
Notes de lecture

Rubrique préparée par Denis Maurel

Université de Tours, LIFAT (Laboratoire d'informatique fondamentale et appliquée)

Jimmy LIN, Rodrigo NOGUEIRA, Andrew YATES. Pretrained Transformers for Text Ranking: BERT and Beyond. Morgan & Claypool publishers. 2021. 308 pages. ISBN : 978-1-636-39230-1.

Lu par **Marilyne LATOUR**

Entreprise ReportLinker

Le but du « text ranking » ou classement de textes est de générer une liste ordonnée de textes extraits d'un corpus en réponse à une requête. Bien que l'utilisation la plus courante du classement de textes soit pour la recherche d'informations, d'autres applications du traitement automatique des langues (TAL) peuvent également en bénéficier. Ce livre donne un aperçu du classement de textes avec les architectures de réseaux de neurones appelées transformateurs, dont BERT (Bidirectional Encoder Representations from Transformers) est l'exemple le plus connu. Ce livre fournit une synthèse des travaux existants en tant que point d'entrée pour les praticiens qui souhaitent mieux comprendre comment appliquer les transformateurs aux problèmes de classement de textes et les chercheurs qui souhaitent poursuivre leurs travaux dans ce domaine. Il couvre un large éventail de techniques modernes, regroupées en deux grandes catégories : les modèles de transformateurs qui effectuent le reclassement dans des architectures multi-étages (multi-stage architectures) et les modèles qui effectuent directement le classement (dense retrieval – DR – techniques). Deux thèmes principaux transparaissent dans ce livre : les techniques permettant d'aborder le compromis entre l'efficacité (le résultat, la qualité) et l'efficacité (la latence des requêtes, le modèle et la taille de l'index).

Dans le chapitre 1, les auteurs introduisent la notion de *deep learning* (*apprentissage profond* ou *apprentissage en profondeur*) à travers plusieurs modèles : les modèles basés sur la représentation (cette classe de modèles apprend des représentations vectorielles des requêtes et des documents qui peuvent être comparées au moment du classement pour calculer les scores de pertinence requête document), les modèles basés sur les interactions (cette classe de modèles capture explicitement les « interactions » entre les termes de la requête et les termes du document), les modèles hybrides (qui se composent d'un des éléments des deux premiers modèles).

Le chapitre 2 commence par caractériser plus formellement le problème de classement de textes, en énumérant explicitement les hypothèses sur les caractéristiques des données en entrée et en sortie. Les auteurs adoptent la

perspective de l'accès à l'information en se concentrant spécifiquement sur le problème du classement de textes en fonction de leur pertinence par rapport à une requête particulière. Du point de vue de l'évaluation, ce livre se concentre sur ce que l'on appelle communément le paradigme de Cranfield, une approche de l'évaluation orientée système des systèmes de recherche d'informations (SRI).

Le chapitre 3 met en évidence les aspects de BERT qui sont importants pour expliquer ses applications au classement de textes mais n'est pas un tutoriel du modèle. La formulation la plus simple et la plus directe du classement de textes consiste à convertir la tâche en un problème de classification de textes, puis à ordonner les textes à classer en fonction de la probabilité que chaque élément appartienne à la classe souhaitée. Pour les problèmes d'accès à l'information, la classe souhaitée comprend des textes qui sont pertinents pour le besoin d'informations de l'utilisateur, et cette approche peut être utilisée en tant que classification de pertinence. Plus précisément, l'approche consiste à entraîner un classifieur pour estimer la probabilité que chaque texte appartienne à la classe « pertinente », puis au moment du classement (c'est-à-dire de l'inférence) à trier les textes en fonction de ces estimations. Cette approche représente une réalisation directe du *Probability Ranking Principle*, qui stipule que les documents doivent être classés par ordre décroissant de la probabilité estimée de pertinence par rapport au besoin d'informations. C'est notamment la façon dont a été utilisé BERT dans sa première application en 2019.

