

## CURRICULUM VITAE

**ABEILLÉ Anne**

née le 13 septembre 1962 à Paris (11e)

[abeille@linguist.jussieu.fr](mailto:abeille@linguist.jussieu.fr)

<http://www.llf.cnrs.fr/Gens/Abeille>

### Etudes

1983 : Licence de Philosophie à Paris 1.

1984-85 : Licence et Maîtrise de Lettres Modernes à Paris 7.

1983-1987 : ENS Fontenay aux roses.

1986 : Agrégation de Lettres Modernes.

1987 : DEA de Linguistique à Paris 7 (UFRL)

1991 : Thèse de doctorat en linguistique à Paris 7 (mention TH avec félicitations)

1999 : Habilitation à diriger les recherches (Paris 7).

### Postes occupés

1987-88 : Research assistant à l'University of Pennsylvania (Philadelphie)

1988-91 : AND à Paris 7 (UFRL)

1991-94 : Maître de Conférences au département de Sciences du Langage (Paris 8)

1994-2000 : Maître de Conférences à Paris 7 (UFRL)

depuis septembre 2000 : Professeur de linguistique à Paris 7 (UFRL), 1<sup>ère</sup> classe depuis 2005.

Délégation CNRS du 1<sup>er</sup> septembre 2007 au 1<sup>er</sup> mars 2008

Prix Pierre Larousse pour les sciences du langage (1993)

Membre Junior de l'Institut Universitaire de France (1996-2001).

Médaille de bronze du CNRS en 1995 et d'argent en 2007.

### I Activités d'enseignement

Depuis 1994, j'enseigne à tous les niveaux (de la licence au Doctorat), à Paris 7 (UFRL).

#### 1. Enseignements statutaires

Cours magistral de Licence 2: Linguistique française, obligatoire pour les étudiants en 2e année de Lettres, et MASS (option linguistique). Introduction à la syntaxe du français. (27h)

Cours magistral (et TD) de Licence 3 (étudiants de Sciences du Langage, de Linguistique Informatique, et de l'INALCO): Grammaires génératives : Introduction à LFG. (60h)

Cours magistral (et TD) de Master 1 (étudiants de Sciences du langage, Phonétique et Linguistique Informatique): Nouvelles Théories syntaxiques : Introduction à HPSG. (48h)

Cours de Master 2 (étudiants de Sciences du langage, de Phonétique et de Linguistique Informatique) : Syntaxe formelle et comparée (le thème change chaque année) (30h).

Séminaire pour doctorants: Lectures en HPSG, avec Danièle Godard (les thèmes changent selon les années). (30h)

#### 2. Encadrement doctoral

Depuis 1999, j'ai fait soutenir 6 doctorats à l'UFRL (Paris 7) en syntaxe comparée et en linguistique informatique : M-H. Candito (1999), A. Kupsc (2000), L. Clément (2001), K. Gerdes (2002), Y.

Park (2006), S. Barrier (2006). Les 4 premiers sont actuellement Maîtres de conférence, le 5<sup>e</sup> est post-doc (CNET Lannion) et le dernier ingénieur dans le privé.

Je dirige actuellement 6 doctorants : Gabriela Bilbiie (AM), Anna Gazdik (Bourse co-tutelle), F. Henri (Bourse co-tutelle), Kil-So Ko, Oscar Marchena (Contrat CIFRE), G. Walther (AMN) et une HDR (P Samvelian). J'ai été membre extérieur (ou rapporteur) de 5 jurys de thèses et de 4 jurys d'HDR (Philippe Blache, Léa Nash, Guy Perrier, Mireille Bilger).

## II Activités de recherche

Je suis membre du laboratoire de Linguistique Formelle (équipe SSF) depuis 2000, après avoir appartenu au LADL (1988-1992) puis à l'équipe universitaire TALaNa (U. Paris7).

### 1. Travaux de recherche

J'ai travaillé dans trois directions principales:

- (a) la définition de modèles formels pour la syntaxe des langues naturelles, en participant à l'élaboration du modèle dit des Grammaires d'arbres adjoints lexicalisées (ou LTAG), puis à l'enrichissement et au développement du modèle dit des Grammaires syntagmatiques guidée par les têtes (ou HPSG), avec la notion de composition d'arguments ou de tête « faible »,
- (b) un travail d'analyse linguistique, sur la syntaxe du français et la grammaire comparée, en particulier sur les prédicats complexes dans les langues romanes, sur la déficience syntaxique dans des langues variées (contrainte de "légèreté") et sur les constructions coordonnées,
- (c) la production de ressources linguistiques basées sur les travaux théoriques et empiriques précédents, avec en particulier la réalisation d'un corpus électronique syntaxiquement annoté de près d'un million de mots (corpus arboré), et la rédaction d'une « grande grammaire du français ».

