

Actes du colloque

INFOL@NGUES 2016

<http://infolangues.olympie.in>



La 8^{ème} édition du colloque : INFOL@NGUES 2016

L'Informatique au service des Langues

21, 22 et 23 avril 2016

Béja, Tunisie



La 8^{ème} édition du colloque : INFOL@NGUES 2016

L'Informatique au service des Langues

21, 22 et 23 Avril 2016

Béja, Tunisie

Préface

La sélection d'articles publiés dans le présent recueil constitue les actes de la 8^{ème} édition du colloque INFOLANGUES 2016 organisé à Béja du 21 à 22 avril 2016.

Le colloque INFOLANGUES ambitionne comme à l'accoutumée de regrouper dans un même lieu scientifique, chercheurs, informaticiens et linguistes, pour échanger des idées et des approches dans le domaine du traitement des langues. Donc c'est un défi de rassembler une panoplie de chercheurs travaillant sur plusieurs thèmes abordés dans ce colloque, tels que Linguistique computationnelle, Linguistique pour la reconnaissance de l'écriture, Langue et informatique, Dictionnaires numériques, Fouille de données textuelles, Traduction automatique (TA) et automatisée (TAO), Ontologies, Evaluation des modèles de langues par les TICs, Enseignement virtuel, Recherche d'information.

Le colloque, comme il faut le souligner, attire de plus en plus des chercheurs nationaux et internationaux. Dans cette édition nous avons accepté 27 papiers rédigés par des informaticiens et des linguistes. Nous avons invité plusieurs Professeurs et Docteurs, spécialistes dans le domaine du Traitement des Langues Naturelles (TALN), pour présenter l'état d'avancement de leurs recherches au cours des journées du colloque. Nous avons programmé un atelier animé par le professeur Mourad Ghassan de l'Université Libanaise sur des logiciels sur la traduction automatique.

Remerciement

Nos vifs remerciements vont tout d'abord aux auteurs pour leurs communications, aux relecteurs, membres du comité scientifique, dont les rapports d'évaluation circonstanciés et constructifs ont contribué à améliorer significativement la qualité des articles.

Nos remerciements vont également à tous les invités du colloque qui ont donné le plus par leurs implications et présence. Nous citons entre autres le Prof. Mourad Ghassan, le Prof. Sadok Ben Yahya et le Dr. Faiza Derbel.

Nous remercions aussi toute l'équipe du comité d'organisation pour leur travail efficace et leur mobilisation continue.

Enfin nous remercions spécialement pour leur soutien financier et aides, l'Université de Jendouba, l'Institut Supérieur des Langues Appliquées et d'Informatique de Béja, le Laboratoire de recherche en Technologies de l'Information et de la Communication & Génie Electrique à l'ENSIT, l'Association Actes Académiques et l'usine Kromberg & Shubert.

Comités du colloque

Infol@ngues 2016 est organisé par l'Institut supérieur des Langues Appliquées et d'informatique de Béja en collaboration avec les deux Laboratoires de recherches LaTICE, Université de Tunis et LLTA, Université de Sfax.

Comité Scientifique

Président :

- Prof. Abdelmajid Ben Hamadou, Université de Sfax, Tunisie.

Membres :

Eric.Wehrli, Université de Genève, Suisse.
Mohamed Jemni, Université de Tunis, Tunisie.
Mourad Elloumi, Université de Tunis El Manar, Tunisie.
Ezzedine Zagrouba, Université d'El Manar, Tunisie.
Mohamed Gamoudi, Université de la Manouba, Tunisie.
Sadok Ben Yahya, Université de Tunis El Manar, Tunisie
Mohamed Jabir, Université de Carthage, Tunisie.
Yahya Slimani, Université de la Manouba, Tunisie.
Faiez Gargouri, Université de Sfax, Tunisie.
Rim Faiz, Université de Carthage, Tunisie.
Ghassan Mourad, Université Libanaise, Liban.
Jacqueline Bacha, Université de Jendouba, Tunisie.
Lamia Hadrich Belguith, Université de Sfax, Tunisie.
Tahar Labbassi, Université de Tunis, Tunisie.
Sami Faiz, Université de la Manouba, Tunisie.
Mohamed Bouattour, Université de Sfax, Tunisie.
Ricco Rakotomalala, Université de Lyon, France.
Max Selberstein, Université de Franche-Comté, France.
Akila Baklouti, Université de Sfax, Tunisie.
Imed Ben Amar, Université de la Manouba, Tunisie.
Faiza Derbel, Université de la Manouba, Tunisie.
Afef Kacem, Université de Tunis, Tunisie.
Ramzi Farhat, Université de Jendouba, Tunisie.
Karim Chine, Université de Cambridge, Royaume-Uni.
Slim Mesfar, Université de Carthage, Tunisie.
Aymen Khelifi, ENS Lyon, France
Asma Bouhafis, Université de Carthage, Tunisie.
Imed Mehnan, Université de Jendouba, Tunisie.
Sofien Haboubi, Université de Tunis, Tunisie.
Tarek Hamrouni, Université de la Manouba, Tunisie.
Rami Belkaroui, Université de Tunis, Tunisie.
Soumaya Cherichi, Université de Tunis, Tunisie.
Dhouha Jemal, Université de Tunis, Tunisie.
Sondess Missaoui, Université de Tunis, Tunisie.
Ameni Kacem, Université de Tunis, Tunisie.

