N° 12 | 2019
Usages, formes et enjeux de la «poésie numérique»

Langue, genre et automatisation

le projet En française dans la texte

Cécile BABIOLE

Anne LAFORET

Édition électronique :
URL : https://komodo21.numerev.com/articles/revue-12/2521-langue-genre-et-automatisation
DOI : numerev_2350
Date de publication : 01/12/2019

Cette publication est sous licence **CC BY-NC-ND** (Attribution - No commercial - No derivatives).

Pour **citer cette publication** : BABIOLE, C., LAFORET, A. (2019) Langue, genre et automatisation. *Komodo 21*, (12). https://doi.org/10.34745/numerev_2350

En française dans la texte est un projet artistique traitant du rapport de la langue française au genre. L'article revient sur la motivation du projet : faire apparaître et dénoncer le sexisme de la langue française, construit historiquement au XVII^{e} siècle, à l'origine d'une invisibilisation des femmes à travers l'institution de la primauté du masculin dans la langue. Il montre comment, aujourd'hui, les outils de traitement automatique du langage prolongent cette discrimination sexiste dont il analyse certains mécanismes qui peuvent passer inaperçus, par exemple les biais de genre dans le traducteur automatique de Google.

Mots-clés :
Genre, internet, algorithme, langue française, sexisme, Féminin

---

Abstract

En française dans la texte is an artistic project dealing with French language and gender. Its outputs are texts which are edited and then automatically translated from common French language into a version of French whose grammatical gender is feminine only. In French, most words (nouns â□□ including those of inanimate objects â□□, adjectives, pronouns, determiners...) are gendered in masculine or feminine forms. This article is based on our experience as artists with that project. We analyze how sexist gender rules have been set up by the Académie française, a preeminent council for matters pertaining to the French language created by Cardinal Richelieu, the chief minister to King Louis XIII in 1635. Since the seventeenth century, those sexist gender rules have persisted. Nowadays, natural language processing tools reproduce and amplify sexist biases, such as gender prediction in Google Translate.

Keywords

gender, internet, algorithm, French language, sexism, feminine

Ce texte part de notre expérience en tant qu'artistes avec le projet *En française dans la texte* [1], un projet critique sur le thème « langue française et genre ». Il tente d'en

analyser le contexte, c'est-à-dire le sexisme de la langue française, un sexisme construit historiquement au XVII[e] siècle et renforcé par les technologies numériques d'aujourd'hui.

Notre réflexion et notre pratique artistique s'inscrivent dans un renouveau des féminismes depuis quelques années, qui s'exprime sous différentes formes. Tout récemment, par exemple, les campagnes sur les réseaux sociaux comme #metoo et #balancetonporc ont largement rendu publique l'ampleur du phénomène de harcèlement sexiste, mais aussi les réticences spécifiquement françaises à les entendre et à agir en conséquence.

Le projet artistique *En française dans la texte* a pour origine notre volonté de lutter contre le paternalisme inscrit au cœur de la langue française, en traduisant « en française », c'est-à-dire entièrement au féminin, des textes provenant de différents horizons, grâce à des algorithmes complétés par des corrections manuelles. C'est ainsi que les traductions perturbent sensiblement les messages originaux, en créant un sentiment d'inquiétante étrangeté proche du trouble que génère l'expression poétique. Cette littérature « en française » produit aussi un certain humour en dévoilant à l'improviste, comme des lapsus, l'origine ou la parenté de certains mots ou expressions, par exemple « il n'aurait pas le culot de venir ici » devient en française « elle n'aurait pas la culotte de venir ici ». Enfin, notre proposition rend obsolète tout le débat autour de l'écriture dite inclusive.

Ce processus de traduction s'inscrit dans la poésie numérique telle que Kenneth Goldsmith la définit dans *L'écriture sans écriture â□□ du langage à l'âge numérique*, c'est-à-dire une poésie ouverte aux multiples régimes d'écriture de l'Internet, une poésie dont les moteurs sont l'appropriation, la copie, la compilation, l'agrégation de matériaux hétérogènes et déjà existants. Une poésie à mille poètes, humains et non-humains (numériques, incluant toutes les opérations conduites par les ordinateurs).

Le programme de traduction en française génère automatiquement une nouvelle prosodie, les phrases en française sont bousculées par des hiatus et des allitérations : *sa amour* (au lieu de *son amour*), *elle la la donne* (au lieu d'*il le lui donne*)… Les mots gagnent une expressivité nouvelle par l'aspect déconcertant et insolite des formes féminines substituées et des néologismes. Ce jeu avec la langue est un acte performatif, un acte de langage, qui agit sur le monde et sème littéralement un « trouble dans le genre [2] ». C'est un des propres de la poésie que de produire des effets de révélation du monde.

1. Genre linguistique : histoire d'une discrimination construite

L'histoire du genre linguistique ne s'est pas faite seulement par l'usage des locuteur-rices, elle est surtout le résultat d'une lutte idéologique (incarnée par l'Académie

française puis prolongée de manière implicite par les GAFA[3] pour imposer la primauté du masculin au détriment du féminin et pour invisibiliser et discriminer les femmes dans l'espace social. C'est l'histoire d'une discrimination construite par des règles et par des techniques.

1.1. Grammaire : introduction de la primauté du masculin au XVII^{e} siècle

La langue française comporte deux genres : tous les substantifs désignant indifféremment des entités animées et inanimées sont soit féminins, soit masculins. La règle de base prévoit que les déterminants, les adjectifs, les participes passés, les pronoms s'accordent en genre et en nombre avec les substantifs auxquels ils se rapportent ou auxquels ils se substituent (par exemple, *un pantalon blanc*, *des chaussettes blanches*).

