

Analyser la valeur de la femme dans la bible avec word2vec ?

Victoria Constantini

April 1, 2024

Un grand merci pour l'aide précieuse de :

Philippe Blanc

Sébastien Travadel

Benoit Gschwind

Abstract

Le développement de modèles de langage s'est accompagné de critiques et mises en garde quant à l'incorporation de biais, sexistes entre autres, lors de leur apprentissage. Ces biais sont appris parce qu'ils existent dans le corpus d'entraînement. Se dessine alors, en miroir, la possibilité d'étudier les biais sexistes dans un texte choisi en exploitant les facultés des modèles de langage. Dans cet article, cette démarche a été explorée en choisissant la bible comme objet d'étude, et donc comme corpus d'entraînement d'un modèle Word2vec. Word2vec nous permet de saisir les termes et concepts associés à 'femme' dans le texte mais sans préciser la nature de la relation. L'usage d'un modèle de langage nous donne des pistes solides, mais insuffisantes pour en déduire la valeur de la femme dans un texte. Seule une étude humaine, appuyée théoriquement sur le concept linguistique du noyau de sens, nous permettra d'acquérir la profondeur d'analyse requise pour comprendre la place de la femme dans l'oeuvre.

I Introduction

L'avènement de word2vec en 2013 s'est accompagné de réserves quant à l'incorporation de biais, sexistes notamment, dans les plongements textuels des mots. Ainsi, l'article de Tolga Bolukbasi ([1]).s'inquiète par exemple de la proximité entre les mots 'nurse' et 'woman' ou 'football' et 'man' et introduit plusieurs méthodes pour quantifier ces proximités. Par ailleurs, l'article de Kaytlin Chalon et al. ([2]) propose d'évaluer les biais sexistes contenus dans différents textes en faisant varier les corpus d'entraînement. Fort de ces articles, il semble naturel de s'interroger sur la possibilité d'étudier plus généralement la valeur de la femme dans un texte à l'aide de word2vec. Nous allons mener cette étude sur un ouvrage qui a forgé les cultures et normes sociales françaises : la bible.

I.1 Word2vec

Cette section consiste en un bref condensé des notions de word2vec utiles pour la suite de cet article. Il est inutile à la lecture pour les utilisateurs confirmés du modèle de langage de Tomas Mikolov ([4], [5]).

I.1.1 Principe général

Word2vec est un algorithme d'apprentissage non-supervisé qui apprend des représentations vectorielles

de mots. La caractéristique fondamentale de ce modèle de langage, qui lui a valu son succès, est sa faculté à conserver les relations entre les mots. Par exemple, nous avons l'égalité vectorielle : $\vec{\text{homme}} - \vec{\text{femme}} + \vec{\text{roi}} = \vec{\text{reine}}$. L'entraînement de word2vec que nous allons aborder maintenant s'appuie sur l'hypothèse linguistique suivante énoncée par J.R. Firth : "You shall know a word by the company it keeps.". Autrement dit, le sens d'un mot peut être compris par les contextes dans lesquels il est utilisé.

I.1.2 Entraînement

Le modèle utilise une architecture CBOW (Continuous bag of words). Après avoir attribué lors de l'initialisation des vecteurs aux mots de manière aléatoire, le modèle s'entraîne sur un corpus de la façon suivante. Pour chaque mot est sélectionné un contexte qui est constitué d'un certain nombre de mots avant et après le mot cible. Le réseau de neurones reçoit alors en entrée le contexte et doit prédire en sortie le mot cible.

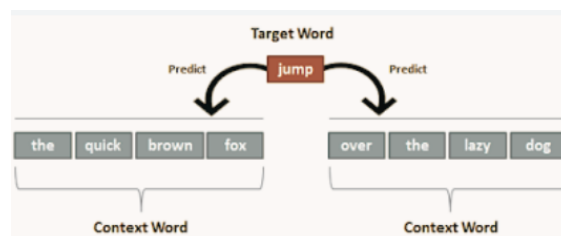
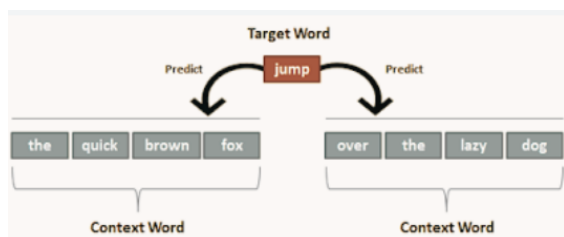


Illustration de l'entraînement sur un exemple choisi

I.1.3 Paramètres

Il existe plusieurs paramètres du modèle à choisir pour optimiser sa performance et sa robustesse.

