



LES 20 ANS DE L'ATILF



Orientations scientifiques et fondements de la mise en place et de la structuration de l'ATILF

25 janvier 2021

Jean-Marie.Pierrel@atilf.fr

En guise d'introduction

Focalisation sur les dix premières années de l'ATILF

- ▶ Rappel de la situation en 2000
- ▶ Structuration du projet scientifique
- ▶ Quelques défis et réussites
- ▶ Ouverture des données : la science ouverte avant l'heure
- ▶ Pour un véritable esprit de laboratoire
- ▶ Quelques faits marquants des dix premières années

L'INaLF à la fin des années 90

- ▶ Le CNRS s'interroge sur le devenir de l'INaLF
 - Le dernier volume du TLF a été publié en 1994
 - l'INaLF c'est alors :
 - 2 unités, réparties sur plusieurs sites (Nancy, Paris, Besançon) avec une direction essentiellement parisienne
 - USR *Institut National de la Langue Française*
 - UPR *Trésor Général des Langues et Parlers Français*
 - Et un certain nombre d'unités de recherche linguistique (URL) associées (Saint Cloud, Villette entre autres)
- ▶ Cette structure peu claire, l'INaLF étant à la fois le tout et la partie, est remise en question
- ▶ L'INaLF est devenue une FRE (Formation de Recherche en Evolution) à la fin des années 1990

Les perspectives envisagées début 2000

- ▶ L'INaLF deviendrait une *Fédération de recherche*, l'ILF, formée d'UMR autonomes
- ▶ La question du maintien sur Nancy d'une unité se trouve clairement posée
 - D'aucuns verraient bien les activités de recherche de l'INaLF s'installer à Lille ou à Lyon
 - Des acteurs politiques lorrains se battent pour un maintien sur Nancy où le TLF a vu le jour
- ▶ La solution qui émerge est celle d'une UMR avec l'université Nancy 2
- ▶ Une question se pose alors : quel projet scientifique pour une telle UMR ?

Mes premières implications

- ▶ Fort de 25 ans de travaux sur le traitement automatique de la langue orale au LORIA, je siège à ce moment-là au Comité national de la recherche scientifique
 - en section 7 Informatique
 - au conseil du département SPI
 - et section 34 Sciences du langage
- ▶ Nommé au comité d'évaluation de l'INaLF
- ▶ Rapport sur l'INaLF et son informatique
 - y sont soulignées les forces et faiblesses de la composante nancéienne
 - Et, comme dans tout rapport, j'y émets des recommandations pour l'avenir

Mon implication dans la création de l'ATILF

- ▶ Fin 1999-début 2000, rencontre avec Alain Péraube, DSA chargé de la section 34 à l'INSHS
 - Le CNRS serait d'accord pour la création d'une UMR avec l'Université Nancy 2
 - et souhaite mettre en place mes recommandations pour l'orientation de cette UMR
 - mais à une condition : que j'accepte de prendre la direction de cette UMR
- ▶ Or je suis Professeur à l'UHP et membre du LORIA
 - j'y ai mon équipe de recherche et ai dirigé ce laboratoire pendant 4 ans lorsqu'il s'appelait le CRIN (1994-1997)
- ▶ Nancy 2 de son côté veut un directeur issu de ses rangs
- ▶ Finalement CNRS et Nancy 2 converge vers une codirection avec Bernard Combettes

Périmètre de cette nouvelle UMR ATILF

- ▶ L'ensemble des personnels CNRS nancéiens de l'INALF
 - 3 chercheurs
 - Plus de 50 ITA
- ▶ + l'antenne de Besançon de l'INALF
 - 4 ITA
- ▶ Et l'équipe d'accueil Landisco (Langue, Discours et Cognition) de Nancy 2 accompagnée de l'ensemble des forces en sciences du langage de Nancy 2 dont par exemple le laboratoire de phonétique
 - Environ 25 enseignants-chercheurs
 - 1 secrétaire gestionnaire

Structuration du projet de laboratoire

Trois principaux objectifs sont affichés

- ▶ Faire émerger un programme scientifique pour une recherche de qualité alliant linguistique et informatique
 - recherche fondamentale
 - recherche finalisée autour de la notion de ressources linguistiques informatisées (textes, dictionnaires, lexiques et outils)
- ▶ Une mission de service auprès de la communauté scientifique et de la société pour valoriser les connaissances sur le français au travers du web
- ▶ Une visibilité interne et externe accrue à travers une plus grande valorisation des productions du laboratoire