C'est au XVII^{e} siècle que naît une nouvelle règle qui déroge à la précédente et institue la primauté du masculin sur le féminin. Elle est l'œuvre de l'Académie française (créée par Richelieu en 1635 pour normaliser et perfectionner la langue). En 1647, le grammairien et académicien Claude Favre de Vaugelas la justifie par le postulat suivant : « le genre masculin étant le plus noble, il doit prédominer chaque fois que le féminin et le masculin se trouvent ensemble. » Mais cette nouvelle règle ne fut pas unanimement, ni immédiatement acceptée. Ainsi, Gilles Ménage (autre grammairien du XVII^{e} siècle) rapporte une conversation avec Madame de Sévigné :

> S'informant sur ma santé, je lui dis : Madame, je suis enrhumé. Je la suis aussi, me dit-elle. Il me semble, Madame, que selon les règles de notre langue, il faudrait dire : je le suis. Vous direz comme il vous plaira, ajouta-t-elle, mais pour moi, je croirais avoir de la barbe au menton si je disais autrement [4].

D'ailleurs, en ancien français comme en latin, il était permis d'accorder l'adjectif en genre et en nombre avec le nom qu'il qualifie le plus proche. Cette règle, qui reprend du service aujourd'hui, est nommée règle de l'accord de voisinage, ou de proximité, comme dans la formulation *des sondages et statistiques électorales* au lieu de *des sondages et statistiques électoraux* qui est la forme « correcte » toujours préconisée aujourd'hui par l'Académie. Elle fut d'un usage commun jusqu'à la Révolution française. Racine appliquait la règle de voisinage de manière habituelle, par exemple avec l'alexandrin « mais le fer, le bandeau, la flamme est toute prête [5] » dans *Iphigénie* (1674) ou bien « Armez-vous d'un courage et d'une foi nouvelle [6] » dans *Athalie* (1691).

Malgré tout, la primauté du masculin sur le féminin finit par s'imposer définitivement au XVIII^{e} siècle en vertu de la tautologie : « le genre masculin est réputé plus noble que le féminin à cause de la supériorité du mâle sur la femelle [7] », rappelée par le

grammairien et académicien Nicolas Beauzée en 1767.

Aujourd'hui, même si l'influence de l'Académie s'émousse, le masculin s'impose encore, contre toute logique, dans des expressions comme *quelle heure est-il ?* » au lieu de *quelle heure est-elle ?* qui serait pourtant la forme attendue puisque le mot *heure* est du féminin. Sans parler des cas où le masculin est utilisé à la place de ce qui serait un neutre dans d'autres langues, par exemple, *il fait beau* où le *il* est purement arbitraire…

Tout récemment, l'Académie française continue à raisonner de manière sexiste et erronée en prétendant que « le masculin a valeur générique ou non marquée, le féminin est la forme marquée [8] ». En réalité, comme le fait remarquer Éliane Viennot et ses co-autrices dans l'ouvrage collectif *L'académie contre la langue française* [9], les académiciens confondent masculin et racine d'un mot. Ils pensent sans doute que *coiffeuse* est formé d'après *coiffeur,* alors qu'en fait, les mots dérivent d'une racine commune et non d'une forme masculine à laquelle il faudrait retrancher ou rajouter quelque chose. À partir de n'importe quelle racine, le français est apte à faire des substantifs masculins et féminins, des verbes, des adjectifs… Par exemple : la racine *coiff-* permet de créer : *coiffeur*, *coiffeuse*, *coiffer*… Cette méconnaissance de la langue est étonnante de la part de l'Académie, mais n'est pas si étrange puisque la composition actuelle de cette institution ne comprend aucun.e linguiste, aucun.e agrégé.e de grammaire, aucun.e historien.ne de la langue [10].

1.2. Vocabulaire : dépréciation du féminin – Galant, galante

Pour des raisons qui ne font que traduire le sexisme de la société, le féminin est souvent dépréciatif. C'est fréquemment le cas quand il désigne une qualité ou une fonction. Il suffit de comparer les formes masculines et féminines pour saisir la différence sémantique : *galant*, *galante*, un homme galant est courtois, une femme galante est une prostituée ; *professionnel*, *professionnelle*, un professionnel connaît bien son métier, une professionnelle est une prostituée ; *couturier, couturière*, un couturier est un créateur, une couturière est une petite main, etc. Dans le domaine de la terminologie des métiers et fonctions, jusqu'au XX^{e} siècle, les termes comme *la présidente* ou *la capitaine* ne désignaient pas celle qui exerce la fonction de présidente ou de capitaine, mais l'épouse du président ou du capitaine, c'est-à-dire la subordination au statut marital. En matière de désignations métaphoriques, l'exemple des noms d'oiseaux est flagrant : pour quelques coqs ou paons, combien de poules, poulettes, poules pondeuses, poules mouillées, poules de luxe, mères-poules, cocottes, cailles, oies (blanches), pies (jacassantes), bécasses, dindes ? Toute la basse-cour y passe.

1.3. « Madame le Président » â Invisibilité linguistique des femmes : une invisibilité délibérée

Un des usages fondamentaux du latin, des langues romanes qui en sont issues et en conséquence du français, est de parler des femmes au féminin et des hommes au masculin.

C'est à partir du XVIII[e] siècle que le français est délibérément masculinisé. C'est tout particulièrement visible dans les noms de métiers et de fonctions. Il est à noter que le XVII[e] siècle est la période à laquelle un certain nombre de femmes de lettres comme Mmes de Lafayette, de Scudéry, de Sévigné, de Villedieu, Deshouillères, de Coligny de la Suze et bien d'autres, commencent à accéder à la notoriété, malgré l'interdiction qui est faite aux femmes de poursuivre des études secondaires et supérieures. Cette reconnaissance nouvelle des femmes dans le monde des lettres, jusqu'alors chasse-gardée exclusive des hommes, explique le refus des académiciens de tolérer désormais les termes féminins désignant les activités qu'ils ne voulaient pas partager avec les femmes : « [i]l faut dire cette femme est poète, est philosophe, est médecin, est auteur, est peintre ; et non poétesse, philosophesse, médecine, autrice, peintresse, etc. [11] » écrivait ainsi Andry de Boisregard (*Réflexions sur l'usage présent de la langue françoise, 1689*). Si les académiciens prirent la peine de les interdire, c'est que ces formes étaient bien en usage courant jusqu'au XVII[e] siècle. Pour preuve, la langue juridique conserve encore aujourd'hui des traces de cet usage ancien des formes féminines. Par exemple, les termes *demanderesse* ou *défenderesse* ont réussi à passer outre l'interdiction de l'Académie, sans doute parce que la langue juridique est encore plus conservatrice que cette dernière.

