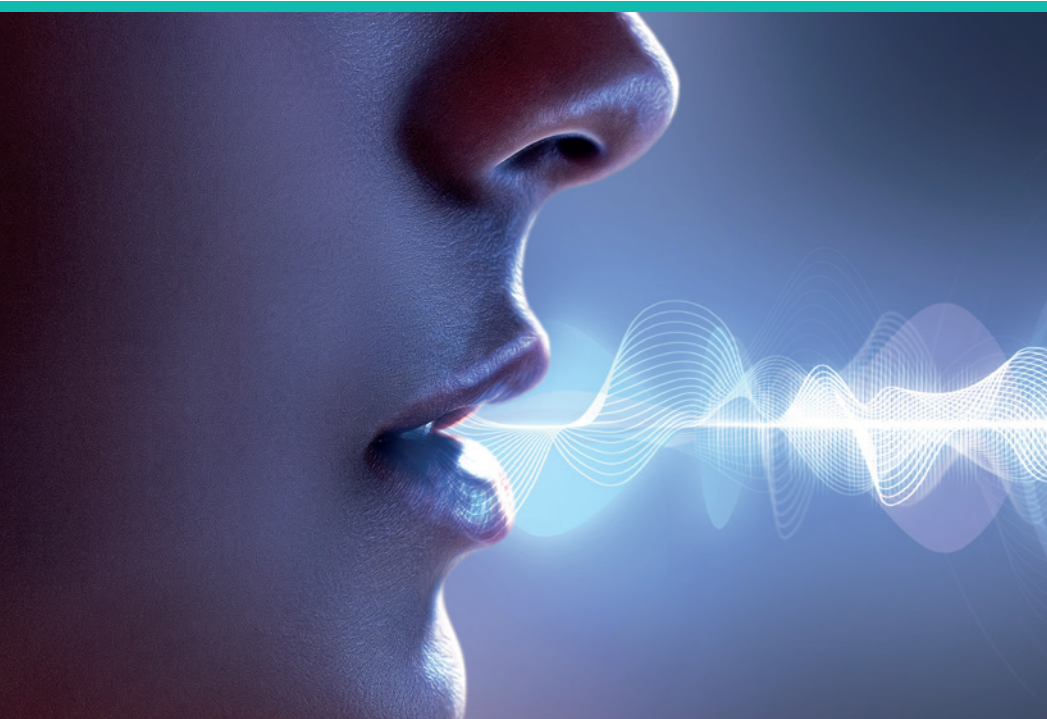


JOURNÉE GRAND PUBLIC / MARDI 12 JANVIER 2021

CNRS Michel-Ange, 3 rue Michel-Ange, Paris



Intelligence artificielle et technologies des langues :

l'ordinateur passe la barrière de la langue



GDR Groupement
de recherche
TAL Traitement automatique
des langues

9h30	Introduction B. Daille, directrice du GdR TAL
9h45	Fouille de publications scientifiques sur le COVID B. Favre
10h30	Ironie, sarcasme et sentiments : les algorithmes dans les subtilités de la langue F. Benamara
11h15	pause
11h30	Mon ordinateur est-il un bon psy ? Le TAL au service du diagnostic médical M. Amblard
12h15	La traduction automatique à l'heure des réseaux de neurones F. Yvon
Repas	
14h15	Le dialogue virtuel au service de la formation M. Ochs
15h00	Quand la science-fiction inspire les technologies de la langue F. Landragin
15h45	Les algorithmes à la chasse aux infox V. Claveau

Introduction

> Par **Béatrice Daille, directrice du GdR TAL / beatrice.daille@univ-nantes.fr**