Les principales difficultés rencontrées

- ▶ Réunion de deux communautés qui n'avaient pas l'habitude de travailler ensemble
 - La disposition des locaux est assez révélatrice !
- ▶ Côté CNRS
 - Un déséquilibre entre 50 ITA et 3 Chercheurs
- ▶ Côté Université
 - Un déséquilibre entre 25 enseignants-chercheurs et 1 IATOS
- ▶ Un manque de structuration en équipes de taille raisonnable
 - Beaucoup de projets individuels
- ▶ Et pour ma part une double extériorité
 - Scientifique : Pr en informatique
 - Institutionnelle : Membre du LORIA et de l'UHP

Structuration du projet de laboratoire (1)

- ▶ Au vu de l'existant, des éléments de développement proposés dans mon rapport et des apports du LanDisCo, convergence vers une structuration scientifique autour de trois axes de recherche
 - Etudes synchroniques
 - Etudes diachroniques
 - Usages, modèles et traitements informatiques
- ▶ Auxquels s'ajoute des services d'accompagnement
 - Bases de données et ressources linguistiques
 - Secrétariat et services généraux
 - Moyens informatiques
 - Documentation
- ⇒ Cette structure de laboratoire va petit à petit s'unifier et converger vers ce qu'elle est aujourd'hui

Structuration du projet de laboratoire (2)

- ▶ Une volonté affichée d'une plus grande ouverture sur l'environnement local du laboratoire
 - vers la recherche universitaire en sciences du langage et plus généralement en Lettres et sciences humaines
 - Vers la composante « informatique linguistique » du LORIA
 - PPF « Gestion maintenance et distribution de ressources linguistiques informatisées »
 - projet CPER « Ingénierie des langues, du document et de l'Information scientifique, technique et culturelle » avec L'INIST et le LORIA
- Une plus large ouverture vers l'accueil et la formation de doctorants

Structuration du projet de laboratoire (3)

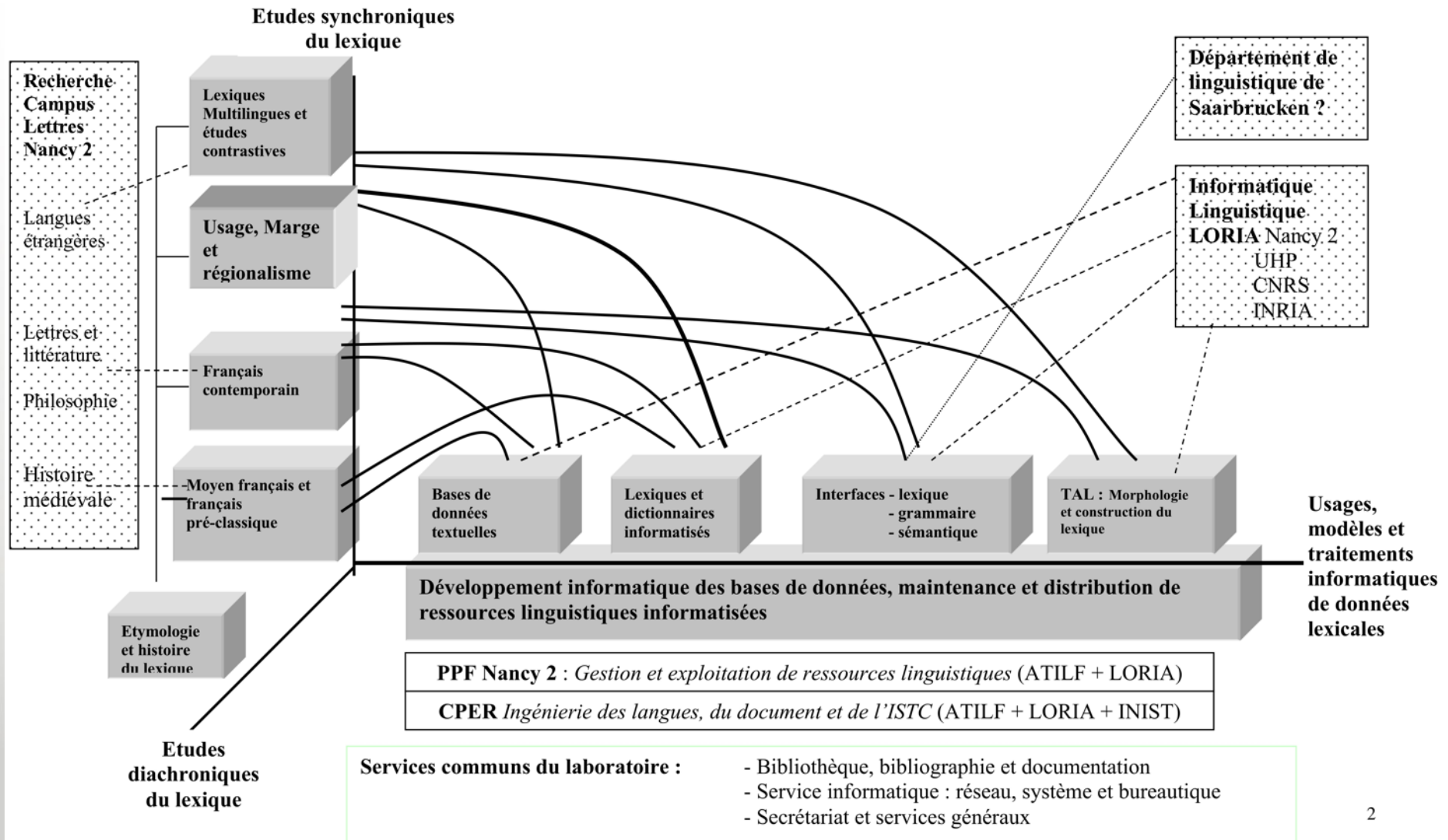
- ▶ Une volonté d'une ouverture vers l'international
 - mise en place et consolidation de relations scientifiques avec des laboratoires étrangers
 - Université de Chicago,
 - Fonds National de la Recherche Scientifique Suisse
 - Etc.
 - choix de normes de codage de ressources linguistiques compatibles avec les standards internationaux de fait que sont ceux de la TEI (Text Encoding Initiative)
 - étude du lexique français dans un cadre multilingue, en particulier au travers des recherches comparatives sur des lexiques multilingues.