Dans le chapitre 4, les auteurs présentent un certain nombre de techniques basées sur des transformateurs préentraînés qui opèrent sur les représentations *textuelles* des requêtes et des documents. Celles-ci peuvent être caractérisées comme des techniques d'interrogation et d'expansion de documents, qui ont une riche histoire dans la recherche d'informations, remontant à plusieurs décennies. La section 4.1 commence par un bref aperçu, mais les auteurs ne se veulent pas exhaustifs. Au lieu de cela, ils se concentrent uniquement sur les préalables nécessaires pour comprendre l'expansion des requêtes et des documents dans le contexte des modèles de transformateurs préentraînés. La discussion sur l'expansion des requêtes et des documents dans ce chapitre se déroule comme suit : remarques générales puis dans la section 4.2, techniques d'expansion des requêtes en utilisant des commentaires de pseudo-pertinence qui tirent parti des modèles basés sur les transformateurs. Les auteurs présentent ensuite quatre techniques d'expansion de documents : doc2query, DeepCT, HDCT et DeepImpact. Toutes ces techniques se concentrent sur la manipulation de représentations terminologiques (*i.e.* textuelles) des requêtes et des textes du corpus. Enfin, en dernière section de ce chapitre, les auteurs portent leur attention sur les techniques qui manipulent les requêtes et les représentations textuelles qui ne sont pas basées directement sur le document textuel.

Le chapitre 5 traite des transformateurs pour générer des représentations de textes qui conviennent au classement dans un cadre supervisé (*representation learning*). Les auteurs commencent ce chapitre en identifiant les liens entre les problèmes de pertinence et de similarité textuelle. En particulier, dans une perspective de classement, le principal défi reste le problème de l'estimation de la relation entre deux morceaux de texte. De la même manière que la recherche par

mot-clé nécessite des index et une infrastructure associée pour prendre en charge le classement *top-k* en utilisant des correspondances exactes sur un grand corpus, le classement *top-k* en termes d'opérations de comparaison vectorielle simples telles que les produits internes sur des représentations denses nécessite également une infrastructure dédiée. Les auteurs traitent de ce problème de la recherche du plus proche voisin (*nearest neighbor search*) dans la section 5.2. Comme pour les techniques de reclassement neuronal, il est utile de discuter du développement historique en termes de modèles pré-BERT et post-BERT dans la section 5.3. La section 5.4 introduit la notion de *bi-codeurs* : cette section se concentre sur les *bi-encodeurs* « simples », où chaque texte du corpus est représenté par un seul vecteur, et le classement est basé sur des opérations de comparaison simples telles que les produits scalaires. La section 5.5 présente des techniques qui améliorent la conception de base du *bi-codeur* de deux manières : chacun des textes du corpus peut être représenté par plusieurs vecteurs et le classement peut être effectué à l'aide de comparaisons plus complexes entre les représentations. Ces techniques visent des compromis entre efficacité et rendement différents par rapport aux *bi-codeurs* « simples ». La section 5.6 traite des techniques de récupération dense qui tirent parti de la distillation des connaissances. Enfin, les auteurs discutent des défis ouverts et des spéculations sur ce qui nous attend.

Le chapitre 6 conclut sur des questions de recherche ouvertes. Bien que les architectures de transformateurs et les techniques de pré-entraînement soient des innovations récentes, de nombreux aspects de leur application au classement de textes sont relativement bien compris et représentent des techniques établies. Cependant, il reste encore de nombreuses questions de recherche, et donc en plus de jeter les bases des transformateurs préentraînés pour le classement des textes, ce livre tente également de prévoir où peut se diriger le domaine.

L'ouvrage de Rodrigo Nogueira, Jimmy Lin et Andrew Yates est une étude fouillée, précise et extrêmement bien documentée du classement de textes (*text ranking*). La couverture de l'ensemble du domaine, ainsi que la mise en perspective des techniques utilisées rendent la lecture très attrayante. Nous recommandons donc chaudement la lecture de ce livre à tout chercheur en traitement automatique des langues, quel que soit son domaine de spécialité.

Anders SØGAARD. Explainable Natural Language Processing. Morgan & Claypool publishers. 2021. 108 pages. ISBN : 978-1-636-39215-8.