- 'window size' (taille de la fenêtre) : nombre de mots voisins à gauche et à droite d'un mot cible donné qui sont considérés comme son contexte.
- 'vector size' (taille des vecteurs) : dimensionnalité des représentations vectorielles des mots apprises par le modèle.
- epoch : nombre d'itérations complètes sur l'ensemble du corpus d'entraînement.



Dans l'exemple : window size = 4

I.2 Etude de biais sexistes : méthode

Les approches visant à étudier les biais sexistes sont diverses et variées, comme en témoignent les travaux des articles [1], [2], [3]. Ces méthodes s'efforcent principalement de déterminer si les préjugés observés dans le monde réel se retrouvent également incorporés

dans le word2vec. Cependant, elles peuvent être limitées dans leur capacité à fournir une évaluation objective de la représentation des femmes dans le texte. En effet, plutôt que d'examiner directement la nature des biais présents dans le texte, ces approches se concentrent sur la vérification de l'existence de préjugés préalablement définis. Seule une méthode émerge comme étudiant effectivement les biais à partir du texte lui-même : la méthode par analogie ([1]).

I.2.1 Méthode par analogie

L'analogie consiste à trouver des relations de similarité entre des éléments de manière systématique. La tâche analogique standard consiste à partir de 3 mots (par exemple, 'homme', 'femme', et 'roi') et trouver un quatrième mot 'x' qui résout l'équation suivante : 'homme' est à 'roi' ce que 'femme' est à 'x'. Donc ici, nous avons $x = \text{'reine'}$.

Dans l'article ([1]), la tâche d'analogie est modifiée afin de générer de manière systématique des couples d'analogies pertinents. Le principe est le suivant : étant donné deux mots, par exemple 'homme', 'femme', nous voulons générer une paire de mots, 'x' et 'y', tels que 'homme' à 'x' et 'femme' à 'y' soit une bonne analogie.

L'entrée dans le générateur d'analogies de l'article est une direction initiale, qui correspond à la différence normalisée

$\vec{a} - \vec{b}$ entre les deux mots de départ (donc ici $(\vec{\text{homme}} - \vec{\text{femme}})/(\vec{\text{homme}} - \vec{\text{femme}})$). Nous attribuons ensuite un score à toutes les paires de mots \vec{x}, \vec{y} avec la métrique suivante :

$$S_{ab}(\vec{x}, \vec{y}) = \begin{cases} \cos(\vec{a} - \vec{b}, \vec{x} - \vec{y}) & \text{si } \|\vec{x} - \vec{y}\| < \delta \\ 0 & \text{sinon.} \end{cases}$$

Dans l'équation, δ est un seuil de similarité, qui sert à avoir des paires de mots sémantiquement cohérentes. La métrique de score est d'autant plus élevée que la différence entre \vec{x} et \vec{y} est parallèle à la direction de départ.

II Expérience

II.1 Création du corpus d'entraînement

Le corpus d'étude choisi est la bible de Jérusalem. Un premier traitement du corpus pour permettre une analyse correcte est de remplacer l'ensemble des prénoms de femme dans la bible par le mot 'femme'. En effet, les figures féminines importantes sont systématiquement appelées par leur prénom et ne pas prendre cette précaution entraînerait une analyse de la femme 'lambda', quelconque. On prend la même précaution pour 'homme' en remplaçant tous les prénoms masculins. Seulement, le corpus ainsi constitué ne

fait que 2000 pages, ce qui est insuffisant pour entraîner correctement un modèle Word2vec. On décide d'y adjoindre un corpus 'neutre', qui, idéalement, ne doit pas comporter de biais sexistes. On choisit de considérer les textes de loi (code civil, code du travail, code de l'éducation, etc.) et un dictionnaire pour élargir le vocabulaire du modèle.

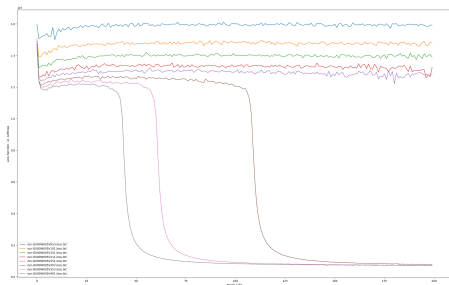
II.2 Entraînement du modèle

Maintenant que le corpus est optimal, il s'agit d'entraîner le modèle avec des paramètres qui le rende performant et robuste.

