

Curriculum vitae (août 2023)

Contenu et remarque initiale : ce CV long est composé d'une synthèse de carrière (début p. 1), d'une section consacrée à mes activités d'enseignement (p. 3), d'une section dédiée à mes activités de recherche (p. 7), d'une section décrivant mes responsabilités collectives hors enseignement et recherche (p. 17). Il se termine par des annexes listant mes publications et mes encadrements. Afin de mettre en évidence mes activités les plus récentes, les parties purement antérieures à 2019 ont été mises en gris.

Nom d'usage : Sébillot **Prénom** : Pascale

1. Synthèse de la carrière

Diplômes

- ingénieur informatique INSA Rennes 1984
- doctorat informatique INSA Rennes 1989
- HDR Université de Rennes 1 2002

Situation professionnelle

- PRU d'informatique à l'INSA Rennes depuis septembre 05 (1^{re} classe en septembre 10, classe EX1 en septembre 15, classe EX2 depuis septembre 20) ; recherche dans l'équipe LinkMedia de l'IRISA, UMR 6074 ; section CNU : 27
 - Demi-délégation au CNRS depuis septembre 21 (suite à ma prise de fonction de directrice adjointe de l'Irisa au 1^{er} janvier 2021)
- Auparavant MC à l'Université de Rennes 1 (1993-2005) et MC à l'IUT de Vannes (1990-1993)

Enseignement

Mise en place de formations

- En septembre 06, **responsabilité de la refonte complète du Master 2 recherche en informatique MRI** co-habilité par 8 établissements d'enseignement supérieur de Bretagne. Proposition d'un tronc commun et de 10 parcours thématiques, soit un total de 54 unités d'enseignement
- En février 07, **co-responsabilité (50%) de la mise en place**, en 3^e, 4^e et 5^e années informatique INSA Rennes, **d'un parcours Traitement et modélisation de masses de données (TMMD)**. Définition de 20-25% des enseignements de 4 semestres

Responsabilités de formations (uniquement les plus récentes ci-dessous)

- De septembre 16 à septembre 21, **responsable de la 5^e année du département Informatique** de l'INSA Rennes (70 à 75 étudiants par an)
- De septembre 06 à août 11, co-responsable (95% jusqu'en septembre 09, 50% ensuite) du Master 2 recherche en informatique MRI : 50 à 60 étudiants, 4 sites, 8 établissements, 54 UE, 110 intervenants
- De février 07 à août 17, co-responsable (50%) du parcours Traitement et modélisation de masses de données (TMMD) de l'INSA

Fort investissement dans la mise en place d'enseignements

- En particulier dans le domaine du stockage et du traitement d'informations volumineuses à mon arrivée à l'INSA, et dans celui du traitement automatique des langues depuis 15-16
- Nombreux cours créés (responsabilité assurée) : 25 (15 créations, 10 co-crétions)

Enseignements : de bac+1 à bac+5, sous toutes les formes de pédagogie

Recherche

Thématiques : traitement automatique des langues (TAL), structuration automatique de collections de documents multimédias, couplage contenus multimédias et bases de connaissance

Encadrements

Thèses (co)dirigées et soutenues	Thèses (co)dirigées et en cours	Post-doctorats	M2 recherche
17	2	1	27

Publications (pour les conférences uniquement sélection sur articles longs et hors workshops)

Chapitres internationaux	Chapitres nationaux	Revue internationale	Revue nationale	Conférences internationales	Conférences nationales
5	3	9	6	46	27

