mail, parfois un nom, c'est tout ce que je sais d'eux. À présent, il y a beaucoup moins de contributeurs, ce projet connaît un certain ralentissement. Il faut dire que le dictionnaire est bien plus fourni qu'autrefois, même s'il y a sans doute encore beaucoup à faire pour les domaines spécialisés comme la médecine, la biologie ou la chimie. Les nouveaux mots qu'on ajoute maintenant concernent surtout les sciences, ou bien des vieilleries peu utilisées.



4 dictionnaires orthographiques proposés

Quant à moi, on pourrait croire que je suis un passionné d'orthographe et de grammaire, et que je m'amuse à faire des concours de dictée. Pas du tout. Pour tout dire, la grammaire ne m'intéresse que parce que je conçois un correcteur grammatical, c'est tout. Ce que j'aime avant tout, ce sont les livres, la littérature et l'informatique. Il y a une douzaine d'années, je récupérais sur le Web pas mal de textes anciens introuvables en librairie, et je les mettais en forme pour mon usage.

En 2005, quand j'ai découvert l'existence d'OpenOffice.org, j'ai immédiatement été impressionné par Writer, qui permettait de concevoir des textes de manière bien plus cohérente et propre que Word. Par ailleurs, Word me fâchait par son incapacité à relire correctement les anciens documents conçus avec lui (j'ai commencé avec Word 6). Il fallait souvent refaire certaines mises en page. Et le format binaire ne permettait guère de retrouver ses petits si le document était

corrompu. La qualité de Writer et le format ouvert de documents sont les raisons pour lesquelles j'ai migré vers OOo.

Mais il y avait quand même un aspect de Writer qui était en deçà de son concurrent : le correcteur orthographique. Il était très lacunaire. C'est la raison pour laquelle j'ai commencé à m'intéresser à la question. Finalement, j'ai d'abord cherché à remplir mes besoins d'utilisateur. En 2006-2007, j'ai retroussé mes manches et j'ai d'abord amélioré le dictionnaire, j'ai repris les différentes versions disponibles sur le Net, et j'ai conçu un site web pour recevoir les propositions des utilisateurs. En 2008, j'ai fini par réécrire toutes les règles du dictionnaire pour normaliser les données (avant ça, c'était vraiment le bordel), et j'ai posé des étiquettes grammaticales dessus, avec l'idée que ça servirait un jour à celui ou celle qui aurait l'idée saugrenue de faire un correcteur grammatical, puis j'ai recréé le site web (parce que le premier était mal foutu). À l'époque, je n'avais pas du tout l'intention de concevoir un correcteur grammatical. Je ne me sentais pas assez fou pour me lancer là-dedans.

J'avais bien sûr essayé LanguageTool, mais il ne me convenait pas du tout, car il y avait vraiment trop de faux positifs. En 2010, j'ai tenté d'améliorer LanguageTool pour le français, mais j'y ai finalement renoncé à cause d'une histoire de typographie et de mon désamour pour Java et le XML. C'est alors que j'ai découvert Lightproof, capable d'interroger les dictionnaires dans Writer. Tiens, tiens, intéressant, me suis-je dit, et si je faisais un petit test ? Au commencement, j'ai pas mal galéré pour diverses raisons plus ou moins complexes, mais j'ai eu assez vite un correcteur typographique auquel j'ai ajouté quelques règles simples concernant la grammaire. Encore une fois, je ne faisais que satisfaire mes propres exigences d'utilisateur. Puis, comme ça a plu, j'ai continué à améliorer le moteur interne du correcteur, peu à peu, en ajoutant des mécanismes plus complexes et en polissant peu à peu les rugosités du logiciel.

