

**Traitement Automatique des Langues
Naturelles et Rencontres des Étudiants
Chercheurs en Informatique pour le
Traitement Automatique des Langues
(TALN-RÉCITAL 2013)**

**Les Sables d'Olonne, France
17-21 June 2013**

Volume 1: TALN 2013

Editors:

**Emmanuel Morin
Yannick Estève**

ISBN: 978-1-63266-079-4

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2013) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2014)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Table des matières

Conférenciers invités	1
<i>Améliorer la traduction des langages morphologiquement riches</i> Alexander Fraser	1
<i>Recherche d'Information et Traitement Automatique des Langues Naturelles</i> Josiane Mothe	2
Charte Ethique et Big Data	3
<i>La Charte Éthique et Big Data : pour des ressources pour le TAL (enfin !) traçables et pérennes</i> Karën Fort et Alain Couillault.....	3
Articles longs	5
<i>Analyse Automatique de la Morphologie Nominale Amazighe</i> Fatima Zahra Nejme, Siham Boulaknadel et Driss Aboutajdine	5
<i>Apprentissage symbolique et statistique pour le chunking : comparaison et combinaisons</i> Isabelle Tellier et Yoann Dupont.....	19
<i>L'utilisation des POMDP pour les résumés multi-documents orientés par une thématique</i> Yllias Chali, Sadid A.Hasan et Mustapha Mojahid.....	33
<i>Sélection non supervisée de relations sémantiques pour améliorer un thésaurus distributionnel</i> Olivier Ferret.....	48
<i>Groupement de termes basé sur des régularités linguistiques et sémantiques dans un contexte cross-langue</i> Marie Dupuch, Thierry Hamon et Natalia Grabar	62
<i>WoNeF : amélioration, extension et évaluation d'une traduction française automatique de WordNet</i> Quentin Pradet, Jeanne Baguenier-Desormeaux, Gaël de Chalendar et Laurence Danlos	76
<i>Approches statistiques discriminantes pour l'interprétation sémantique multilingue de la parole</i> Bassam Jabaian, Fabrice Lefèvre et Laurent Besacier	90

<i>Identification automatique des relations discursives « implicites » à partir de données annotées et de corpus bruts</i> Chloé Braud et Pascal Denis.....	104
<i>Apprentissage d'une hiérarchie de modèles à paires spécialisés pour la résolution de la coréférence</i> Emmanuel Lassalle et Pascal Denis.....	118
<i>Apprentissage supervisé pour l'identification de relations sémantiques au sein de structures énumératives parallèles</i> Jean-Philippe Fauconnier, Mouna Kamel, Bernard Rothenburger et Nathalie Aussenac-Gilles	132
<i>Techniques de TAL et corpus pour faciliter les formulations en anglais scientifique écrit</i> Marie-Paule Jacques, Laura Hartwell et Achille Falaise	146
<i>Construction d'un large corpus écrit libre annoté morpho-syntaxiquement en français</i> Nicolas Hernandez et Florian Boudin.....	160
<i>Vers un treebank du français parlé</i> Anne Abeillé et Benoit Crabbé.....	174
<i>L'apport du faisceau dans l'analyse syntaxique en dépendances par transitions : études de cas avec l'analyseur Talismane</i> Assaf Urieli et Ludovic Tanguy	188
<i>Un modèle segmental probabiliste combinant cohésion lexicale et rupture lexicale pour la segmentation thématique</i> Anca Simon, Guillaume Gravier et Pascale Sébillot	202
<i>Traitements d'ellipses : deux approches par les grammaires catégorielles abstraites</i> Pierre Bourreau.....	215
<i>Chunks et activation : un modèle de facilitation du traitement linguistique</i> Philippe Blache.....	229
<i>Extraction de lexiques bilingues à partir de corpus comparables par combinaison de représentations contextuelles</i> Amir Hazem et Emmanuel Morin.....	243
<i>Découverte de connaissances dans les séquences par CRF non-supervisés</i> Vincent Claveau et Abir Ncibi.....	257
<i>Annotation automatique d'un corpus d'apprenants d'anglais avec un jeu d'étiquettes modifié du Penn Treebank</i> Thomas Gaillat	271