Animation de la recherche

- Animation et développement de l'activité TAL à l'IRISA (seule permanente du domaine à Rennes jusqu'en 05)
- Depuis 19 : **membre du bureau du GDR TAL ; co-responsable du groupe de travail** Intermodalité et multimodalité de janv. 19 à déc. 20 ; en 18 : membre du comité des animateurs des (8) axes de réflexion du pré-GDR TAL, en charge avec B. Fabre (LIS) de l'axe Intermodalité et multimodalité.
- De 13 à 20 : **membre du comité permanent (CPerm) de la conférence TALN**, en charge du suivi de l'organisation de la conférence
- Mai 18 : **membre du comité d'organisation** de CORIA-TALN-RECITAL 2018 à Rennes, 250 participants, et **co-présidente du comité de programme de TALN 2018**
- 03-09 : **membre du conseil d'administration** de l'Association pour le traitement automatique des langues (ATALA), élue en 03, réélue en 06
- Plusieurs responsabilités ou co-responsabilités de contrats (dont récemment : un Labex CominLabs à 5 partenaires ; une Cifre ; une ARED Région Bretagne (1/2 bourse) et un AI4SDA (1/2 bourse du programme doctoral *AI for semantic data analytics* de l'ANR) ; nombreuses participations dont, récemment, au projet JPICH READ-IT, à l'IPL Inria iCODA, aux ANR ARCHIVAL et JCJC A4LL (membre de l'*advisory board* pour cette dernière) et à un contrat avec Le Télégramme (journalisme augmenté)

Éléments principaux de visibilité

- Plusieurs collaborations internationales contractualisées ou non, dont parmi les plus récentes, Dublin City University et KU Leuven
- **Rédacteur en chef de la revue TAL** (traitement automatique des langues) de 14 à 23, et **membre de son comité de rédaction** depuis 09
- Comités de programme : participations régulières à des CP de conférences de mes domaines de recherche, dont depuis 19 : 9 conférences internationales (IJCAI, ACL, ECIR...) et 6 nationales
- Jurys de thèses et HDR : thèses (hors encadrements) : rapporteur : 21, expert international : 1, présidente : 14, examinatrice : 10 ; HDR : rapporteur : 6, examinatrice : 4
- Participations régulières à des campagnes d'évaluation (MediaEval, TREC Vid, CLEF HIPE)

Responsabilités collectives

Implication locale

- **Directrice-adjointe de l'UMR IRISA** depuis janv. 21 (plus de 850 personnes, 37 équipes)
- **Vice-présidente du Comité d'orientation scientifique de l'IRISA** depuis sept. 13 (comité de dix membres, en charge de la prospective scientifique à moyen et long terme de l'IRISA)
- **Responsable de la composante INSA Rennes de l'IRISA** de oct. 16 à août 21 (50 membres dont 23 permanents)
- **Membre du conseil de l'ED MathSTIC** de 17 à 22 ; **membre du comité des thèses de l'ED MathSTIC** de juin 15 à avril 21 (inscriptions et demandes de soutenance pour les thèses en informatique du site de Rennes, environ 150 dossiers par an à étudier à 2)
- Membre élu du Conseil scientifique de l'INSA Rennes de 11 à 14 et de 18 à 22 ; membre élu du Conseil du département informatique de l'INSA de 06 à 17 et depuis 21
- Membre du comité des projets de l'IRISA de 98 à 02 ; membre du Conseil de la composante INSA

Rennes de l'IRISA de 11 à 21, présidente d'oct. 16 à août 21

- Présidente du bureau de la section 27 de l'INSA Rennes de 10 à déc. 20, interlocutrice des RH pour tout ce qui a trait aux recrutements d'enseignants-chercheurs et d'ATER de cette section du CNU (en charge, en particulier, de l'organisation des comités de sélection)
- de 94 à 04 : membre du bureau (assesseur) de la commission de spécialistes de la 27^e section de l'Université de Rennes 1
- Offre de formation (quadriennal 08-11) : part très active prise d'une part à l'élaboration du master 2^e année recherche en informatique de Bretagne, mais également à la rédaction des dossiers d'habilitation du master informatique de l'université de Rennes 1, établissement porteur au sein duquel ce master 2 était une des spécialités, et du master recherche en informatique dans l'offre masters de l'INSA Rennes (appréciation « A » de l'AERES).

