
Préface

En mémoire de Jean Véronis

Alors que nous finissons la mise en forme de ce numéro, la communauté du traitement automatique des langues est endeuillée par la disparition accidentelle de Jean Véronis survenue le 8 septembre 2013. Jean, en tant que président de l'ATALA, a marqué la revue TAL de manière décisive en poussant à la conversion de celle-ci, alors imprimée et distribuée par un éditeur, en revue en libre accès sur Internet, incitant ainsi notre communauté à partager ouvertement les résultats de ses recherches. Ce numéro est un numéro non thématique, ordinairement appelé « *Varia*. » Rappelons que cette dénomination est due à Jean qui avait eu le premier l'idée d'un tel numéro annuel regroupant des articles sur des sujets variés du traitement automatique des langues. Au nom du comité d'édition et des membres de notre communauté, nous disons ici notre douleur à la suite de cette grande perte.

Dans ses diverses activités, le comité de rédaction est de temps en temps sollicité pour prendre position sur certains aspects de la vie éditoriale scientifique. Ainsi, le comité de rédaction a récemment signé une pétition pour l'accès ouvert aux résultats de la recherche¹, réaffirmant ainsi les principes que la revue TAL, sous l'impulsion de Jean Véronis, s'est donné lors de son passage à l'accès ouvert.

Comme il est de coutume, cet éditorial de numéro non thématique donne des statistiques sur l'ensemble des articles soumis à notre revue et publiés sur une certaine période de temps : nous considérons les dix derniers numéros des trois dernières années, du début de 2010 jusqu'à ce numéro *Varia* de 2013 inclus. Les statistiques des deux numéros à venir de 2013 (*Entités nommées* et *Réseaux sociaux et TAL*) ne sont pas encore connues sauf pour le nombre de soumissions au numéro *Entités nommées* (treize soumissions). Le tableau 1 donne les taux de sélection par numéro et par volume. La ligne du total synthétise ces chiffres sur l'ensemble des dix numéros considérés.

Le taux de sélection sur ces numéros s'élève à 32,2 % en moyenne. Notre revue est donc assez sélective avec un article publié sur trois soumis. Le principe suivi par le comité de rédaction est d'accepter les articles selon leur qualité, indépendamment du nombre d'articles. On constate, pour les numéros considérés, que le nombre d'articles

1. <http://iloveopenaccess.org/signataires/>

publiés par numéro peut varier fortement avec un minimum de trois articles seulement pour le numéro 53-1 *Varia* et un maximum de neuf articles pour le numéro 52-3 *Ressources libres*. Le nombre de soumissions par numéro varie presque du simple au double, ici de onze à vingt-cinq.

Par tradition, les statistiques sur l'origine géographique des articles considèrent le pays du premier auteur : hors de France ou pas. La langue de publication fait aussi l'objet de comptages. Rappelons la règle énoncée dans tous les appels à communications de la revue : « Les soumissions en anglais ne sont acceptées que pour les auteurs non francophones. » Le tableau 2 donne des résultats pour les mêmes numéros que précédemment. Au total, les trois quarts des articles sont publiés en français (25,4 % publiés en anglais) et presque la moitié des articles (40 %) ont un premier auteur hors de France. Les résultats étant donnés en pourcentages, on remarquera que les pourcentages par volume et les pourcentages totaux ne sont pas des moyennes des lignes précédentes puisque les dénominateurs, c'est-à-dire le nombre d'articles publiés, changent à chaque fois. Comme pour le taux de sélection, on constate des fluctuations importantes par numéro : certains numéros n'ont aucun article en anglais (51-3 *Opinions et sentiments*, 52-1 *Varia* et 53-1 *Varia*) et d'autres sont majoritairement en anglais (53-2 *Temps et espace*, 53-3 *Erreurs*). Peut-on en conclure que l'internationalisation des comités spécifiques aux numéros thématiques – chose à laquelle notre comité de rédaction apporte tout son soin en accord avec les rédacteurs invités – augmente régulièrement depuis une dizaine de numéros l'attractivité de notre revue en dehors de la francophonie ?