Cela dit, cela va vous paraître bizarre, mais j'éprouve un doute de nature philosophique sur la pertinence de concevoir un correcteur grammatical. Je ne juge pas qu'un correcteur grammatical soit inutile, mais à force de plancher sur la question, la grammaire française a commencé à me paraître inutilement compliquée et incohérente. On écrit par exemple : Je commence, tu commences, il commence. Mais on écrit : Je finis, tu finis, il finit. Il est étonnant qu'on juge utile de distinguer la deuxième personne du singulier au premier groupe, mais au deuxième groupe on préfère distinguer la troisième personne du singulier. À

l'impératif, dans le deuxième groupe, la graphie de la deuxième personne du singulier est la même que celle à l'indicatif (« finis »), mais au premier groupe la graphie de la deuxième personne du singulier est différente à l'indicatif (« commences ») et à l'impératif (« commence »), ce qui trompe d'ailleurs beaucoup de monde. Pire : au deuxième groupe, la première et la deuxième personnes du singulier (« finis ») ont la même graphie qu'un participe passé. Encore une belle occasion de semer la confusion.

On va me rétorquer que c'est notre « héritage », que ça vient des origines de la langue, que l'étymologie, c'est important. Oui, mais c'est un argument creux. D'abord, qui connaît l'origine de la variation des graphies des conjugaisons ? Pas grand-monde, je parie. Un coup d'œil sur la question sur Wikisource. Les anciens étaient-ils parfaitement logiques et cohérents ? C'est très discutable. La langue n'a pas évolué de manière uniforme. Un autre exemple : habiter vient du latin *habitare*, c'est pour ça qu'il y a un h au début du mot. *Habitare* dérive lui-même du mot *habere*, qui signifie avoir. Ah, tiens, le h a disparu sur avoir. Autrement dit, l'histoire préserve et altère les graphies très diversement. Pourtant, préserver le h sur avoir aurait été bien utile, car ça éviterait que certaines formes verbales de avoir soient identiques ou semblables à d'autres mots sans rapport avec ce verbe, comme a, as, avons, aura, ais (qui n'est pas une forme conjuguée de avoir). Absurde de rajouter h à avoir, pensez-vous ? Pourtant, on a autrefois ajouté des lettres aux graphies des mots pour mieux les distinguer. On a ajouté un d à pied (parce que ça vient de *pedis*), un g à doigt (parce que ça vient de *digitus*), etc. Personnellement, j'aimerais bien que avoir retrouve son h...

Le français est plein de confusions, d'ambiguïtés et de bizarreries. Il y a tellement de choses à retenir. Savez-vous à quoi ressemble la somme de la connaissance sur la grammaire française ? À un pavé de 1600 pages écrit en petites lettres qui s'appelle *Le Bon Usage de Grevisse* (au format poche, ça ferait plus de 3500 pages, je pense). Et encore, on n'y trouve pas tout.



Couverture de la 15^e
édition du Bon usage

Récemment, une de mes amies s'indignait que ses enfants dussent apprendre par cœur les pluriels irréguliers de caillou, genou, hibou, etc. Pourtant ces mots ne

dérivent pas de mots plus anciens contenant des x. Ce x n'est dû qu'à une écriture abrégée employée il y a longtemps, où un X remplaçait "us". Au lieu d'écrire chous, certains écrivaient choX. (référence sur Wikipédia). Personnellement, il me paraît bien plus grave de confondre "on" et "ont", "à" et "a", "se" et "ce", que de se tromper sur le pluriel de caillou ou d'écrire "tu commence". Mais le français est si plein de choses à retenir qu'on voit régulièrement des gens ne pas se tromper sur des questions accessoires et écrire des phrases dont la syntaxe fait mal aux yeux.

Voilà pourquoi j'éprouve un doute sur la pertinence de concevoir un correcteur grammatical en l'état des choses. Je crains d'aider à figer une langue dans toutes ses incohérences et ambiguïtés. (Mais ceux qui vendent des correcteurs y trouvent probablement leur compte.) Il me semblerait plus utile que les experts se réunissent pour concevoir un français avec le moins possible d'incohérences, d'ambiguïtés et d'irrégularités. Je comprends que c'est pour certains un scandale de toucher à la langue. C'est pourtant ce que font souvent ceux qui créent des langages de programmation, quand ils veulent les améliorer. Ils modifient la syntaxe, ils ajoutent du vocabulaire, font les modifications qu'ils jugent utiles. Résultat : un langage plus lisible, moins ambigu et plus cohérent.

De toute façon, si l'on ne fait rien, le français évoluera. De manière incohérente probablement, comme jusqu'à présent. Et on appellera ça notre culture.