II.2.1 Qualité de l'apprentissage

On commence par évaluer la qualité de l'apprentissage en affichant l'évolution de la 'loss function' au cours de l'entraînement pour des modèles de différents paramètres. En théorie, il aurait fallu regarder ces fonctions en faisant varier tous les paramètres (window size et vector size ici puisqu'on affiche la loss function selon le nombre d'epoch) pour trouver les paramètres optimaux, mais il était trop long de procéder ainsi. On a donc décidé de fixer la window size à 5, donc un contexte de 10 mots sachant qu'un verset comporte généralement entre 15 et 25 mots. Le graphe suivant

affiche les loss functions de plusieurs modèles, qui diffèrent seulement par le choix de dimension des vecteurs.



Loss function pour des modèles avec des paramètres différents de vector size.

On peut retrouver la vector size du modèle dans le nom de celui-ci : `runG00W005Vdimension du vecteur`

Seuls les modèles qui occasionnent une forte décroissance apprennent. On constate qu'il y a donc une taille minimale de dimension de vecteurs de 300 pour que l'apprentissage se déroule correctement.

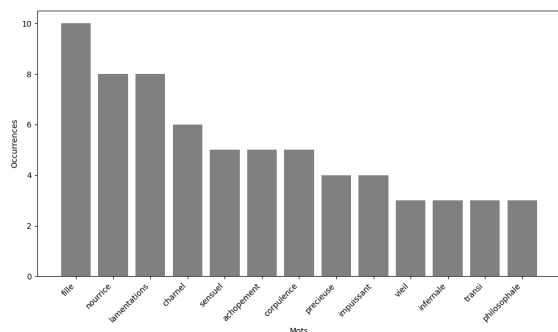
On choisit donc les paramètres de modèle suivant :

- window size = 5
- vector size = 353
- epoch = 1000 (qu'on trouve graphiquement : il faut choisir une abscisse située après la forte décroissance)

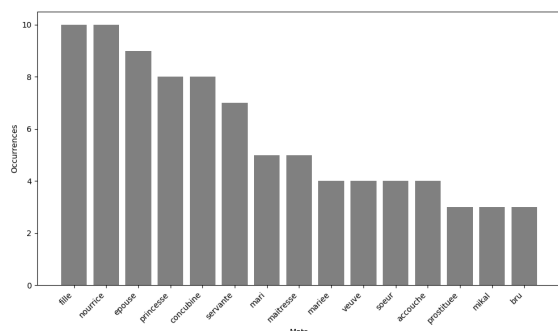
II.2.2 Robustesse du modèle

On cherche maintenant à évaluer la robustesse du modèle choisi. Pour ce faire, on précède l'entraînement par un mélange aléatoire du corpus et on répète cette opération 10 fois. On obtient donc 10 modèles, qui diffèrent uniquement par le shuffle du corpus avant l'entraînement. On veut évaluer en priorité si le modèle est stable sur la notion de femme, et si possible sur la notion d'homme. Une première méthode, assez stricte, consiste à faire l'intersection pour les 10 modèles des 10 ensembles constitués des mots les plus similaires à femme, puis à homme. On obtient les ensembles mots-similaires-femme = {fille, nourrice} et mots-similaires-homme = {}. Au-delà d'une dissymétrie de stabilité entre homme et femme sur laquelle nous reviendrons plus tard, nous constatons que les résultats sont assez décevants avec cette mesure. On essaye une méthode moins stricte pour vérifier si notre modèle est effectivement catastrophique. On affiche un histogramme des occurrences de chaque mot présent dans les 10 ensembles et de leurs occurrences. Le résultat obtenu est plus correct, avec 7 mots qui apparaissent au moins 5 fois parmi les 10 ensembles pour le modèle entraîné uniquement sur le corpus neutre, et 8 mots qui apparaissent au moins 5 fois parmi les 10 ensembles pour le modèle

entraîné sur le corpus neutre et la bible.



Occurrences des mots après entraînement sur le corpus constitué seulement des textes neutres



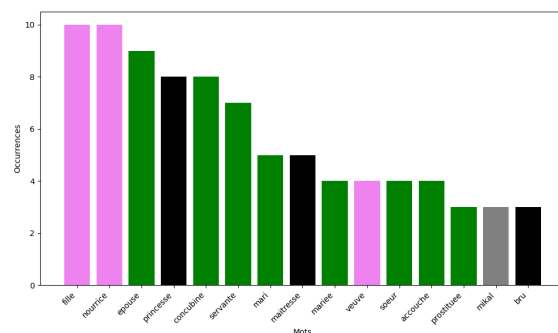
Occurrences des mots après entraînement sur le corpus constitué des textes neutres ET de la bible

II.3 Biais corpus neutre vs bible

En observant le graphe des occurrences neutres, on constate que l'absence de biais dans le corpus neutre était un espoir assez vain. Il convient donc de distinguer les biais du corpus neutre des biais de la bible pour éviter de fausser l'analyse.