Lu par **Marie CANDITO**

Université Paris Cité / LLF, UMR CNRS Paris Cité

Le TAL neuronal a drastiquement accru l'opacité des modèles. Les tentatives pour comprendre le comportement d'un système de TAL, ou plus généralement un modèle neuronal, sont devenues un domaine de recherche, dont le nom même est fluctuant : l'auteur utilise les termes explainable (neural) NLP (TAL (neuronal) explicable), avec explainable

explicitement synonyme de interpretable, plus que model explainability (explicabilité des modèles), ainsi que de model explanation (explication de modèles), terme que nous retiendrons. L'émergence de ce domaine s'accompagne d'une redondance, entre chercheurs issus de traditions scientifiques légèrement différentes, qui peuvent présenter les mêmes idées sous des termes différents. D'où la proposition d'une taxonomie des méthodes d'explication de modèles neuronaux utilisés en TAL, avec l'objectif de faciliter les rapprochements entre différents travaux et de pointer les problèmes qui demeurent ouverts dans ce domaine, ce qui, pour l'auteur, devrait permettre in fine d'accélérer les recherches dans ce domaine. Cet objectif est important à souligner, car il explique pourquoi l'ouvrage n'est pas très pédagogique : il s'agit d'un ouvrage pour des chercheurs du domaine, destiné à mieux éclairer leurs recherches.

Dans le chapitre d'introduction, l'auteur commence par présenter deux distinctions usuelles concernant les méthodes d'explication. La première distingue les méthodes « locales » *versus* « globales » : une méthode est dite globale si elle nécessite d'utiliser forcément tout un échantillon d'exemples. Au contraire, une méthode sera dite locale si elle peut fournir des explications pour des exemples individuels. Il s'agira typiquement d'expliquer les sorties pour un exemple particulier (par exemple pourquoi tel verbatim a été classé comme négatif). Une méthode globale cherchera plutôt à mettre au jour des caractéristiques globales d'un modèle (par exemple, s'il comporte des biais, quelle est la densité des représentations qu'il construit, etc.).

La deuxième distinction concerne les méthodes d'explication intrinsèque (prédictions conjointes des sorties du modèle et des explications) *versus* « *post hoc* » (qui utilisent des techniques orthogonales aux modèles à expliquer). L'auteur va ensuite motiver une nouvelle taxonomie. La section 1.2 décrit rapidement huit taxonomies existant dans la littérature très récente, et en pointe les incohérences et incomplétudes. Là encore, cette partie s'adresse à qui connaît déjà très bien les travaux en TAL explicable.

L'auteur ne retient pas comme caractéristique première la forme des explications (par exemple, explication sous forme de visualisation graphique, de coefficients, de règles), qui est en réalité orthogonale aux méthodes, une même méthode pouvant fournir des explications sous différentes formes. La nouvelle taxonomie est ensuite présentée rapidement : l'auteur reprend la distinction locale et globale, mais plutôt que la distinction « intrinsèque » *versus* « *post hoc* », il choisit la distinction entre méthodes « avant » *versus* méthodes « arrière », faisant référence à la propagation avant des réseaux, *versus* la rétropropagation : les méthodes arrière utilisent essentiellement les gradients de la fonction de perte. Les méthodes « avant » sont subdivisées en trois sous-cas, selon qu'elles expliquent des représentations intermédiaires, des représentations continues ou des représentations discrètes. Cette subdivision n'est pas étanche et est destinée à structurer la présentation des méthodes. Ceci donne au final $2 \text{ (local/global)} * 4 = 8$ catégories de méthodes.

Le chapitre 2 introduit très rapidement les architectures principales du TAL neuronal (classification non linéaire, encodage récurrent et encodage à l'aide de *transformers*) et détaille les caractéristiques des huit catégories de méthodes.

Ensuite, chacune des huit catégories donne lieu à un chapitre (chapitres 3 à 10). Pour chaque chapitre, l'auteur liste diverses méthodes, en présentant rapidement la technique formelle, puis un certain nombre de travaux de TAL ayant utilisé ladite méthode. La description technique est très succincte, la valeur ajoutée est clairement dans la liste des travaux cités pour chaque type de méthode.

Les méthodes « arrière » (utilisant une propagation arrière, pour calculer les gradients ou calculer des valeurs de « pertinence » de chaque neurone) :

- locales (applicable à un exemple précis) : la technique de base consiste simplement à étudier le gradient de la perte par rapport à l'exemple d'entrée (*vanilla gradients*), pour en déduire les traits de l'entrée expliquant le mieux la sortie. (Bien que l'auteur n'utilise pas ce terme, on retrouve le concept de *feature attribution*, qui prend un exemple en entrée et assigne un score aux traits selon la contribution que ce trait apporte dans la sortie du modèle.) Avec la *layer-wise relevance propagation*, des valeurs de pertinence de *chaque* neurone sont rétropropagées de couche en couche ;