Comité d'organisation

Président :

- Dr. Faouzi Mhamdi

Membres :

Ali Mansouri
Youcef Mezrigui
Lamia Mrad
Bedis Ben Ezedine
Rim Ktari
Mounir Argoubi
Ridha Djebali
Mondher Ebdeli
Asma Riyahi
Asma Gharbi
Anouar Mekni
Samir Ouechen
Bachar Mekni
Akrem Melki
Imen Romdhani
Maha Maaroufi
Olfa Bhiri
Iheb Djibi
Abdel Hafidh Mgaydiya
Kais Bouzidi
Kaouther Zouaghi
Naima Bjaoui
Fatma Mejri
-

Table de matières

Colloque invités

<i>Quel nouveau paradigme en sciences humaines à la lumière de l'informatique ?</i> Ghassan Mourad	2
Extraction Efficace des concepts formels « stables» , Sadok Ben Yahya	3
Constructing New Learning Zones in Electronic Learning Environments : Making the Peadagogical shift, Faiza Derbel	4

Apprentissage Automatique du Langage

<i>Retrieving Passages for Question Answering.</i> Nouha Othman	6
<i>From Page to Screen :A Study of Hitchcock's Sabotage as a cinematic adaptation of Corad's The Secret Agent,</i> Asma Chahed	7
<i>Time-course of the construction of discourse-level representation in L1 and L2 ,</i> Marwa Mekni	8

Traduction Automatique

<i>الترجمة الآلية ومعالجة اللغات حاسوبيا</i> Aicha Zaidi	9
<i>Dictionnaire électronique et grammaires locales pour la traduction automatique des adjectifs.</i> Tawfik Khil	10
<i>Les problèmes de traduction automatique des constructions à verbes supports entre l'arabe et le français.</i> Hamdi Raja	11
<i>Traduction automatique des collocations nom+adjectif,</i> Siwar Mabrouk	12
<i>Causes des problèmes de traduction automatiques français-arabe des temps verbaux,</i> Fadhel Thouraya	13
<i>Traduction et Ordinateur !</i> Abdessattar Bouzid	14

Réseaux Sociaux

<i>Filtrage multicritères pour une recommandation personnalisée des tweets,</i> Dorsaf Ben Hadj Salem	15
<i>Optimal Sentiment Analysis Facebook' Statuses Mining.</i> Radhia Toujani	16
<i>Les écrits électroniques : faire du face à face avec de l'écrit</i> Najeh Elouni	17

Langue et Informatique

<i>Technology-enhanced formative assessment in English as a Foreign Language through Google Drive.</i> Asma Galoul	18
<i>An insight onto the utility of corpora in research studies : the example of nafs.</i> Abir Medini	19
<i>Recherche d'information juridique guidée par les ontologies.</i> Ines Berazaga	20
<i>Lignes directives pour le développement des jeux éducatifs ciblant les apprenants souffrant de retard mental.</i> Rihab Bessi	21

Traitement Automatique de la Langue

<i>L'expansion nominale dans un article scientifique et un texte littéraire : Étude comparative.</i> Hela Saidani	22
<i>Fonctionnement de la construction savoir (que) : étude et traitement au moyen du logiciel AntConc.</i> Rym Salem	23
<i>La complexité du traitement automatique des paraphrases.</i> Ifaoui Imen	24
<i>Résolution d'anaphores : Conception et évaluation des performances.</i> Afef Selmi	25
<i>Reconnaissance et extraction automatique des noms propres : une étude en corpus politique journalistique.</i> Yaccoub Ghérissi	26

Invités du colloque

Prof. Ghassan Mourad

Le professeur Ghassan Mourad, est un Enseignant-chercheur à l'université Libanaise et au Centre des Sciences du Langage et de la Communication depuis 2003. Il est le directeur du Centre des Sciences du Langage et de la Communication depuis 2007. Il est le Coordinateur de l'équipe LILAS (Linguistique, Ingénierie, Langue, Apprentissage, Sémiotique) depuis 2008. Entre 1997 et 2014 le Prof. Ghassan Mourad a été un Chercheur associé à l'équipe de recherche LaLIC (Langage, Logique, Informatique, Cognition; PARIS IV – SORBONNE). Il a encadré plusieurs thèses à l'Université libanaise et en cotutelle avec l'Université de Tunis. Il a publié plusieurs livres et auteur d'une vingtaine d'articles spécialisés publiés par des revues et des magazines scientifiques.

Prof. Sadok Ben Yahia

Sadok Ben Yahia is a Professor at the Computer Sciences department of the Faculty of Sciences of Tunis since October 2009. He obtained his Habilitation to Lead researches in Computer Sciences from the University of Montpellier in April 2009.