Concernant les grammaires d'arbres adjoints, j'ai défini, en collaboration avec Aravind Joshi et Yves Schabes, une variante mieux adaptée aux descriptions syntaxiques, en ajoutant l'opération de substitution, et la contrainte de lexicalisation, qui impose d'avoir pour chaque structure syntaxique élémentaire une ancre lexicale non vide (Abeillé, Joshi, Schabes 1988). Si l'on autorise plusieurs ancres lexicales, on peut représenter de façon originale les expressions idiomatiques qui ont une sémantique non compositionnelle mais une syntaxe relativement régulière et flexible. J'ai développé une grammaire électronique du français, basée sur ce modèle, et utilisée en analyse et en génération (Abeillé 1991, 2002). Face à la lourdeur du travail d'implémentation, et à la redondance du modèle qui oblige à manipuler des milliers de structures élémentaires, j'ai défini, en collaboration avec M-H. Candito, la notion de « métagrammaire » qui permet une factorisation linguistiquement intéressante et extrêmement utile en pratique (Abeillé et Candito 2000), qui a donné lieu à plusieurs thèses à Paris 7, à Nancy 2 et à l'Université de Pennsylvanie.

Concernant la grammaire syntagmatique, j'ai cherché, en collaboration avec D. Godard et I. Sag, à dégager les grands types de structures syntaxiques du français et les types de têtes lexicales qui leur correspondent. Nous avons d'abord travaillé sur les structures à auxiliaires, et plus généralement à plusieurs prédicats, qui posent des problèmes pour la détermination de la tête syntaxique ; nous avons proposé la notion de partage d'arguments qui permet notamment de régler la question de « montée » des clitiques sans avoir besoin de règles de déplacement (Abeillé & Godard 1996, 2002 ; Abeillé et al. 1997, 1998). Nous avons également travaillé sur les types de phrases (relatives, interrogatives) et sur la structure du groupe nominal et du groupe adjectival, en proposant une asymétrie entre ajouts (lexicaux ou légers) à gauche de la tête et ajouts (syntagmatiques) à droite de la tête (Abeillé & Godard 1997, 1999). En collaboration avec des sémanticiens, j'ai également travaillé sur les adverbes de degré qui posent des problèmes

intéressants de polymorphisme syntaxique et d'interface syntaxe-sémantique, notamment pour rendre compte de la diversité des catégories qu'ils modifient (Abeillé & Godard 2003,2004, Abeillé et al 2004). Depuis trois ans, je travaille sur les constructions mettant en jeu des conjonctions de coordination, pour distinguer celles qui sont véritablement des constructions coordonnées et celles qui relèvent d'autres structures. J'ai redéfini les conjonctions dites de coordination, comme des têtes « faibles », c'est-à-dire comme héritant de la plupart des traits syntaxiques de leurs compléments (Abeillé 2003, 2005).

Depuis plusieurs années, j'ai également cherché à confronter certains résultats obtenus sur le français aux faits observables dans d'autres langues, en particulier aux langues romanes. Concernant les prédicats complexes, j'ai proposé, avec D. Godard, de distinguer ce qui concerne le partage d'arguments (propriétés de monoclausalité partagées par plusieurs langues) et ce qui concerne la structure en constituants, qui peut varier de langue à langue (Abeillé et Godard 2002, 2003). Pour le français et l'italien, nous avons proposé un nouveau type de structure en constituants : structure "plate" où les deux verbes sont au même niveau que les compléments. Concernant l'ordre des mots, nous avons mis à jour une contrainte de légèreté (qui bloque la mobilité de certains compléments et modifieurs), qui s'applique à différentes catégories et à des types de langues variés (danois, coréen, espagnol etc). Cette notion amène à réexaminer la distinction entre mot et syntagme, et à critiquer la théorie X-barre selon laquelle les mots projettent forcément des syntagmes (Abeillé et Godard 2000, 2004). Elle amène à redéfinir la notion de « déficience » syntaxique, en ajoutant, à côté des formes dites « faibles », des formes dites « légères ».

