# LEXICOGRAPHIE NFORMATIQUE

BILAN ET PERSPECTIVES



23 au 25 JANVIER 2008 NANCY, ATILF Campus Lettres et Sciences Humaines

Colloque international à l'occasion du 50° anniversaire du lancement du projet du Trésor de la Langue Française











# LE BILAN



# Organisation Délégations Partenaires

## Comité scientifique

Jean-Paul Chauveau, ATILF/CNRS - Nancy Université
Thierry Fontenelle, Microsoft Speech and Natural Language Group, Redmond, USA
Hiltrud Gerner, ATILF/CNRS - Nancy Université
Benoît Habert, Université de Paris X, directeur de l'ILF
Franz-Josef Hausmann, Université d'Erlangen, Allemagne
Evelyne Jacquey, ATILF/CNRS - Nancy Université
Robert Martin, Membre de l'Institut
Jean-Marie Pierrel, Université Henri Poincaré, directeur de l'ATILF, Nancy
Claude Poirier, Université Laval, TLFQ, Québec
Pierre Rézeau, CNRS, Strasbourg
Isabelle Turcan, ATILF/CNRS - Nancy Université
Mathieu Valette, ATILF/CNRS - Nancy Université

# Comité d'organisation

Pascale Baudinot, ATILF/CNRS - Nancy Université Annie Bernardoff, ATILF/CNRS - Nancy Université Viviane Berthelier, ATILF/CNRS - Nancy Université Danièle Bujon, ATILF/CNRS - Nancy Université William Del Mancino, ATILF/CNRS - Nancy Université Laurent Gobert, ATILF/CNRS - Nancy Université Jean-Marc Voirin, ATILF/CNRS - Nancy Université

# **Communications et participants**

Soumissions reçues : 53 Soumissions sélectionnées : 24 Conférences invitées : 6

Pays représentés : Allemagne, Belgique, Canada, Chypre, Corée du Sud, Danemark, États-Unis, France, Italie, Pays-Bas, Roumanie, Royaume-Uni, Suisse