Pour ce faire, nous allons nous servir des histogrammes de la section précédente. En comparant l'évolution des occurrences

des mots entre les entraînements avec et sans la bible, on peut espérer saisir la différence de biais présents dans le corpus neutre et dans la bible. En effet, on considère que des mots dont les occurrences sont très présentes sans et avec la bible correspondent à des biais présents dans les deux corpus. Ceux dont les occurrences diminuent lors d'un entraînement avec la bible correspondent à des biais du corpus neutre et ceux dont les occurrences augmentent lors d'un entraînement avec la bible correspondent à des biais du texte saint. Ces considérations sont résumées dans l'histogramme suivant, dont les différentes couleurs englobent les différents cas.



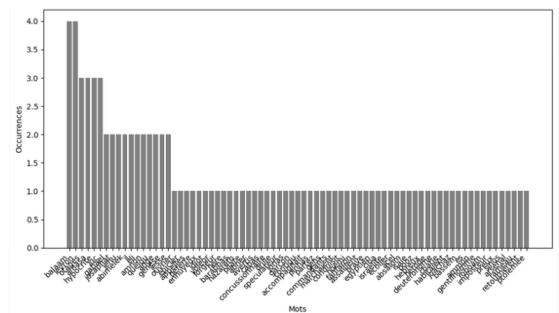
Distinction sur l'origine des biais : corpus neutre ou la bible.

Légende :

- rose : biais présent dans le corpus neutre et dans la bible
- vert : biais présent dans la bible
- noir : anomalie
- gris : prénom oublié qui n'a pas été retiré

Les mots associées à la couleur noire constituent des cas où ce raisonnement ne peut s'appliquer, puisque leurs occur-

rences augmentent après l'entraînement dans la bible alors que ces termes n'existent pas dans le texte saint.



Occurrences des mots similaires à HOMME après entraînement sur le corpus constitué des textes neutres ET de la bible

II.4 Application de la méthode d'analogies

Nous appliquons la méthode détaillée plus haut (tirée de [1]) avec le word2vec entraîné sur la bible en cherchant les 10 meilleurs couples d'analogies. On obtient parmi d'autres les résultats suivants : ('ange', 'moyeu'), ('range', 'potion'), ('imposteur', 'aspect'), ('menaçant', 'habillée'), etc. Les résultats sont donc complètement absurdes et incohérents.

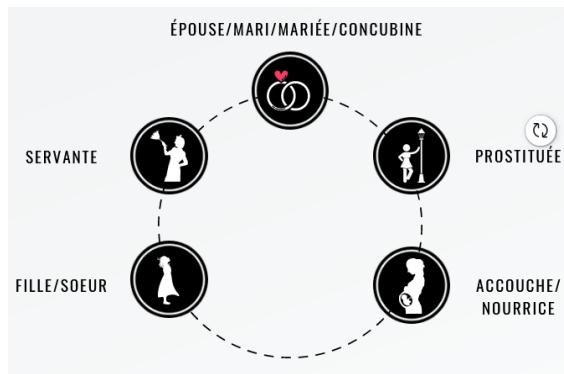
La raison de cet échec tient dans l'instabilité du mot homme, qui était apparue en demi-teinte dans la partie précédente dans la dissymétrie des ensembles de mots les plus similaires associés à 'femme' et 'homme'. Lorsqu'on s'intéresse à l'histogramme des occurrences de mots les plus similaires associés à homme, nous obtenons la figure suivante :

L'illisibilité de la figure confirme l'instabilité des mots les plus similaires à homme, avec de nombreux termes qui apparaissent une seule fois. Les mots les plus stables associés à homme ne sont mentionnés que quatre fois sur dix parmi les listes de mots plus similaires. Nous constatons donc que le modèle Word2vec n'arrive pas du tout à appréhender la notion de 'homme'. Une des causes pourrait être la multiplicité des contextes dans lesquels ce terme est utilisé : il y a environ 72 000 occurrences de 'homme' dans le corpus contre 14 000 de 'femme'. D'autant plus que le terme 'homme' signifie à la fois 'humain mâle' et 'humain' plus généralement.