<i>GLÀFF, un Gros Lexique À tout Faire du Français</i> Franck Sajous, Nabil Hathout et Basilio Calderone	285
<i>Constitution d'une ressource sémantique arabe à partir de corpus multilingue aligné</i> Authoul Abdul Hay et Olivier Kraif	299
<i>Identification, alignement, et traductions des adjectifs relationnels en corpus comparables</i> Rima Harastani, Béatrice Daille et Emmanuel Morin.....	313
<i>Utilisation de la similarité sémantique pour l'extraction de lexiques bilingues à partir de corpus comparables</i> Dhouha Bouamor, Nasredine Semmar et Pierre Zweigenbaum.....	327
<i>Inférences déductives et réconciliation dans un réseau lexico-sémantique</i> Manel Zarrouk, Mathieu Lafourcade et Alain Joubert.....	339
<i>Regroupement sémantique de relations pour l'extraction d'information non supervisée</i> Wei Wang, Romaric Besançon, Olivier Ferret et Brigitte Grau.....	353
<i>Sémantique des déterminants dans un cadre richement typé</i> Christian Retoré.....	367
<i>Détection de zones parallèles à l'intérieur de multi-documents pour l'alignement multilingue</i> Charlotte Lecluze, Romain Brixtel, Loïs Rigouste, Emmanuel Giguët, Régis Clouard, Gaël Lejeune et Patrick Constant.....	381
<i>Un système de traduction de verbes entre arabe standard et arabe dialectal par analyse morphologique profonde</i> Ahmed Hamdi, Rahma Boujelbane, Nizar Habash et Alexis Nasr	395
<i>Extension dynamique de lexiques morphologiques pour le français à partir d'un flux textuel</i> Benoît Sagot, Damien Nouvel, Virginie Mouilleron et Marion Baranes	407
<i>Fouille de règles d'annotation partielles pour la reconnaissance des entités nommées</i> Damien Nouvel, Jean-Yves Antoine, Nathalie Friburger et Arnaud Soulet.....	421
<i>Segmentation de textes arabes en unités discursives minimales</i> Iskandar Keskes, Farah Beanamara et Lamia Hadrach Belguith.....	435
<i>Un cadre d'apprentissage intégralement discriminant pour la traduction statistique</i> Thomas Lavergne, Alexandre Allauzen et François Yvon.....	450
<i>Annotation sémantique pour des domaines spécialisés et des ontologies riches</i> Yue Ma, François Lévy et Adeline Nazarenko.....	464
<i>Pré-segmentation de pages web et sélection de documents pertinents en Questions-Réponses</i> Nicolas Foucault, Sophie Rosset et Gilles Adda.....	479

Articles courts.....507

TALN Archives : une archive numérique francophone des articles de recherche en Traitement Automatique de la Langue

Florian Boudin.....507

Similarités induites par mesure de comparabilité : signification et utilité pour le clustering et l'alignement de textes comparables

Pierre-Francois Marteau et Gildas Ménier515

ProLMF version 1.2. Une ressource libre de noms propres avec des expansions contextuelles

Denis Maurel et Béatrice Bouchou Markhoff.....523

Vers un décodage guidé pour la traduction automatique

Benjamin Lecouteux et Laurent Besacier531

La La prédiction avec des règles peu coûteuses, utile pour la TA statistique des forums ?

Johanna Gerlach, Victoria Porro, Pierrette Bouillon et Sabine Lehmann.....539

Édition interactive d'énoncés en langue des signes française dédiée aux avatars signeurs

Ludovic Hamon, Sylvie Gibet et Sabah Boustila.....547

ANCOR, premier corpus de français parlé d'envergure annoté en coréférence et distribué librement

Judith Muzerelle, Anaïs Lefevre, Jean-Yves Antoine, Emmanuel Schang, Denis Maurel, Jeanne Villaneau et Iris Eshkol.....555

Segmentation Multilingue des Mots Composés

Elizaveta Loginova-Clouet et Béatrice Daille564

Gestion des terminologies riches : l'exemple des acronymes

Ying Zhang et Mathieu Mangeot.....572

Ngrammes et Traits Morphosyntaxiques pour la Identification de Variétés de l'Espagnol