Evolution de la structure du laboratoire

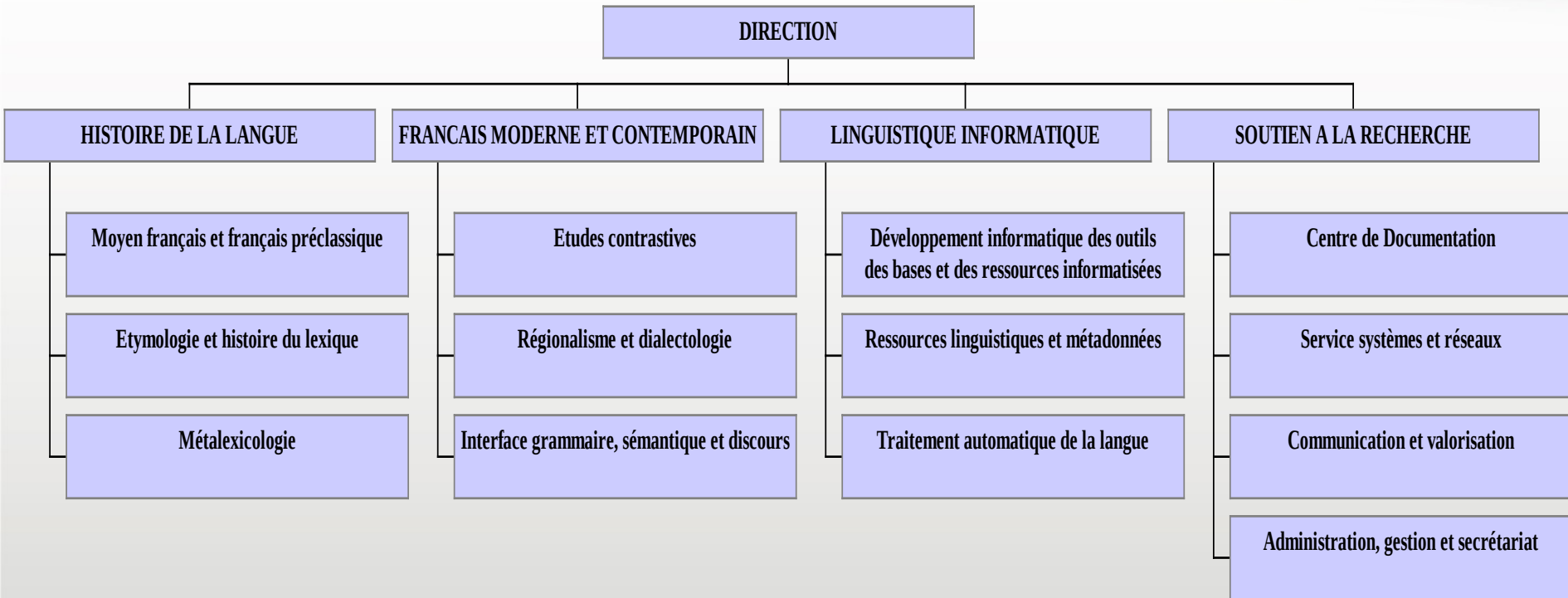
La structure de laboratoire va petit à petit s'unifier et converger au cours de 10 premières années vers ce qu'elle est aujourd'hui

- Une première approche du projet en 2000
- Structuration en 2002
 - Pour l'évaluation à mi-parcours
- Structuration en 2004
 - Pour le 2^{ème} quadriennal
- Structuration en 2006
 - Avec l'intégration de l'équipe Didactique des langues (CRAPEL)
- Structuration en 2010
 - Les 5 équipes actuelles émergent