En France, la lutte contre l'invisibilité linguistique des femmes est récente : il faut attendre 1984 pour qu'une commission relative à la féminisation des noms de métiers [12] soit mise en place sous l'impulsion de Yvette Roudy, alors ministre des Droits de la femme (de 1981 à 1986). L'instigatrice de cette commission était Anne-Marie Houdebine, sa présidente, Benoite Groult.

Malgré les préconisations ministérielles favorisant la féminisation dans les textes officiels, les résistances restent fortes. En 2014, l'Académie française continue de refuser des formes telles que *professeure*, *auteure*, *ingénieure*, *procureure* etc, qu'elle considère comme de « véritables barbarismes [13] » alors que ces formes sont la norme au Québec.

À l'Assemblée nationale, en octobre 2014, alors que la présidente de séance, Sandrine Mazetier (députée PS), demande au député LR Julien Aubert de respecter la présidence et le règlement de l'Assemblée en l'appelant « Madame la présidente », il persiste à employer les mots « Madame le président ». Sanctionné par le bureau de l'Assemblée et privé à ce titre d'un quart de son indemnité parlementaire pour un mois (soit 1378 euros), il dépose un recours en justice en janvier 2015.

1.4. Le langage épicène : le « péril mortel » pour les Immortels

Plusieurs propositions de langue épicène, c'est-à-dire permettant de nommer les deux genres sans discrimination, ont vu le jour récemment, comme l'emploi de parenthèses, mais cette forme discrimine la marque du féminin en la plaçant entre parenthèses, c'est-à-dire en position inférieure ; à l'opposé, l'emploi du *E* majuscule (*motivéEs*), est la forme la plus militante, utilisée dans des contextes féministes et LGBTQI [14], mettant en avant la marque du féminin ; l'emploi du trait d'union ; l'emploi du point médian,

semblable au trait d'union mais plus discret, ainsi que nous l'utilisons dans cet article.

Ces pratiques typographiques pour démasculiniser la langue écrite n'ont pas encore abouti à une norme commune. De plus, elles provoquent des réactions disproportionnées de la part de nombreuses personnalités, en particulier des ministres, des journalistes et des membres de l'Académie française qui déclarent en 2017 « devant cette aberration "inclusive", la langue française se trouve désormais en péril mortel [15] ».

2. Renforcement de la discrimination par la technologie

Les formes de sexisme et de discriminations à l'œuvre dans les relations sociales et inscrites dans la langue sont conséquemment présentes dans les outils numériques utilisés au quotidien, en particulier sur Internet : moteurs de recherches, plateformes et réseaux sociaux…

2.1. Les nouvelles formes de misogynie en ligne

Le cyber-harcèlement sexiste est un fait avéré, même s'il n'est pas reconnu par les instances modératrices des plateformes, ni par la justice. Selon des études récentes citées par le collectif Féministes contre le cyberharcèlement :

> Une jeune femme sur cinq déclare avoir été victime d'au moins un cyberharcèlement d'ordre sexuel depuis l'âge de quinze ans et une adolescente sur quatre déclare être victime d'humiliations et de harcèlement en ligne concernant son attitude (notamment sur son apparence physique ou son comportement amoureux ou sexuel). Ce que ces études ne disent pas en revanche, c'est que les femmes qui subissent d'autres formes de discriminations en raison d'un handicap, de leur origine, leur couleur de peau, leur religion, leur orientation sexuelle, leur identité de genre sont encore plus exposées à ce type de violences [16].

Contrairement à l'espace physique de la rue, du lieu de travail ou du lieu de vie, les agressions misogynes en ligne prennent la forme de textes écrits, souvent associés à la propagation d'informations et de photographies personnelles sans consentement (*doxing*).

Insultes et autres menaces fusent facilement, pouvant émaner d'un très grand nombre de personnes à la fois, à travers des campagnes de harcèlement coordonnées. Ces actes misogynes sont minimisés autant par les plateformes publiques d'échanges et de discussions sur lesquelles elles ont lieu, que par la police et la justice, et plus

généralement l'opinion publique, et par conséquent sont perpétrés souvent en toute impunité. Par exemple, malgré de nombreuses polémiques au cours des années, une plateforme comme Twitter n'a pas pris de disposition pour empêcher le cyber-harcèlement de ses utilisateur·rices, sauf en de très rares cas, et de manière souvent ponctuelle.

Ce fut le cas par exemple avec le bannissement de Twitter en janvier 2016 de Milo Yiannopoulos, ex-journaliste (du site américain néoconservateur Breitbart News) pour sa campagne de harcèlement contre l'actrice africaine-américaine Leslie Jones à l'occasion de la sortie de la nouvelle version de la comédie fantastique *S.O.S Fantômes* avec un casting féminin.

Cette parole misogyne se retrouve à tous les échelons de la société, ainsi le candidat puis président Trump adopte sciemment une façon de communiquer sur Twitter sexiste et raciste. Ses tweets envers Barack Obama alors président et Hillary Clinton, sa rivale dans l'élection présidentielle américaine, en sont des exemples [17]. Donald Trump sait pertinemment qu'ils seront « retweetés » et largement cités dans la presse, autant par ses partisans que par ses opposants outrés par sa façon de s'exprimer.

Pour attirer l'attention sur le fait que les pratiques de harcèlement sont courantes et par la même occasion attirer les trolls [18], Sarah Nyberg a crée le compte @arguetron [19], un bot qui envoie toutes les 10 minutes des affirmations progressistes et qui reçoit donc des tonnes d'immondices sexistes.

En ce qui concerne le cyber-harcèlement en France, il semble que peu de plaintes aboutissent. Par exemple, la militante féministe Caroline de Haas a porté plainte en février 2016 pour diffamation et injures publiques et provocation au viol suite à des milliers de messages sur Twitter et Facebook en réaction à son tweet le 7 janvier 2016 qui disait : « Ceux qui me disent que les agressions sexuelles en Allemagne sont dues à l'arrivée des migrants : allez déverser votre merde raciste ailleurs ». Un an plus tard, le verdict tombe, un non-lieu car le juge a annoncé ne pas pouvoir identifier les personnes, le procureur aurait pu imposer à Facebook et Twitter de fournir les adresses IP mais ne l'a pas fait.

