

51



Le bulletin d'information de l'UPS 2259

focus

Mars 2019

► Sommaire

Infos du mois

Jeudis du campus
21 Mars à 13h15

Vie du CNRS

**Aux origines du
CNRS...**

Service audiovisuel et communication

Mise en ligne
« Le temps d'une thèse...

Documentation AGH

**ScanR le moteur
de recherche...**

Catalogue AGH

Nouvelles acquisitions

Web informatique

**L'analyse lexicale en
TAL...**

► Info du mois

Jeudis du campus – 21 Mars à 13h15

Conférence : Les systèmes miniers dans le désert d'Atacama. Un laboratoire international en sciences humaines et sociales, par Nicolas Richard (chercheur au CREDA)

LES SYSTÈMES MINIERS
DANS LE DÉSERT D'ATACAMA
UN LABORATOIRE INTERNATIONAL
EN SCIENCES HUMAINES ET SOCIALES
présenté par Nicolas Richard, anthropologue et historien, chercheur au CNRS

21 mars 2019
13h15 à 14h

LES JEUDIS DU CAMPUS
Salle de conférence, bâtiment L - Campus de Villejuif
lesjeudisducampus-pvj@cnrs.fr

Le désert d'Atacama est un laboratoire privilégié, à l'échelle mondiale, pour l'étude de la problématique minière. D'une part, il permet d'appréhender le phénomène avec une profondeur historique et archéologique tout à fait exceptionnelle (exploitations pré-inca, Inca, coloniales et contemporaines) et constitue un des seuls points du globe où l'histoire de l'occupation humaine repose principalement sur l'exploitation minière. D'autre part, il est le théâtre d'une très large variété typologique d'exploitations minières (de très grande à très petite échelle, métalliques et non métalliques, en surface ou souterraines, etc.) qui organisent une histoire sociale, technique et culturelle complexe, inscrite au centre de circulations régionales et mondiales. De même, il s'agit d'un carrefour de cultures et de savoirs-faire miniers d'horizons divers qui impliquent une pluralité d'acteurs et de temporalités. Enfin, haut-lieu de l'exploitation minière contemporaine (cuivre, lithium, or, nitrates, argent...), il constitue un terrain stratégique et privilégié pour observer les formes des capitalismes miniers sous contraintes politique et environnementale et permet de réfléchir, dans un contexte global de course aux ressources minières, aux stratégies des acteurs globaux et aux conditions d'adaptation des sociétés locales.

Nous espérons vous voir nombreux dans la salle bâtiment L jeudi 21 Mars à 13h15.

► Vie du CNRS

Aux origines du CNRS... Conférence de Denis GUTHLEBEN, attaché scientifique au comité pour l'histoire du CNRS



Depuis 80 ans, nos connaissances
bâtissent de nouveaux mondes

Pourquoi et comment le CNRS a-t-il vu le jour ? Pour moderniser et pérenniser la recherche en l'organisant et en lui donnant les moyens de son existence. Ce sont ses premières années et celles qui ont précédé sa création que Denis Guthleben a décidé de nous raconter dans cette conférence.

<https://80ans.cnrs.fr/ressource/aux-origines-du-cnrs-conference-de-denis-guthleben/>

► Service audiovisuel/communication

Mise en ligne « Le temps d'une thèse : Jhonnatan Rangel, linguiste »

Jhonnatan Rangel travaille comme linguiste sur les variations linguistiques et les langues menacées de disparition, notamment sur le cas du numte ode ou ayapaneco dans l'état de Tabasco, au Mexique.

L'Ayapaneco ne se parle que dans le village d'Ayapa, situé dans l'Etat de Tabasco au Mexique. Cette langue peut également être appelée Soque Ayapaneco, car elle appartient à la famille Mixe Soque. Les locuteurs appellent leur langue Numde Ode ou Numde Odi. Cela signifie vrai voix ou vrai parler.



https://www.canal-u.tv/video/cnrs_ups2259/le_temps_d_une_these_jhonnatan_rangel_linguistique.49037

► Centre de documentation AGH



Connaissez-vous ScanR ?