Currently, He is leading a group of researchers in Tunis, whose research interests include Formal concept analysis, Minimal transversal extraction and Ontology engineering. In the course of his scientific career, He has created more than 150 peer reviewed publications, including 7 editions of international journals special issues, 35 International Journal publications and more than 100 papers at international peer-reviewed conferences. In addition Sadok BEN YAHIA is an accepted and recognized member of the international research community of Data Mining. He has co-organized 2 international conferences (CLA2006, and EGC'2010) and various workshops and is regular member in the Program committee of the most important conference on the area of Data Mining (ICDM, CIKM, FUZZ-IEEE, EGC) and related areas. He acted as the co-track chair of Computational intelligence at the CIAA'2015 (<http://ciia2015.ensma.fr/?lng=fr>) and the co-program chair of the 12th edition of CLA conference (<http://cla2015.isima.fr/>).

Dr. Faiza Derbel

Dr. Faiza Derbel is Assistant Professor of English and Applied Linguistics at Faculté des Lettres, Arts et Humanités, Manouba. She teaches graduate courses in various areas of Applied Linguistics and supervises related research.

Her research interests include Computer-assisted Language Learning (CALL), ICT-supported teaching/learning, digital pedagogy and how they relate to Second Language Learning, assessment and teacher education. She regularly presents locally and internationally on issues related to ELT pedagogy in Tunisia. Faiza Derbel is founding member of Tunisia TESOL. She occupied the positions of President-Elect, President and is currently Past President.

Les participants du colloque

Retrieving Passages for Question Answering

Nouha Othman¹, Rania Othman¹ and Rim Faiz²

¹ Institut Supérieur de Gestion de Tunis, LARODEC, Université de Tunis, Tunisia

² IHEC Carthage, LARODEC, Université de Carthage, Tunisia

othmannouha@gmail.com, rania.othman@gmx.com, rim.faiz@ihec.rnu.tn

Abstract : The aim of question answering (QA) systems is to directly return accurate and short answers to questions asked in a natural language from a massive document collection. Retrieving passages remains a crucial but also a challenging task in QA. In this paper, we propose a novel passage retrieval approach using n-grams. This approach relies on a new measure of similarity between a passage and a question for the selection and ranking of the different passages. More precisely, our measure is based on the dependency degree of n-gram words of the question in the passage. We validate our approach by the development of the "SysPex" system. Our experiments showed promising results and significant improvement over a previous similar system.

Keywords: Question Answering, Passage Retrieval, N-grams, Similarity measure.

From Page to Screen:
A Study of Hitchcock's *Sabotage* as a cinematic adaptation of Conrad's *The Secret Agent*

Asma Chahed

Faculté des Lettres et des Sciences Humaines, Université de Sousse

Miss-English21@hotmail.com

Abstract: Most of Conrad's writings were adapted to screen because of their importance which results from their analysis of contemporary issues that touch every individual. *The Secret Agent* is one of them. Nevertheless, due to its anarchistic violence and political terrorism, this novel was not met with huge success upon its publication. Its cinematic adaptation contributed to attracting critics' attention towards the work as a whole that was considered afterwards as the most brilliant depiction of the post 9/11 novel. In light of Hutcheon's theory of adaptation, the current paper studies the mechanisms of filmic adaptation and its construction through context, characters, themes, imagery, intertextuality and audience manipulation and the imposition of a sense of trauma through public unrest and panic in order to examine Hitchcock's degree of innovation.

Through choosing Conrad's *The Secret Agent*(1907) as his source of adaptation, in "Sabotage" (1936), Hitchcock highlights the intimate link between literature and cinema and proves Tagre's saying that "cinema is still playing second fiddle to literature". Nevertheless, fiction relies on reader's imagination as opposed to cinema which relies on the visual and aural. In other words, both fields complement each other. The film semiotician Christian Metz highlights the need for connection between literature and cinema through affirming that cinema is superior to literature because it broadens its analysis through being different and that is why according to him: "there is a reason for the possibility as well as for the necessity of adaptations" (Hutcheon 3).

Time-course of the construction of discourse-level representations in L1 and L2

Marwa Mekni Toujani
Higher Institute of Languages, Tunis, Tunisia

marwaISLT@gmail.com

Abstract: This paper is grounded within the field of experimental psycholinguistics and more precisely within the area of discourse processing. It aims at investigating the speed with which readers are sensitive to anomalous information in text with respect to prior context. Two experiments based on word-by-word self-paced reading methodology will examine the time-course of the construction discourse-level representations in L1 and L2 reading. Those experiments will be run on a windows Laptop using PsychoPy. While experiment 1 will use spatial anomaly, experiment 2 will use character-attribute anomaly. The present paper aims at raising the awareness of Arabic software programmers to develop software that are able to run similar experiments using the Arabic Script to allow Arabic psychological research that can be run on computers to be easily implemented.

Key words: Discourse processing, time-course, self-paced reading, PsychoPy, SPSS.

الترجمة الآلية ومعالجة اللغات حاسوبيا
الأستاذة: عائشة زايدي
جامعة باجي مختار – عنابة-

wardanounou2@gmail.com

المخلص:

يتناول هذا البحث مجالاً مهماً من مجالات معالجة اللغة بواسطة الحاسوب، وهو مجال الترجمة الآلية، وهذه الأخيرة لها أهمية وتاريخ وعوامل نجاح وفشل، وسيعرض البحث لهذه الجوانب .