Implication nationale

- **Membre du CNU 27 de 11 à 19** (nommée de nov. 11 à nov. 15 et élue de nov. 15 à nov. 19)
- En 13 : membre de l'instance d'évaluation de la PES
- En 18 : membre du comité de visite HCERES du LIP6 ; en 13 : membre du comité de visite AERES du LIMSI
- Membre de 16 comités de sélection (hors INSA) depuis 19 : Aix-Marseille, Angers, Avignon, Caen (présidente), Grenoble, Lannion, Le Mans (x2), Nancy, Nantes, Paris-Saclay (x2), Rennes ENS, Toulouse, Vannes (x2) ; également 4 fois experte ou membre de commissions d'audition repyramidage (hors INSA) depuis 19 : Angers, Paris-Sud, Rennes et Rennes 2

2. Activités pédagogiques

2.1. *Présentation de l'activité d'enseignement*

Enseignements effectués : l'ensemble de mes enseignements depuis 1988 est résumé ci-dessous :

Lieux : IUT de Vannes ; IFSIC/Université de Rennes 1 (maintenant ISTIC) ; INSA Rennes

Formations (appellations LMD) : IUT Statistique et Traitement Informatique des Données 1^{re} et 2^e années ; IUT Informatique 1^{re} et 2^e années ; L1 PM ; L1 Biologie ; L2 MPI ; L2 SM ; INSA 1^{re} et 2^e années ; L3 et M1 d'informatique ; IUP 2 Génie Biologique ; DIIC 1 et 2 (Diplôme d'ingénieur) ; IUP-Miage1, 2 et 3 ; INSA 3^e et 4^e années ; M1 Sciences du Numérique et Sport (SNS) de l'EUR (école universitaire de recherche) Digisport ; M2pro-CCI (Compétence complémentaire en informatique) ; M2pro-ISA (Informatique et ses applications) ; M2pro-MITIC (Méthodes informatiques et technologies de l'information et de la communication) ; M2R informatique ; INSA 5^e année.

Matières : Initiation à l'algorithmique et à la programmation, Programmation logique, Programmation objet, Méthodes de programmation, Algorithmique des graphes, Systèmes d'exploitation, Bases de données relationnelles, Bases de données avancées et systèmes d'information modernes, Logique et bases de données, Intelligence artificielle, Traitement automatique des langues, Acquisition de connaissances à partir de données, Recherche d'information.

Forme : CM, TD, Cours-TD, TP et projets.

Autres : Maîtresse-auxiliaire en Mathématiques en seconde et première F2 en lycée ; formation continue.

Mise en place de formations

- En septembre 2006, j'ai pris la **responsabilité de la refonte complète du Master 2 recherche en informatique** co-habilité par 8 établissements d'enseignement supérieur de Bretagne (INSA de Rennes, Université de Rennes 1 (IFSIC et ENSSAT Lannion), SUPÉLEC, TELECOM Bretagne (Rennes et Brest), ENS Cachan antenne de Bretagne, ENIB, Université de Bretagne-Sud, Université de Bretagne-Occidentale), afin de préparer la nouvelle habilitation 2008-2011. La modification fondamentale apportée à la maquette précédente de ce diplôme a été un passage d'une organisation en filières assez généralistes à celle de parcours thématiques plus focalisés. Cette réorganisation a nécessité la mise en place sur les 4 sites géographiques (Rennes, Brest, Lannion et Vannes) d'une réflexion (plusieurs dizaines de réunions) à laquelle ont participé les enseignants-chercheurs des établissements concernés, ainsi que les chercheurs de leurs laboratoires (IRISA, LISyC, VALORIA, équipe « Sécurité des systèmes d'information et réseaux » du campus de Rennes de Supélec, et départements « Informatique » et « Réseaux, Sécurité et Multimédia » de Télécom Bretagne). Cette

réflexion a abouti à la proposition d'un tronc commun et de 10 parcours thématiques, ce qui correspond à un total de 54 unités d'enseignement