Participants: 120

# « Lexicographie et informatique : bilan et perspectives »





# **Organisation**







## **Soutiens**















# Lescons Armin

# Programme

# Mercredi 23 Janvier 2008

	Mercredi 23 Janvier 2008
10h15	Accueil des participants
10h30 - 11h15	Conférence d'ouverture par le Directeur de l'ATILF Jean-Marie PIERREL : Informatisation et valorisation sur le Net : une deuxième vie pour le TLF
	1. Valorisation des ressources lexicales Président de séance : Robert Martin
11h15 - 11h45 11h45 - 12h15	Pascale Bernard et Christiane Jadelot (ATILF CNRS & Nancy-Université): <i>TLF et TLFi. Naissance et évolution d'un dictionnaire</i> Jean-Luc Manguin (CRISCO CNRS & Caen-Université), Lonneke Van der Plas et Jörg Tiedemann (Alfa-Informatica, Université de Groningen): <i>Le traitement automatique</i> : <i>un moteur pour l'évolution des dictionnaires de synonymes</i>
	Repas
	2.1 Lexicographie et méthodologie Présidente de séance : Isabelle Turcan
14h00 - 14h45	Conférencier invité : Robert Martin (Membre de l'Institut) : L'apport méthodologique du TLF et les orientations d'aujourd'hui
14h45 - 15h15	Pierluigi LIGAS (Université de Vérone): Définition et exemple: quelle complémentarité? L'illustration du concept dans le «Dictionnaire alphabétique et analogique du français des activités physiques et sportives» (à paraître, 2009)
15h15 - 15h45	Paolo FRASSI (Université de Vérone) : Métalangage et définition de substantif dans le TLF : le cas des "entre-crochets"
Pause	
	2.2 Lexicographie et méthodologie
16h15 - 16h45 16h45 - 17h15	Mélanie Petit (Université d'Orléans): L'intégration de l'information prosodique en lexicographie: nouveaux supports, formats de présentation et techniques de discrimination Corinne Feron (Université du Maine, Le Mans) et Danielle Coltier (Université du Maine, Le Mans – CELTED, Metz): De la description linguistique à la description lexicographique:
17h15 - 17h45 18h30	l'exemple des adverbiaux de phrase du type en + lexème Aude GREZKA et Françoise MARTIN-BERTHET (LDI CNRS - Univ. Paris 13) : Un dictionnaire des verbes  Pécention à l'Hêtel de Ville de Napou
181130	Réception à l'Hôtel de Ville de Nancy
	Jeudi 24 Janvier 2008 matin
	3.1 Contacts linguistiques, emprunts, étymologie Présidente de séance : Hiltrud Gerner
9h00 - 9h45	Conférencier invité : Claude POIRIER (TLFQ – Université de Laval, Québec) : Pour un vrai Trésor du français : proposition de mise en relation du TLFi et de la BDLP
9h45 - 10h15	David Trotter (University of Wales Aberystwyth) : L'anglo-normand et le TLF, dans le passé et dans l'avenir
10h15 - 10h45	Claude FREY (Université Paris 3) : De la préface du TLF à l'idéologie francophone : pratiques lexicographiques et description du français en Afrique
Pause	
	3.2 Contacts linguistiques, emprunts, étymologie Président de séance : Jean-Paul Chauveau
11h00 - 11h30	Myriam Benarroch (Université Paris-Sorbonne) : Le traitement des emprunts au portugais dans le TLFi
11h30 - 12h00	Nadine Steinfeld et Marta Andronache (ATILF CNRS & Nancy-Université): Quoi de neuf du côté de la lexicographie étymologique? La méthode utilisée dans le cadre du projet TLF-Etym pour distinguer les emprunts au latin de l'Antiquité de ceux faits au latin médiéval
12h00 - 12h30	Veronika Lux (INIST CNRS), Agnès Tutin (LIDILEM/Université Grenoble 3): Extraction de collocations à partir du champ syntagme du TLFi: application aux noms transdisciplinaires des écrits scientifiques