إن الثورة الرقمية أصبحت متزايدة أكثر من ذي قبل في إنتاج آلات تستحق أن توصف بالكائنات الميكانيكية، والتي سيكون توسعها في المستقبل أن تكون قادرة على فهم اللغة وتوصيفها، وأن تكون قادرة على القراءة بنفسها، وفهم ما تقرأه، بل تتجاوز ذلك كله إلى التفاعل مع العالم المادي، والتعامل مع وسائل الإعلام بجميع أشكالها.

الكلمات المفتاحية: الترجمة ، الآلية ، اللغة ، الرقمنة ، الذكاء الاصطناعي

Dictionnaire électronique et grammaires locales pour la traduction automatique des adjectifs

Tawfik khil

Faculté des Lettres et Sciences Humaines de Sfax, Université de Sfax
Laboratoire Langage et Traitement Automatique

khiltawfik@yahoo.fr

Résumé : Cette contribution vise à exposer les principes cruciaux pour une traduction juridique automatique admise de l'arabe vers le français. Par des enchaînements théoriques et pratiques, nous assisterons à une démonstration d'une traduction juridique postulant la double performance ; la qualité linguistique et la sécurité juridique. Pour y arriver, trois étapes seront proposées. La perspective méthodologique de ces trois étapes est le traitement du terme juridique d'abord dans sa langue source et dans son système juridique de départ. Ensuite, son traitement dans le transfert vers la langue cible et le système juridique d'arrivée. Cette méthodologie sera le résultat d'une forte coopération de deux piliers indispensables au traducteur juridique ; Le juriste comme un appui du traitement traductif inter-juridique et le linguiste comme un appui du traitement traductif inter-linguistique.

Mots clés : Traduction juridique, Langue source, Langue cible, Systèmes juridiques, Décodage, encodage.

Les Problèmes de Traduction Automatique des Constructions à Verbes Supports entre l'Arabe et le Français

HAMDI Raja

Faculté des Lettres et Sciences Humaines de Sfax, Université de Sfax
Laboratoire Langage et Traitement Automatique

hamdiraja74@yahoo.fr

Résumé : Cet article porte sur la traduction automatique des constructions à verbe support français en arabe à l'aide des systèmes de traduction automatique. En effet, les constructions à verbe support prennent une importance prépondérante dans les systèmes de traitement automatique de langage naturel en particulier dans les systèmes de traduction. Leur traitement informatique est une tâche difficile puisqu'elles posent de sérieux problèmes du fait de leurs contraintes syntaxiques et sémantiques.

Bien que ces expressions constituent une difficulté pour de nombreuses applications en TAL, c'est à coup sûr dans le domaine de la traduction automatique que l'absence d'un traitement adéquat de ces expressions se fait sentir. Autrement, le problème des collocations aux verbes supports est plus complexe à traiter dans sa globalité ce qui peut causer problème lors des tâches de traitement des langues. Les constructions à verbe support sont sources d'erreur des systèmes de traduction automatique. La raison pour laquelle ces systèmes ont des résultats peu satisfaisants lorsqu'ils rencontrent ces groupements usuels, est qu'ils ne les considèrent pas comme un problème spécifique avec des solutions spécifiques, ainsi, ils manquent de ressources concernant ce phénomène linguistique et n'ont généralement pas les informations linguistiques nécessaires pour remonter jusqu'au sens, ni pour ensuite générer la bonne expression à partir du sens. Les systèmes de traduction automatique devraient donc partir de l'idée que l'opération de traduction des constructions à prédicats nominaux est une mise en relation d'équivalence entre leurs schémas d'arguments respectifs en L1 et L2 et doivent prendre en considération les spécificités de chaque système linguistique, à savoir les informations relatives aux compatibilités entre un nom prédicatif et les verbes supports qu'il sélectionne, le degré d'appropriation de ceux-ci, ainsi que les propriétés morphologiques du prédicat et leur impact sur les reformulations par paraphrase.

Mots clés : traitement automatique de langage naturel, constructions à verbes supports, les systèmes de traduction automatique, problèmes de traduction automatique, paraphrase.

Traduction Automatique des Collocations Nom + Adjectif

Siwar Mabrouk

Faculté des Lettres et Sciences Humaines de Sfax, Université de Sfax
Laboratoire Langage et Traitement Automatique

siwar.mabrouk@yahoo.fr

Résumé : Cet article a pour objectif de mettre en exergue le phénomène collocatif dans les systèmes de traduction automatique.

En fait, caractérisées par l'ambiguïté de leur frontière et leur complexité, les collocations constituent une problématique dans la traduction automatique ainsi que dans la traduction humaine. Cela rend l'existence d'un traitement adéquat de ce type de combinaisons, aussi variées, est quasiment impossible, sachant que les collocations varient d'une communauté linguistique à une autre, voire d'une personne à une autre. Ajoutant à cela le caractère spécifique de ces expressions, conçues comme peu décodables et peu prédictibles, ce qui rend difficile leur traitement dans les systèmes de traitement automatique de langage naturel, essentiellement dans les systèmes de traduction.