Avez-vous déjà entendu parler de cette application web proposée par la Sous-direction des Systèmes d'information et études statistiques du ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche ?

Il s'agit d'un outil indexant les structures de recherche et d'innovation en France, quel que soit le domaine d'intérêt ou le statut, public ou privé, de l'entité : unité de recherche, entreprises, associations...

ScanR présente sous forme structurée, et rend accessible en les combinant, les informations extraites des sites web de ces structures et des données issues de source institutionnelle ouvertes.

Ces données permettent de décrire et caractériser par leurs publications, thèses, collaborations, projets scientifique, organisation de colloques plus de 15 000 acteurs français.

ScanR permet donc d'une part de se renseigner sur une structure donnée, d'autre part de connaître les entités d'un domaine de recherche particulier mais aussi, d'augmenter la visibilité de sa propre unité.

Bien sûr l'outil est perfectible, il s'agit d'ailleurs d'une version bêta, mais il est possible à notre niveau, d'en améliorer les performances.

En effet, chacun au sein de son unité y participe en versant ses publications dans HAL, ou même en structurant les informations délivrées par son site web, en particulier sa page d'accueil, puisque ce sont là des sources principales de cet outil.

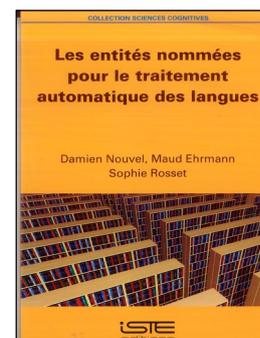
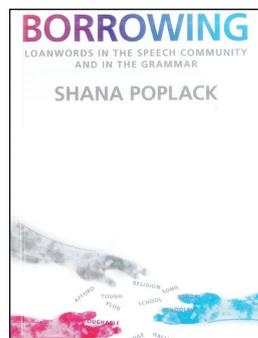
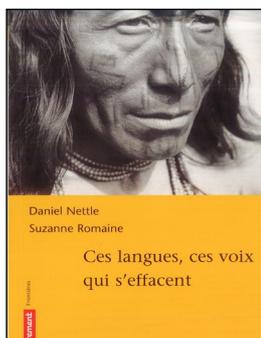
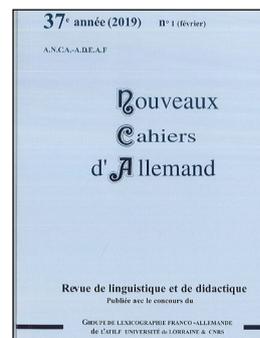
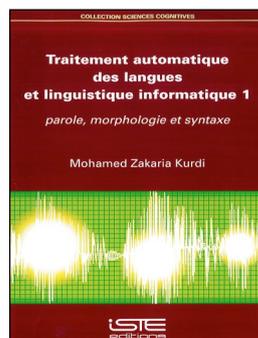
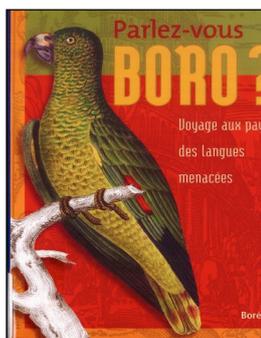
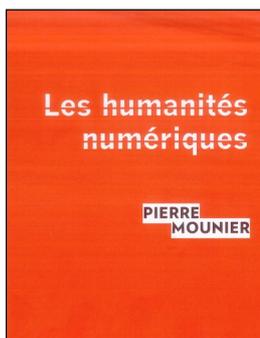
Plus d'info sur

<https://scanr.enseignementsup-recherche.gouv.fr/>

► 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.