Marcos Zampieri, Binyam Gebrekidan Gebre et Sascha Diwersy580

L'apport des Entités Nommées pour la classification des opinions minoritaires

Amel Fraisse, Patrick Paroubek et Gil Francopoulo588

Trouver les mots dans un simple réseau de co-occurrences

Gemma Bel-Enguix et Michael Zock.....596

Analyse statique des interactions entre structures élémentaires d'une grammaire

Guy Perrier.....604

<i>Influence des annotations sémantiques sur un système de détection de coréférence à base de perceptron multi-couches</i>	
Eric Charton, Michel Gagnon et Ludovic Jean-Louis	612
<i>Traduction automatique statistique pour l'arabe-français améliorée par le prétraitement et l'analyse de la langue</i>	
Fatiha Sadat et Emad Mohamed	620
<i>Expériences de formalisation d'un guide d'annotation : vers l'annotation agile assistée</i>	
Bruno Guillaume et Karèn Fort.....	628
<i>Repérer des toponymes dans des titres de cartes topographiques</i>	
Catherine Dominguès et Iris Eshkol-Taravella	636
<i>Extraction des relations temporelles entre événements médicaux dans des comptes rendus hospitaliers</i>	
Pierre Zweigenbaum et Xavier TANNIER	643
<i>Similarité de second ordre pour l'exploration de bases textuelles multilingues</i>	
Nikola Tulechki et Ludovic Tanguy	651
<i>Apprentissage d'une classification thématique générique et cross-langue à partir des catégories de la Wikipédia</i>	
François-Régis Chaumartin	659
<i>Apprentissage supervisé sur ressources encyclopédiques pour l'enrichissement d'un lexique de noms propres destiné à la reconnaissance des entités nommées</i>	
Nadia Okinina, Damien Nouvel, Nathalie Friburger et Jean-Yves Antoine	667
<i>Convertir des analyses syntaxiques en dépendances vers les relations fonctionnelles PAS-SAGE</i>	
Patrick Paroubek, Munshi Asadullah et Anne Vilnat.....	675
<i>Résolution d'anaphores et traitement des pronoms en traduction automatique à base de règles</i>	
Sharid Loáiciga.....	683
<i>Lexiques de corpus comparables et recherche d'information multilingue</i>	
Frederik Cailliau, Ariane Cavet, Clément de Groc et Claude de Loupy	691
<i>Utilisation de contextes pour la correction automatique ou semi-automatique de réclamations clients</i>	
Philippe Suignard et Sofiane Kerroua.....	699
<i>SegCV : traitement efficace de CV avec analyse et correction d'erreurs</i>	
Luis Adrián Cabrera-Diego, Juan-Manuel Torres-Moreno et Marc El-Bèze	707

<i>Recherche et utilisation d'entités nommées conceptuelles dans une tâche de catégorisation</i> Jean-Valère Cossu, Juan-Manuel Torres-Moreno et Marc El-Bèze	715
<i>Un corpus d'erreurs de traduction</i> Guillaume Wisniewski, Anil Kumar Singh, Natalia Segal et François Yvon	723
<i>Une méthode d'évaluation des résumés basée sur la combinaison de métriques automatiques et de complexité textuelle</i> Samira Walha Ellouze, Maher Jaoua et Lamia Hadrach Belguith	731
<i>Segmentation thématique : processus itératif de pondération intra-contenu</i> Abdessalam Bouchekif, Géraldine Damnati et Delphine Charlet	739
<i>Recherche et visualisation de mots sémantiquement liés</i> Alexander Panchenko, Hubert Naets, Laetitia Brouwers, Pavel Romanov et Cédric Fairon	747
<i>Un analyseur morphologique étendu de l'allemand traitant les formes verbales à particule séparée</i> Jean-Philippe Guilbaud, Christian Boitet et Vincent Berment	755
<i>Construction et exploitation d'un corpus français pour l'analyse de sentiment</i> Marc Vincent et Grégoire Winterstein.....	764
<i>Résolution d'anaphores appliquée aux collocations : une évaluation préliminaire</i> Luka Nerima et Eric Wehrli.....	772
<i>Aide à l'enrichissement d'un référentiel terminologique : propositions et expérimentations</i> Thibault Mondary, Adeline Nazarenko, Haïfa Zargayouna et Sabine Barreaux.....	779
Démonstrations	787
<i>DAnIEL : Veille épidémiologique multilingue parcimonieuse</i> Gaël Lejeune, Romain Brixtel, Charlotte Lecluze, Antoine Doucet et Nadine Lucas ..	787
<i>Lexique multilingue dans le cadre du modèle Compreno développé ABBYY</i> Elena Kozlova, Maria Gontcharova et Tatiana Popova	789
<i>Inbenta Semantic Search Engine : un moteur de recherche sémantique inspiré de la Théorie Sens-Texte</i> Manon Quintana	791
<i>FMO : un outil d'analyse automatique de l'opinion</i> Jean-Leon Bouraoui et Marc Canitrot	793
<i>Corriger, analyser et représenter le texte Synapse Développement</i> Patrick Séguéla et Dominique Laurent	795