# « Lexicographie et informatique : bilan et perspectives » Lorraine – Nancy Janvier 2008

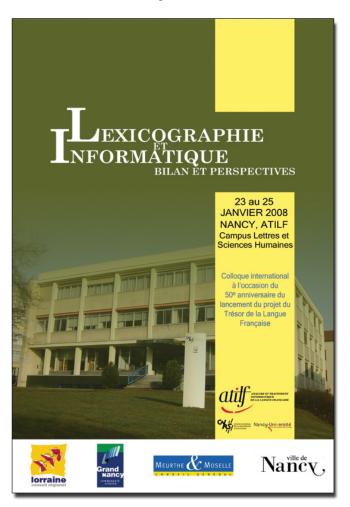


	Jeudi 24 Janvier 2008 après-midi	
	4.1 Ressources lexicales et traitement automatique Présidente de séance : Évelyne Jacquey	
14h00 - 14h45	Conférencier invité : Alain Polguere (OLST, Université de Montréal, Québec) : La lexicographie explication	
14h45 - 15h15	combinatoire à l'épreuve de l'informatisation Benoît SAGOT (INRIA – Projet Alpage) et Laurence DANLOS (Université Paris 7) : Méthodologie lexicographique de constitution d'un lexique syntaxique de référence pour le français	
15h15 - 15h45	Denis Maurel (Université François Rabelais, Tours/LI) : <i>Prolexbase</i> : une base de données lexicale de noms propres pour le TAL	
Pause		
	4.2 Lexicographie et dictionnaires nationaux	
16h15 - 16h45	Monica Busuloc et Florin Vasilescu (Institut linguistique lorgu lordan/Al. Rosetti, Bucarest, Roumanie) : Le Trésor de la langue roumaine	
16h45 - 17h15	Henrik Lorentzen et Lars Traps-Jensen (Society for Danish Language and Literature Copenhague): The Dictionary of the Danish Language online: From Book to Screen – an Beyond	
17h15 - 17h45	Seong Heon Lee et Chai-Song Hong (Université de Séoul, Corée du Sud) : Le Dictionnaire électronique du coréen contemporain : le Dic Sejong – ses caractéristiques et son intérêt	
	Vendredi 25 Janvier 2008	
	5.1 Lexicographie et enseignement Président de séance : Mathieu Valette	
9h00 - 9h45	Conférencier invité : Étienne Brunet (BCL, Université de Nice) : L'exploitation statistique des bases lexicographiques	
9h45 - 10h15 10h15 - 10h45	Serge VERLINDE, Jean BINON et Ann BERTELS (Université Catholique de Louvain, Belgique) : La Lexicographie au service de l'apprentissage / enseignement des combinaisons de mots Cécile FABRE (UMR 5263/CNRS, Université Toulouse – Le Mirail) : Utilisation pédagogique du TLFi – Ce que des étudiants peuvent apprendre d'un dictionnaire informatisé	
	Pause	
	5.2 Exploitation de ressources informatisées	
11h00 - 11h30	Jacqueline LEON (UMR 7597/CNRS, Histoire des Théories Linguistiques, Paris 7): Automatisation du langage, premiers corpus informatisés et lexicographie dans les années	
11h30 - 12h00	Mick Grzesitchak et Évelyne Jacquey (ATILF CNRS & Nancy-Université) et Fabienne Baider (Université de Chypre) : Annotation sémantique : profilage textuel et lexical	
12h00 - 12h30	Évelyne Jacquey et Christiane Jadelot (ATILF CNRS & Nancy-Université) : Noms d'objets imprimés : ambiguïté lexicale sémantique et proxémie	
Repas		
	6. Ressources dictionnairiques Président de séance : Jean-Marie Pierrel	
14h00 - 14h45	Conférencier invité : Russon Wooldridge (Trinity College, Université de Toronto, Ontario) : "Caught in the Web of	
14h45 - 15h15	Words": la lexicographie et la Toile Pascale Bernard, Geneviève Flechon et Marie-Josèphe Mathieu (ATILF CNRS & Nancy-	
15h15 - 15h45	Université): Le Supplément du Trésor de la langue française  Philippe Grea (Université Paris 10, Nanterre) et Sylvain LOISEAU (LIMSI/CNRS): "Le dictionnaire comme genre ou comme ressource: relations sémantiques et représentations on graphe"	
15h45 - 16h15	en graphe" Conclusions	



# Supports de communication

# • Affiche du colloque



 Mot d'accueil du colloque (inséré dans le programme)







La Direction du Département SHS du CNRS et la Présidence de Nancy-Université, La Direction du laboratoire ATILF, Le Comité scientifique et l'équipe d'organisation

Out le plaisir de vous accueillir au colloque international "Lexicographie et informatique : bilan et perspectives" à l'occasion du 50 anniversaire du lancement du projet du Trésor de la Langue Française.

Ils vous souhaitent un agréable séjour et d'excellents travaux. Ils remercient pour leur soutien les partenaires sans lesquels ce rendez-vous scientifique n'aurait pu avoir lieu :