II.5 Insuffisance des traitements informatiques

Nous nous retrouvons ainsi dépourvus de méthodes conventionnelles d'analyse des biais sexistes, en partant du corpus plutôt

que des biais présents dans le monde réel. Le modèle de langage Word2Vec nous a permis, à travers l'outil des mots les plus similaires, d'identifier les termes considérés comme étant proches de "femme".



Mots proches de femme selon Word2vec

Bien que ces mots fournissent une première piste pour notre projet, une étude plus approfondie est nécessaire afin d'éviter de nous limiter à une simple liste de termes sans explication des notions et relations sous-jacentes.

III Discussion

III.1 Revenir au texte

Grâce aux démarches antérieures, nous identifions les mots étroitement liés à 'femme', mais sans connaître véritablement la nature de cette relation. Détaillons cette distinction sur un exemple. Nous savons que 'prostituée' est proche de femme, mais cela signifie-t-il que toutes les femmes de la bible sont des prostituées ? ou que les prostituées ne

sont plus considérées comme des femmes ? Quelles sont les caractéristiques partagées par 'prostituée' et 'femme' responsables de leur rapprochement par le word2vec ? Simplement leur genre ? ou une proximité des deux avec le concept de 'tentation' ? Au-delà du fait que ces questions demeurent sans réponse, nous projetons ici nos propres biais et présupposés dans la démarche de recherche. C'est parce que nous supposons que la prostituée et la femme sont assimilées à la tentation dans la bible que nous voulons évaluer si ce rapprochement précis est effectivement pertinent. Pour éviter d'introduire nos propres conclusions hâtives dans l'étude, il est nécessaire de revenir au texte lui-même, pour révéler objectivement la nature des proximités révélées par le word2vec.

Pour extraire uniquement les versets pertinents dans notre analyse, nous utilisons un algorithme qui extrait uniquement les phrases contenant un des mots considérés comme proches de femme pour word2vec. Certains mots ont des occurrences extrêmement nombreuses : le mot mère est, par exemple, associé à (par exemple, le mot 'fille' a 302 occurrences). Pour faciliter le traitement des versets, nous optons pour une méthode de clustering qui permettra de distinguer les différentes thématiques associées à chaque mot.

Dans l'ensemble, les clusters semblent cohérents mais quelques réserves doivent

être émises pour certains cas particuliers. Par exemple, le verset : 'Je les rassasiais et ils devenaient adultères; ils se précipitaient à la maison de la prostituée' est dans le même cluster que 'Comme ton coeur était faible - oracle du Seigneur Yahvé - en commettant toutes ces actions dignes d'une véritable prostituée!', et dans un cluster différent que celui de 'Ils y allèrent, se rendirent à la maison d'une prostituée nommée femme et ils y couchèrent'. Pour comprendre ces erreurs, rappelons d'abord le principe de clusterisation, en s'intéressant plus précisément au fonctionnement de l'algorithme k-medoids que nous avons utilisé.

L'algorithme K-Medoids est une méthode de clustering qui cherche à regrouper des données similaires en clusters tout en utilisant des points de données spécifiques, appelés "médoïdes", comme représentants de chaque cluster. Pour l'utiliser sur nos versets, il a fallu transformer ces derniers en point de données. La méthode choisie, assez répandue, consiste à représenter une phrase par la moyenne des vecteurs des mots qui la constituent. Le vecteur obtenu est supposé être représentatif de la phrase et contenir son sens. Seulement, cette méthode a fait l'objet de quelques critiques :

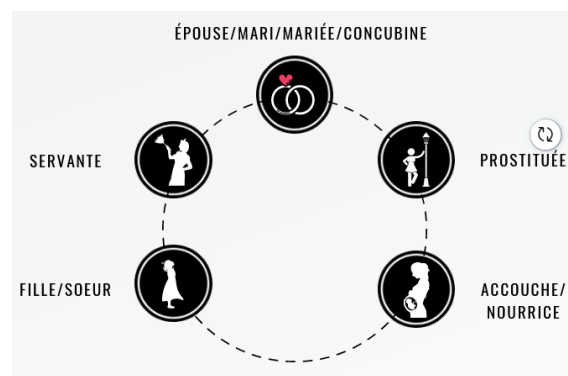
- perte de l'ordre des mots
- perte de la pondération des mots : Tous les mots contribuent de manière égale à la représentation finale, ce qui

peut ne pas être souhaitable car certains mots peuvent être plus importants que d'autres pour la signification globale du texte.