Le Traitement Automatique des Langues (TAL) est une discipline de l'informatique traitant et produisant des données langagières. Le TAL regroupe un nombre important de chercheurs et d'industriels qui collaborent étroitement et mettent en œuvre des techniques et des outils de l'intelligence artificielle conçus pour la langue. Ces dernières années, le TAL a connu de nombreuses évolutions, principalement dues à la grande vague de l'apprentissage avec des réseaux de neurones profonds. Les applications phares du TAL comme la traduction automatique, le dialogue homme-machine, la reconnaissance de la parole ont fait des progrès considérables et engrangés de beaux succès académiques et industriels. Citons le prix Turing - équivalent du Nobel pour l'informatique - co-décerné en 2019 au canadien Y. Bengio, pour ses travaux d'IA appliqués à la langue. Les équipes de recherche française travaillant sur la langue écrite, orale ou signée se sont constituées en réseau, le GDR TAL, pour répondre aux défis scientifiques à venir et aider à concevoir des applications novatrices, comme celles présentées lors de cette journée.

Béatrice Daille est professeure en informatique à l'université de Nantes et membre du LS2N (Laboratoire des Sciences du Numérique de Nantes) au sein de l'équipe « Traitement automatique des langues naturelles » qu'elle a dirigée de 2002 à 2016. Ses recherches portent sur la définition d'algorithmes facilitant l'accès à l'information dans des textes dans des formats, des domaines et des langues variées. Elle est directrice du réseau de l'INS2I des équipes de recherche travaillant en technologies des langues, le GDR TAL, depuis janvier 2020, après avoir œuvré à sa création.

Pour aller plus loin :

<https://gdr-tal.ls2n.fr/>

Fouille de publications scientifiques sur le COVID

> Par Benoit Favre, LIS/CNRS UMR 7020, Aix-Marseille Université / benoit.favre@lis-lab.fr

Au plus fort de l'épidémie de COVID-19, le NIH (National Institutes of Health) a recensé la publication ou pré-publication de 2 500 articles de recherche sur le sujet par semaine. Ces articles traitent de sujets divers comme les résultats de recherche fondamentaux sur les vaccins, les pratiques des services hospitaliers s'occupant des malades, ou encore l'impact du confinement sur notre société. Comment le corps médical peut-il « digérer » une telle quantité d'information afin d'en extraire des connaissances pertinentes pour la prise en compte au quotidien de l'épidémie ? Cette présentation montrera par quels moyens le traitement automatique du langage peut offrir une aide précieuse pour la veille scientifique, notamment grâce à des systèmes automatiques qui lisent, organisent et résumant les articles publiés au jour le jour dans le monde entier.

Benoit Favre est maître de conférences habilité à diriger les recherches au Laboratoire d'Informatique et Systèmes (LIS/CNRS UMR 7020) à Aix-Marseille Université. Il a obtenu sa thèse à l'université d'Avignon et a fait des séjours de recherche à UC Berkeley, USA et QUT Brisbane, Australie. Ses recherches portent sur l'interaction entre traitement du langage naturel et apprentissage automatique. Il est particulièrement intéressé par le traitement de la langue dans son contexte large, comme par exemple dans le cadre du langage multimodal. Il est membre du comité de lecture de la revue TAL, membre invité du comité permanent de la conférence Traitement Automatique du TALN).

Ironie, sarcasme et sentiments : les algorithmes dans les subtilités de la langue

> Par Farah Benamara, Maître de conférences en informatique à l'université Paul Sabatier, rattachée au laboratoire IRIT / farah.benamara@irit.fr

Le langage évaluatif désigne toute expression du langage qui permet d'exprimer nos points de vue, opinions, émotions, souhaits à propos d'une personne, objet ou événement. Il est omniprésent dans les médias d'aujourd'hui (blogs, commentaires, forums, réseaux sociaux, etc.), ce qui rend crucial le développement d'outils pour extraire, synthétiser et comparer les opinions exprimées par les internautes. Les applications sont multiples, telles que la prédiction des résultats de votes, l'analyse de la satisfaction clients ou encore l'aide à la modération de contenus. Dans ce contexte, l'exposé présentera les travaux sur la détection de l'ironie et du sarcasme ainsi que la détection de messages haineux.