- globales (nécessitant un échantillon d'exemples) : les méthodes globales arrière sont majoritairement fondées sur l'idée que des modèles plus parcimonieux seront mieux interprétables. Aussi, l'auteur inclut-il ici des méthodes d'élagage de modèles (où une partie des poids sont supprimés). On suppose que toute méthode d'explication sera plus parlante si elle est appliquée sur un modèle élagué. L'idée la plus ancienne est de supprimer les poids correspondant à de faibles dérivées partielles. L'élagage est fait soit après apprentissage (*post-hoc unstructured pruning*), soit par réapprentissage de candidats modèles élagués (*lottery tickets*), soit conjointement à l'apprentissage (*dynamic sparse training*).

Les méthodes « avant » (utilisant une propagation avant au sein du réseau complète ou partielle) :

- expliquant des représentations intermédiaires :

- locales : étude des *gates* ou des poids d'attention pour un exemple donné,

- globales : étude de quel *gate* ou quelle tête d'attention code telle ou telle information. Élagage de *gate* et de têtes d'attention ;

- expliquant des représentations continues :

- locales : l'auteur classe ici d'une part, les études des vecteurs statiques de mot (corrélation avec jugements humains de similarités, analogies entre mots...) et, d'autre part, l'étude de la dynamique des valeurs d'activation, au sein d'une séquence. La distinction d'avec les représentations intermédiaires n'est pas très claire ici,

- globales : les méthodes qui analysent tout un nuage de points, chaque point étant une représentation continue d'un exemple de l'échantillon (comme un nuage de mots). L'analyse peut être un *clustering*, ou bien la corrélation avec un autre nuage de points, issus d'un autre système ou de mesures humaines (par exemple des points issus d'enregistrements fMRI). Sont également classées ici les sondes

linguistiques (des classifieurs appris sur des représentations issues de réseaux, pour prédire des propriétés linguistiques). Les *concepts activation vectors* dépassent l'attribution de traits en permettant une attribution de concepts. Il s'agit de généraliser les exemples d'entrée, en entraînant des classifieurs prédisant des concepts prédéfinis, puis de quantifier pour chaque concept dans quelle mesure il explique la prédiction de telle ou telle classe ;

– expliquant des sorties discrètes :

- locales : l'auteur classe ici les méthodes utilisant des jeux de données linguistiques difficiles ou d'intérêt, comme des constructions linguistiques particulières. Se retrouve classée ici, par exemple, la technique pour tester la capacité à encoder l'accord sujet verbe, utilisant directement les probabilités d'un modèle de langue du verbe s'accordant *versus* la forme violant l'accord. La taxonomie montre ses limites, ainsi le lien avec les autres sondes linguistiques citées *supra* n'est pas fait. Une autre technique célèbre classée ici est LIME (*Local Interpretable Model-agnostic Explanations*), qui permet d'expliquer un modèle boîte noire, au moyen de classifieurs locaux, intrinsèquement interprétables (comme des arbres de décision), entraînés sur des paires entrée et prédiction du modèle à expliquer. Enfin, sont citées ici les techniques pour identifier les « exemples influents », c'est-à-dire ceux dont la suppression à l'apprentissage a le plus d'impact sur le modèle appris,

- globales : la technique d'*uptraining* peut être utilisée pour entraîner un modèle *m'*, intrinsèquement interprétable, sur un grand volume de données annotées *via* le modèle *m* à expliquer. L'interprétation du modèle *m'* servira d'interprétation de *m*.

L'auteur cite également des techniques d'analyse des performances du modèle sur différents jeux de données, ce qui peut aider à caractériser le comportement d'un modèle.

Le chapitre 11 traite des techniques pour évaluer les méthodes d'explication. Le chapitre 12, intitulé « *Perspectives* », commence par des observations générales, mais très techniques, sur les différentes catégories de méthodes, qui ne sont pas vraiment des perspectives.

Dans la dernière (courte) section du chapitre, l'auteur prend du recul, en abordant la question des motivations du TAL explicable, avec un parallèle entre l'explication de décisions « neuronales » et l'explication de décisions humaines. Selon l'auteur, parmi les motivations pour une explicabilité des modèles neuronaux (analyse d'erreurs, maintenance des modèles, amélioration de l'efficacité, détection de la vulnérabilité à des attaques), la motivation la plus citée est le « droit à l'explication ». L'auteur s'interroge sur les motivations de ce droit à l'explication : d'un point de vue « moral », qu'est-ce qui le justifie ? L'auteur semble sceptique quant au bien-fondé de ce droit, en défendant rapidement l'idée que, malgré leurs limites, les méthodes d'explication des modèles neuronaux permettent finalement un niveau d'explicabilité que n'ont pas les décisions humaines elles-mêmes.