Ce sont ces défauts qui sont très probablement responsables des quelques dysfonctionnement sur les versets. Cette méthode est plus efficace sur des grands paragraphes pour obtenir un sens global, que sur des phrases courtes comme les versets, qui nécessitent une représentation plus précise. Par manque de temps, nous nous contenterons de faire une lecture exhaustive des versets.

III.2 Aide de la linguistique

Nous allons maintenant analyser la valeur de la femme dans la bible à l'aide d'un outil de linguistique : le noyau de sens. Le mot femme est associée à différents mots et concepts qui ont été mis en exergue avec les mots les plus similaires obtenus par l'outil word2vec.



Mots proches de femme selon Word2vec

Pour saisir pleinement la valeur de la femme dans le texte, il est nécessaire d'accéder au dénominateur commun de tous ces termes, à la partie du sens qui reste invariante lors des modifications et qu'on appelle noyau de sens. Pour confirmer le choix d'un mot/d'une image en tant que noyau de sens, il faut s'atteler à la démarche suivante, détaillée et mise en application par les linguistes Bernard Victorri et Catherine Fuchs ([6]) : étant donné ce noyau de sens, commun à toutes les valeurs de l'expression, quels sont les paramètres variables qu'il faut spécifier pour obtenir chacune de ces valeurs ? Nous allons procéder ainsi par la suite. Les analyses des textes bibliques s'appuient sur les écrits de Micheline Gagnon ([8]), Cathy McClive ([9]) et Jean-Jacques Daban ([7]).

Le corps de la femme est d'abord perçu comme une possession. On rappelle la définition de posséder : disposer en maître de (quelque chose). On peut, pour corroborer cette idée, relever le verset : '1 Corinthiens 7, 4 La femme ne dispose pas de son corps, mais le mari'. Le corps féminin est objectifié, et répond donc à des fonctions précises (définition d'un objet : 'Chose solide ayant unité et indépendance et répondant à une certaine destination.' (Le Robert)). Une des fonctions probantes du corps féminin est la procréation. Conformément aux conceptions de L'Ancien Testament, la recherche de la

fécondité est, pour Israël, une question de vie ou de mort. Car Israël est un peuple élu et vit sous la promesse. 'Isaïe 61, 9 Leur race sera célèbre parmi les nations, et leur descendance au milieu des peuples; tous ceux qui les verront les reconnaîtront comme une race que Yahvé a bénie'. 'Isaïe 65, 23 [...] ils seront une race de bénis de Yahvé, et leur descendance avec eux'. Et c'est à travers la femme qui enfante que peut s'accomplir la promesse. La mère prend la valeur d'une source de vie, à tel point qu'elle est parfois désignée par métonymie, "ventre". D'où le choix d'une représentation du noyau de sens de la femme comme un réceptacle, ayant pour fonction première d'accueillir la semence masculine.



Le réceptacle comme noyau de sens de la femme

Le réceptacle doit vérifier certaines conditions de 'qualité' pour s'assurer de la 'qualité' de la descendance. Un réceptacle de qualité doit être 'propre', ce qui à travers les critères énoncés par la bible se

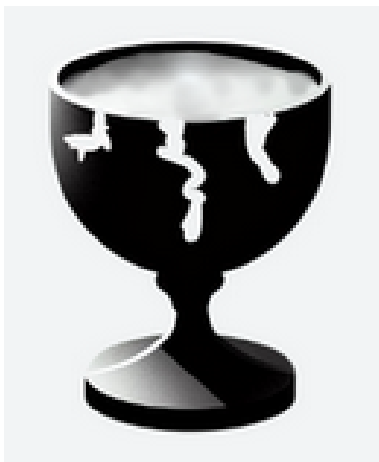
traduit par la virginité de la femme avant son mariage.



La virginité représentée comme un réceptacle propre

C'est une notion absolument cruciale dans l'Ancien Testament, et qui pourtant a échappé à Word2vec. L'importance de la virginité est perceptible dans les règles suivantes. Si la virginité d'une femme non-fiancée est prise par un homme, la femme peut uniquement engendrer une descendance avec cet homme : "Deutéronome 22, 28 Si un homme rencontre une jeune fille vierge qui n'est pas fiancée, la saisit et couche avec elle, pris sur le fait, Deutéronome 22, 29 l'homme qui a couché avec elle donnera au père de la jeune fille 50 pièces d'argent; elle sera sa femme, puisqu'il a usé d'elle, et il ne pourra jamais la répudier". Si une femme promise à un homme se voit prendre sa virginité par un autre, elle est condamnée à mort : "Deutéronome 22, 23 Si une jeune fille vierge est fiancée à un homme, qu'un autre homme la rencontre dans la