En effet, si nous examinons la traduction automatique des collocations, nous constatons que celles-ci sont considérées comme les principales sources d'erreurs essentiellement par rapport à d'autres phénomènes linguistiques, ce qui rend la qualité de la traduction automatique proposée est peu admissible et même mauvaise dans certains cas. Certes, cela revient au fait que les systèmes automatiques considèrent les collocations définies comme des groupements usuels, comme de simples mots avec espaces. Or, imposant un problème particulier aux systèmes de traduction automatique, implique certainement des solutions particulières telles que l'implémentation d'une base de données suffisante qui permet de déterminer le sens de la collocation à traduire ainsi que de trouver la bonne traduction ou son équivalent dans la langue cible.

De ce fait, et afin de garantir une traduction de bonne qualité de la collocation de la langue source, le système de traduction automatique devrait respecter les spécificités linguistiques de la langue source L 1 et de la langue cible L 2.

Mots clés : collocations – traduction automatique – systèmes de traduction automatique – difficultés automatiques.

Causes des problèmes de traduction automatique français-arabe des temps verbaux

Thouraya Fadhel

Faculté des Lettres et Sciences Humaines de Sfax, Université de Sfax
Laboratoire Langage et Traitement Automatique

fadhel.thouraya@gmail.com

Résumé : Cet article vise à montrer que l'on peut prédire en amont le type du problème qui peut surgir lors de la traduction (automatique ou non) des temps verbaux dans le couple français-arabe. La redécouverte des systèmes temporels propres aux deux langues aidera à trouver des solutions systématiques et automatisables pour chaque type de problème et pourra, par conséquent, contribuer à l'amélioration des traducteurs automatiques.

Mots clés : Traduction automatique, temps verbaux, français, arabe, problème.

Traduction et ordinateur!

Bouzid Abdessattar
Faculté des Lettres et Sciences Humaines de Sfax, Université de Sfax
Laboratoire Langage et Traitement Automatique

abdessattar.bouzid08@hotmail.fr

Résumé : Les langues naturelles, le langage et les linguistiques deviennent un enjeu international qui engage toute l'espèce humaine. Après la deuxième guerre mondiale, la communication de masse se développe dans un monde multilingue. En effet, la démocratisation des voyages entre les nations, le progrès scientifique et la création de nouvelles technologies nous mettent en contact avec les langues et les linguistiques les plus diverses.

Dans cette relation étroite entre langues et technologies, la traduction revêt, comme l'a montré Roman Jakobson, une valeur importante au regard de la linguistique. Donc, cette opération touche les différents types de texte (journalistique, littéraire, politique, économique ou bien juridique), et s'étend et se diversifie pour prendre de nouvelles formes à travers de nouvelles technologies, à savoir la traduction automatique, qui constitue la seule médiation non seulement entre les langues, mais également entre les cultures. Ce qui mentionne donc son rôle bien déterminant.

Dans cette perspective, s'inscrit notre projet d'étude qui s'intitule "*Traduction et ordinateur!*". Nous commencerons, tout d'abord, par étudier la notion de traduction automatique en mettant l'accent sur sa définition exacte et ses différents enjeux. Ensuite, nous allons établir une petite application portant sur deux langues: le français et l'arabe. Enfin, nous terminerons par démontrer la valeur chère de la traduction automatique sur la recherche et l'enseignement.

Filtrage Multicritères pour une Recommandation Personnalisée des Tweets

Dorsaf Ben Hadj Salem et Asma Bouhafs
Institut des Hautes Etudes de Commerce de Carthage

dorra.liaa2@gmail.com, asma_bouhafs@yahoo.com

Résumé: Ce travail de recherche décrit une approche de recommandation des tweets pertinents qui intègre différents critères du microblog Twitter. Nous avons exploité l'évolution importante qu'a réalisée Twitter en 2014 au niveau de ses fonctionnalités pour améliorer nos résultats de recommandation. En effet, l'objectif principal de cette approche est d'offrir des résultats qui répondent aux besoins d'informations de l'utilisateur correspondant à ses intérêts personnels. Pour atteindre cet objectif nous avons proposé un système de recommandation personnalisé en mettant au point une nouvelle mesure pour calculer le score de classement des tweets utiles. A fin d'évaluer la performance de notre système, nous avons appliqué un filtrage multicritères et nous avons utilisé un corpus pour étudier l'impact de l'association des critères Retweets, Favoris, Hashtags et URLs sur les résultats de la recommandation.

Optimal Sentiment Analysis Facebook' Statuses Mining

Radhia Toujani and Jalel Akaichi

Universite de Tunis Institut Superieur de Gestion, Université de Tunis

toujaniradia@gmail.com, j.akaich@gmail.com

Abstract: Recently, it seems to mark that the widespread awareness of the research on sentiment analysis and opinion mining raise due to the rise of machine learning methods in natural language processing and information retrieval. In the current work, we aim to apply machine learning and sentiment classification to analyze Tunisian users' statuses updates on Facebook during the Tunisian Election". To achieve this purpose, we propose the combination of machine learning and metaheuristic to ensure the optimality of Support Vector Machine parameter selection. In this combination, we present a model selection for Support Vector Machine via Tabu Search strategy. Sentiment analysis mainly deals with the evaluation type of opinions which imply positive or negative sentiments. However, in this paper the proposed sentiment classification method focus also on fuzzy opinion. That's why we propose to follow Fuzzy Support Vector Machine during the classification process. This later is compared with the basic Support Vector Machine referring to the standard classification performance metrics used in previous information retrieval and text classification studies: accuracy, precision, recall, and F-measure. Besides divers average training error for the proposed Fuzzy Support Vector Machine are generated through the application of Cplex solver.