Pour ce qui est de l'avenir proche, le volume 55 de la revue, à paraître en 2014, inclura de nouveau un numéro non thématique. L'appel est lancé et la soumission d'ores et déjà possible sur le site <http://tal-55-1.sciencesconf.org>.

De fait, le comité de rédaction, en parallèle à l'utilisation de la plateforme *sciencesconf* s'interroge sur le moyen le plus pratique de pouvoir soumettre des articles tout au long de l'année en dehors des appels explicites. Ces soumissions seraient alors traitées pour être publiées, si acceptées, dans le numéro *Varia* suivant.

Le numéro présent contient les articles retenus lors de l'appel non thématique clos à la mi-décembre 2013 portant sur tous les aspects du traitement automatique des langues. Il a suscité vingt et une soumissions dont quatre en anglais. À l'issue du processus de sélection habituel à deux tours, sept articles ont été retenus pour publication, dont un en anglais. Un article sélectionné en première lecture n'a pas été resoumis pour le second tour de sélection. Le taux de sélection s'établit donc à 33 %, ce qui correspond à la moyenne sur les dix derniers numéros.

L'un des articles de ce numéro résume, en un peu plus de trente pages, le contenu de la thèse ayant obtenu le prix de thèse de l'ATALA pour l'année 2012. Deux autres articles offrent l'originalité d'être accompagnés de données ou de programmes. Les lecteurs trouveront sur le site de la revue des archives accompagnant ces articles. Cet usage se répand peu à peu dans notre communauté à la suite de conférences comme CICLing ou ACL où les organisateurs incitent les auteurs à partager leur programmes

Intitulé	Vol.	N°	Année	Soumis	Acceptés	% acceptés
Varia	51	1	2010	20	5	25,0 %
Multilinguisme	51	2	2010	16	4	25,0 %
Opinions et sentiments	51	3	2010	25	5	20,0 %
Sous-total	51		2010	61	14	23,0 %
Varia	52	1	2011	22	6	27,3 %
Morphologie	52	2	2011	13	5	38,5 %
Ressources libres	52	3	2011	17	9	53,0 %
Sous-total	52		2011	52	20	38,4 %
Varia	53	1	2012	15	3	20,0 %
Temps et espace	53	2	2012	11	6	54,5 %
Erreurs	53	3	2012	11	5	45,5 %
Sous-total	53		2012	37	14	37,8 %
Varia	54	1	2013	21	7	33,3 %
Total			Dix derniers n^{os}	171	55	32,2 %

Tableau 1. Taux de sélection aux appels de la revue TAL sur les dix derniers numéros de la période 2010-2013

Intitulé	Vol.	N°	Année	% 1 ^{er} auteur hors France	% en anglais
Varia	51	1	2010	40,0 %	20,0 %
Multilinguisme	51	2	2010	50,0 %	25,0 %
Opinions et sentiments	51	3	2010	0,0 %	0,0 %
Pourcentages par volume	51		2010	28,6 %	14,3 %
Varia	52	1	2011	33,3 %	0,0 %
Morphologie	52	2	2011	60,0 %	40,0 %
Ressources libres	52	3	2011	44,4 %	22,2 %
Pourcentages par volume	52		2011	45,0 %	20,0 %
Varia	53	1	2012	0,0 %	0,0 %
Temps et espace	53	2	2012	50,0 %	66,6 %
Erreurs	53	3	2012	80,0 %	60,0 %
Pourcentages par volume	53		2012	50,0 %	50,0 %
Varia	54	1	2013	28,6 %	14,3 %
Pourcentages totaux			Dix derniers n^{os}	40,0 %	25,5 %

Tableau 2. Proportion des articles publiés d'un premier auteur hors de France et proportion des articles publiés rédigés en anglais sur les dix derniers numéros de la période 2010-2013

et leurs données et accordent à de telles soumissions un crédit supplémentaire. Le comité se félicite de telles initiatives venant des auteurs, mais s'interroge aussi sur les capacités de notre serveur à accueillir des données importantes de façon pérenne.