Pour conclure sur l'ouvrage, la présentation de la taxonomie et des méthodes relevant de chaque catégorie est rapide et technique, et s'adresse à un public connaissant bien le domaine. Plutôt qu'une taxonomie cherchant à assigner une seule catégorie à chaque méthode, une classification multifacette serait peut-être mieux adaptée. En particulier, typer les méthodes d'explication en fonction des objectifs généraux (Quels traits expliquent une sortie ? Quels concepts expliquent une sortie ? Quels concepts sont encodés dans des paramètres ?) aiderait la lectrice ou le lecteur à aborder plus facilement ce domaine foisonnant.

Beata BEIGMAN KLEBANOV, Nitin MADNANI. Automated Essay Scoring. Morgan & Claypool publishers. 2021. 294 pages. ISBN : 978-1-636-39224-0.

Lu par **Laurie ACENSIO**

Lexiane Formation (Paris)

Considéré comme étant une alternative à l'évaluation manuelle des enseignants, la notation automatisée des essais permet de fournir aux apprenants une évaluation instantanée. En complément des techniques d'apprentissage automatique, cet ouvrage aborde les techniques de traitement du langage (TAL) davantage adaptées pour la notation automatisée des productions écrites. Des travaux scientifiques à travers des cas pratiques (tests de langue, tâches ouvertes, questions à réponse élaborée) sont exposés dans un cadre universitaire et industriel en mettant en évidence les aspects multidimensionnels (conceptuel, méthodologique et technique) pour automatiser l'évaluation des compétences de l'expression écrite au sein d'un système de notation.

L'ouvrage est structuré autour de cinq parties : l'introduction, les bonnes pratiques pour construire un système de notation automatisé simple associé à des cas d'expérimentation, un état de l'art scientifique sur les différents modèles de notation, l'implémentation et l'évaluation des systèmes, les méthodes d'argumentation (rétroaction, analyse des contenus et discours), puis une dernière partie est consacrée aux discussions et aux perspectives de recherche.

La première partie aborde l'approche historique de ce domaine de recherche à travers les travaux de Page qui a développé le premier système de notation automatisée nommé PEG (*Project Essay Grade*). La technique est basée selon une approche simple : une phase d'entraînement et une phase de notation en utilisant un ensemble de coefficients de corrélation pour attribuer un score comparé par la suite à une notation humaine. Acquis par l'entreprise Measurement Inc, le PEG a progressivement intégré de nombreuses caractéristiques intrinsèques liées à la qualité de l'écriture (fluidité, diction, grammaire) en utilisant des techniques d'analyse sémantique et syntaxique. Peu utilisé à ses débuts, cet outil l'est massivement aujourd'hui au sein des écoles et universités américaines, notamment lorsque les productions écrites se sont informatisées à partir des années 1990. Ainsi, la composante arbitraire réelle ou perçue de l'évaluation manuelle s'est atténuée progressivement afin de garantir davantage d'objectivité à travers des critères d'évaluation instrumentés par des outils d'aide à la notation. Néanmoins, il apparaît

que la notation automatisée reste un outil d'assistance lors du jugement d'évaluation humain, mais ne peut se substituer pleinement à celui-ci.

La deuxième partie décrit les différentes étapes pour construire un système de notation informatisée : la collecte de données à partir d'un corpus écrit (par exemple des résultats de tests de langue comme le TOEFL¹ ou ESOL²), la phase de modélisation dont le jugement humain s'avère être déterminant, les critères d'évaluation ainsi que les expérimentations. L'automatisation du processus de notation implique de définir des attributs observables et des critères d'évaluation. Or, il apparaît que les approches statistiques basées sur une corrélation élevée entre les attributs ne sont pas forcément pertinentes au niveau de la qualité de l'écriture. De plus, l'analyse textuelle (fréquence et longueur des mots) et celle au niveau de la surface de texte (grammaire et orthographe) s'avèrent rapidement insuffisantes pour évaluer les compétences associées à l'expression écrite. Progressivement, de nouveaux critères linguistiques ont été mis en place pour s'adapter à la complexité des réponses, l'un des auteurs les plus cités dans l'ouvrage propose une analyse multicritère pour déterminer un score global (ou holistique) lors de l'analyse de la production écrite (style, clarté, argumentation, cohérence et pertinence).