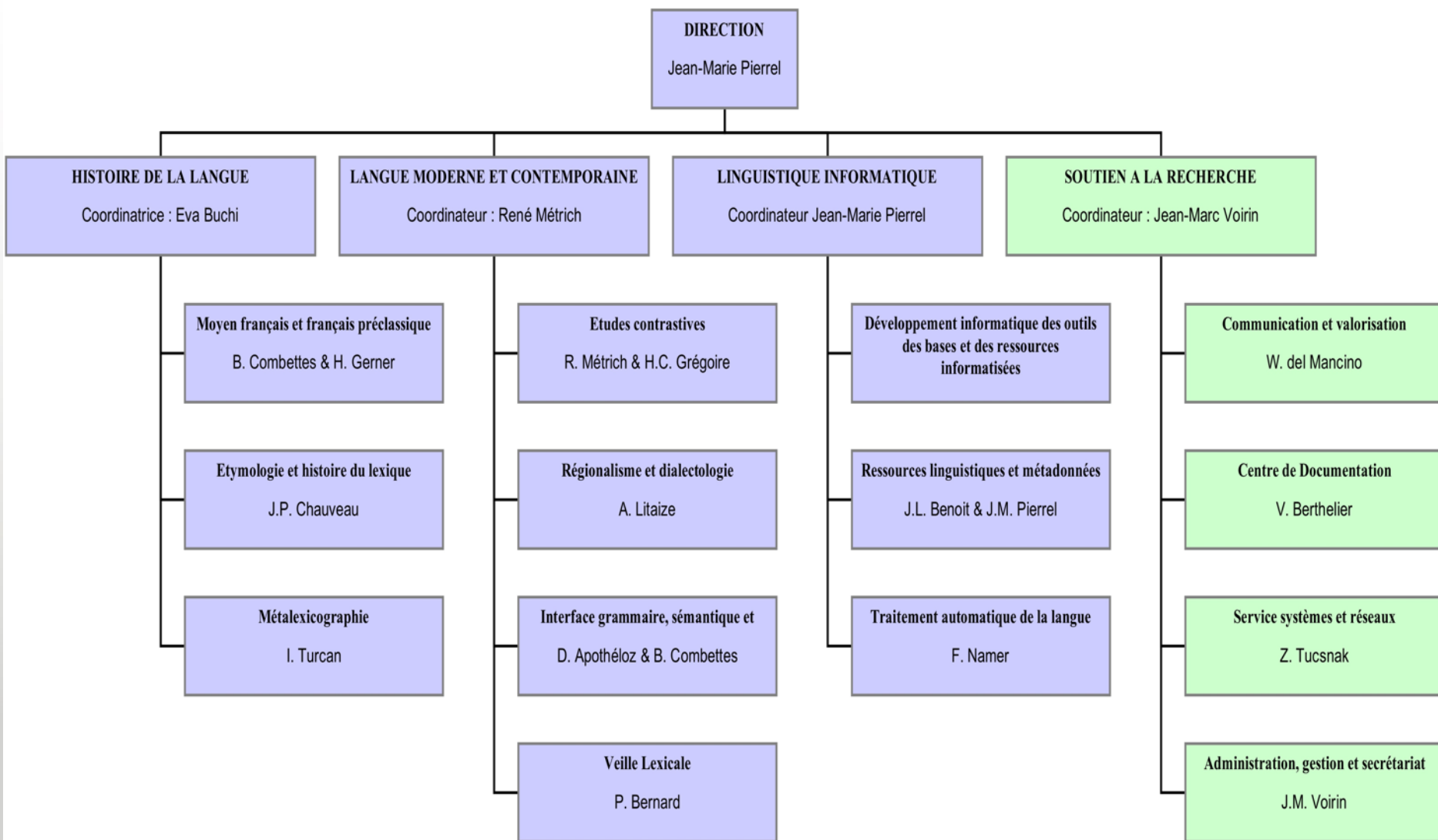
Une première approche du projet en 2000



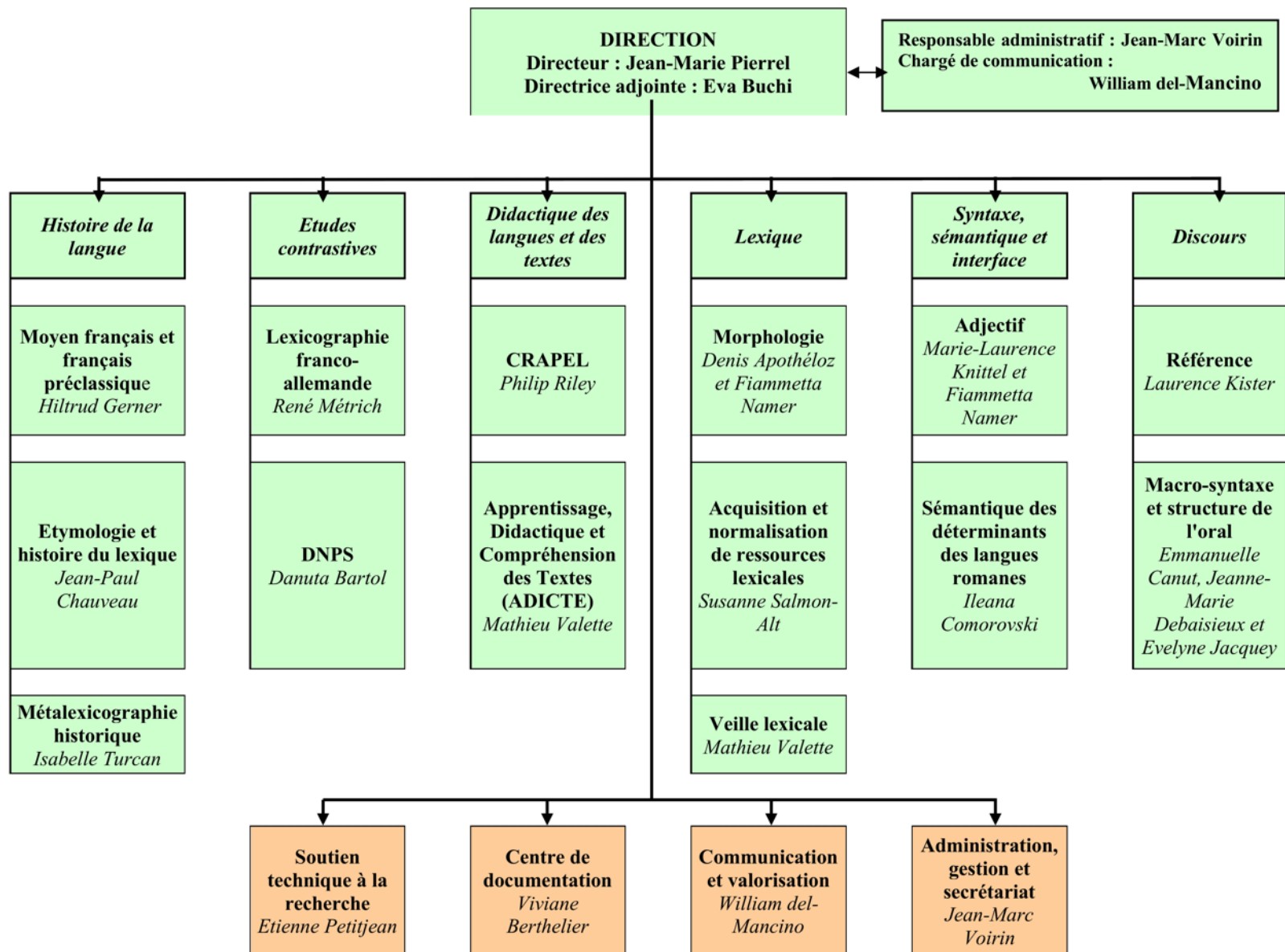
La structure en 2002



La structure en 2004

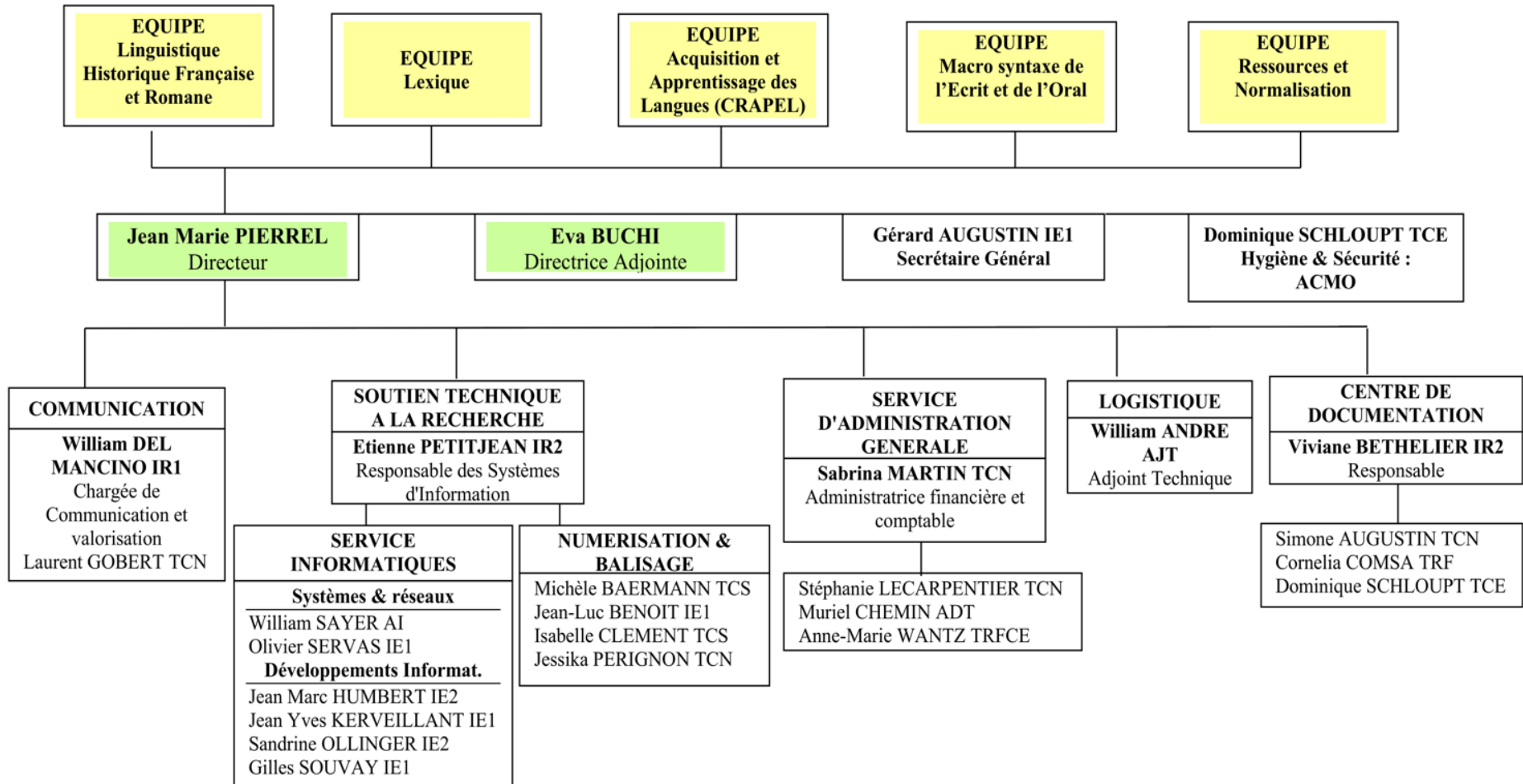


La structure en 2006



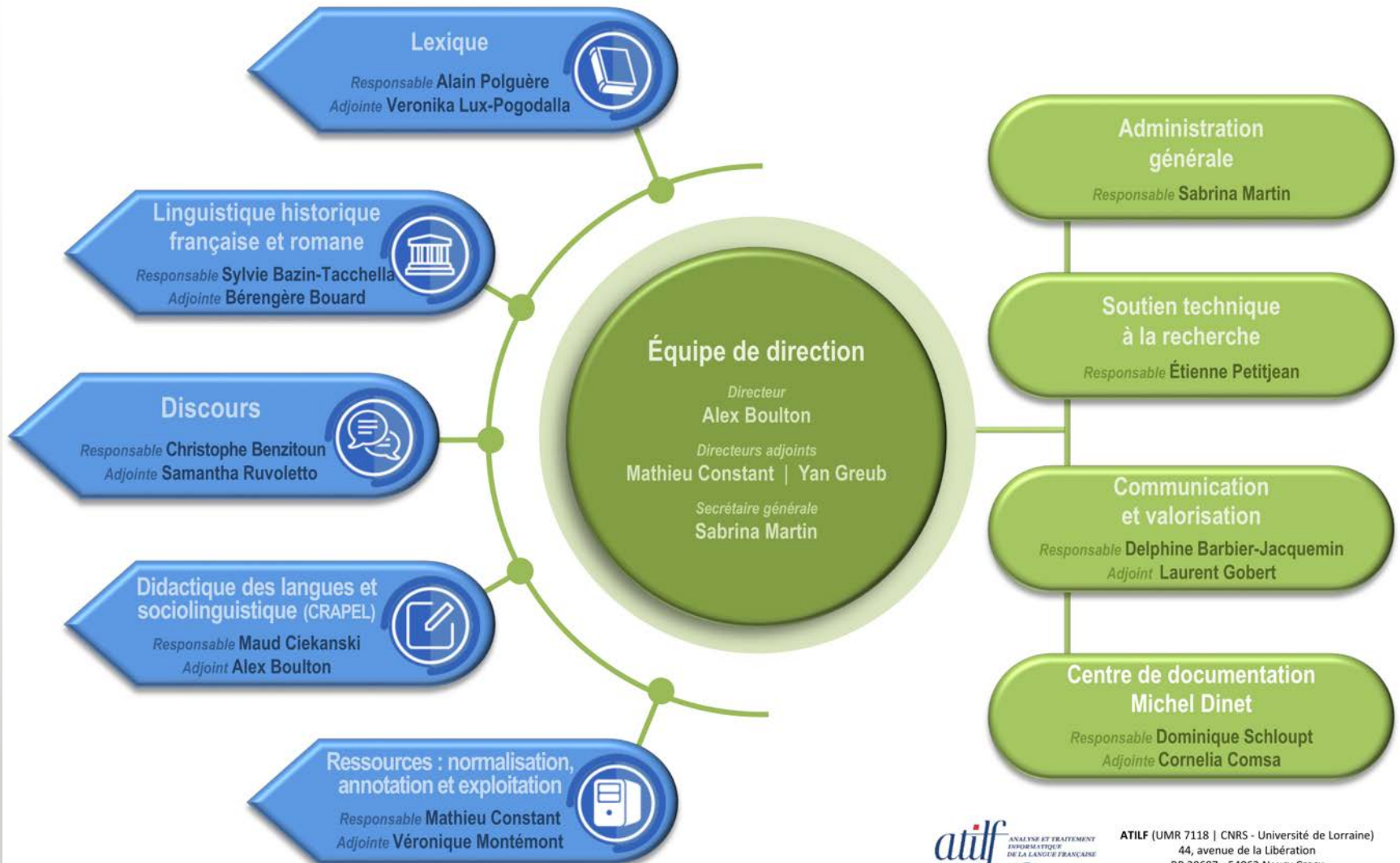
La structure en 2010

ORGANIGRAMME de l'UMR 7118 CNRS-Nancy Université



La structure aujourd'hui

Organigramme structurel



Quelques défis et réussites

- ▶ Evolution du projet de DMF
- ▶ Valorisation scientifique du TLF et de Frantext
- ▶ Meilleure insertion de l'ATILF dans son environnement au travers de projets coopératifs
- ▶ Ouverture des données : La science ouverte avant l'heure
- ▶ Du CNRTL à ORTOLANG
- ▶ Un laboratoire où chacun est heureux et fier de travailler

Evolution du DMF

- ▶ En 2000 : 1^{er} volume après plusieurs années de travail