2.2. Les algorithmes sont-ils sexistes ?

Le langage utilisé en ligne charrie donc stéréotypes et préjugés qui concernent le genre (mais aussi la race, l'orientation sexuelle, la classe sociale, etc). Ces mots, phrases, documents deviennent la matière première traitée par les algorithmes en ligne. Ces textes deviennent des données qui sont classées, triées, agrégées afin d'être utilisées sous de multiples modes.

Pour exploiter sur le plan sémantique le vaste volume de données produites chaque jour (55 000 gigabytes par seconde de trafic internet selon le site Internet Live Stats [20]), des technologies ont dû être inventées. L'une d'elles, particulièrement en expansion actuellement, est le « *machine learning* » (ou apprentissage automatique), qui est un

des domaines de l'intelligence artificielle.

2.2.1. Un exemple de méthode de machine learning : *word embeddings*

Afin de créer du sens à partir des *mégadonnées* ou *big data*, c'est-à-dire ces très larges ensembles de données hétérogènes non structurées (archives numérisées, données scientifiques, commentaires et autres textes issus de *chats* et des réseaux sociaux…), il faut développer des méthodes beaucoup plus complexes que les algorithmes utilisés pour des données structurées (déjà répertoriées et classées dans des bases de données). Des modèles sont alors conçus afin de réduire cette complexité et la transformer en unités d'informations utilisables de manière opérationnelle par les machines.

Cet apprentissage automatique, terme français pour traduire l'expression *machine learning*, renvoie à des techniques variées où l'intervention humaine est plus ou moins présente. On parle d'apprentissage supervisé ou non-supervisé en fonction de l'implication humaine dans le processus. La supervision (humaine) concerne la phase d'entraînement des machines afin de constituer des modèles à partir d'exemples validés. À l'inverse, dans le cas d'un apprentissage non supervisé, la nature des données d'entraînement n'est pas connue, l'algorithme permettant l'émergence d'une structure depuis les données mêmes.

Le recours aux *word embeddings* est une des techniques les plus efficaces du traitement automatique du langage naturel. Il s'agit d'une méthode d'apprentissage automatique de la langue par les machines, basée sur une représentation des relations entre les mots par des vecteurs dans un espace 3D. La proximité sémantique des mots est représentée par leur proximité spatiale. Cette représentation géométrique est ensuite beaucoup plus facile à manipuler par les machines pour reconnaître les entités-mots, les classer, les étiqueter. Les proximités représentées par les vecteurs reflètent les préjugés sociaux, eux-mêmes inscrits dans le *big data* : ainsi le terme *femme* est plus proche de celui d'*infirmière* que celui de *médecin*, et le terme *homme* est plus proche de celui de *médecin* que d'*infirmier*.

2.2.2. La notion de biais

C'est ainsi que le *machine learning* relaye les stéréotypes et les distorsions (ou biais) extraites du *big data*. Or, ce type de développement biaisé sert de base à de nombreuses applications comme le moteur de recherche de Google ou son traducteur automatique.

Aylin Caliskan[21] a montré comment, dans le cas de la traduction d'une langue non genrée vers une langue genrée, le traducteur automatique prédit le genre en fonction des proximités spatiales générées par les *word embeddings*. En turc, langue non genrée, « o bir doktor » est traduit par Google Translation par « il est médecin » (au lieu d'« il ou elle est médecin ») et « o bir hemsike » est traduit par « elle est infirmière [22] ».

Le moteur de recherche de Google s'appuie également sur la méthode des *word embeddings* pour répondre aux requêtes. Ainsi, lorsqu'il doit afficher la liste des doctorants en informatique d'une université [23], il ordonne les résultats en classant d'abord les étudiants puis les étudiantes, car le mot *informaticien* est plus proche spatialement du mot *homme* que du mot *femme*, biais manifeste.

En mars 2016, Microsoft lance sur Twitter Tay [24], un robot de conversation qui emprunte l'identité d'une adolescente et qui a pour but de devenir plus intelligente au fil de la conversation. Au bout de quelques heures, Microsoft ferme le compte suite aux propos extrêmes tenus par le bot (racistes, sexistes, homophobes, négationistes...). En effet, le programme d'intelligence artificielle fonctionne par mimétisme et met dans sa bouche le vocabulaire et les tournures utilisées généralement en ligne, et donc stéréotypées. De plus, Tay a fait l'objet d'une campagne très suivie de détournement de ses capacités conversationnelles. Partisans de Donald Trump, soutiens du mouvement anti-féministe GamerGate et autres trolls de tous bords se sont associés pour transformer l'intelligence artificielle en un compte Twitter de robot néo-nazi.

Ces trois exemples montrent bien comment les algorithmes relaient en les systématisant le sexisme et le racisme de la société.

Des chercheurs travaillent sur des manières de supprimer les préjugés liés au genre reproduits et renforcés par ces outils. Toutefois, la réelle volonté des GAFA de corriger ces biais est bien faible car ces entreprises commerciales n'ont pas pour but la poursuite du bien commun ni la justice sociale. De plus, l'égalité femme-homme n'est pas au cœur de leurs pratiques : par exemple, Google vante sa politique d'égalité dans l'entreprise alors que le ministère du travail américain vient justement de l'accuser de pratiquer une disparité salariale « extrême ».

Par ailleurs, l'élaboration des algorithmes, la collecte et le tri des données, la supervision des phases d'apprentissages des machines ne sont pas expertisés de manière indépendante. Une grande opacité semble donc régner sur des processus qui influencent notre accès à l'information. Ces processus nous sont présentés comme transparents et purement techniques alors qu'ils façonnent de manière biaisée notre rapport au monde.

3. Le projet *en française dans la texte*

Nous avons développé le projet *En française dans la texte* pour prendre le contre-pied des masculinistes qui imposent leur loi depuis des centaines d'années. Il s'agit d'une opération de décolonisation de la langue française, de démasculinisation radicale qui institue un seul genre, le féminin.