La troisième partie constitue la majeure partie de l'ouvrage avec une description des modèles et des techniques utilisées : la régression linéaire ainsi que l'analyse sémantique latente (*Latent Semantic Analysis*) sont les techniques les plus implémentées dans les logiciels. Plus récemment, les techniques d'apprentissage profond (*Deep Learning*) démontrent une nette amélioration des résultats, dont les travaux de Taghipour qui exploite un réseau neuronal pour améliorer la précision des scores holistiques en comparant les résultats avec d'autres techniques (classification hiérarchique et classification de texte bayésienne). Néanmoins, la complexité de traitement des réseaux de neurones est un obstacle pour expliciter les scores prédits, notamment à travers les commentaires afin d'indiquer les points d'amélioration pour l'apprenant.

Les chapitres suivants abordent les fonctionnalités génériques, les différents types d'essais (l'écriture argumentative, l'écriture narrative et l'écriture de réflexion), la phase de production d'un système de notation à travers un exemple d'architecture illustré par un schéma global et des extraits de code source.

La quatrième partie aborde les systèmes de rétroaction (*feed-back*) notamment en relation avec le chapitre 5 de la partie précédente concernant la qualité de l'écriture argumentative et la qualité de réflexion. Puis, la notation automatisée du contenu et du discours est abordée à travers quatre chapitres dans la perspective d'une personnalisation des rétroactions. Cette phase de rétroaction pédagogique est une étape déterminante constatée à partir d'une méta-analyse dont l'ensemble des travaux démontre que cette forme de révision a des effets particulièrement positifs sur la qualité d'écriture des apprenants lors de l'évaluation des reformulations (ou corrections). Néanmoins, la rétroaction doit être explicitée afin de s'assurer une

1 *Test of English as a Foreign Language.*

2 *English for Speakers of Other Languages.*

bonne compréhension de l'apprenant. Il apparaît que les corpus annotés nécessaires et qui impliquent d'être construits dans un domaine de connaissance spécifique sont peu disponibles pour optimiser cette étape d'évaluation.

La cinquième partie finalise l'ouvrage en mettant en évidence les défis de la notation automatique des essais, dont l'évaluation de l'écriture dans plusieurs langues, la standardisation des tests, la validité des textes, l'interprétabilité du modèle, la recherche de l'équité lors de l'attribution des scores, l'omniprésence et l'évolution constante de la technologie. Il apparaît que la principale utilité de la notation automatisée des essais est le gain de temps pour les évaluateurs, l'élimination de préjugés humains et la différence de perception afin d'assurer l'équité dans la notation.

Cet ouvrage démontre que la notation automatisée des essais est un sujet de recherche actif s'appuyant sur de nombreuses références bibliographiques essentiellement issues de revues internationales, mais sans la présence d'équipes de recherche en France. Il peut s'adresser à des chercheurs en TAL intéressés au domaine d'application liée à l'éducation impliquant des questions de recherche tout autant en didactique qu'en linguistique. En effet, la rétroaction pédagogique est un défi considérable afin d'identifier les erreurs et, par conséquent, de mieux cibler les points d'amélioration de l'apprenant qui ne sont pas détaillés lors de l'attribution d'un score de type holistique. Actuellement, les travaux restent focalisés sur les techniques utilisées pour améliorer la qualité de prédiction des scores holistiques avec notamment le potentiel de l'apprentissage profond (ou réseaux neuronaux). Néanmoins, ils abordent peu les enjeux de l'IA explicable afin d'aboutir à des systèmes de notation automatisés transparents et impartiaux. Les techniques de TAL se sont imposées du fait de la complexité croissante des tâches d'évaluation d'apprentissage notamment à travers les questions ouvertes ou les questions à réponse élaborée. Les problématiques sont d'ordre syntaxique, lexical et sémantique afin de détecter principalement des erreurs de l'apprenant, mais demeurent largement étudiées préalablement dans d'autres contextes applicatifs (par exemple l'aide au diagnostic médical). Néanmoins, les enjeux de ce domaine de recherche impliquent de prendre en considération le style et le type d'écriture (argumentation, réflexion, narratif et créatif) lors du traitement du contenu textuel soulevant ainsi des questions originales liées au processus cognitif humain associées à une représentation du monde et des connaissances. En effet, l'évaluation de la qualité argumentative d'une production écrite soulève des défis relativement nouveaux consistant à prendre en compte la créativité et la subjectivité humaine. Au-delà du cadre éducatif, cet axe de recherche aborde des questions communes avec l'extraction et l'analyse automatique des arguments (*argument mining*) ou des opinions (*opinion mining*) au sein de corpus textuels.