Key words: Social network, social network analysis, opinion mining, sentiment classification, machine learning, natural language processing, metaheuristic.

Les écrits électroniques : faire du face à face avec de l'écrit

Najeh Elouni^{1,2}

¹ Faculté des Lettres et Sciences Humaines de Sfax, Université de Sfax
Laboratoire Langage et Traitement Automatique

² Université de Bourgogne Centre Interlangues Texte Image Langage (EA 4182)

najhelouni@gmail.com

Résumé : Le présent travail met l'accent sur l'association de divers codes sémiotiques linguistiques et extralinguistiques dans les forums de discussion. Il part d'un corpus fortement marqué par l'emploi de procédés d'écriture habituellement utilisés dans une communication en présentiel. Ces procédés non-verbaux servent souvent à exprimer les émotions des utilisateurs. Ils entraînent une certaine similarité entre les conversations des corpus en présentiel et en ligne et amènent à s'interroger sur le rapport particulier existant entre les écrits médiatisés par ordinateur et la communication orale donnant ainsi une dimension orale aux écrits électroniques. Nous avons relevé divers procédés permettant la mise en scène et la théâtralité dans notre corpus dont nous citons notamment les émoticônes, les interjections, les formes acronymiques voire des procédés graphiques comme l'écriture en capitales et la ponctuation expressive. Nous avons rassemblé ces différents procédés langagiers comme étant appartenant à une même famille permettant de « faire du face à face avec de l'écrit ». Cette réflexion permet d'identifier nettement la conversation en face à face comme cadre de référence dans lequel s'inscrit la communication médiatisée par ordinateur, avec des caractéristiques qui ne sont plus seulement stylistiques mais interactionnelles et communicationnelles.

Mots-clés : CMO, écrit électronique, oral, en présentiel, en ligne, émotion, émoticône.

Technology-enhanced formative assessment in English as a Foreign Language through Google Drive

Asma Maaoui
Higher Institute of Studies Applied In Humanities of Tunis
University of Tunis

asma.maaoui@gmail.com

Abstract: Formative Assessment (FA) based on Information and communication-technologies (ICT) can enhance student learning. Drawing on theories of Assessment for Learning and self-efficacy, this study investigates the value of conjoining English as a Foreign Language (EFL) FA with tools available through Google Drive to promote language learning. A case study based on an interview with four learners of EFL in higher education in Tunisia aimed at studying their experiences with FA using Google Apps. Data were also collected from three teachers using focus group interviews. The results indicate that technology use for FA purposes motivated the learners and made them more engaged in collaboration, peer and self-assessment, as well as immediate feedback. Online platforms offered by Google also fostered the students' thinking about their future learning plans in the light of their assessment outcomes.

An insight onto the utility of corpora in research studies: the example of *nafs*

Abir Medini
University of Hamburg

abir.medini@gmail.com

Abstract: New insights for reflexive pronouns have been gained from various studies led so far (Steinbach 2002, Koenig & Siemund 2000). They show the developmental use of *itself* starting from its use as intensifier to become a reflexive marker and more frequently a non-reflexive marker; This phenomenon has been substantiated through data analysis of the British National Corpus (Siemund 2010). Astonishingly, the identity of the reflexive marker *nafs* in Arabic has shown to shift throughout time. 8274 examples were detected in the Arabic Gigaword Corpus, in which *nafs* is used non-reflexively out of 92127 examples where *nafs* occurs always in various functions. The findings reveal the move of *nafs* from a grammatical function to another, similarly to the way the English reflexive marker behaves. The two languages not only supplement each other but mutually support progress in research of language contact.

Keywords : language contact, corpus linguistics, middle domain

Recherche d'information juridique guidée par les ontologies

Ines Berrazega¹, Rim Faiz² et Ghassan Mourad³

LARODEC, ISG de Tunis B.P.1088, 2000 Bardo, Tunisie
LARODEC, IHEC de Carthage, 2016 Carthage Présidence, Tunisie,
Centre des Sciences du Langage et de la Communication Université Libanaise

ines_berrazega@yahoo.fr, Rim.Faiz@ihec.rnu.tn, ghassan.mourad@ul.edu.lb

Résumé : Avec l'essor des technologies de l'information et de la communication, nous constatons une croissance considérable de la quantité des documents juridiques quotidiennement créés et diffusés. Cette croissance est accompagnée par un besoin de mettre en place des systèmes de recherche d'information adaptés à la nature de ces documents de spécialité. A cet égard, le développement des méthodes d'annotation sémantique et d'indexation des contenus juridiques se fut de plus en plus nécessaire. De telles méthodes pourraient exploiter les techniques du web sémantique, notamment les ontologies de domaine. Nous abordons dans cet article la problématique de l'exploitation des ontologies juridiques dans le processus de recherche d'information à partir des documents juridiques.

