

Société de Linguistique de Paris  
École Pratique des Hautes Études (EPHE), IV<sup>e</sup> section

La séance aura lieu  
**en présence** à l'EPHE, en Sorbonne, escalier E, 1er étage, **salle Gaston Paris**  
et pourra être suivie **à distance** par Microsoft Teams :

[https://teams.microsoft.com/l/meetup-join/19%3ameeting\\_OWQ3NjM1MWUtYTE3Yy00NzcwLWE3NGEtOWVkn2M2ZjdiNjh%40thread.v2/0?context=%7b%22Tid%22%3a%22970d5a38-c648-47a9-b305-9feb33e86cce%22%2c%22Oid%22%3a%22a462e874-e658-48b0-9956-bb09a021842b%22%7d](https://teams.microsoft.com/l/meetup-join/19%3ameeting_OWQ3NjM1MWUtYTE3Yy00NzcwLWE3NGEtOWVkn2M2ZjdiNjh%40thread.v2/0?context=%7b%22Tid%22%3a%22970d5a38-c648-47a9-b305-9feb33e86cce%22%2c%22Oid%22%3a%22a462e874-e658-48b0-9956-bb09a021842b%22%7d)

**Séance du samedi 19 février 2022 (17h-19h)**

**Le traitement automatique des langues au service de la linguistique :  
quelques travaux à l'interface entre linguistique et informatique**

Benoît Sagot (Inria, Paris)

Le traitement automatique des langues (TAL) est un champ de recherche multidisciplinaire, à l'interface entre linguistique et informatique, dont l'enjeu est de permettre l'analyse, la production et la transformation informatiques de données textuelles. Ses applications ont un impact croissant sur notre quotidien, notamment depuis que le succès récent des réseaux de neurones a fait réapparaître dans le discours scientifique et médiatique le terme d'« intelligence artificielle ». La traduction automatique, les agents conversationnels et l'analyse automatique de grands volumes de textes sont trois exemples parmi d'autres. Toutefois, les applications du TAL ne sont pas toutes du ressort de l'innovation technologique : le TAL, que l'on a pu qualifier de linguistique appliquée, contribue à faire avancer des disciplines comme la médecine ou l'histoire, mais également, dans un retournement remarquable, la linguistique, qui devient alors un domaine d'application du TAL.

Dans cet exposé, après une brève présentation du TAL et de l'évolution remarquable des techniques informatiques qu'il a convoquées au fil des décennies, j'essayerai de montrer deux exemples de travaux où des questions linguistiques ont pu être étudiées grâce à des outils ou des approches relevant du TAL proprement dit et pas seulement de l'outillage informatique. Je montrerai tout d'abord comment des notions relevant de la théorie de l'information ont permis de proposer plusieurs façons de quantifier la notion de complexité morphologique, permettant la comparaison entre analyses concurrentes de systèmes morphologiques et l'émergence de classes flexionnelles à partir des données brutes. Je montrerai ensuite comment une approche contemporaine de la traduction automatique permet de jeter un regard nouveau sur l'évolution du français et de ses graphies au fil du XVII<sup>ème</sup> siècle. Je profiterai de ce que cette approche fait usage de réseaux de neurones et de ce que l'on appelle l'apprentissage profond (*deep learning*) pour présenter au préalable les idées générales qui se cachent derrière ces méthodes mathématico-informatiques complexes. Je présenterai ainsi la notion de modèle de langue, qui fait régulièrement son apparition jusque dans la presse généraliste, et qui est en effet au cœur du TAL contemporain. Je montrerai notamment que de tels modèles retrouvent dans une certaine mesure, à partir de grands volumes de textes bruts et d'aucun présupposé d'aucune sorte, la hiérarchie d'analyse allant de la morphosyntaxe à la syntaxe et à la sémantique.

Je conclurai en mentionnant d'autres travaux où le TAL semble pouvoir contribuer à apporter un éclairage nouveau à des questions linguistiques.

---

**Séances ultérieures de la SLP**

**Samedi 19 mars 2022** : Sylvain KAHANE & Olivier BONDÉELLE « Les particules verbales du wolof et leur combinatoire syntaxique et topologique »

**Samedi 9 avril 2022** : Olga ARTYUSHKINA « Les reduplications nominales et la mise en saillance dans les langues slaves orientales »

**Samedi 14 mai 2022** : Hélène et André WLODARCZYK « Vers la linguistique conceptuelle - théorie et expérimentations interactives »

**Samedi 18 juin 2022** : Pierre LARRIVÉE et Cecilia POLETTO « Ordre des mots, changement syntaxique et micro-indicateurs dans deux langues romanes »

**Samedi 19 novembre 2022** : Sebastian FEDDEN « Les langues papoues : synchronie, diachronie, diversité »

**Samedi 10 décembre 2022** : Alex FRANÇOIS « Tectonique lexicale : Innovations locales et convergence aréale dans les structures sémantiques »