 - Plusieurs centaines de pages allant de la lettre AA à la lettre AH
 - Le CNRS demande d'abandonner ou de profondément restructurer ce projet
- ▶ En mai 2001, nouvelle orientation
 - Décision d'une [version électronique évolutive](#) permettant des ajouts et des corrections.
 - A partir d'une première version (corpus de 13 lexiques et d'un lexique complémentaire), addition progressive de données nouvelles ainsi que des approfondissements de l'analyse.

Valorisation scientifique du TLF et de Frantext

- ▶ Un constat en 2001 :
 - Frantext et le TLFi n'avaient pas l'audience qu'ils méritaient auprès de la communauté scientifique, nationale et internationale
 - Manque de publications scientifiques sur et autour de ces 2 grandes ressources
- ▶ Au cours des 3 premières de l'ATILF :
 - Plusieurs publications sur nos ressources et leurs exploitations : TAL, JADT, TALN, LREC et trois participations à des ouvrages collectifs
- ▶ Sur les autres thèmes : renforcement de nos contributions dans des revues des colloques et des workshops

Une meilleure insertion de l'ATILF dans son environnement

► Local :

- structuration d'un PPF interne à Nancy 2
« Gestion et exploitation de ressources linguistiques »
- Coopérations ATILF-LORIA

► Régional :

- projet du CPER Lorrain « Ingénierie des Langues, du Document et de l'Information Scientifique, Technique et Culturelle » en partenariat avec l'INIST et le LORIA
- participation forte au projet de MSH et structuration de l'axe 2 de la MSH « Langues Textes et Documents » qui servira de support pour le CPER suivant

Une meilleure insertion de l'ATILF dans son environnement

► National :

- Membre de la fédération CNRS ILF
- Plate-forme nationale de ressources linguistiques
 - En 2005, création du CNRTL
- Mise en place de projet coopératifs dans le cadre de réponses à des appels d'offre dès le début des années 2000
 - 2001 Ananas
 - 2002-2003 projets Technolange TILT, EASY, Lexitec et RNIL
 - 2006-2007 ANR ELICO, DETCOL, SOURENCYME, DeRom
 - 2008 Projet ERC Histoire du lexique politique Français (Cité de Dieu)

Une meilleure insertion de l'ATILF dans son environnement

► international :

➤ Participation à la TEI

- Le laboratoire, membre de la TEI. et en association avec l'INIST et le LORIA, centre européen support

➤ Projets INTERA et OLAC sur un réseau de méta-données

➤ Coopération avec le FNRS (Fond National suisse de la Recherche Scientifique) : projet FEW