# « Lexicographie et informatique : bilan et perspectives »

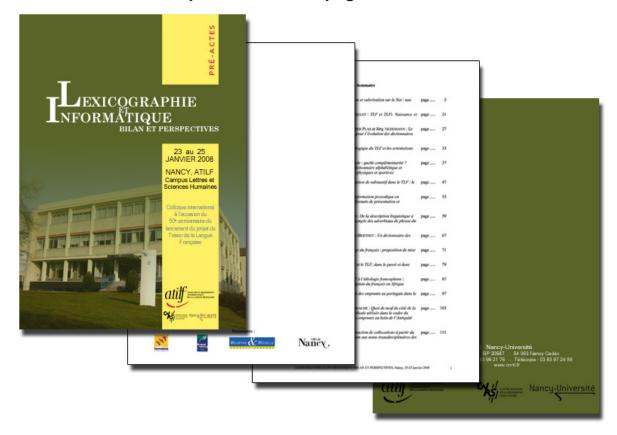


# Lorraine - Nancy Janvier 2008

Programme du colloque : dépliant 4 pages



Pré-actes du colloque : reliure 224 pages



# Lexicogis aveni Security Declare van

# Articles de presse

© Le Républicain Lorrain, Jeudi le 24 Janvier 2008. / Lorraine / Actualité Droits de reproduction et de diffusion réservés

## Les 50 ans d'un Trésor

Le Trésor de la Langue Française conçu à Nancy a 50 ans et mérite bien qu'on célèbre son anniversaire. Avec 300 000 connexions par jour, son site internet est un immense succès.

Ca pourrait être une histoire qu'on raconte aux enfants... Il y a cinquante ans tout juste, des chercheurs en lexicologie et en langue française s'interrogeaient. Ils étaient confrontés à une double problématique : voilà comment on parle le français, et voici comment on doit le parler. Ce qui n'est pas la même chose. C'est de ce double questionnement, sous l'impulsion du recteur Paul Imbs, qu'est née l'idée follement ambitieuse - et très gaullienne - d'écrire le Trésor de la Langue Française. Il s'agissait de concevoir un dictionnaire original tenant compte de l'utilisation actuelle des mots, dont la définition était enrichie de documents puisés dans la littérature, les citations, les différents usages...La mission fut confiée au centre nancéien de recherche pour le Trésor de la Langue Française. Un travail de bénédictin pour créer ce qui, plus que jamais, est un dictionnaire de référence. On célèbre cette semaine à Nancy le jubilé du TLF, que Jean-Marie Pierrel, le directeur du laboratoire « Analyse et traitement informatique de la langue française » (ATILF) a l'habitude de résumer en quelques chiffres qui frappent l'esprit : « 16 volumes, 100 personnes pendant 30 ans, 270 000 définitions ». Depuis 2002, on en trouve une version informatisée sur le web, qui reçoit 300 000 connexions quotidiennes. Et en 2002 et 2004, le TLFi a vu le jour sous la forme d'un CD-Rom. A partir de ce mercredi, une centaine de spécialistes de la langue française venus de douze pays vont participer à un colloque international. Il s'agit notamment d'envisager l'évolution de la langue française et de l'outil qu'est le dictionnaire à la lumière des progrès de l'informatique. « On vit aujourd'hui une profonde révolution par la pénétration de l'informatique dans tous les secteurs. Les rapports de l'homme du dictionnaire changent complètement », souligne Jean-Marie Pierrel, qui pilote les recherches du laboratoire.

Les possibilités du TLF sont vertigineuses et d'autres pays s'en inspirent, comme la Roumanie ou la Corée qui vient de lancer un grand projet de dictionnaire électronique du coréen moderne. Celles du portail lexical du centre national des ressources textuelles et lexicales le sont tout autant. L'inventivité des chercheurs ne connaît pas de limites. Ils travaillent sur un programme baptisé *Dilan*, permettant de trouver « *les mots qu'on a sur le bout de la lanque* » !

M. R.

# « Lexicographie et informatique : bilan et perspectives »





Samedi 26 Janvier 2008, © L'Est Républicain / ÉVÉNEMENTS Droits de reproduction et de diffusion réservés

# Tous les mots mènent à Nancy

Les dictionnaires de demain se conçoivent sur le campus lettres, dans le laboratoire ATILF-CNRS.



Jean-Marie Pierrel, directeur du laboratoire ATILF.