Mais rassurez-vous, je n'ai aucunement l'intention d'imposer mes idées, et le correcteur grammatical essaiera de faire respecter les règles actuelles. □

Actuellement, Grammalecte est disponible pour les suites bureautiques libres (LibreOffice, Apache OpenOffice, OOo4kids et OOoLight), j'ai cru comprendre que la prochaine étape était de couper le cordon et de l'adapter pour d'autres logiciels.

Avant de répondre à cette question, une remarque sur les suites bureautiques : je ne sais pas du tout ce qui se passe du côté d'OOo4Kids et OOoLight. Je pensais que le développement de ces logiciels avait cessé. Je crois savoir qu'ils utilisent Python 2.6, et je ne fournis plus de nouvelles extensions pour cette version de Python depuis assez longtemps. C'est déjà assez contrariant de fournir une version utilisable par OpenOffice (qui n'intègre que la version 2.7 de Python). Le problème, ce n'est pas OpenOffice, c'est cette version de Python dont le module d'expressions régulières est un peu bogué, ce qui rend Grammalecte moins efficace et génère parfois des faux positifs indépendants de mon contrôle. En fait,

je teste tout avec LibreOffice, puis l'extension est convertie pour OpenOffice.

Mais, oui, la prochaine étape, c'est de *désimbriquer* Grammalecte de l'écosystème LibreOffice/OpenOffice, notamment pour pouvoir greffer le correcteur grammatical sur Firefox et Thunderbird. Ça fait longtemps que j'y songe, mais il y a pas mal de prérequis à cela. Il faut refondre et réorganiser une très grosse partie du code, transformer toutes les données, optimiser pas mal de choses, écrire les fonctionnalités qui manqueront après s'être détaché de LibreOffice/OpenOffice, améliorer certains points de la correction grammaticale, convertir en JavaScript (le langage de programmation des navigateurs), concevoir une interface adaptée, et j'en oublie certainement. En bref, il y a des eaux tumultueuses à traverser avant de pouvoir reprendre une navigation sereine. C'est pourquoi j'ai monté une campagne de financement participatif pour pouvoir m'y consacrer sereinement.

The image shows a crowdfunding page for Grammalecte. The main header features the project name "Grammalecte" and the subtitle "Correcteur grammatical open source". A progress bar indicates that 51% of the 15,000 € goal has been reached, with 7,676 € collected. The page shows 20 days remaining and a "Contribuer" button starting at 5 €. The creator, Olivier R., is listed as being from Lorraine, with 1 project created and 2 projects supported. The main content area displays the Grammalecte logo and the text "GRAMMALECTE CORRECTEUR GRAMMATICAL OPEN SOURCE".

Et tu ne comptes pas t'arrêter aux logiciels mozilliens. Quel est ton objectif ultime ?

Produire une extension pour Firefox et Thunderbird fait déjà partie de mon but « ultime », c'est déjà à mes yeux une très importante finalité en elle-même, mais en effet ce n'est pas tout.

Séparer Grammalecte de Writer a aussi pour dessein de bâtir une application

autonome, un serveur capable de renvoyer les erreurs à toute autre application qui lui transmettrait du texte, ce qui permettrait à ces applications de proposer des corrections grammaticales. Charge à elles de concevoir l'interface. Après, idéalement, j'aurais aimé revoir complètement la gestion des ressources lexicales, refaire le site web du dictionnaire de fond en comble, mais ce n'est pas indispensable et ça demanderait beaucoup de travail. Alors j'ai préféré être plus raisonnable en proposant de concevoir divers outils annexes.

Parmi ceux-ci, il y a notamment un assistant pour proposer de nouveaux mots à la base de données en ligne, pour simplifier toute la procédure. Il y a aussi un outil pour détecter les répétitions et compter les mots en les regroupant par lemme. Je prévois aussi d'améliorer le « lexicographe » afin de fournir sur les mots toutes les données dans la base, comme le champ sémantique auxquels ils appartiennent, leur indice de fréquence, leur origine étymologique et toute information potentiellement utile.

