

48



Le bulletin d'information de l'UPS 2259

focus

Novembre 2018

Sommaire

Infos du mois

Jeudis du campus

Vie de l'unité

**Chercheur(e)s
à l'ouvrage**

Service audiovisuel et communication

**Projection
le 4 décembre**

Documentation AGH

**Au Llacan : croissant
pour tout le monde !**

Catalogue AGH

Nouvelles acquisitions

Web informatique

**Apprentissage profond
pour le TAL (1)**

Info du mois

Jeudis du campus – Vendredi 16 novembre à 13h15 (exceptionnellement)

Logo CNRS, Délégation Paris-Villejuif • Culture, Langues, Textes • CLAS de Villejuif presentent

I FILM I

6 PORTRAITS DES MÉDAILLÉS CNRS 2018

Médaille de Bronze : Judit Gervain, Nicolas Minc
Médaille d'argent : Jérôme Gaillardet, Patrick Le Galès
Cristal du CNRS : Thibault Carpentier, Sébastien de Raucourt

16
nov
2018
13h15
à 14h
Exceptionnellement
le vendredi

LES JEUDIS DU CAMPUS Rencontres - Débats - Conférences - Films

Salle de conférence, bâtiment L - Campus de Villejuif
lesjeudisducampus-pvj@cnrs.fr

Projection de 6 portraits des médaillés CNRS 2018

- **Médaille de Bronze :**
Judit Gervain (Laboratoire psychologie de la perception)
Nicolas Minc (Institut Jacques Monod)
- **Médaille d'argent :**
Jérôme Gaillardet (Institut de physique du globe de Paris)
Patrick Le Galès (Centre d'études européennes et de politique comparée)
- **Cristal du CNRS :**
Thibault Carpentier (Laboratoire Sciences et Technologies de la Musique et du Son)
Sébastien de Raucourt (Institut de physique du globe de Paris)

Rendez-vous à la Salle de conférences du bâtiment L **le vendredi 16 novembre à 13h15**

► Vie de l'unité

Chercheur(e)s à l'ouvrage



Le centre de documentation A.-G. Haudricourt de l'UPS 2259 - Cultures, Langues, Textes a le plaisir de vous inviter une nouvelle fois dans le cadre de *Chercheur(e)s à l'ouvrage* à la présentation par Konstantin Pozdniakov de *L'écriture de l'île de Pâques*, par P. Horley et K. Pozdniakov

Konstantin Pozdniakov est enseignant-chercheur au LLACAN et enseigne notamment le wolof à l'INALCO. Dans son exposé, il nous présentera l'histoire et les problèmes de l'étude de l'écriture Rongorongo qui reste non-déchiffrée. K. Pozdniakov a été commissaire de l'exposition « Les bois parlants », sur le Rongorongo, au Musée Champollion à Figeac et qui s'est tenue du 30 juin au 4 novembre 2018. Il a participé à ce titre à l'édition du catalogue publié par Actes Sud et qu'il va nous présenter. La présentation sera suivie d'une discussion libre avec l'auteur autour d'un café.

Ouvert à tou(te)s et entrée libre. **Judi 15 novembre de 13h15 à 14h15** dans la Salle de conférences, rdc du Bâtiment D. Venez nombreux!

► Service audiovisuel/communication

Projection le 4 décembre

« La fabrique de l'océan Indien » !



Le documentaire « La fabrique de l'océan Indien » (environ 40min.), produit avec le soutien de l'UMR « Orient et méditerranée » et « Laboratoire de Médiévisiologie Occidentale de Paris » et en coproduction avec CNRS Images sera projeté le **mardi 4 décembre à 18h** au cinéma *Le grand Action*, (5 rue des Ecoles Paris 5^{ème})

L'océan Indien n'est pas le fruit d'une découverte extraordinaire mais plutôt de successions de rencontres. Sa présence est également signalée depuis bien longtemps dans les cartes. Elles ont ainsi contribué à donner à l'océan Indien un sens et une cohérence. L'histoire de l'océan Indien est intimement liée à la représentation du monde. Ce film est l'adaptation du livre *La fabrique de l'océan Indien*, dirigé par Emmanuelle Vagnon et Eric Vallet, édité par les Publications de la Sorbonne.

Avec la participation (par ordre alphabétique) : Dejanirah Couto, Jean-Charles Ducène, Catherine Hofmann, Eric Rieth, Pierre Schneider, Emmanuelle Vagnon, Eric Vallet.

L'entrée est libre, dans la limite des places disponibles. La projection sera suivie d'un cocktail. Si vous souhaitez venir, merci de vous inscrire à cette adresse : ferlita@vjf.cnrs.fr.

Au LLACAN : Croissant pour tout le monde !



Un nouveau chapitre s'est ouvert au sein du fonds du LLACAN, un chapitre issu d'un projet de recherche spécifique concernant les langues du Croissant !