L'analyse lexicale en TAL

(Traitement Automatique de la Langue)

#TAL #analyse_lexicale

L'analyse lexicale est indispensable pour toute chaîne de traitement en TAL. Elle a pour but d'identifier les unités du vocabulaire du texte (en anglais Atomic Linguistic Unit ALU), en partant des séquences de caractères qui le constituent. Ces ALUs sont des éléments non-analysables qui peuvent être des morphèmes, des mots simples, des mots composés ou des expressions discontinues. L'analyse lexicale contient une étape de normalisation et de segmentation qui permet de standardiser les différentes variantes typographiques, de reconnaître les valeurs numériques et de segmenter, selon un séparateur, les phrases puis les formes. Elle permet également la reconnaissance des formes simples et des formes composées ainsi que celles de leurs structures et de leurs caractéristiques afin de représenter toutes les hypothèses lexicales et les ambiguïtés potentielles. Elle fait appel, principalement, à des dictionnaires électroniques. Ces dictionnaires sont constitués des entrées lexicales auxquelles sont associées diverses informations linguistiques. Ces

dernières sont définies selon les phénomènes linguistiques à résoudre et le besoin applicatif pour lequel le dictionnaire a été établi. Il existe donc plusieurs types de dictionnaires électroniques qui sont classés selon la nature des informations linguistiques associées à leurs entrées lexicales. Ils sont évalués selon leur robustesse c'est-à-dire leur capacité de couvrir le vocabulaire d'un texte et la précision et l'exhaustivité des descriptions employées pour décrire les entrées lexicales. En effet, les dictionnaires électroniques sont des formalismes qui offrent les outils permettant la mise en place d'une description des différents usages d'une forme. En d'autres termes, comme le montre l'exemple ci-dessous, pour une forme ambiguë, on distingue les différents emplois possibles selon le contexte et on les associe à des informations linguistiques distinctes. Quelle que soit la langue, les ambiguïtés sont réparties à plusieurs niveaux d'analyse et un dictionnaire électronique devrait fournir les informations permettant de les décrire avec précision.

Forme	Signification	Exemple	Informations associées
roy	Chef souverain de certains États, celui qui règne souverainement sur un pays, sur un royaume	le roy a voulu, pour le bien commun, faire les traduire en français.	roi,NC+SENS=1
roy	Être en mesure de disposer de quelque chose.	il n'avait roy d'iretage ne chevanche nulle	roi,NC+SENS=2

Tableau 1. Les emplois de la forme moyen français « roy »

Les dictionnaires électroniques sont mis en place en utilisant des lexiques et des analyses morphologiques auxquels on associe diverses informations linguistiques. Les lexiques sont généralement constitués d'une forme canonique et d'une liste des variantes associées tandis que l'analyse morphologique relie la forme canonique et les formes orthographiques par suppression et substitution de morphèmes.

Finalement, l'analyse lexicale permet de regrouper les formes ayant une caractéristique commune en ajoutant à chaque entrée lexicale diverses informations linguistiques telle la partie de discours, des informations de flexion comme le genre et le nombre pour les noms et comme la personne, le genre, le nombre, le mode et la voix pour les verbes, et des traits sémantiques pour caractériser leur signification.

Le service informatique de l'UPS met en place des modules variés permettant l'analyse automatique de nombreux phénomènes lexicaux pour diverses langues.

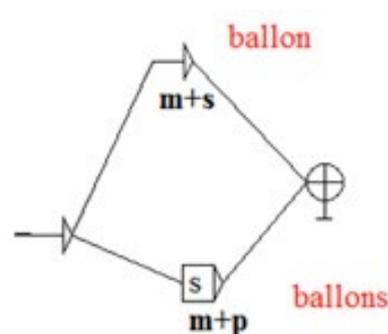


Figure 2. Graphe du paradigme « Ballon »

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

Directeur de publication : Bernard Weiss - Responsable éditoriale : Céline Ferlita

Création graphique et mise en page : Emmanuelle Seguin et Isabelle Michel

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