Ce choix de réduire les deux genres à un seul, le féminin, outrepasse les solutions préconisées par le langage épicène, c'est-à-dire utilisant une terminologie générique

désignant indifféremment le féminin et le masculin, ou par l'écriture inclusive qui tente de représenter de manière équitable les deux genres. C'est là notre manière de refuser la binarité du genre (masculin/féminin), de la faire exploser au profit de tous les possibles. Une lecture attentive de nos traductions en française est parfois nécessaire pour démêler les relations entre des personnages tous désignés au féminin et pour imaginer leurs identités potentielles.

Après plusieurs siècles de prééminence du masculin, et prenant acte du retard français en matière d'égalité femme/homme dans tous les domaines de la société (travail, politique, culture, loisirs…) et de la langue, nous trouvons légitime de promouvoir aujourd'hui ce systématisme inversé. Quand, enfin, la langue française ne discriminera plus les femmes, nous songerons à participer à l'invention d'une nouvelle forme de langue non genrée.

Le projet consiste à traduire au féminin, à adapter, à éditer des textes pré-existants venant de différents domaines : fiction, documentaire, textes théoriques…

Les textes créés ont en commun un processus de production algorithmique qui a pour effet de révéler l'arbitraire du genre, et de générer des trouvailles littéraires (*la marché aux bestiales*, *la droite de vote*, *les factrices économiques*…). Notre projet s'inscrit ainsi dans une forme de littérature algorithmique ou de poésie numérique.

L'algorithme de féminisation utilisé est complété par des relectures et des corrections manuelles. Nous occupons ainsi une double fonction d'autrices et de petites mains que nous avons expérimentée à travers deux réalisations : *A votée* et *Wikifémia*.

3.1. Processus de traduction en française

Le projet *En française dans la texte* donne lieu à une production textuelle en française qui fait l'objet de performances, de publications, d'expositions… Ces traductions nécessitent le développement d'outils de travail spécifiques.

3.1.1. Les règles de féminisation

Nous nous sommes longuement interrogées sur l'envergure à donner à notre féminisation. Devrions-nous utiliser les formes féminines existantes dans le français actuel ? Ou bien inventer des formes féminines en nous basant sur les règles préconisées par le guide d'aide à la féminisation des noms de métiers, titres grades et fonctions publié en 1999 [25] ? Ou enfin créer des néologismes de toutes pièces ? Après avoir testé ces différentes possibilités, nous avons opté pour une féminisation franche, mais qui préserve la lisibilité.

Par ailleurs, nous avons opté pour une féminisation des mots, pas des choses. Il ne s'agit pas d'un monde sans hommes mais d'une langue sans masculin : *un homme* devient en française *une homme*, pas *une femme*. Nous développerons éventuellement plus tard d'autres modalités. Nous nous réservons le droit de faire évoluer ces règles

dans un sens plus radical ou plus créatif.

Nous avons intitulé ces règles la bonne usage en référence au *Bon usage* de Grevisse en vigueur dans toutes les écoles primaires depuis plusieurs générations.

La bonne usage

â Règle générale :

Substituer systématiquement les formes féminines aux formes masculines.

â Application de la règle aux substantifs :

Remplacer systématiquement les formes masculines par les formes féminines quand elles existent. Si aucune forme féminine n'existe, utiliser les formes masculines précédées d'un déterminant féminin par exemples : *une institutrice*, *une parapluie*, *une petite sac en papier*, *ma grande-père*…

â Application de la règle aux participes présents :

Les participes présents sont arbitrairement au masculin singulier en français, il faut donc utiliser arbitrairement le féminin singulier en française, par exemple, *nous travaillons en chantante.*

â Application de la règle aux participes passés :

Les participes passés sans complément d'objet direct antécédent sont arbitrairement au masculin singulier en français, il faut donc utiliser arbitrairement le féminin singulier en française, par exemple, *a votée.*

3.1.2. Le programme de traduction automatique

Le programme informatique de traduction est développé en Python. Python est un langage de programmation orienté objet, multiplateforme, placé sous une licence libre, et particulièrement adapté à l'analyse lexicale. Ce script fait appel à un fichier qui substitue les formes masculines aux formes féminines. Ce dictionnaire est établi en application de *la bonne usage*. Nous l'avons constitué à partir de la base de données Lexique 3.81 [26], outil libre et gratuit développé par l'Université de Savoie. Il est en perpétuelle expansion et reçoit au fur et à mesure de nouvelles entrées au gré de nos traductions.

Nous avons aussi pris le parti de créer des néologismes. Ainsi le terme *ordinateur* est traduit par *ordinatrice* sur le modèle des noms de machines de l'atelier de mécanographie (la tabulatrice, l'interclasseuse ou encore la calculatrice…). C'était d'ailleurs l'un des noms préconisés en 1955 par Jacques Perret, professeur de philologie latine à la Sorbonne, dans une lettre à Christian de Waldner, président d'IBM France, qui l'interrogeait pour avoir son avis sur le nom à donner aux premières machines qu'IBM

s'apprêtait à construire en France [27]. Actuellement, le script ne corrige pas encore automatiquement les erreurs liées aux homonymes orthographiques que nous corrigeons manuellement : « il répondit sur un ton sévère » est traduit automatiquement par « elle répondit sur une ta sévère »…

3.2. Notre double rôle d'autrices et de petites mains

Nous accordons une attention particulière à ce travail de relecture et de correction manuelle. Nous sommes conscientes de la manière dont les systèmes automatisés sont basés sur l'exploitation du travail de *petites mains*.

L'automatisation ne remplace pas le travail humain, mais le déplace loin des lieux visibles de la technologie car le discours se concentre sur l'automatisation des fonctions cognitives par les machines. Ce travail est invisibilisé, à cause du statut subalterne des personnes qui le font, mais aussi parce que, pour les besoins des machines, les tâches sont découpées dans des unités si petites qu'elles ne sont pas toujours reconnues comme du « vrai » travail. Ainsi, lorsque les internautes remplissent des formulaires ou écrivent des commentaires et qu'il leur est demandé de prouver qu'ils « ne sont pas des robots », les opérations qu'ils font, comme cliquer sur les photos de voiture, servent à améliorer le système de reconnaissance d'images de Google, les internautes servant gratuitement d'experts pour calibrer les algorithmes. Sur la plateforme Mechanical Turk d'Amazon dont le sous-titre est ironiquement « l'intelligence artificielle artificielle », les *turkers* peuvent réaliser, pour quelques centimes, des tâches qui sont nommées des Hits (Human Intelligence Tasks) qui sont trop complexes pour être réalisées actuellement par des algorithmes, comme écrire des légendes de photos, traduire quelques phrases d'une langue à l'autre, etc. À ces deux types de travail non qualifiés, il faut ajouter d'autres pratiques sous le nom de « digital labor » : modérations de plateformes de l'internet, annotations de base de données, numérisations de grands volumes de livres, etc. [28]. Hétéromation est le processus dans lequel le travail des humains et des machines est imbriqué de telle sorte qu'il est difficile de les différencier [29].