Farah Benamara est Maître de conférences habilitée à diriger des recherches en informatique à l'université Paul Sabatier depuis 2005. Elle a d'abord obtenu le titre d'ingénieur en informatique en 1999, un master option image et langage en 2000, puis une thèse en traitement automatique du langage en 2004. Ces travaux portent sur le développement de modèles pour la compréhension du langage naturel écrit, en particulier sur l'analyse du langage évaluatif, l'analyse du discours et la détection des intentions.

Mon ordinateur est-il un bon psy ? Le TAL au service du diagnostic médical

> Par **Maxime Amblard**, Université de Lorraine - Loria/IDMC / maxime.amblard@univ-lorraine.fr

Dans le secteur de la santé mentale, l'analyse automatique de la parole du patient peut jouer un rôle important. À l'interface de la psychiatrie-psychopathologie, la linguistique, la sémantique formelle et les sciences du numérique, nous étudions les Troubles du Langage et de la Pensée (TLP) pour définir un modèle cognitif et sémantique des Troubles du Discours (TDD). L'IA et le traitement automatique des langues sont utilisés pour produire des indices proposés aux professionnels de santé pour la détection précoce, le diagnostic et le suivi des patients. Au travers de plusieurs projets, nous nous focalisons sur le contenu langagier et sur les enregistrements vidéos. Les objectifs sont à chaque fois de recueillir des données, valider des modèles théoriques et proposer des solutions logicielles pour les médecins.

Maxime Amblard est maître de conférences en informatique à l'Université de Lorraine. Il a réalisé une thèse en informatique à l'Université de Bordeaux 1 après une formation en informatique et mathématiques et est habilité à diriger des recherches depuis 2016. Il est responsable du master international en Traitement Automatique des Langues à l'Institut des Sciences du Digital, Management et Cognition. Ses recherches se concentrent sur l'analyse, la modélisation et le traitement des dialogues. Il est responsable de l'Action exploratoire Inria ODIM et est le co-responsable du projet franco-allemand Inria-DFKI MePheSTO. Ces deux programmes de recherche s'intéressent à l'analyse des interactions entre patients souffrant de schizophrénie et cliniciens afin de produire des outils d'analyse automatique des interactions.

Pour aller plus loin :

- <https://lejournal.cnrs.fr/articles/mieux-diagnostiquer-la-schizophrénie>
- <https://www.francoculture.fr/emissions/la-methode-scientifique/table-ronde-sur-lactualite-des-sciences?s=09>
- <https://www.inria.fr/fr/schizophrénie-analyser-le-discours-pour-mieux-diagnostiquer>
- <https://www.inria.fr/fr/mephesto-lia-au-service-de-la-detection-des-troubles-psychiatriques>

La traduction automatique à l'heure des réseaux de neurones

> Par **François Yvon**, Directeur de Recherche, LIMSI-CNRS / francois.yvon@limsi.fr

Cette intervention vise à faire un état des lieux concernant les derniers développements dans le domaine de la traduction automatique. Nous analyserons en particulier les avancées récentes obtenues grâce à l'adoption de nouvelles architectures d'apprentissage profond (deep learning). Nous discuterons également des nouvelles thématiques qui émergent sous l'impulsion des grands acteurs de l'Internet, en particulier la traduction non supervisée (sans corpus parallèle) et la traduction multilingue (depuis et vers de multiples langues). En nous appuyant sur les résultats expérimentaux des évaluations internationales, nous essaierons de montrer que la question de la traduction automatique n'est pas entièrement résolue, ce qui incite à explorer plus avant d'autres questions de recherche.