Philippe BARBAUD. L'instinct du sens – Essai sur la préhistoire de la parole. Éditions *Des auteurs des livres*. 2021. 342 pages. ISBN : 978-2-9570999-9-3.

Lu par **Georgeta CISLARU**

Paris Nanterre / MoDyCo

Cet ouvrage se propose de répondre à la question, complexe, des origines du langage, que l'auteur articule à la non moins complexe question des origines du sens. Malgré cette double complexité, le texte se laisse lire avec aisance, en raison à la fois de l'avant-propos, qui pose explicitement les jalons du raisonnement, et de la clarté du style. Les deux grands axes que l'auteur met en avant sont la référenciation et l'énonciation, en tant que terreau de la construction du sens et moteurs de l'émergence du langage articulé, et l'abstraction, en tant qu'artefact culturel sous-tendu par les capacités mémorielles de l'humain, rendant possibles l'encodage morphologique, la grammaticalisation et les relations grammaticales, entre autres.

Le langage sert à parler du monde, mais, avant tout, il sert à vivre, son essence même est de signifier, selon E. Benveniste³. C'est pourquoi la question des origines du langage n'a pour lui aucun sens : « *le langage est aussi ancien, ou aussi primordial, que la signification elle-même, et l'on ne saurait imaginer un homme qui ne posséderait pas la faculté fondamentale de donner un sens aux choses, c'est-à-dire de parler* »⁴. Dans son ouvrage, P. Barbaud s'emploie à donner du sens au chemin parcouru par l'humain depuis ses origines, sens qui se trouve sémiotisé par le langage.

L'ouvrage est constitué d'un avant-propos et de trois chapitres détaillés qui s'appuient sur des publications issues de nombreux champs des sciences du langage⁵. Les enjeux et les positionnements théoriques sont explicitement formulés dès l'avant-propos, ce qui facilite la lecture.

Le premier chapitre esquisse l'évolution de 2,5 millions d'années et la renvoie simienne pour les placer dans une perspective darwinienne en termes d'adaptation, de sélection naturelle et d'évolution. L'instinct animal de la communication régit des comportements réflexifs en réponse à des signaux, tandis que le délitement de la

3 Benveniste É., « La forme et le sens dans le langage », *Problèmes de linguistique générale II*, Paris, Gallimard, p. 215-238, 1974.

4 Mosès S., « Émile Benveniste et la linguistique du dialogue », *Revue de métaphysique et de morale* ; n° 32 (4), p. 509-525, 2001.

5 Comme cela arrive souvent, le fait d'impliquer, au fil de l'argumentation, des problématiques diverses conduit parfois à des « points aveugles », par exemple à ne pas convoquer les travaux de Willy Van Langendonck pour défendre une conception catégorielle des noms propres qui ont un rôle à jouer dans l'interpellation comme étape de l'émergence du langage.

rengaine, relevant de l'hérédité régressive, et l'exaptation des organes phonatoires répondent à la pression adaptative exercée par la nécessité de faire sens, donnant lieu à ce que l'auteur appelle la « naissens » de la parole. L'auteur s'intéresse plus particulièrement à la période transitoire constituée d'imitations de sons de la nature, de babils et autres émergences phoniques précédant l'invention des premiers mots (et, donc, des premiers symboles). C'est la curiosité humaine et le discernement de référents pertinents, détachés perceptivement et émotionnellement du contexte environnant, qui permettent cette invention, qui font aussi que, selon l'auteur (suivant en cela P. Ricœur), le sens est inséparable de la référence.

La thèse principale du deuxième chapitre est l'évolution, grâce à l'hérédité régressive, de la rengaine animale en signes dotés de sens. Les protolangues seraient ainsi des produits de l'esprit-cerveau. L'auteur distingue quatre phases du langage : interjective, vocative, objective et énonciative. Dans le processus évolutif, la production des outils comme la taille des galets s'accompagnant d'interjections exprimant des émotions (douleur, joie, surprise) et des interpellations de l'autre.