Je travaille désormais sur la syntaxe de la coordination, pour montrer que sont en jeu des constructions variées. Je propose en particulier de distinguer le sous-constituant introduit par une conjonction (que j'appelle syntagme conjoint) et l'ensemble de la construction qui peut relever de la coordination (si les contraintes de parallélisme sont observées) ou non (Abeillé 2003, 2005). Dans cette optique, j'ai analysé les conjoints à prosodie incidente tels que (*Il est parti et vite*, ou *Fantômette, car c'était elle*, ...) comme des constructions non coordonnées (le syntagme conjoint est un ajout). J'ai également travaillé, avec R. Borsley, sur les constructions corrélatives (*Plus je lis, plus j'apprends*) pour montrer qu'elles appartiennent à deux types de construction selon les langues : soient elles sont coordonnées (comme en français : *Plus je lis et plus j'apprends*), soit elles sont subordonnées comme en anglais (Abeillé et Borsley 2006, 2007).

Mes travaux en syntaxe appliquée ont pour but la mise à disposition de ressources linguistiques informatisées. J'ai travaillé, avec plusieurs doctorants, à la réalisation des grammaires électroniques de large couverture, pour le français (FTAG), ou pour d'autres langues (allemand, coréen) selon les mêmes principes (Abeillé 2002). Depuis 1997, et grâce à ma nomination à l'IUF, j'ai coordonné également l'annotation linguistique d'un corpus français de référence (1 Million de mots extraits du journal *Le Monde*), avec marquage des mots composés, désambiguation des catégories lexicales, et marquage des frontières de constituants ainsi que de leurs fonctions grammaticales (Abeillé et al. 2003). Alors qu'un corpus équivalent a été réalisé pour l'anglais dès 1994 (Penn Treebank) aucun corpus de ce genre n'était disponible pour le français. Ce corpus arboré (*treebank*) est décrit sur le site web du laboratoire (<http://llf.cnrs.fr/>) et permet d'extraire directement des constructions, avec leurs fréquences, et leurs contextes. Il est distribué gratuitement à des fins de recherche et utilisé par une quarantaine d'équipes françaises et étrangères.

Je participe également au projet de *Grande Grammaire du Français*, qui a démarré à l'automne 2002, coordonné par A. Delaveau et D. Godard, et soutenu par l'ILF. Il s'agit de faire le point sur les avancées des 30 dernières années en syntaxe française, et de les mettre à la disposition du public, de façon assez proche de la *Grande grammatica Italiana di consultazione*, coordonnée

par L. Renzi et G. Salvi pour l'italien. A la différence de cette dernière, qui se basait essentiellement sur les travaux en grammaire générative, le projet français pourra aussi bénéficier de l'appui des corpus, pour une prise en compte de la variation. Je suis responsable des chapitres "Prédicats Complexes" et "Coordination et Juxtaposition" et auteur (ou co-auteur) de plusieurs sections dans six autres chapitres (Chapitre Phrase, Chapitre Nom, Chapitre Détermination, Chapitre Adverbe, Chapitre Proforme, Chapitre Ordre des mots).

## 2. Projets scientifiques

Je participe actuellement aux projets suivants :

- 2002-08: *Grande Grammaire du Français*, projet piloté par A. Delaveau et D. Godard et soutenu par la fédération ILF (CNRS) qui fait collaborer une cinquantaine d'auteurs pour un ouvrage de 7,2 millions de signes (20 chapitres). Je suis auteur (ou co-auteur) de 17 sections, responsable de deux chapitres et membre du comité de Direction.

- 2006-10 : PRO-GRAM (La prosodie dans la grammaire), projet piloté par J-M. Marandin, soutenu par l'ANR, qui fait collaborer les équipes LLF, ARP (Paris 7), LPL et DELIC (Aix).

- Depuis 1997 : Je coordonne le développement d'un corpus électronique annoté pour le français (1 Million de mots extraits du journal *Le Monde*) : le corpus arboré de Paris 7. Ce corpus a été marqué à la fois pour les catégories morphosyntaxiques associées aux mots simples et aux mots composés, mais aussi pour les constituants majeurs auxquels sont associées des fonctions grammaticales. Je coordonne actuellement le développement d'une interface d'interrogation à distance (par Internet) pour utilisateurs linguistes, avec le soutien du CNRTL (Nancy).