François Yvon est chercheur au CNRS et conduit ses recherches au sein du groupe « Traitement du Langage Parlé » du LIMSI à Orsay, laboratoire qu'il a dirigé de juillet 2013 à décembre 2019. Ses travaux récents s'intéressent principalement à la traduction automatique par des méthodes statistiques et neuronales - et plus généralement à l'apprentissage automatique appliqué à des données langagières multilingues, qu'elles soient écrites ou orales. Précédemment, François Yvon a enseigné l'informatique à l'Université Paris-Sud, ainsi qu'à Télécom Paris. Il est titulaire de l'Habilitation à Diriger des Recherches (obtenue en 2006), d'une thèse en informatique (obtenue en 1996). Il est également diplômé de l'Ecole Nationale de la Statistique et de l'Administration Economique (ENSAE) et de l'Ecole Polytechnique.

Quand la science-fiction inspire les technologies de la langue

> Par **Frédéric Landragin, Dr. CNRS, Lattice / frederic.landragin@ens.fr**

L'ordinateur parlant HAL de 2001, l'Odyssée de l'espace (Kubrick, 1968) a fasciné nombre de spectateurs, dont certains sont devenus par la suite chercheurs en intelligence artificielle. Une manière de boucler la boucle, car HAL, imaginé en collaboration avec les chercheurs de l'époque, préfigurait les chatbots et robots de compagnie que l'on voit se développer ces dernières années. Autre exemple (beaucoup moins réaliste), le Babel fish de la science-fiction a donné son nom à un traducteur automatique réel. Dans cette présentation, nous passerons en revue quelques inventions stimulantes de la science-fiction, certaines bien connues comme le droïde de protocole C3PO dans Star Wars, d'autres moins connues mais tout aussi amusantes, comme le comprimé informatico-traducteur à avaler, traducteur automatique universel d'une efficacité redoutable qui, au-delà de sa formule magique, pose des questions essentielles sur les enjeux de la traduction automatique et du TAL en général.

Frédéric Landragin est directeur de recherche au CNRS, spécialisé en linguistique et en traitement automatique des langues. Ses recherches portent sur la compréhension automatique du langage et sur le dialogue entre humain et machine. Il travaille dans le laboratoire Lattice (Langues, Textes, Traitements Informatiques, Cognition) affilié à l'ENS Paris et l'Université Sorbonne Nouvelle. Passionné de science-fiction, il entretient une activité de vulgarisation scientifique, notamment pour les éditions du Béliat' avec des essais sur la linguistique, le TAL et l'IA en science-fiction.

Pour aller plus loin :

- Landragin F. (2020), « Comment parle un robot ? Les machines à langage dans la science-fiction », Éditions du Béliat', collection Parallaxe.
- Landragin F. (2018), « Comment parler à un alien ? Langage et linguistique dans la science-fiction », Éditions du Béliat', collection Parallaxe.
- Landragin F. (2013), « Dialogue homme-machine. Conception et enjeux », Éditions Hermès-Lavoisier.
- Landragin F. (2004), « Dialogue homme-machine multimodal », Éditions Hermès-Lavoisier.
- <https://lejournal.cnrs.fr/billets/que-valent-les-traducteurs-automatiques-de-la-science-fiction>

Le dialogue virtuel au service de la formation

> Par Magalie Ochs, Laboratoire d'Informatique et des Systèmes (LIS), Aix Marseille Université / magalie.ochs@lis-lab.fr

Aujourd'hui, les personnages virtuels sont de plus en plus utilisés dans le domaine de la formation. Ils sont utilisés pour la formation des compétences sociales des individus. Ils permettent de les immerger dans un environnement de réalité virtuelle et de simuler une gestion d'équipe, un harcèlement scolaire, des phobies sociales. Nous présenterons l'agent virtuel qui forme les médecins à l'annonce d'événements indésirables graves. Nous avons développé une plateforme de réalité virtuelle permettant aux médecins de simuler une situation d'annonce d'événements indésirables graves face à un patient virtuel autonome. A travers un casque de réalité virtuelle, le médecin peut dialoguer avec une patiente virtuelle humanoïde réalistes (expressions d'émotion, gestes de douleur). D'autres recherches sont en cours pour la formation à la prise de parole en public face à une audience virtuelle et la sensibilisation aux discriminations sociales.