Nancy recèle un trésor convoité par des millions d'êtres humains à travers le monde : le Trésor de la langue française. Peu de Nancéiens savent que la capitale des ducs de Lorraine est une capitale mondiale de la lexicographie depuis près d'un demisiècle. Et que c'est ici que se conçoivent les dictionnaires de demain. Cette semaine, d'éminents spécialistes de la langue française de douze pays ont participé à un colloque international organisé par le laboratoire ATILF (Analyse et traitement informatique de la langue française), dans ses locaux du campus lettres sur le thème « lexicographie et informatique, bilan et perspectives »

# 600.000 connexions par jour

Même s'il est gratuitement accessible via internet, le CD-rom du Trésor de la Langue française (TLF) se vend comme des petits pains. Chaque jour, près de 600.000 internautes se connectent au portail lexical du Centre national de ressources textuelles et lexicales (cnrtl.fr) adossé au laboratoire ATILF- CNRS Nancy-Université. « Un tiers des utilisateurs sont chercheurs, un tiers des professionnels de l'écriture et un tiers des particuliers », détaille le directeur de labo Jean-Marie Pierrel. « Les États-Unis sont dans le top 3 tous les jours ». L'École de Nancy de la lexicographie, adoptée par l'Académie Française, fait même des émules. D'ailleurs, les représentants coréens et roumains participant au colloque ont pour mission de mettre au point un Trésor de la langue de leur pays sur le modèle français. L'aventure, « très gaullienne dans l'esprit » a commencé ici même il y a cinquante ans, « En 1957, le Littré entrait dans le domaine public. Le recteur Paul Imbs a alors eu pour mission de le remplacer par une œuvre entièrement originale tant par sa structure et sa conception que par ses fondements », résume Jean-Marie Pierrel. Pas un hasard si le TLF est « le reflet le plus fidèle de l'usage de la langue française aujourd'hui dans la littérature, la presse, au 20 heures... Les articles doivent pouvoir expliquer tous ces usages. Une révolution! » Cent personnes se sont consacrées à cette tâche incommensurable dès 1960, dans le tout nouveau Centre de recherche pour un Trésor de la langue française crée pour l'occasion. En 1971, le premier tome est publié. Il fauda attendre 1994 pour que le seizième et dernier volume paraisse. Le TLF regroupe 100.000 mots avec leur histoire, 270.000 définitions, 430.000 exemples d'au-teurs). Une référence mondiale. Et une formidable base de données boostée par la révolution informatique et numérique accessible par le portail du CNRTL. Grâce à d'astucieux logiciels, cet outil permet d'explorer la langue sans limite et de traquer le mot juste. « On peut même créer n'importe quel dictionnaire à thème en un simple clic, celui des noms d'arbres par exemple ».

# École de Nancy

Fascinant. Et même rageant pour les éditeurs qui peuvent légitimement s'inquiéter de cette redoutable concurrence. Le scientifique s'en défend : « C'est bien la preuve que la recherche publique profite directement aux citoyens », justifie Jean-Marie Pierrel. « Il est normal que les citoyens récoltent le fruit de notre travail qu'ils financent avec leurs impôts ». La gratuité serait indispensable. Il y va de l'intérêt supérieur de la francophonie : « Stratégiquement, le libre accès s'imposait. Surtout dans un contexte d'hégémonie anglo-saxonne. » S'agissant du traitement informatique de la langue, les Nancéiens n'ont rien à envier aux Américains. Il est même fort probable que les moteurs de recherche du futur s'inspireront des travaux du Labo ATILF. « Sur Google au- jourd'hui, si vous tapez "élevage cheval" tous les sites où les mots élevage et cheval apparaissent, s'afficheront. Ceux qui se contentent d'afficher "éleveur de chevaux" vous échapperont. Les logiciels que nous mettons au point permettent d'éviter cet écueil », explique le directeur de ce labo linguistique issu de l'informatique. « Une autre révolution ». slabidi@estrepublicain.fr

Saïd LABIDI