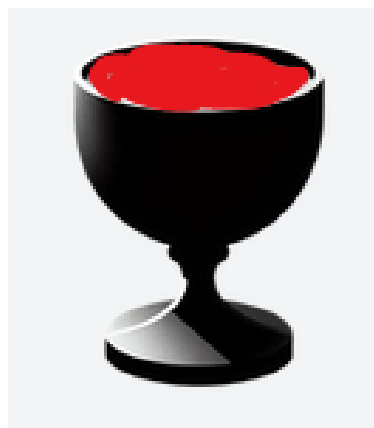
ville et couche avec elle, Deutéronome 22, 24 vous les conduirez tous deux à la porte de cette ville et vous les lapiderez jusqu'à ce que mort s'ensuive: la jeune fille parce qu'elle n'a pas appelé au secours dans la ville, et l'homme parce qu'il a usé de la femme de son prochain". Pour les prêtres dont la descendance doit être la plus pure, les critères sont les suivants : "Ezéchiel 44, 22 Ils ne prendront pas pour femme une veuve, ni une femme répudiée, mais une vierge de la race d'Israël; Ainsi, les femmes répudiées et les veuves qui ont déjà été en contact avec un homme ne sont plus suffisamment pures pour engendrer la descendance d'un prêtre. On a aussi : "Lévitique 21, 14 La veuve, la femme répudiée ou profanée par la prostitution, il ne les prendra pas pour épouses; c'est seulement une vierge d'entre les siens qu'il prendra pour épouse: Lévitique 21, 15 il ne profanera point sa descendance, car c'est moi, Yahvé, qui l'ai sanctifiée". La prostitution est associée à la profanation, dont la définition est la suivante : Action d'avilir (= Faire perdre de sa valeur, déprécier), de dégrader (quelque chose de vénérable, de précieux (CNRTL). On peut donc associer à la femme prostituée l'image d'un réceptacle dégradé, ébréché.



La prostitution associée à un réceptacle ébréché

L'impureté de la femme pendant les règles, déclarée par exemple dans le verset suivant : "Lévitique 15:19-23 : La femme qui aura un flux, un flux de sang en sa chair, restera sept jours dans son impureté. Quiconque la touchera sera impur jusqu'au soir." s'inscrit dans une impureté plus générale des émissions génitales masculines et féminines. Les tabous menstruels et séminaux dans le Lévitique 15, 18-19 et 20-18, fonctionnent comme un moyen d'assurer la procréation d'enfants sains. Protéger le sacré consistait donc à interdire l'acte sexuel en dehors des périodes optimales pour la conception, c'est-à-dire pendant des émissions pathologiques, après une pollution nocturne, immédiatement après l'acte sexuel quand la semence est considérée comme faible, et pendant la menstruation. On peut associer à cette interdiction l'explication imagée suivante : lorsque le réceptacle est plein de sang de menstruation, il est inapte à accueillir

la semence masculine donc inapte à accomplir sa fonction première.



Les règles qui remplissent le réceptacle de sang

Dans cet optique de procréation, la stérilité et la fécondité sont des enjeux de taille pour la femme biblique. Le premier qu'on pourrait représenter par un 'défaut de conception' du réceptacle est en fait assimilée à une malédiction divine dans les textes saints.