<i>Une interface pour la validation et l'évaluation de chronologies thématiques</i> Xavier Tannier, Véronique Moriceau et Erwan Le Flem	797
<i>CasSys Un système libre de cascades de transducteurs</i> Denis Maurel et Nathalie Friburger	799
<i>iMAG : post-édition, évaluation de qualité de TA et production d'un corpus parallèle</i> Lingxiao Wang et Ying Zhang.....	801
<i>Technologies du Web Sémantique pour l'exploitation de données lexicales en réseau (Lexical Linked Data)</i> David Rouquet	803
<i>Adaptation de la plateforme corporale ScienQuest pour l'aide à la rédaction en langue seconde</i> Achille Falaise.....	805
<i>Démonstrateur Apopsis pour l'analyse des tweets</i> Sébastien Peña Saldarriaga, Damien Vintache et Béatrice Daille.....	807
<i>L'analyse des sentiments au service des centres d'appels</i> Frederik Cailliau et Ariane Cavet.....	809
<i>TTC TermSuite alignement terminologique à partir de corpus comparables</i> Béatrice Daille et Rima Harastani.....	812
Liste des auteurs.....	814
Liste des mots clés.....	817

Traitement Automatique des Langues Naturelles et Rencontres des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues

(TALN-RÉCITAL 2013)

**Les Sables d'Olonne, France
17-21 June 2013**

Volume 2: RÉCITAL 2013

Editors:

**Florian Boudin
Loïc Barrault**

ISBN: 978-1-63266-079-4

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2013) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2014)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Table des matières

Articles longs	1
<i>Acquisition de lexique bilingue d'expressions polylexicales : Une application à la traduction automatique statistique</i>	
Dhouha Bouamor	1
<i>Quelques variations sur les mesures de comparabilité quantitatives et évaluations sur des corpus comparables Français-Anglais synthétiques</i>	
Guiyao Ke.....	15
<i>Inférence grammaticale guidée par clustering</i>	
Noémie-Fleur Sandillon-Rezer.....	28
<i>Améliorer l'extraction et la description d'expressions polylexicales grâce aux règles transformationnelles</i>	
Aurélie Joseph	42
<i>Construction de corpus multilingues : état de l'art</i>	
Manuela Yapomo	56
<i>Développement de ressources pour l'entraînement et l'utilisation de l'étiqueteur morpho-syntaxique TreeTagger sur l'arabe</i>	
Dhaou Ghoul	69
<i>Détection de polarité d'opinions dans les forums en langue arabe par fusion de plusieurs SVM</i>	
Amel Ziani, Nabih Azizi et Yamina Tlili Guiassa	83
<i>État de l'art des méthodes d'extraction automatique de termes-clés</i>	
Adrien Bougouin.....	96
<i>Influence de l'étiquetage syntaxique des têtes sur l'analyse en dépendances discontinues du français</i>	
Ophélie Lacroix	110
<i>Une approche linguistique pour l'extraction des connaissances dans un texte arabe</i>	
Houda Saadane	124
<i>Extraction des mots simples du lexique scientifique transdisciplinaire dans les écrits de sciences humaines : une première expérimentation</i>	
Sylvain Hatier	138
<i>Vers une identification automatique du chiasme de mots</i>	
Marie Dubremetz	150

<i>Représentation des connaissances du DEC : Concepts fondamentaux du formalisme des Graphes d'Unités</i>	
Maxime Lefrançois.....	164
<i>Vers un système générique de réécriture de graphes pour l'enrichissement de structures syntaxiques</i>	
Corentin Ribeyre	178
<i>État de l'art de l'induction de sens : une voie vers la désambiguïsation lexicale pour les langues peu dotées</i>	
Mohammad Nasiruddin	192
<i>Génération des corpus en dialecte tunisien pour la modélisation de langage d'un système de reconnaissance</i>	
Rahma Boujelbane	206
<i>Détection automatique des sessions de recherche par similarité des résultats provenant d'une collection de documents externe</i>	
Simon Leva et Nicolas Faessel	217
<i>Une approche mixte morpho-syntaxique et statistique pour la reconnaissance d'entités nommées en langue chinoise</i>	
Zhen Wang.....	231
Liste des auteurs.....	244
Liste des mots clés.....	245