En tout état de cause, et tant qu'autrices et artistes, nous endossons et revendiquons le double rôle de néo-académiciennes-programmeuses ET de petites mains [30].

3.3. Choix des textes à traduire

Le travail sur le corpus de textes à traduire a nécessité plusieurs étapes. Nous avons pratiqué une série de tests sur différents matériaux textuels (fictions, essais, articles de journaux, etc), et nous nous sommes concentrées sur des champs plus spécifiques (féminismes, approches critiques des technologies…).

Ces expérimentations linguistiques ont orienté nos choix artistiques et éditoriaux, et nous ont conduit, après avoir abordé un texte de fiction écrit par un auteur unique, à finalement privilégier un texte documentaire et dont l'écriture est collective.

L'encyclopédie Wikipédia nous intéresse particulièrement comme texte source car il s'agit d'un écosystème, d'un monde en miniature, et d'un laboratoire de la production du savoir universel. Son organisation structurée et ses protocoles nous ont semblé propices à accueillir notre traitement algorithmique de féminisation et nos propres protocoles. Nous avons eu plus d'audace à modifier, perturber, déconstruire un tel texte, ressenti plus de liberté à nous l'approprier, peut-être aussi parce que nous sommes également, par ailleurs, contributrices de Wikipédia et connaissons cette construction de l'intérieur.

Notre appropriation du texte va bien au delà d'une simple traduction automatique, il s'agit, dans un premier temps, d'un travail de sélection, d'édition et d'adaptation « humaine » afin de construire un récit.

3.4. Réalisations

Jusqu'à présent, nous avons travaillé à partir de deux textes, un texte de fiction par un auteur de science-fiction reconnu, et un texte documentaire à partir d'articles écrits collectivement dans le cadre de l'encyclopédie en ligne Wikipédia.

3.4.1. *A votée*

Notre première traduction a été établie d'après une nouvelle d'Isaac Asimov sur les rapports entre élections, technologies et genre, qui résonnait avec la campagne présidentielle française de 2017 et avec l'actualité technologique (algorithmes prédictifs, *machine learning*, etc). Cette nouvelle a servi de texte-source pour une performance.

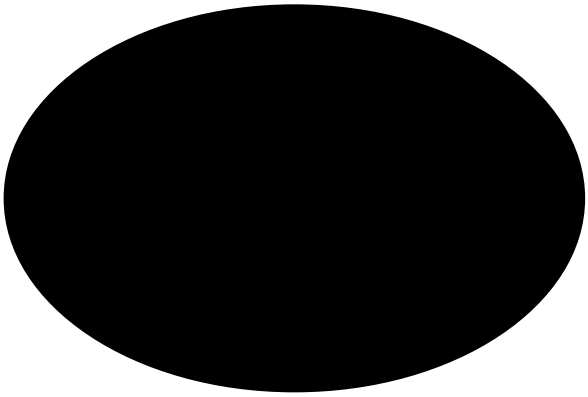
À titre d'exemple, voici le début de la performance :

> C'était la grande jour. La Jour de l'Élection ! À la début, ça avait étée une année semblable à toutes les autres. Peut-être une peu plus mauvaise parce que c'était une année présidentielle mais, somme toute, pas pire qu'une autre année présidentielle. Les politiciennes parlaient de la grande corps électorale et de l'énorme cerveau électronique qui était à sa service. La presse analysait la situation à l'aide d'ordinatrices industrielles et multipliait les allusions à ce qui allait se produire. Commentatrices et éditorialistes en contradiction les unes avec les autres énuméraient les États et les comtés critiques.

La performance consiste en une lecture du texte en française par une comédienne, accompagnée d'un surtitrage en vidéo qui fait apparaître l'orthographe étonnante du texte en française [31]. De plus, sur l'écran, apparaissent des commentaires informatifs, critiques, ironiques qui viennent ponctuer la lecture.

Pour la performance nous avons fait appel à la comédienne Coraline Cauchi [32] qui a la

tâche ardue de faire passer de manière naturelle et fluide un texte très difficile à dire car il remet en cause des automatismes linguistiques ancrés depuis l'enfance.



Doc. 1 â□□Performance *A votée* présentée le 2 mars 2017 à l'université d'Orléans.

3.4.2. *Wikifémia*

Wikifémia propose de traduire en française et de mettre en scène des biographies de femmes remarquables figurant dans l'encyclopédie en ligne Wikipédia. Bien qu'elle soit basée sur un principe de neutralité, l'encyclopédie en ligne Wikipédia n'échappe pas aux stéréotypes : dans sa version francophone, 16% seulement des biographies sont consacrées à des femmes. Avec le projet *Wikifémia*, nous souhaitons faire connaître plus largement des personnalités historiques invisibilisées, alors qu'elles ont contribué au développement de la pensée, de la culture, de la science, de la politique, au même titre que les hommes, du moins quand elles n'en ont pas été empêchées. Nous produisons ainsi de nouvelles narrations ouvertes pour compenser des représentations monolithiques quasi exclusivement masculines.

Wikifémia est conçue comme un ensemble de plusieurs productions parallèles et complémentaires : une série de performances, des outils en ligne, des *workshops* et éditathons [33], une publication papier à paraître ultérieurement.

Les performances prennent comme point de départ un article de Wikipédia et en déplient les hyperliens. Nous choisissons de ne retenir que les liens concernant des femmes. Ainsi nous mettons en lumière les réseaux et les relations que ces femmes entretiennent. Les aiguillages que constituent ces hyperliens structurent le récit. Nous générons ainsi une narration non linéaire qui relie les protagonistes dans une perspective féministe. Dans le même temps, nous souhaitons mettre en valeur le processus d'élaboration des articles par les contributeur·rices de Wikipédia.