J'ai été responsable des projets suivants:

- 2001-03: Projet CORFRANS (*Corpus électronique français annoté pour la syntaxe*), financé par l'Aupelf-Uref, piloté par LLF en collaboration avec le LATL (Univ. de Genève) et le RALI (Univ. de Montréal), dans le domaine de l'annotation morphosyntaxique corpus français journalistiques contemporains.

- 2000-04: Coopération internationale avec l'équipe d'A. Joshi à l'*Institute for Research in Cognitive science* (IRCS) à l'Université de Pennsylvanie (Philadelphie) dans le cadre du contrat quadriennal de Paris 7 : échange de chercheurs et doctorants dans le domaine des ressources linguistiques et des grammaires d'arbres adjoints (TAG).

- 2003-04 (avec A Kupsc) « Grammaires formelles comparées du français et du polonais »: projet de coopération internationale CNRS-IPIPAN avec l'Académie des Sciences de Varsovie. J'ai travaillé en particulier sur l'analyse syntaxique et l'implémentation des pronoms clitiques et des constructions relatives dans les deux langues.

- 2000-02: Projet *Evaluation et enrichissement de corpus annotés*, soutenu par le CNRS (projet APN), qui fait collaborer Paris 7 (LLF), avec les équipes d'Aix (LPL et Corpus), de Nancy (LED, Loria) et du LIMSI, autour de l'utilisation des corpus français annotés.

- 2002-05 (avec D. Godard) "Grammaire typologique des Formes faibles", projet 10 de la fédération de Typologie (CNRS), rassemblant une quinzaine de chercheurs de cinq laboratoires (LLF, SILEX, SFL, LACITO, Dynamique du langage).

J'ai participé aux projets suivants:

2000-02 : Projet : *Préférences et analyse syntaxique du français* (dans le cadre de l'action Cognitive), piloté par J. Pynte (Aix) rassemblant 4 équipes françaises, anglaise et américaine (LPL, LLF, CUNY, Dundee).

2000-02: Projet: RLT (Ressources linguistiques pour les TAG), piloté par E. de la Clergerie (Inria, Rocquencourt), financé par l'INRIA (ARC) qui rassemble 3 équipes françaises (Attol-INRIA, Calligramme-LORIA, LLF).

2000-03 : projet PICS coordonné par Henriette de Swart et Francis Corblin : « Formal semantics and French data », qui a fait collaborer des linguistes français et hollandais et débouché sur la publication du livre *Handbook of French semantics*. J'ai travaillé dans le groupe Adverbes avec Jenny Doetjes et Henriette de Swart.

2003-04 : Projet EASy (*Evaluation des analyseurs syntaxiques*), piloté par P. Paroubek (LIMSI) qui rassemble une vingtaine d'équipes participantes française et étrangères, dans le cadre du programme Technolangue du MRT. J'ai coordonné la conversion et la mise à disposition d'un corpus de référence annoté.

2004-06 : projet Alliance « Grammatical Interfaces », resp. D. Arnold et D. Godard, collaboration entre LLF (CNRS, Paris 7) et University of Essex.

2006-07 : Mosaïque (Modélisation de formalismes syntaxiques de haut niveau), projet piloté par L. Clément, soutenu par l'Inria (ARC), qui fait collaborer plusieurs équipes françaises autour de l'implémentation de formalismes syntaxiques (LORIA, LPL, LLF, Attol, Modyco).

### 3. Relations avec le monde industriel

Je suis responsable des relations avec les partenaires industriels intéressés par la distribution du corpus de Paris 7 annoté pour la syntaxe (*French treebank*), et des contrats de licence associés. Le corpus a été vendu à 3 partenaires : Rhetorical Systems (Edinbourg), Toshiba (Cambridge), Microsoft (Seattle).

### 4. Publications

Six Ouvrages, de nombreux articles de revue et chapitres dans des ouvrages collectifs.

1993. *Les nouvelles syntaxes : grammaires d'unification et analyse du français*, Coll. Linguistique, Armand Colin, Paris.

2000. Ed. (avec O. Rambow), *Tree Adjoining grammars : Formalism, linguistic analysis and Processing*, Stanford : CSLI Publications.

2002. *Une grammaire électronique du français*, Coll. Sciences du langage, CNRS Editions, Paris.

2003. Ed. *Treebanks : building and using parsed corpora*, Kluwer, Dordrecht.

2007. *Les grammaires d'unification*, Hermès. Londres

2007. Le Dictionnaire du Vélib', Editions du Panama

Depuis janvier 2006, je dirige la collection *Langues et Syntaxe* aux Editions Hermès. 2 titres par an. ouvrages parus :

Denis Creissels, *Linguistique générale, une introduction typologique*, 2 volumes.