Magalie Ochs est Maîtresse de Conférences, Habilitée à Diriger des Recherches, en Informatique au Laboratoire d'Informatique et des Systèmes (LIS) à Aix Marseille Université. Elle travaille depuis plus de 10 ans sur l'intégration d'une forme d'Intelligence Artificielle Socio-Emotionnelle dans les systèmes interactifs humanoïdes (avatars, personnages virtuels, robots humanoïdes). L'objectif est d'intégrer dans les machines une capacité de sociabiliser avec l'utilisateur et de simuler des émotions. Après un doctorat à Orange Labs pour le développement de systèmes de dialogue empathique, M. Ochs a travaillé dans différents laboratoires en France (LIP6, Télécom ParisTech) et à l'étranger (NII, Tokyo). Aujourd'hui ses travaux de recherche se concentrent sur le développement de plateformes de réalité virtuelle pour la formation des compétences sociales des individus à travers des systèmes de simulations avec des personnages virtuels autonomes.

Les algorithmes à la chasse aux infox

> Par Vincent Claveau, CNRS, IRISA, Rennes / vincent.claveau@irisa.fr

Les fausses informations sont très présentes sur le Web et les réseaux sociaux, avec des conséquences potentiellement graves pour notre société. D'un point de vue informatique, ces fake news peuvent prendre différentes formes, et leur détection nécessite alors de disposer d'outils adaptés à chacune de ces formes. Nous présenterons comment le traitement automatique de langues est utilisé pour la détection de certaines fake news, parfois à l'aide d'analyse des images et vidéos associées. Nous montrerons également les nouvelles menaces que le dévoiement d'algorithmes d'apprentissage profond (deep learning) font émerger, et les contre-mesures qui se mettent en place.

Vincent Claveau a obtenu son doctorat d'informatique de l'université de Rennes 1 en 2003. Après un post-doctorat à l'université de Montréal et un court passage à l'Inserm à Paris, il obtient un poste de chargé de recherche au CNRS à l'IRISA à Rennes. Il obtient son habilitation à diriger des recherches en 2020. Ses domaines de recherches principaux sont la recherche d'information (RI), le traitement automatique des langues (TAL), avec un intérêt plus spécifique pour l'extraction d'information à partir de textes, mais aussi à partir de documents multimédias. Il est rédacteur en chef de la revue RIDOWS, relecteur et membre de comités de programmes de nombreux journaux et conférences en TAL et RI.



Le GDR TAL est une unité fédérative thématique du CNRS qui s'intéresse à la modélisation informatique de la langue sous toutes ses formes : écrite, orale, signée, à ses algorithmes et ses applications. Les communautés centrales du GDR TAL sont celles du traitement automatique des langues, du traitement automatique du langage parlé et de la recherche d'informations. Les communautés secondaires sont le traitement automatique du document, le web sémantique, les neurosciences et les sciences cognitives, et toutes les communautés de recherche où la question linguistique est d'importance.

Le TAL a un fort impact sociétal au cœur des développements en sciences des données avec des applications dans le médical, l'éducation, le droit, le journalisme, le handicap. Le GDR TAL mène une prospection scientifique sur les grands enjeux du TAL, réfléchit à l'animation scientifique de la communauté en vue d'améliorer sa stratégie scientifique, son attractivité et sa visibilité. Le GDR TAL fédère les équipes de la recherche française et est l'interlocuteur privilégié de l'INS2I.

Le GDR TAL possède aussi un club des partenaires visant à une meilleure connaissance mutuelle entre les milieux académique et industriel et à dynamiser les relations, les échanges, les osmose et les transferts entre recherche institutionnelle et acteurs industriels.

Le GDR TAL organise des journées thématiques ouvertes à tous, des écoles d'été pour ses chercheurs, des groupes de lecture et soutient des manifestations dans toute la France.

<https://gdr-tal.ls2n.fr>