L'avant-signes interjectif aurait ainsi contribué à installer un habitus phonatoire, pour ensuite implémenter le sens dans la cognition humaine, en passant par l'interjection vocative suivie de ce que l'auteur appelle la conquête du référent objectal détaché du moi et qui deviendra l'autre. C'est donc dans l'altérité que le sens émerge et, avec lui, le langage tel qu'on le connaît. La prégnance et la saillance⁶, en tant que principes de discrimination perceptuelle (système DISPER), constituent des éléments déterminants de la stabilisation du référent objectal. Le passage du référent au signifié implique le passage d'une mémoire autobiographique à une mémoire sémantique.

S'ensuit l'émergence de la première et de la deuxième personne, communes à toutes les langues, comme pendant de la fonction conative du langage : moi s'énonce et interpelle l'autre, en faisant ainsi évoluer les interjections vers des formes pronominales, puis des noms propres. Dans cette optique l'autre est le non-allocutaire, la non-personne qui peut, de ce fait, être intégrée à la catégorie des personnes. L'émergence des toponymes viendrait par ailleurs configurer les dyades *moi-ici* (et *eux-là-bas*). La diversité des référents saillants conduit à la généralisation des noms communs. En dénommant, l'humain présapiens confère « *son signifié descriptif à un signifiant référentiel appartenant à l'univers du "ça" dérivé de celui de la "non-personne"* ». Pour reprendre les termes de l'auteur, la référence argumentale vient compléter la référence temporelle induite par le dessein. En parallèle, aurait eu lieu une implémentation psychocérébrale de la symbiose entre la perception et la phonation, ouvrant ainsi la voie aux développements articulatoires, du monosyllabique au polysyllabique. L'auteur fait coïncider l'émergence de l'articulation monosyllabique avec la maîtrise du feu, intervenue il y a ± 500 000 ans, qui a permis à l'humain présapiens de dépasser sa condition animale.

6 Que l'on retrouve, par exemple, dans certaines théories de la référenciation et de l'anaphore.

Dans le troisième chapitre, l'auteur tient à dissocier les dynamiques de l'émergence de la parole de la description de l'état actuel du langage. On perçoit néanmoins, dans les analyses et les spéculations qui sont proposées en marge de l'encodage syntaxique, des éléments permettant de comprendre des préférences structurelles, comme l'ordre SOV dans la phrase.

La récursivité apparaît comme un moteur de l'évolution et de la structuration des langues. La récursivité gouverne la mémoire et, grâce à des outils lexicaux tels les hyperonymes, domine la grammaire. Dans le même esprit, P. Barbaud met en avant le rôle de la mémoire plutôt que celui de la grammaire universelle dans l'enseignement et l'apprentissage des langues. C'est le même mécanisme qui entre en jeu lorsque des lexèmes commencent à fonctionner comme des morphèmes, à commencer par les classificateurs propres à plusieurs langues asiatiques, amérindiennes, etc. La morphologie est ainsi, à la différence du lexique⁷, le produit artefactuel du système et non du cerveau-esprit ; il s'agit de la manifestation d'une évolution culturelle qui favorise l'abstraction et qui entraîne dans le sillage la grammaticalisation.

En filigrane, on reconstitue une représentation du langage et des langues ancrés dans une sémantique se détachant des formalismes. Pour l'auteur, la combinaison de signifiants est régie par des mécanismes indépendants de ceux qui régissent la combinaison des signifiés qui leur sont associés, tandis que la structure n'est pas garante de sens.

Au terme des 327 pages riches en notes et commentaires, complétées par un glossaire et deux index, on apprécie la diversité des références bibliographiques, couvrant plusieurs champs de la linguistique et des disciplines connexes comme la psychologie, les sciences cognitives, etc. Des encadrés, des notes de fin de chapitre et des schémas accompagnent le propos, de sorte que certaines pages donnent à l'ouvrage une valeur ressource indéniable. Que l'on soit ou non en accord constant avec le propos développé par l'auteur, on apprécie également la réflexion de P. Barbaud, qui n'hésite pas à scruter sous des angles inédits des aspects ponctuels ou des affirmations plus générales issus des différents travaux cités.

7 Dont elle reste partie prenante, comme objet de mémoire collective.