

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

**(TALN-RECITAL 2014)**

**Long Papers**

**Marseille, France  
1-4 July 2014**

**Volume 1 of 4**

**ISBN: 978-1-63439-515-1**

**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© (2014) 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](mailto: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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# Table des matières

## Articles longs

### Fouille de données et TAL

[O – RI.1] Influence des marqueurs multi-polaires dépendant du domaine 1  
pour la fouille d’opinion au niveau du texte

*Morgane Marchand, Olivier Mesnard, Romaric Besançon, Anne Vilnat*

[O – RI.3] Influence des domaines de spécialité dans l’extraction de 13  
termes-clés

*Adrien Bouguin, Florian Boudin, Béatrice Daille*

[O – RI.4] Etiquetage en rôles événementiels fondé sur l’utilisation d’un 25  
modèle neuronal

*Emanuela Boros, Romaric Besançon, Olivier Ferret, Brigitte Grau*

### Sémantique

[O – S1.1] Utilisation de représentations de mots pour l’étiquetage de rôles 36  
sémantiques suivant FrameNet

*William Léchelle, Philippe Langlais*

[O – S1.4] Cross-lingual Word Sense Disambiguation for Predicate La- 46  
belling of French

*Lonneke Van Der Plas, Marianna Apidianaki*

### Parsing 1

[O – P1.1] Améliorer l’étiquetage de "que" par les descripteurs ciblés et 56  
les règles

*Assaf Urieli*

[O – P1.2] Jouer avec des analyseurs syntaxiques 67

*Eric Villemonte De La Clergerie*

### Lexique 1

[O – L1.1] Principes de modélisation systémique des réseaux lexicaux 79

*Alain Polguère*

[O – L1.2] Un modèle pour prédire la complexité lexicale et graduer les 91  
mots

*Nuria Gala, Thomas François, Delphine Bernhard, Cédrick Fairon*

[O – L1.3] Annotations et inférences de relations dans un réseau lexico-sémantique: Application à la radiologie	103
<i>Lionel Ramadier, Manel Zarrouk, Mathieu Lafourcade, Antoine Micheau</i>	

## Gestion des erreurs en TAL

[O – E.1] Correction automatique par résolution d’anaphores pronominales	113
<i>Maud Pironneau, Éric Brunelle, Simon Charest</i>	

[O – E.2] Peut-on bien chunker avec de mauvaises étiquettes POS ?	125
<i>Iris Eshkol, Isabelle Tellier, Yoann Dupont, Ilaine Wang</i>	

[O – E.3] Normalisation de textes par analogie: le cas des mots inconnus	137
<i>Marion Baranes, Benoît Sagot</i>	

## Modèles linguistiques

[O – F.1] Une évaluation approfondie de différentes méthodes de compositionnalité sémantique	149
<i>Antoine Bride, Tim Van de Cruys, Nicholas Asher</i>	

[O – F.2] Génération de textes : G-TAG revisité avec les Grammaires Catégorielles Abstraites	161
<i>Laurence Danlos, Aleksandre Maskharashvili, Sylvain Pogodalla</i>	

## Méthodes numériques pour le TAL

[O – N1.1] Apprentissage partiellement supervisé d’un étiqueteur morpho-syntaxique par transfert cross-lingue	173
<i>Guillaume Wisniewski, Nicolas Pécheux, Elena Knyazeva, Alexandre Allauzen, François Yvon</i>	

[O – N1.3] Construire un corpus monolingue annoté comparable	184
<i>Nicolas Hernandez</i>	

[O – N1.4] Vers une approche simplifiée pour introduire le caractère incrémental dans les systèmes de dialogue	196
<i>Hatim Khouzaimi, Romain Laroche, Fabrice Lefevre</i>	

## Lexique 2

[O – L2.1] La base lexicale Démonette : entre sémantique constructionnelle et morphologie dérivationnelle	208
<i>Nabil Hathout, Fiammetta Namer</i>	

[O – L2.2] Explorer le graphe de voisinage pour améliorer les thésaurus distributionnels 220

*Vincent Claveau, Ewa Kijak, Olivier Ferret*

[O – L2.3] Réduction de la dispersion des données par généralisation des contextes distributionnels : application aux textes de spécialité 232

*Amandine Périnet, Thierry Hamon*

[O – L2.4] Extraction non supervisée de relations sémantiques lexicales 244

*Juliette Conrath, Stergos Afantenos, Nicholas Asher, Philippe Muller*

## Traduction Automatique

[O – T.1] Modèles de langue neuronaux: une comparaison de plusieurs stratégies d'apprentissage 256