Les sept articles regroupés dans ce numéro sont présentés *infra*. Les deux premiers articles traitent de la morphologie des termes et de l'identification des mots composés. Les troisième et quatrième articles font la charnière entre traitement des mots et traitement des textes en considérant le calcul sémantique et la désambiguïsation lexicale en contexte. Le cinquième article explore une technique d'extraction d'information à partir de textes. Enfin, les deux derniers articles traitent, chacun sous un angle différent, de la lisibilité des textes.

1) « Analyse morphologique non supervisée en domaine biomédical », Vincent Claveau et Ewa Kijak ;

2) « Stratégies discriminantes pour intégrer la reconnaissance des mots composés dans un analyseur syntaxique en constituants », Matthieu Constant, Anthony Sigogne et Patrick Watrin ;

3) « Évaluer et améliorer une ressource distributionnelle : protocole d'annotation de liens sémantiques en contexte », Clémentine Adam, Cécile Fabre et Philippe Muller ;

4) « Désambiguïsation lexicale de textes : efficacité qualitative et temporelle d'un algorithme à colonies de fourmis », Didier Schwab, Jérôme Goulian et Andon Tchechmedjiev ;

5) « Une méthode d'extraction d'information fondée sur les graphes pour le remplissage de formulaires », Ludovic Jean-Louis, Romaric Besançon et Olivier Ferret ;

6) « Les apports du TAL à la lisibilité du français langue étrangère », Thomas François et Cédric Fairon ;

7) « Learning Domain-Specific, L1-Specific Measures of Word Readability », Shane Bergsma et David Yarowsky.

Pour finir, redisons qu'en complément des articles, la revue TAL publie des notes de lecture de livres et des résumés de thèses. Nous encourageons chaudement nos lecteurs à nous envoyer de telles informations, en se mettant en contact avec Denis Maurel pour les notes de lecture et Fiammetta Namer pour les résumés de thèses. Que tous deux trouvent ici nos chaleureux remerciements pour leur travail.

Béatrice Daille
LINA, université de Nantes
Beatrice.Daille@univ-nantes.fr

Éric de la Clergerie
ALPAGE, INRIA, Paris
Eric.De_La_Clergerie@inria.fr

Yves Lepage
IPS, université Waseda, Japon
Yves.Lepage@waseda.jp

Jean-Luc Minel
MoDyCo, université Paris-Ouest,
Nanterre La Défense
Jean-Luc.Minel@u-paris10.fr

Merci aux relecteurs spécifiques de ce numéro :

Marianna Apidianaki, LIMSI, Orsay
Florian Boudin, LINA, université de Nantes
Marine Carpuat, Conseil national de recherches, Ottawa, Canada
Marc El-Bèze, LIA, université d'Avignon
Iris Eshkol, LLL, CNRS, Orléans
Núria Gala, LIF, université d'Aix-Marseille
Pablo Gamallo Otero, Universidade de Santiago de Compostella, Espagne
Vincent Guigue, LIP6, université Pierre & Marie Curie, Paris
Thomas Lebarbé, LIDILEM, université Stendhal-Grenoble 3
Nadine Lucas, GREYC, CNRS, Caen
Daniel Luzzati, LIUM, université du Maine, le Mans
Aurélien Max, LIMSI, université Paris-Sud, Orsay
Yannick Parmentier, LIFO, université d'Orléans
François Rioult, GREYC, université de Caen Basse-Normandie
Fatima Sadat, université du Québec à Montréal, Canada
Benoît Sagot, ALPAGE, INRIA, Paris
Juan-Manuel Torres-Moreno, LIA, université d'Avignon
Leo Wanner, Universitat Pompeu Fabra, Barcelone, Espagne
Guillaume Wisniewski, LIMSI, université Paris-Sud, Orsay

ainsi qu'aux membres du comité de rédaction de la revue (voir sa composition sur notre site).