Mots clés : Recherche d'information, ontologies juridiques, web sémantique, indexation

Lignes directives pour le développement des jeux éducatifs ciblant les apprenants souffrant de retard mental

Ramzi Farhat¹, Rihab Bessi² et Hanene Rejeb³

^{1,2} Laboratoire de recherche en Technologies de l'Information
et de la Communication & Génie Electrique

³ Laboratoire I3M (Information Milieux Médias, Médiations) de Nice, Université de Nice

ramzi.farhat@esstt.rnu.tn, bessi_rihab@hotmail.fr, rejebimmo@yahoo.fr

Résumé : Notre travail de recherche porte sur l'usage des jeux éducatifs pour les personnes souffrants d'un retard mental. Essentiellement, nous nous sommes penchés sur l'incidence de la présence ou l'absence des éléments multimédias dans les jeux éducatifs pour cette catégorie d'apprenants. Dans ce cadre, nous avons conçu et évalué un ensemble de versions d'un même jeu éducatif comportant chacune une combinaison différente d'éléments multimédia. Ceci nous a permis de définir un ensemble de lignes directives pour toute personne souhaitant développer des jeux éducatifs pour des apprenants souffrants d'un retard mental.

Mots clés : jeux éducatifs, retard mental, ligne directives, multimédia

L'expansion nominale dans un article scientifique et un texte littéraire : Étude comparative

Hela SAIDANI

Institut Supérieur des Langues de Tunis (ISLT)
Laboratoire Langage et Traitement Automatique

helasaidani1991@gmail.com

Résumé : La langue et l'informatique sont devenues indissociables de nos jours. Effectivement, grâce aux collaborations entre linguistes et informaticiens, le traitement d'un corpus donné est devenu automatique, d'où le « Traitement Automatique du Langage » ou TAL, d'ailleurs, ce qui permet une analyse plus rapide et, surtout, précise. Les corpus sur lesquels les chercheurs travaillent sont multiples. Nous nous proposons de travailler sur le Syntagme Nominal qui peut être étendu par une expansion. L'expansion est, traditionnellement, un adjectif, un syntagme prépositionnel, un syntagme nominal apposé, une subordonnée relative adjectivale ou une conjonctive complétive.

Les types de textes utilisent différemment la structure en question. Chaque texte a des spécificités quant au choix des expansions des noyaux nominaux. Nous pouvons trouver des textes qui comportent des noms avec les différentes expansions mentionnées ci-dessus, comme il existe des textes dont les auteurs font un choix, une sélection des expansions nominales.

Nous suggérons de travailler sur deux corpus différents ; l'un est littéraire intitulé *La Parure* de Guy de Maupassant, et l'autre scientifique dont le titre est *Une étoile à la luminosité étrange surprend les scientifiques* et ce, afin de voir le fonctionnement dudit SNE et pour relever certaines régularités qui lui sont propres.

Pour ce faire, nous aurons essentiellement recours à deux logiciels: *Antidote* et *AntConc*, et nous insérerons, par la suite, les données que nous obtiendrons dans le tableur de Microsoft

Excel, sachant que, dans ce tableur, il existe deux sous-menus, le *tableau croisé dynamique* et le *sous-total*, qui permettent d'obtenir des résultats par rapport à ce que nous voulons relever.

Nous rendrons également compte des différentes structures du SNE et nous verrons combien d'expansions le noyau nominal peut avoir. Nous ferons un descriptif des expansions nominales en partant des cas les plus réguliers vers les cas les moins réguliers.

**Fonctionnement de la construction savoir (que) :
Etude et traitement au moyen du logiciel AntConc**

Rym Salem

Faculté des Lettres et Sciences Humaines de Sfax, Université de Sfax
Laboratoire Langage et Traitement Automatique

rymme.salem@gmail.com

Résumé: Notre objectif consiste à examiner le fonctionnement de la construction savoir (que) au moyen du logiciel AntConc. En fait, *savoir que* n'est qu'un outil linguistique parmi d'autres qui expriment la modalité épistémique. *Savoir que* se distingue par certaines propriétés qui lui assurent un comportement particulier.

La complexité du traitement automatique des paraphrases

Ifaoui Imen

Faculté des Lettres et Sciences Humaines de Sfax, Université de Sfax
Laboratoire Langage et Traitement Automatique

Imen_ifaa@live.fr

Résumé : La langue jouit d'une richesse à la fois qualitative en recelant une variété de mots et de formules différentes désignant la même réalité et quantitative en offrant des formulations variées renvoyant à une même réalité. L'étude de l'un des mécanismes de la langue, à savoir *la paraphrase* et la mise en œuvre de son traitement automatique par des machines est un domaine d'études riche en applications.

Dans le présent article, nous allons essayer de souligner la complexité de l'étude de la *paraphrase* avec ses différentes formes lexicales, syntaxiques et stylistiques à cause de la présence de l'ambiguïté. Ce phénomène se manifeste dans la multitude d'interprétations possibles pour un seul énoncé.

La quantité d'implicite dans la communication naturelle exige un surplus de travail interprétatif pour un décodeur. Le recours spontané à l'intuition met en place une certaine subjectivité difficile à gérer également pour le récepteur humain et le récepteur machine.