- En février 07, j'ai pris la **co-responsabilité** (à 50%), avec V. Patilea du centre de mathématiques, **de la mise en place, en 3^e, 4^e et 5^e années** au sein du département informatique de l'INSA, du parcours **Traitement et modélisation de masses de données** (TMMD ; précédemment intitulé traitement des données et modélisation mathématique mais restructuré vers des aspects plus *Big Data* en 12-13). Ceci a impliqué une coordination entre des collègues mathématiciens et du département informatique pour bâtir un programme permettant de doter les étudiants des bases nécessaires pour se confronter à des données volumineuses, imparfaites..., et en faire émerger les modèles et connaissances utiles. Cet enseignement de spécialisation représentait environ 20-25% des enseignements dispensés au 2nd semestre de 3^e année, lors des 2 semestres de 4^e année et au 1^{er} semestre de 5^e année.

Mise en place et responsabilité d'enseignements

Depuis 1988, j'ai eu l'occasion de créer un certain nombre d'enseignements. Et j'ai été recrutée en tant que PRU en 2005 à l'INSA Rennes, en particulier, pour renforcer, au département Informatique, les enseignements portant sur le traitement des données, et pour m'investir dans le 1^{er} cycle sur la mise en place d'un enseignement d'initiation à l'informatique portant sur la gestion des bases de données, à destination d'étudiants se destinant ou non par la suite à l'informatique.

Actuellement, un premier domaine que j'enseigne porte donc sur les **bases de données relationnelles**. Il se décline, avec des objectifs différents, en 3 versions puisque je l'enseigne, en l'adaptant au public visé, à 3 populations d'étudiants : depuis 06-07, enseignement dispensé à l'ensemble des étudiants de 2^e année de l'INSA Rennes ; depuis 12-13, aux étudiants entrant à l'INSA en 3^e année INFO ; et depuis cette année universitaire 22-23, à des étudiants de M1 Sciences du Numérique et Sport (SNS) de l'EUR Digisport. Le second domaine dans lequel je m'investis est le **traitement automatique du langage naturel (TAL)**, plus proche de mes activités de recherche. Lors de la remise à plat des enseignements des 3 années du département INFO en 15-16, il a été collégialement décidé de mettre en place 2 modules de TAL de 26h de cours et TP en 3^e et 4^e année, à destination de la moitié des étudiants de ces promotions (parcours *Data science* et Médias et interactions) et portant sur ce domaine et son intrication avec l'apprentissage automatique. Après une première version de **ces deux modules, dont je suis responsable** (3 intervenants, 2 en CM pour le 1^{er} et 3 pour le 2nd), nous avons souhaité accentuer la place dédiée aux modèles neuronaux qui deviennent de plus en plus prégnants dans cette discipline et que les étudiants doivent donc maîtriser. **En 21-22, nous avons donc revu l'ensemble des cours (et transparents) et TP du module de 3^e année et en 22-23 l'ensemble de ceux de 4^e année**, pour coller au plus près aux modèles les plus récents. Ces modifications ont été faites avec le souci de toutefois conserver une part importante aux aspects linguistiques et TAL plus « traditionnels » afin que les étudiants que nous formons dans cette matière ne tombent pas dans un travers consistant à appliquer des modèles d'apprentissage disponibles sur étagère sans réfléchir au matériau traité, du texte, interprétable et sur lequel il est possible de réfléchir afin d'ajuster au mieux les traitements requis par la tâche visée. Pour ces enseignements, des supports sous forme de transparents ont été produits, et sont distribués aux étudiants et disponibles sous Moodle. Outre ces cours, je cite ci-dessous les matières dont j'ai eu la responsabilité, distinguant les cas où la mise en place de l'enseignement s'est faite à partir d'un cours préexistant ou non. J'indique entre parenthèses les années concernées (où j'ai, entre autres, assuré le CM). La mention *polycopié* réfère un document rédigé.