La première performance de la série propose une narration augmentée autour d'une personnalité oubliée de la fin du XIX^{e} siècle, Madeleine Pelletier. Cette performance prend pour point de départ l'article de Wikipédia qui lui est consacré. Madeleine Pelletier (1874-1939) devient en 1906 la première femme médecin diplômée en psychiatrie en France. Elle est également connue pour ses multiples engagements politiques et philosophiques et fait partie des féministes les plus engagées à l'époque

des premières luttes pour le droit de vote des femmes à la fin du XIX^{e} et au début du XX^{e} siècle.

Le texte de la performance est composé d'une sélection de fragments d'articles réécrits en française. Il est construit sur la base de l'arborescence des liens depuis l'article Madeleine Pelletier, sur plusieurs niveaux d'hyperliens. Nous avons pré-selectionné une centaine d'articles consacrés à des femmes ou à des organisations de femmes (journaux, associations, organisations politiques, problématiques féministes). Nous abordons différents aspects de la vie de ces femmes, mêlant anecdotes, citations, histoire des mentalités et commentaires critiques. Le travail d'édition a consisté à agencer ces différents éléments afin de construire une progression dramatique.

Voici un extrait du texte de la performance *Wikifémia* :

> Article La Fronde
>
> La Fronde a pour originalité de ne pas être seulement une journal destinée aux femmes, c'est la première quotidienne française conçue, rédigée, administrée, fabriquée et distribuée exclusivement par des femmes journalistes, rédactrices, typographes, imprimeuses, colporteuses. La Fronde est fondée par Marguerite Durand en 1897 et paraît jusqu'en 1903.
>
> Article Marguerite Durand
>
> Marguerite Durand (1864-1936) est une journaliste, féministe et actrice française, fondatrice de la journal La Fronde.
>
> "La Figaro en 1896 m'avait chargée d'écrire une article sur la Congrès internationale de la condition et des droites de la femme que des obstructions malveillantes, des quolibets et des chahuts d'étudiantes signalaient bruyamment à l'attention publique. Je fus frappée par la logique de la discours, la bien-fondée des revendications et la maîtrise, qui savait dominer l'orage et diriger les débats, de la présidente Maria Pognon."
>
> Article Maria Pognon
>
> Maria Pognon (1844-1925) est une journaliste et oratrice française, socialiste, féministe et franche-maçonne.
>
> "Les hommes avancent, de classe en classe, jusqu'à la poste de Directrice ; pourquoi les femmes ayante prouvée par leur travail, des capacités égales à celles des hommes, sont-elles exclues de toutes les emplois rémunératrices ? Nous attendons la réponse !"
>
> Retour à l'article Marguerite Durand

> “Je refusai d’écrire l’article de la Figaro. Mais l’idée m’était venue d’offrir aux femmes une arme de combat, une journal qui devait prouver leurs capacités en traitante non seulement de ce qui les intéressait directement, mais des questions les plus générales et leur offrir la profession de journaliste active.”

La performance consiste en un tissage de voix reprenant le principe des hyperliens de Wikipédia. La parole est distribuée entre une comédienne, Coraline Cauchi, des performeuses-opératrices, Cécile Babiole et Anne Laforet, et des voix de synthèse. Ces différents régimes de parole permettent de mettre en relief notre travail d’édition sur les articles et notre réflexion critique sur l’encyclopédie.

Les recherches faites autour de notre projet *En française dans la texte* nous ont amenées à nous interroger sur l’évolution de la littérature algorithmique. Depuis l’Oulipo, la littérature algorithmique, en changeant d’échelle, a aussi changé de nature : il ne s’agit plus de jouer avec les mots dans le périmètre familier d’un dictionnaire de langue française, mais d’une pratique qui se confronte à la masse en expansion du big data. La maîtrise des outils de traitement automatique du langage met en jeu de nombreux facteurs économiques et techniques (capacité à utiliser et développer des programmes informatiques sophistiqués), culturels (prédominance de l’anglais dans les modèles de traitement du langage…), qui dépassent largement les frontières de la seule langue française et de la langue tout court.

Bibliographie

BOLUKBASI Tolga, Chang Kai-Wei, ZOU James, SALIGRAMA Venkatesh, KALAI Adam, *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings*, en ligne.

CALISKAN Aylin, BRYSON Joanna, NARAYANAN Arvind, « Semantics derived automatically from language corpora contain human-like biases », *Science*, 356 (6334). p. 183-186.

CARDON Dominique & CASILLI Antonio, *Qu’est-ce que le Digital Labor ?* Bry-sur-Marne, INA, coll. « Études et controverses » 2015.

EKBIA Hamid & NARDI Bonnie, *Heteromation and its (dis)contents : The invisible division of labor between humans and machines*, First monday, 2014, en ligne.

IRANI Lilly, *Justice for « Data Janitors »*, Public books, 2015, en ligne.

GOLDSMITH Kenneth, *L’écriture sans écriture – du langage à l’âge numérique,* Paris, Jean Boîte Éditions, 2018.

O’NEIL Cathy, *Weapons of math destruction*, New York, Crown Books, 2016.

VIENNOT Éliane (dir.), *L'Académie contre la langue française : le dossier «féminisation »*, Donnemarie-Dontilly, Éditions iXe, 2015.

—, *Non, le masculin ne l'emporte pas sur le féminin ! Petite histoire des résistances de la langue française*, Donnemarie-Dontilly, Éditions iXe, 2014.

Notes

[1] http://enfrancaisedanslatexte.fr/ et http://robertelarousse.fr/ (consultation le 11 mai 2018).

[2] Cette formule est un clin d'œil au titre de l'essai de Judith Butler, *Trouble dans le genre*, paru en 1990, qui a montré comment est performé le genre.

[3] Expression communément utilisée pour désigner l'ensemble : Google, Amazon, Facebook et Apple.

[4] Éliane Viennot, *Non, le masculin ne l'emporte pas sur le féminin ! Petite histoire des résistances de la langue française*, Donnemarie-Dontilly, éditions iXe, 2014, p 80.