Ces spécificités de la langue qui interviennent lors de décodage de *la paraphrase* doivent être prises en considération par les outils du TAL pour faciliter son extraction.

Nous allons procéder dans notre travail à une utilisation de la plate-forme *Nooj* et le logiciel du traitement automatique *AntConc* pour faire l'extraction automatique de différentes reformulations paraphrastiques entourant des marqueurs de reformulation paraphrastique (MRP) pour distinguer entre les relations paraphrastiques et non-paraphrastiques en utilisant comme corpus l'œuvre de Montesquieu intitulée *Lettres persanes*.

Ce travail ne vise pas à présenter de manière exhaustive ce que nous connaissons du TAL. Nous envisageons de souligner la complexité de l'étude de *la paraphrase* car la machine ne dispose que de cette connaissance d'arrière-plan, donc, les outils de TAL ne sont pas susceptibles de résoudre complètement le comportement langagier.

Mots clés : la paraphrase, MRP, extraction automatique, Nooj, AntConc, complexité du traitement automatique.

Résolution d'anaphores : Conception et évaluation des performances

Afef Selmi

Faculté des Lettres et Sciences Humaines de Sfax, Université de Sfax
Laboratoire Langage et Traitement Automatique

selmiafef_ling@yahoo.fr

Résumé : On a souvent tendance à relever des contradictions en TAL entre les systèmes qui utilisent des connaissances linguistiques et ceux qui recourent simplement à des indices de surface. Les premiers systèmes font généralement défaut de fiabilité étant donné qu'ils utilisent des connaissances parfois erronées car elles sont calculées automatiquement ou incomplètes lorsqu'elles sont produites manuellement. Les seconds systèmes se basent sur des méthodes d'apprentissage automatique et sur des indices de surface plus accessibles mais inefficaces quant aux cas complexes.

Il faut préciser que cette communication revient tout d'abord sur les raisons de l'opposition précédente.

L'anaphore est définie comme étant la relation linguistique existant entre deux entités textuelles précises et qu'une entité textuelle (l'*anaphore*) renvoie à une autre entité du texte (l'*antécédent*). Il faut noter que la présence d'anaphores diminue considérablement les performances des systèmes de TAL, ce qui explique alors le fait que le problème de leur résolution est étudié depuis longtemps. Seulement le travail dirigé ne s'est intéressé qu'à la résolution de l'anaphore du pronom *il* dans les textes français, l'anaphore la mieux connue et la plus facile à résoudre.

Généralement, la résolution automatique de l'anaphore distingue trois étapes : la distinction des pronoms anaphoriques et impersonnels, la sélection des candidats possibles à l'antécédence et le choix de l'antécédent. Dans chacune de ces étapes, les premiers systèmes conduits dans la littérature utilisaient des connaissances linguistiques complexes qui reflétaient en fait les contraintes syntaxiques et sémantiques qui déterminent l'anaphore. Le calcul automatique de ces connaissances considéré alors comme impossible ou trop peu fiable pour être utilisable, les connaissances linguistiques étaient produites manuellement, ce qui nécessitait donc un travail d'analyse préalable des textes exhaustif.

« **Reconnaissance et extraction automatique des noms propres :**

Une étude en corpus politique journalistique »

Yacoub Ghérissi

Institut Supérieur des Langues de Tunis

gh.yac.2013@gmail.com

Résumé : La reconnaissance des noms propres (désormais Npr) fait partie des problèmes qui se posent au traitement automatique des langues naturelles (TALN).

Pour les études purement linguistiques, focalisées presque exclusivement sur cette partie du discours, il est le plus souvent question de sémantique et des caractéristiques intrinsèques du Npr. Les systèmes informatiques, de leur côté, s'appuient sur la présence de la majuscule et sur des dictionnaires informatiques comportant de longues listes de Npr. Comme nous le verrons, ces listes, même si elles connaissent une mise à jour constante, ne peuvent fonctionner que sur les articles relatifs à la France. En effet, le Npr *Jacques Chirac*, pour un journaliste français qui s'adresse à un lecteur français à propos de la politique française, ne pose ni problème d'interprétation ni de référence. Cependant, pour les Npr étrangers, aucune liste n'est disponible ou même possible à établir.

Notre objectif est de changer de perspective pour tenter de démontrer que pour étudier une classe grammaticale (ici les Npr), il vaudrait mieux, comme le suggère la linguistique distributionnelle, caractériser les distributions d'une unité lexicale plutôt que l'unité en elle-même.

Notre corpus est constitué des articles du mois de janvier 2005 du quotidien français *Le Monde* classés en deux rubriques, « France » et « international ».

Dans le type de discours qui nous intéresse, à savoir le discours politique journalistique, un Npr n'est jamais employé sans être précédé ou suivi d'une expansion qui fournit certaines informations sur la personne en question. Sans cette glose, le lecteur n'est pas en mesure d'identifier le référent du Npr.

Nous tenterons de démontrer comment, grâce au co-texte, c-à-d en caractérisant les contraintes qui pèsent sur les contextes gauche et droit, il est aisé d'aboutir à une caractérisation des Npr qui permettrait aussi bien l'extraction automatique que la traduction.