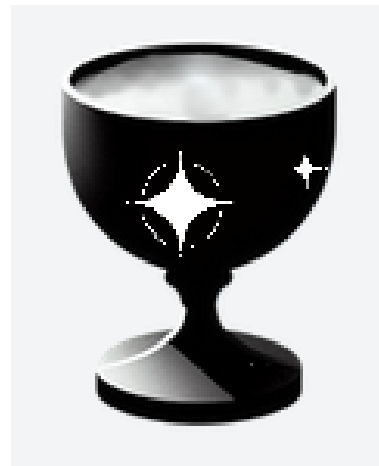


La stérilité comme défaut de conception du réceptacle

La procréation d'une descendance sup-

pose en fait une étroite collaboration avec Yahvé, de qui dépend toute fertilité. Face à la stérilité, deux dénouements fréquents sont présents. Le premier est un recours à des mères porteuses. Les servantes-concubines n'existent que pour donner des enfants à leurs maîtresses ; c'est là leur seule gloire et la seule justification de leur présence dans la famille. On peut citer : 'Genèse 30, 10 Zilpa, la servante de Léa, enfanta à Jacob un fils', "Genèse 36, 12 Eliphaz, fils d'Esau, eut pour concubine Timna et elle lui enfanta Amaleq", '1 Chroniques 2, 46 Ephraïm, concubine de Caleb, enfanta Harân, Moça et Gazèz'. Mais la stérilité prend parfois un sens nouveau et incarne ce qui permet la manifestation de la Toute-Puissance divine à travers la maternité des femmes. Les femmes stériles accèdent fréquemment à des naissances inespérées comme interventions divines, qui manifestent la collaboration singulière des femmes avec Dieu comme source de vie. C'est uniquement parce que Dieu se révèle à elles qu'elles parviennent à s'émanciper d'un destin assigné à partir de données biologiques. Dans les écrits s'accumulent les impossibilités biologiques pour soutenir le point de vue théologique, qui est de faire ressortir comment, dans et à travers l'histoire des épouses stériles, finit par s'inscrire le dessein de Dieu. Par exemple, avant même l'annonce de Dieu à Abraham d'une descendance, Sara est dite 'stérile' (Gn 11, 30), plus tard avancée en âge (18, 11), où il ne reste plus, hu-

mainement parlant, la moindre chance de concevoir un héritier. Ainsi, le 'défaut de conception' du réceptacle est en fait un moyen d'exprimer la toute-puissance du divin, qui parvient, par un coup de baguette magique, à rendre le réceptacle fonctionnel.



Réceptacle d'une femme stérile après l'intervention divine

IV Conclusions

En conclusion, ce projet a étudié la représentation de la femme dans la Bible en exploitant les techniques de traitement automatique du langage naturel, notamment l'algorithme Word2Vec. Cette approche informatique a offert une première exploration fructueuse, en permettant d'identifier un ensemble initial de notions associées à la femme. Toutefois, cette analyse reste préliminaire et nécessite une investigation plus approfondie pour appréhender pleinement la nature des relations entre ces différentes notions et

pour saisir complètement la valeur de la femme dans l'œuvre. L'étude doit se poursuivre avec des outils autre, notamment de linguistiques, pour arriver à une conclusion. Ce projet s'inscrit dans une démarche plus large d'alliance entre les possibilités informatiques et l'intelligence humaine, entre la 'data' et la 'Sophia'.

Une réflexion finale concerne l'appellation des techniques de traitement automatique du langage naturel, souvent désignées comme des "modèles de langage". Selon la définition du dictionnaire Larousse, le langage représente la capacité universelle d'exprimer la pensée et de communiquer au moyen de signes vocaux et graphiques. À la lumière des conclusions de cette étude, des réserves peuvent être formulées quant à la capacité réelle de ces modèles à comprendre le langage humain dans toute sa complexité.

V Perspectives

Une direction possible pour de futures recherches serait de reproduire l'étude en utilisant différents modèles de langage. Cette approche nous permettrait d'évaluer si les avancées technologiques récentes ont permis aux modèles de mieux comprendre un texte et ses subtilités. Peut-être sont-ils désormais capables d'effectuer une analyse précise de la place de la femme dans une œuvre ? En fin de compte, cela nous offrirait une esquisse de réponse à la question conclusive

de cet article : un modèle de langage sait-il effectivement communiquer ?

References

- [1] Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, Adam Kalai, *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings*, 2016.
- [2] Kaytlin Chalon, Alfredo Maldonado, *Measuring Gender Bias in Word Embeddings across Domains and Discovering New Gender Bias Word Categories* in Proceedings of the 1st Workshop on Gender Bias in Natural Language Processing, pages 25–32, Florence, Italy, 2019
- [3] Aylin Caliskan, Joanna J. Bryson, Arvind Narayanan, *Semantics derived automatically from language corpora necessarily contain human biases*, 2017
- [4] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean *Efficient Estimation of Word Representations in Vector Space*, 2013
- [5] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean, *Distributed Representations of Words and Phrases and their Compositionality*, 2013
- [6] Bernard Victorri, Catherine Fuchs, *La*

polysémie- construction dynamique du sens. dans Hermès, pp.131, 1996

- [7] Jean-Jacques Daban, *Femmes et filles dans la bible* dans La lettre de l'enfance et de l'adolescence 2003/1 (n°51) pages 15 à 20
- [8] Micheline Gagnon, *Le rapport stérilité-fécondité dans l'Ancien Testament comme signe annonciateur du mystère virginité-maternité dans le Nouveau Testament*, 1985
- [9] Cathy McClive, *Engendrer le tabou. L'interprétation du Lévitique 15, 18-19 et 20-18 et de la menstruation sous l'Ancien Régime* dans Annales de démographie historique 2013/1 (n° 125), pages 165 à 210