

## Séminaire International de Sémiotique à Paris (2024-2025)

### *Intelligence artificielle générative et nouveaux enjeux sémiotiques.*

#### *Traduction et appropriations créatives*

Séance du 15 janvier 2025

**Carine Duteil-Mougel (Université de Limoges)**

*« Remonter » la génération des discours à partir de la description de la structuration des textes : contributions de la textométrie à la compréhension de l'IA générative. PART 1*

Notre propos entend questionner la démarche des IA génératives à l'aune de la complexité de la communication langagière. Alors qu'une vision de l'information cybernétisée confrontée à « la pensée complexe » telle que la définit Edgar Morin, semble mener à une aporie, nous chercherons, dans une perspective de linguistique appliquée, à la dépasser. La notion d'information, telle que nous la connaissons aujourd'hui, a été développée principalement par Hartley, puis approfondie par Shannon et Weaver. Cette conception de l'information repose sur deux aspects fondamentaux : (i) l'aspect communicationnel : il s'agit de la transmission de messages, intégrée dans une théorie plus large de la communication ; (ii) l'aspect statistique : il concerne la probabilité (ou plus précisément l'improbabilité) d'apparition d'unités élémentaires porteuses d'information, appelées *binary digits* ou *bits*. Cette théorie de l'information a trouvé sa première application dans le domaine même qui l'a vue naître : les télécommunications. Et elle continue d'alimenter les applications en traitement automatique du langage naturel (*NLP*). Or, comme le souligne Morin (2014), l'aspect communicationnel ne rend absolument pas compte du « caractère polyscopique de l'information, qui se présente au regard tantôt comme mémoire, tantôt comme savoir, tantôt comme message, tantôt comme programme, tantôt comme matrice organisationnelle. » (p. 36) ; et « l'aspect statistique ignore, y compris même dans le cadre communicationnel, le sens de l'information, il ne saisit que le caractère probabilitaire-improbabilitaire, non la structure des messages et, bien entendu, ignore tout de l'aspect organisationnel. » (p. 38). De plus, les méthodes distributionnelles, suivies par les IA génératives, qui apprennent à prédire des distributions de probabilités sur les séquences de mots ou d'actions possibles, s'affrontent au désordre, à la « néguentropie » et au processus de réorganisation et de re-conception propre aux systèmes complexes, tels que le langage. Bien que l'on parle d'« émergence sémantique » depuis l'avènement de l'architecture *Transformer* dans les IA, de nombreux défis persistent, notamment en ce qui concerne la détection des nuances sémantiques, la compréhension des intentions pragmatiques et la prise en compte des processus d'interprétation. Des pistes d'amélioration peuvent-elles exister ? Évoquons quelques pistes telles que l'annotation des données d'entraînement avec des informations sémantiques et pragmatiques, telles que les intentions des locuteurs, les émotions, et les relations entre les concepts, afin d'enrichir la compréhension du modèle, voire même la création d'*embeddings* culturels et pragmatiques (en lien avec les pratiques discursives). Est-ce réaliste ? Souhaitable ?

**Julien Longhi (CY Cergy Paris Université)**

*« Remonter » la génération des discours à partir de la description de la structuration des textes : contributions de la textométrie à la compréhension de l'IA générative. PART 2*

L'analyse du discours (AD), déjà largement outillée et souvent assistée par des technologies informatiques, a connu une transformation significative avec l'émergence et la démocratisation de l'intelligence artificielle (IA) et des modèles de langage. Toutefois, à l'instar des étapes précédentes de son développement, l'usage de ces nouvelles technologies doit être accompagné d'une réflexion épistémologique approfondie sur les types de modélisations impliqués. En nous appuyant sur les travaux de Jean-Guy Meunier (2017, 2019, 2020, entre autres) ainsi que sur diverses recherches récentes en linguistique de corpus et en fouille de texte, nous nous proposons de décrire certaines caractéristiques des approches discursives attentives à la description textuelle. Nous identifierons également différentes tendances illustrant plusieurs conceptions de la "génération" des discours.