MAIS DE QUEL CROISSANT PARLONS-NOUS ?

Y aurait-il un rapport avec les croissants au beurre du petit-déjeuner Open Access qui a eu lieu le mois dernier au centre de documentation ?

Ou avec les phases du croissant de lune ?

À moins que ce ne soit en lien avec le « Croissant-fertile » ?

Que nenni !

Nous parlons ici DU Croissant, le seul, le vrai, l'unique ! Celui que constitue l'aire de contact entre la langue d'Oc et la langue d'Oïl, cette zone géolinguistique située au nord du Massif-Central et s'étirant depuis la Charente sur la ligne frontalière entre la Vienne et la Haute-Vienne, l'Indre et la Creuse, le Cher, l'Allier et le Puy-de-Dôme.

Les parlers de cette région sont influencés à la fois par la langue d'Oc (limousin et auvergnat) et la langue d'Oïl (poitevin-saintongeais, berrichon et bourbonnais) et n'ont que très rarement été étudiés dans le détail. Les derniers locuteurs se faisant âgés, il y a péril en la demeure !

Pour y remédier, le projet ANR Croissant se décline en plusieurs axes, visant notamment à :

- Documenter la situation linguistique et sociolinguistique du Croissant

- Contribuer à l'histoire du domaine gallo-roman par l'étude des systèmes linguistiques du Croissant
- Contribuer aux théories du contact par l'étude des phénomènes d'interférences oc-oïl à différentes échelles dans cette zone.

Le projet est porté par Nicolas QUINT (LLACAN) et reçoit la contribution sur le campus de Maximilien GUERIN (LLACAN, chercheur postdoctoral dédié au projet), Amélie DEPARIS (LLACAN, doctorante du projet), Stefano MANFREDI et Sylvain LOISEAU (SEDYL), Marijana PETROVIC (LACITO), Jérôme PICARD (CEH), Céline FERLITA, Franck GUILLEMAIN et Esther LHUILLERY (CLT) ainsi que d'une douzaine d'autres personnes provenant de différentes structures de recherches.

Au centre de documentation, nous avons mis à disposition un espace dans la salle de travail du sous-sol pour accueillir les ouvrages spécifiques de ce projet et permettre leur consultation en accès libre.

Le « fonds Croissant » (le seul au monde de sa catégorie) regroupe déjà actuellement une vingtaine d'ouvrages de référence.

Ce nombre montera rapidement à une trentaine et devrait par la suite s'accroître progressivement.

Venez découvrir ce nouveau fonds en exclusivité au centre de documentation !

<http://llacan.vjf.cnrs.fr/labex/lc4/index.html>

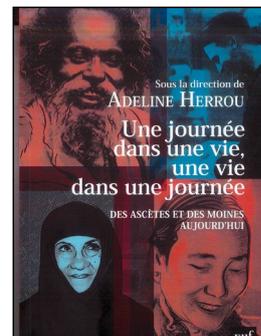
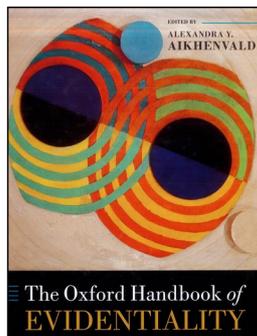
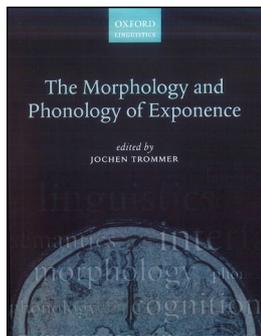
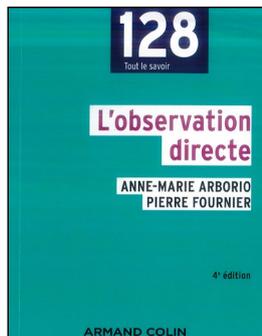
<http://koha.ups2259.vjf.cnrs.fr/cgi-bin/koha/opac-search.pl?q=croissant>

[https://fr.wikipedia.org/wiki/Croissant_\(linguistique\)](https://fr.wikipedia.org/wiki/Croissant_(linguistique))

► Catalogue AGH

Nouvelles acquisitions

Voici une sélection des documents reçus récemment au Centre et acquis par l'UPS ou par les laboratoires partenaires.



N'hésitez pas à faire des propositions tout au long de l'année.

Apprentissage profond pour le TAL (1)

#DeepLearning4NLP

Si le besoin d'automatisation des analyses linguistiques et textuelles se fait de plus en plus sentir par les laboratoires des SHS, les technologies de traitement automatique des langues (TAL), discipline qui cherche à formaliser la capacité langagière humaine (pour plus de détail sur le TAL, consultez Focus 43, avril 2018), connaît un succès grandissant grâce à l'apprentissage profond (Deep Learning).