Elisabeth Delais-Roussarie, Laurence Labrune : *Des sens et des mots*

### III Administration et responsabilités collectives

Depuis plus de dix ans, j'ai assumé des responsabilités administratives, scientifiques et pédagogiques variées, à Paris 7 et dans plusieurs associations scientifiques. De 1999 à 2001, j'étais membre du bureau de l'IUF. De décembre 1999 à décembre 2002, j'ai assuré la direction de l'UFR de linguistique. Je suis membre des commissions de spécialistes de Paris 7 et ENS Lyon (LSH), et membre du CA de l'UFR de Linguistique de Paris 7.

#### Comités de rédaction

J'ai été membre de 1995 à 1997 du comité de rédaction de la revue *Computational Linguistics*, et de 1990 à 1999 du comité de rédaction de la revue TAL.

Membre du Comité de rédaction de *Recherches Linguistiques*, Université de Rabat, Maroc.

Membre de l'*Editorial Board* du *European Student Journal on language and speech* (journal électronique)

#### Invitations d'universités étrangères

Conférencier invité à l'Université d'Edinburgh en 1999, à l'Université de Tübingen et à l'Académie des Sciences de Varsovie en 2000, à l'Université de Sarrebrück en 2002, à l'Université d'Essex, à l'Université Libre de Bruxelles et à l'Université de Bucarest en 2003.

Professeur invité à la Vilem Mathesius School de l'Université Charles de Prague (2000) et à l'Università degli Studi (Venezia) en 2002.

Professeur invité à ESSLI (Colchester) en 1993, au Summer Institute (LSA) en 1997 (Cornell University) et à l'EALING (ENS Ulm) en 2005.

Professeur invité à l'Université Charles de Prague (mai 2007), à l'alliance française de Hong Kong (janvier 2008) et à la maison française d'Oxford (mai 2008).

#### Organisation de conférences

1993: Co-organisation de la conférence TAG+3 à Paris 7

1996: Co-organisation d'un workshop sur la syntaxe des langues romanes (ESSLI) à Prague.

1999: Organisation d'une conférence internationale sur les corpus arborés (treebanks) à Paris 7; membre du comité d'organisation de la conférence TALN (Cargèse).

2000: Organisation de la conférence internationale TAG+5 à Paris 7.

2001: Membre du comité d'organisation de la conférence HPSG (Trondheim), membre du comité scientifique de Formal Grammar Conference (Helsinki).

2002: Membre du comité scientifique du colloque TAG+6, et du Workshop on parsing evaluation (LREC).

2003: Organisation d'une journée d'études sur les formes faibles (Paris 7) avec D Godard.

Co-organisation du 5e Colloque LINC (EACL, Budapest). Membre du comité scientifique du workshop *Evaluation des analyseurs syntaxiques* (TALN). Membre du comité de programme de la Conférence *Mathematics of Language* (MOL 8) (ESSLI) et du Workshop *Multiword Expressions* (ACL).

2004-2005 : Organisation, avec D. Godard, d'une journée d'études sur les prépositions faibles (Paris 7). Membre du comité de programme de la Conférence Formal Grammar (ESSLI) et du Workshop *The linguistic dimensions of Prepositions*

2006 : Organisation d'une conférence internationale : *Coordination and ellipsis* (Paris 7, juin).

Membre du comité de lecture des conférences *Formal grammar* et HPSG.

2007 : membre du comité de lecture de CSSP et TALN.

2008 : Membre du HPSG standing committee et du comité scientifique du CMLF. Organisation d'une conférence internationale : *Elliptical constructions* (Paris Chicago Center)

### **Divers**

De 1995 à 1997, j'étais membre de l'*advisory board* du chapitre européen de l'*Association for computational linguistics* (EACL).

Depuis 2001, je suis membre du jury du prix européen *Beth Prize* (chargé de récompenser chaque année une thèse dans le domaine Langue, Logique et Informatique)

J'ai été expert pour le Fonds suisse de la recherche scientifique, pour l'ULB et pour le NWO (Pays-Bas) pour le programme national Stevin (en traitement automatique du néerlandais) et pour le CRSH canadien.

2006 : Expertises pour les *search committee* de l'Université de Yale et de la KUL (Louvain).