*Quoc-Khanh Do, Alexandre Allauzen, François Yvon*

[O – T.2] Etude de l'impact de la translittération de noms propres sur la qualité de l'alignement de mots à partir de corpus parallèles français-arabe 268

*Nasredine Semmar, Houda Saadane*

[O – T.3] Adaptation thématique pour la traduction automatique de dépêches de presse 280

*Souhir Gahbiche-Braham, Hélène Bonneau-Maynard, François Yvon*

## Traitement de corpus

[O – S2.1] Étude quantitative des disfluences dans le discours de schizophrènes : automatiser pour limiter les biais 292

*Maxime Amblard, Karën Fort*

[O – S2.2] Repérage et analyse de la reformulation paraphrastique dans les corpus oraux 304

*Iris Eshkol-Taravella, Natalia Grabar*

[O – S2.3] Evaluation d'une approche possibiliste pour la désambiguïsation des textes arabes 316

*Raja Ayed, Brahim Bounhas, Bilel Elayeb, Narjès Bellamine, Fabrice Evrard*

## Parsing 2

[O – P2.1] Un analyseur discriminant de la famille LR pour l'analyse en constituants 328

*Benoit Crabbé*

[O – P2.2] **Détection automatique de la structure organisationnelle de documents à partir de marqueurs visuels et lexicaux** 340  
*Jean-Philippe Fauconnier, Laurent Sorin, Mouna Kamel, Mustapha Mojahid, Nathalie Aussenac-Gilles*

[O – P2.3] **Jugement exact de grammaticalité d’arbre syntaxique probable** 352  
*Jean-Philippe Prost*

### Lexique 3

[O – L3.1] **Annotation sémantique et validation terminologique en texte intégral en SHS** 363  
*Mokhtar-Boumeyden Billami, José Camacho-Collados, Evelyne Jacquy, Laurence Kister*

[O – L3.2] **Identification des noms sous-spécifiés, signaux de l’organisation discursive** 377  
*Charlotte Roze, Thierry Charnois, Dominique Legallois, Stéphane Ferrari, Mathilde Salles*

## Articles courts

### Traduction

[P – T.1] **Traduction automatisée d’une oeuvre littéraire: une étude pilote** 389  
*Laurent Besacier*

[P – T.2] **Vers un développement plus efficace des systèmes de traduction statistique : un peu de vert dans un monde de BLEU** 395  
*Li Gong, Aurélien Max, François Yvon*

[P – T.3] **On-going Cooperative Research towards Developing Economy-Oriented Chinese-French SMT Systems with a New SMT Framework** 401  
*Yidong Chen, Lingxiao Wang, Christian Boitet, Xiaodong Shi*

### Lexique 1

[P – L1.1] **Extraction automatique de termes combinant différentes informations: linguistique, statistique et Web** 407  
*Juan Antonio Lossio Ventura, Clement Jonquet, Mathieu Roche, Maguelonne Teisseire*

[P – L1.2] **Analyse automatique d’espaces thématiques** 413  
*Gilles Boyé, Anna Kupsc*

[P – L1.3] **Extraction et représentation des constructions à verbe support en espagnol** 419

Sandra Castellanos

[P – L1.4] **Sous-catégorisation en "pour" et syntaxe lexicale** 425  
*Benoît Sagot, Laurence Danlos, Margot Colinet*

### Étiquetage 1

[P – Et1.1] **Étiquetage morpho-syntaxique pour des mots nouveaux** 431  
*Ingrid Falk, Delphine Bernhard, Christophe Gérard, Romain Potier-Ferry*

[P – Et1.2] **Détection et correction automatique d'entités nommées dans des corpus OCRisés** 437  
*Benoît Sagot, Kata Gábor*

[P – Et1.3] **Méthodes de lissage d'une approche morpho-statistique pour la voyellation automatique des textes arabes** 443  
*Amine Chennoufi, Azzeddine Mazroui*

### Traitement de corpus 1

[P – S1.1] **Évaluation d'un système d'extraction de réponses multiples sur le Web par comparaison à des humains** 449  
*Mathieu-Henri Falco, Véronique Moriceau, Anne Vilnat*

[P – S1.2] **Centrality Measures for Non-Contextual Graph-Based Unsupervised Single Document Keyword Extraction** 455  
*Natalie Schluter*

[P – S1.3] **Détection de périodes musicales d'une collection de musique par apprentissage** 461  
*Rémy Kessler, Nicolas Béchet, Audrey Laplante, Dominic Forest*