➤ Dictionnaire électronique français-allemand des « mots du discours » : avec l'Université d'Heidelberg

Ouverture des données : la science ouverte avant l'heure

- ▶ Le TLFi librement accessible sur le Web
 - Malgré des réticences au départ tant en interne que du côté du CNRS
 - Ouverture officielle du TLFi en mars 2002 (<http://atilf.atilf.fr/tlf.htm>)
- ▶ Un effort de plus grande diffusion de [Frantext](#)
 - Ouverture de Frantext Agrégation dès 2002
- ▶ Début 2004 :
 - TLFi : 110 000 requêtes/jour-ouvrable, plus de 50 000 le WE
 - Frantext : 200 institutions et universités abonnées dont la moitié sont étrangères

Du CNRTL à ORTOLANG

- ▶ 2005, suite à un appel d'offre de la DIST et de l'INSHS : création du CNRTL : www.cnrtl.fr
 - Une grande réussite : son portail lexical
 - Quinze ans après : plus de [700 000 requêtes](#) par jour sur le lexique français
- ▶ 2012, sélection du projet d'Equipex ORTOLANG
 - Seul Equipex validé et piloté en Lorraine
 - Seul Equipex dans le domaine des sciences du langage
 - www.ortolang.fr
 - 484 ressources dont 330 publiées
 - 2086 utilisateurs se sont créés un compte sur ORTOLANG au 13/01/21

Pour un véritable esprit de laboratoire

- ▶ Mise en place des journées annuelles de rentrée dès 2001
- ▶ Mise en place d'AG régulières
- ▶ Mise en place du séminaire de l'ATILF
 - Début en 2001
- ▶ Inform'ATILF
 - 1^{er} numéro en septembre 2003
- ▶ De moments de convivialité tels
 - La « Galette des rois » des 2001
 - Des départs en retraite
 - Des pots de thèses
 - Sans oublier la cafet et la cuisine

Amélioration des conditions de travail

- ▶ En 2001, le bâtiment CNRS était majoritairement dans son jus du milieu des années 60
 - Mise aux normes du bâtiment CNRS pour permettre un accueil de public
 - Travaux de peinture
 - Rééquipement des bureaux
 - Restructuration du centre de documentation
 - Mise en place d'un ascenseur
- ▶ Consolidation des moyens informatiques du laboratoire
 - Serveurs et système de sauvegarde
 - Postes de travail

Quelques faits marquants des 10 premières années (1)

► 2001

- septembre : l'ATILF présent sur le Livre sur la Place

► 2002

- mars : Ouverture sur le Web du [TLFi](#)
- août : validation de 4 projets dans le cadre de l'appel national TechnoLange

► 2003

- 6-9 Novembre : AG de la TEI à l'ATILF

► 2004

- 16/06 : à l'Institut de France, présentation du Dictionnaire de l'Académie en ligne
- 18/11 : lancement du CD-ROM du TLF dans les locaux de CNRS Editions.

Quelques faits marquants des 10 premières années (2)

► 2005

- 01 janvier : l'UHP devient cotutelle de l'ATILF
- 10 mars : ouverture du centre de documentation rénové

► 2006

- 01 janvier : le CRAPEL rejoint officiellement l'ATILF
- Janvier - Février : Présentation de l'exposition «La langue française, langue de l'aventure scientifique» au siège du CNRS
- Ouverture du [site du CNRTL](#)

► 2007

- Eva Buchi Lauréate du prix Albert Dauzat
- Août : Publication du [Dictionnaire de la langue du vin](#)

Quelques faits marquants des 10 premières années (3)

▶ 2008

- 28 mai : signature de la Charte ATILF / Est Républicain
- 23-24 juin : 1er atelier [DEROM](#)

▶ 2009

- Janvier : [LyText](#) accessible à l'ensemble des Lycées de Lorraine sur le serveur de l'Académie Nancy-Metz, et, un an après, convention avec le Ministère de l'Education Nationale pour l'utilisation dans toutes les académies

▶ 2010

- L'ATILF sollicité pour l'organisation de [CILPR 2013](#)

➤ 2011

- Démarrage du projet [RELIEF](#) de création du Réseau Lexical du Français, grâce à un soutien important de la Région Lorraine

En guise de conclusion

Aujourd'hui, l'ATILF peut être fier de ses réussites et s'appuyer sur elles pour construire l'avenir :

- ▶ Le regroupement unifié de l'ensemble des composantes Sciences du Langage sur Nancy
- ▶ Des travaux scientifiques remarquables et pour beaucoup salués au niveau national et international
- ▶ Des bases de données et un Equipex, ORTOLANG, qui font de l'ATILF un laboratoire bien armé dans le cadre du mouvement actuel vers la science ouverte.