En fait, toutes ces choses (les extensions, le serveur et les outils annexes) sont plus liées qu'il n'y paraît. Elles ne sont séparées dans la campagne de financement que pour que celle-ci ait plus de chances d'aboutir. La véritable finalité, c'est de bâtir un écosystème grammatical libre.

Nous pouvons donc soutenir le développement financièrement. Si certains de nos lecteurs souhaitent t'aider d'une autre manière, comment peuvent-ils faire ?

Le point sur lequel il est possible d'aider, c'est la gestion du dictionnaire qui sert de base lexicale au correcteur. Ce n'est malheureusement pas une tâche très enthousiasmante, car c'est répétitif. Mais ajouter les mots qui manquent, les étiqueter, c'est pourtant indispensable. Quand un mot n'est pas identifié, le correcteur est aveugle. Plusieurs fois, j'ai laissé le rôle d'administrateur à des personnes motivées qui ont fait du très bon boulot. Tout le monde peut participer, et si quelqu'un se sent motivé pour administrer, il suffit d'apprendre comment ça fonctionne, se faire la main sur le système et de savoir grosso modo quelle est la politique suivie.

Quant au code, je préfère travailler seul, question de tempérament, mais quand j'aurai fini la réorganisation du projet et que les tests seront mis en place pour éviter les régressions, je serai plus ouvert à la collaboration.

Traditionnellement, nous laissons le mot de la fin à l'interviewé. Y a-t-il une question que tu aurais souhaité qu'on te pose ?

On ne m'a pas encore posé de questions sur le potentiel futur du correcteur, s'il peut encore beaucoup progresser dans la détection des erreurs.

La réponse est oui, il peut encore progresser de manière significative. Il est difficile de faire des prédictions avec une grande fiabilité, mais je suis optimiste sur la distance que celui-ci peut parcourir avant d'arriver au point où il sera difficile d'améliorer les choses sans revoir de fond en comble son fonctionnement.

Pour l'instant, il existe 929 règles de contrôle (qui recherchent les erreurs) et 535 règles de transformation (qui aident les premières à s'y retrouver dans le texte). Ces règles font énormément de choses, mais je n'ai pas encore implémenté nombre de vérifications, parce que c'est parfois compliqué à faire (il faut tester, refaire, vérifier, refaire, revérifier), mais aussi parce qu'il existe nombre d'erreurs auxquelles je n'ai pas pensé. Concevoir les règles de détection, c'est parfois simple, mais ça requiert parfois aussi de l'inventivité.

Pour l'instant, j'ai assez peu travaillé sur certaines erreurs grossières, comme les confusions entre "sa" et "ça", "on" et "ont", "a" et "à", etc. parce que j'en vois peu dans les textes sur LibreOffice. Ce n'est pas le genre d'erreurs qui me vient automatiquement à l'esprit. En revanche, sur le web, ces erreurs sont bien plus fréquentes, et il faudra que je veille à renforcer les contrôles sur ces confusions qui trahissent une méconnaissance assez grave de la grammaire française. Il existe bien sûr déjà des règles pour signaler ces confusions, mais c'est encore à améliorer.

Sans rien changer aux mécanismes internes, il y a encore beaucoup de choses faisables. Mais j'avance prudemment, car la difficulté ce n'est pas de trouver de nouvelles erreurs à signaler, c'est d'en détecter sans se tromper trop souvent. Comme la « devise » de Grammalecte, c'est d'éviter autant que possible les faux positifs, la montée en puissance se fait à un rythme raisonnable, afin de corriger ce qui peut l'être au fur et à mesure et d'éviter d'être submergé par des signalements intempestifs.

Par ailleurs, à l'avenir, va être mis en place un système de désambiguïsation (cf. l'article sur LinuxFR) qui va rendre l'analyse du texte plus sûre et mécaniquement augmenter le taux de détection.

Ensuite, il n'est pas exclu de créer des mécanismes plus complexes, mais c'est une autre affaire. Grammalecte n'en est pas encore arrivé à ce stade.

- Fonctionnalités de Grammalecte : <http://www.dicollecte.org/grammalecte/outils.php>
- Campagne de financement participatif : <http://fr.ulule.com/grammalecte/>