[5] Jean Racine, *Iphigénie*, 1674, Acte III, scène 5.

[6] Jean Racine, *Athalie*, 1691, Acte IV, scène 2.

[7] Éliane Viennot, *Bannir la règle du masculin qui l'emporte sur le féminin*, Slate, 19 mars 2017, en ligne (consulté le 11 mai 2018) ; « La féminisation des noms de métiers, fonctions, grades ou titres – Mise au point de l'Académie française », en ligne (consultation le 11 mai 2018)

[9] Éliane Viennot (dir.), avec Maria Candea, Yannick Chevalier, Sylvia Duverger, Anne-Marie Houdebine, et la collaboration d'Audrey Lasserre, *L'Académie contre la langue française : le dossier « féminisation »*, Donnemarie-Dontilly, Éditions iXe, 2015.

[10] En ligne (consulté le 11 mai 2018). Les académiciens sont écrivain, avocat, haut-fonctionnaire, historien, philosophe, romancier, historien d'art, poète, homme d'Église, homme politique, chef d'État, essayiste, biologiste, scénariste, médecin, journaliste, réalisateur.

[11] Éliane Viennot, *Non, le masculin ne l'emporte pas sur le féminin ! , op. cit.*, p. 52.

[12] En ligne (1) et (2) (consultés le 11 mai 2018).

[13] Art. cit.

[14] LGBTQI : Lesbiennes Gays Bisexuel.les Trans Queer Intersexes

[15] En ligne (consulté le 11 mai 2018).

[16] En ligne (consulté le 11 mai 2018).

[17] « If Hillary Clinton can't satisfy her husband, what makes her think she can satisfy America ? » « Sadly, because President Obama has done such a poor job as president, you won't see another black president for generations! ».

[18] Un.e troll est un.e internaute qui écrit intentionnellement des messages désobligeants, polémiques, provocants, absurdes, de mauvaise foi, voire insultants, et souvent répétitifs, sur des sites communautaires, de conversation ou de réseaux sociaux.

[19] En ligne (1) et (2) (consultés le 11 mai 2018).

[20] En ligne (consulté le 11 mai 2018).

[21] En ligne (consulté le 11 mai 2018).

[22] Aylin Caliskan, conférence « A story of discrimination and unfairness », 27 décembre 2016, 33C3 Hambourg, en ligne (consulté le 11 mai 2018).

[23] Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, Adam Kalai, *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings,* en ligne (consulté le 11 mai 2018).

[24] En ligne (consulté le 11 mai 2018).

[26] En ligne (consulté le 11 mai 2018).

[27] Jacques Perret (1906-1992). Lettre du 16 avril 1955 de J. Perret, professeur à l'université de Paris, à C. de Waldner, président d'IBM France (Archives IBM France). En ligne (consulté le 11 mai 2018).

[28] Lilly Irani, *Justice for « Data Janitors »*, Public books, 2015, en ligne (consultation le 11 mai 2018).

[29] Hamid Ekbia, Bonnie Nardi, *Heteromation and its (dis)contents : The invisible division of labor between humans and machines,* First monday, 2014, en ligne (consulté le 11 mai 2018).

[30] Les « petites mains » sont parfois visibles sur les images des documents numérisés. Voir http://theartofgooglebooks.tumblr.com/ (consultation le 11 mai 2018)

[31] La performance a été montrée le 2 mars 2017 au l'Université d'Orléans, le 5 mai 2017 à la Gaité Lyrique et le 16 juin à l'ENSCI-Les Ateliers à Paris.

[32] Compagnie Serres chaudes : http://serreschaudes.fr/ (consulté le 11 mai 2018).

[33] Mot-valise composé d'édition et de marathon, un éditathon est un atelier court et intensif consacré à la rédaction et à l'apprentissage de l'édition d'articles au sein de Wikipédia.

Auteurs

Cécile Babiole est artiste. Elle est active depuis les années 1980, dans le champ musical d'abord, puis dans les arts électroniques et numériques (voir son site). Elle associe dans ses créations arts visuels et sonores au travers d'installations et de performances qui interrogent les médias, questionnent les technologies et tentent d'en transposer de façon détournée les usages normés dans le champ de la création. Ses derniers travaux s'intéressent à la langue (écrite et orale), à sa transmission, ses dysfonctionnements, sa lecture, sa traduction, ses manipulations (*Conversation au fil de l'eau*, *Leçon de vocabulaire*, *Spell*, *Disfluences*, *Copies non conformes*, *En française dans la texte*, etc.). Son travail a été exposé internationalement (Centre Pompidou, Gaîté Lyrique – Paris, Mutek, Elektra – Montréal, Fact – Liverpool, MAL – Lima, NAMOC – Beijing…). Il a été distingué par de nombreux prix et bourses (Ars Electronica, Locarno, prix SCAM, bourse Villa Médicis hors les murs, Transmediale Berlin, Stuttgart Expanded Media Festival…). Cécile Babiole est par ailleurs membre du collectif d'artistes-commissaires « Le sans titre ».

Anne Laforet est chercheure, enseignante, artiste et critique. Elle est docteure en sciences de l'information et de la communication (voir son site). Sa thèse a été publiée en 2011 éditions Questions théoriques sous le titre *Le Net art au musée. Stratégies de conservation des œuvres en ligne*. Ses thématiques de recherche sont principalement la conservation et la documentation des arts numériques, l'anarchronisme, les relations entre analogique et numérique, l'internet, le logiciel libre et les pratiques artistiques collaboratives. Depuis 2011, elle enseigne à la Haute École des arts du Rhin (HEAR) à Strasbourg. Elle a participé activement au projet de recherche européen *Digital art conservation* à l'Espace Multimédia Gantner et à la HEAR de 2010 à 2012, et a été chercheure associée au laboratoire art audio Locus Sonus de 2011 à 2013 et commissaire de l'exposition « Anarchronismes, machines à perturber le temps » (IMAL, Bruxelles en 2015 ; Espace Gantner, Bourogne en 2016). Depuis 2013, elle collabore régulièrement avec les collectifs Constant et Algolit pour des projets artistiques et éditoriaux.

Copyright

Tous droits réservés.