Depuis les années 2000, la méthodologie de mise en place d'un système TAL se stabilise autour de l'utilisation de corpus (Corpus-based) et l'élaboration d'un manuel d'annotation qui décrit les éléments à prédire par le système. Elle est essentiellement composée de six phases principales qui s'implémentent généralement d'une façon incrémentale et itérative comme le montre la figure suivante.

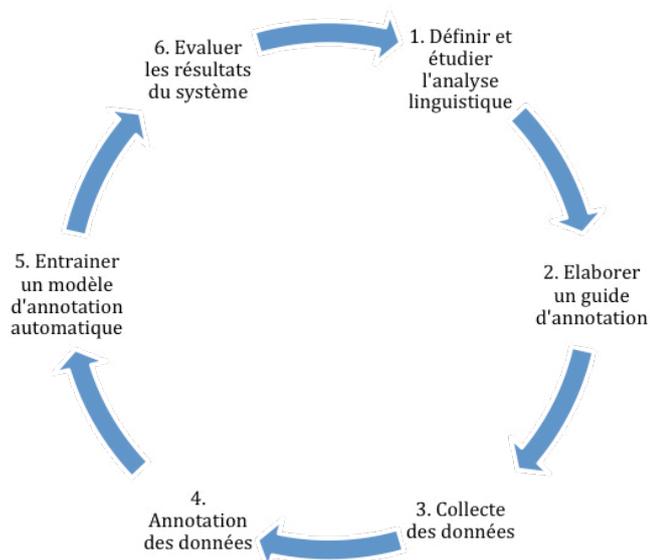


Figure 1. Méthodologie de mise en place d'un système TAL

En plus des compétences linguistiques, ces différentes phases font appel à des batteries d'outils, de méthodes et de pratiques appelées Science des données (Data Science). En effet, les modèles de prédiction et de description ont connu une progression importante en TAL en utilisant des modèles symboliques et des modèles stochastiques. Ces derniers ont surtout bénéficié, depuis les années 2000, des avancées remarquables de l'intelligence artificielle plus précisément des méthodes d'apprentissage automatique

(Machine Learning).

Depuis 2010, les modèles à base de réseaux de neurones dits aussi modèles neuronaux connaissent un succès grandissant pour l'entraînement des modèles en TAL. En effet, l'importante évolution en puissance et en vitesse des mémoires vives et des processeurs graphiques (GPU- Graphics Processing units) a rendu possible la mise en place des réseaux de neurones multicouches. Comme le montre le schéma de la figure ci-dessous, chaque couche est composée de plusieurs neurones, chaque neurone est composé de plusieurs poids qui sont des paramètres utilisés par les fonctions d'activations.

Ces réseaux de neurones profonds peuvent se communiquer entre eux permettant ainsi la conception et la modélisation des architectures avancées. Ces architectures révolutionnaires ont donné pleine satisfaction pour l'automatisation des tâches en TAL. Le service informatique de l'UPS a testé l'apprentissage profond sur plusieurs de ces projets. Cette phase expérimentale donne des résultats plus qu'encourageants. Des réflexions et des explications plus approfondies et détaillées de cette famille de technologies seront présentées dans les mois à venir...

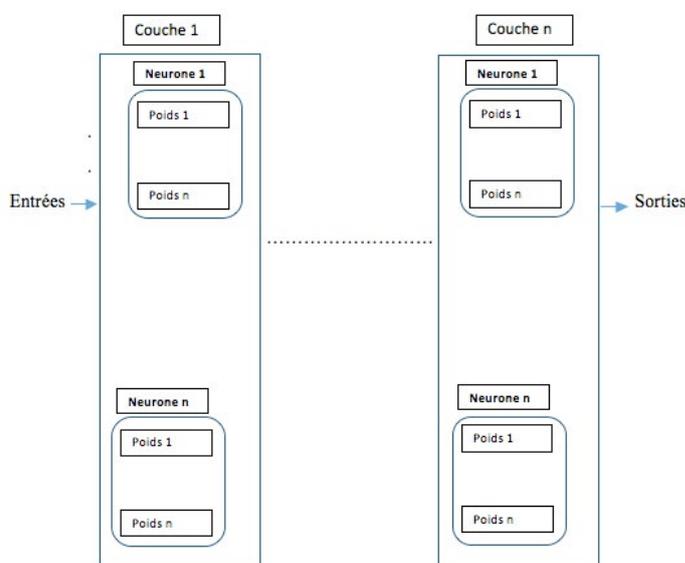


Figure 2. Réseau de neurones profond et ses unités élémentaires

CNRS, UPS 2259, 7 rue Guy Môquet - 94800 Villejuif - Tél : 01 49 58 38 04

Directeur de publication : Fabrice Jecic - Responsable éditoriale : Céline Ferlita

Création graphique et mise en page : Emmanuelle Seguin

<http://www.vjf.cnrs.fr/clt> - [@Ups2259Cnrs](https://twitter.com/Ups2259Cnrs)