[P – S1.4] **AMesure: une plateforme de lisibilité pour les textes administratifs** 467  
*Thomas Francois, Laetitia Brouwers, Hubert Naets, Cédrick Fairon*

### Sentiments

[P – Se.1] **Décomposition des « hash tags » pour l'amélioration de la classification en polarité des « tweets »** 473  
*Caroline Brun, Claude Roux*

[P – Se.2] **Modélisation des questions de l'agent pour l'analyse des affects, jugements et appréciations de l'utilisateur dans les interactions humain-agent** 479  
*Caroline Langlet, Chloé Clavel*

## Outils

[P – Ou.1] **KNG: un outil pour l'écriture facile de cascades de transduc- 485**  
**teurs**  
*Francois Barthelemy*

[P – Ou.2] **Comparaison de deux outils d'analyse de corpus japonais pour 491**  
**l'aide au linguiste, Sagace et Mecab**  
*Blin Raoul*

[P – Ou.3] **Un concordancier multi-niveaux et multimédia pour des corpus 497**  
**oraux**  
*Giulia Barreca, George Christodoulides*

## Étiquetage 2

[P – Et2.1] **Simulation de l'apprentissage des contextes nominaux/verbaux 505**  
**par n-grammes**  
*Perrine Brusini, Pascal Amsili, Emmanuel Chemla, Anne Christophe*

[P – Et2.2] **Impact de la nature et de la taille des corpus d'apprentissage 511**  
**sur les performances dans la détection automatique des entités nommées**  
*Anaïs Ollagnier, Sébastien Fournier, Patrice Bellot, Frédéric Béchet*

[P – Et2.3] **RENAM: Système de Reconnaissance des Entités Nommées 517**  
**Amazighes**  
*Meryem Talha, Siham Boulaknadel, Driss Aboutajdine*

[P - Et2.4] **De la quenelle culinaire à la quenelle politique: identification 525**  
**de changements sémantiques à l'aide des Topic Models.**  
*Ingrid Falk, Delphine Bernhard, Christophe Gérardllep*

## Langue des signes

[P – LS.1] **Grammaire récursive non linéaire pour les langues des signes 531**  
*Michael Filhol*

[P – LS.2] **Vers un traitement automatique en soutien d'une linguistique 537**  
**exploratoire des LS**  
*Rémi Dubot, Arturo Curiel, Christophe Collet*

## Résumé automatique

[P – R.2] **Résumé Automatique Multilingue Expérimentations sur 543**  
**l'Anglais, l'Arabe et le Français**



*Houda Oufaida, Omar Nouali, Philippe Blache*

- [P – R.3] **Porting a Summarizer to the French Language** 550  
*Rémi Bois, Johannes Leveling, Lorraine Goeuriot, Gareth Jones, Liadh Kelly*

## Corpus

- [P – C.1] **Extraction de données orales multi-annotées** 556  
*Brigitte Bigi, Tatsuya Watanabe*

- [P – C.2] **Annotation de la temporalité en corpus : contribution à l'amélioration de la norme TimeML** 562  
*Anaïs Lefevre, Jean-Yves Antoine, Agata Savary, Emmanuel Schang, Lotfi Abouda, Denis Maurel, Iris Eshkol*

- [P – C.3] **Identification automatique de zones dans des documents pour la constitution d'un corpus médical en français** 568  
*Louise Deleger, Aurélie Névéol*

- [P – C.4] **Un schéma d'annotation en dépendances syntaxiques profondes pour le français** 574  
*Guy Perrier, Marie Candito, Bruno Guillaume, Corentin Ribeyre, Karën Fort, Djamé Seddah*

## Traitement de corpus 2

- [P – S2.1] **Analyse argumentative du corpus de l'ACL (ACL Anthology)** 580  
*Elisa Omodei, Yufan Guo, Jean-Philippe Cointet, Thierry Poibeau*

## Lexique 2

- [P – L2.1] **Intégration relationnelle des exemples lexicographiques dans un réseau lexical** 586  
*Veronika Lux-Pogodalla*

- [P – L2.2] **Les couleurs des gens** 592  
*Mathieu Lafourcade, Nathalie Le Brun, Virginie Zampa*

- [P – L2.3] **Induction de sens pour enrichir des ressources lexicales** 598  
*Mohammad Nasiruddin, Didier Schwab, Andon Tchechmedjiev, Gilles Serasset, Hervé Blanchon*

- [P – L2.4] **Un dictionnaire et une grammaire de composés français** 604  
*François Trouilleux*