INSA : 9

Pour tous les enseignements cités, des supports sous forme de transparents ont été produits et sont distribués aux étudiants ou disponibles sous Moodle (sauf en 2^e année (polycopié))

- *Bases de données avancées et systèmes d'information modernes* ; 5INFO ; création du CM et des TP (de 05-06 à 16-17). *Techniques et applications de l'indexation multimédia* ; M2 recherche en informatique ; co-création du CM (de 05-06 à 06-07). *Traitement automatique des langues et interactions langagières* ; 4INFO ; co-création du CM (depuis 16-17). *Bases de données* ; M1 Sciences du Numérique et Sport (SNS) de l'EUR Digisport ; création des Cours-TD et des TP à partir d'un cours préexistant (depuis 22-23). *Acquisition de connaissances à partir de données : méthodes symboliques* ; 4INFO ; co-création du CM et des TP (de 07-08 à 16-17). *Programmation logique* ; 3INFO ; création du CM et co-création des TP (de 12-13 à 20-21). *Traitement automatique*

du langage écrit et oral ; 3INFO ; co-crédation du CM et des TP (depuis 15-16). Bases de données avancées ; 3e année dépt Génie mathématique ; création des CM et TP (de 14-15 à 16-17). Bases de données ; 2e année ; co-crédation des Cours-TD et des TP (depuis 06-07) ; co-rédaction d'un **polycopié**.

IFSIC/Univ. Rennes 1 (maintenant ISTIC ; appellations LMD des formations) : 12

- Algorithmique et programmation ; L1 PM ; création du CM à partir d'un cours préexistant (de 90-91 à 93-94). Algorithmique et programmation ; L2 SM ; création du CM à partir d'un cours préexistant ; co-rédaction d'un **polycopié** (de 95-96 à 97-98). Programmation logique ; IUP-Miage 1, 2, puis 3 ; création des CM, TD et TP à partir d'un cours préexistant (de 94-95 à 97-98, et de 00-01 à 04-05). Bases de données ; IUP 2 Génie biologique ; création des CM, TD et TP (de 01-02 à 04-05). Logique et intelligence artificielle ; DIIC 2-LSI ; co-crédation des CM et TP (de 94-95 à 97-98). Algorithmique et complexité ; M2pro-CCI ; création des CM et TD à partir d'un cours préexistant ; création des TP (en 93-94, puis de 96-97 à 97-98). Mise à niveau en méthodes de programmation ; M2pro-ISA ; création des CM et TP (de 94-95 à 97-98). Intelligence artificielle ; IUP-Miage 3 ; création du CM à partir d'un cours préexistant (03-04). Bases de données en M2pro-MITIC ; co-crédation des CM, TD et TP à partir d'un cours préexistant (03-04). Bases de données en M2pro tronc commun ; co-crédation des CM, TD et TP à partir d'un cours préexistant (04-05). Indexation et bases de données ; M2pro MITIC ; création des CM (de 01-02 à 04-05) et TP (de 03-04 à 04-05). Techniques et applications de l'indexation multimédia ; M2R informatique ; co-crédation du CM (de 03-04 à 04-05).

IUT de Vannes : 4

- Langage C ; STID 1 ; création des CM et TP à partir d'un cours préexistant (de 89-90 à 92-93). Bases de données relationnelles ; STID 2 ; création des CM, TD et TP (de 86-87 à 92-93). Système ; STID 2 ; création des CM et TP ; rédaction d'un **polycopié** (de 89-90 à 92-93). Logique et bases de données ; Informatique 2 ; création des CM, TD et TP (de 88-89 à 91-92)

2.2. Présentation synthétique des enseignements par niveau, par type de formation et par nature

Ayant listé en section 2.1. l'ensemble de mes activités d'enseignement, je me focalise ici sur les seuls cours principaux que je dispense actuellement à l'INSA Rennes. Les enseignements sont à destination d'étudiants en formation initiale. Sauf cas particulier du 2nd semestre 19-20 et de la *quasi* totalité de l'année 20-21, l'ensemble des enseignements se déroule en présentiel. Les volumes indiqués sont ceux me concernant pour l'année d'enseignement la plus récente (volume par étudiant, même pour les matières où j'assure plusieurs groupes).

- Bases de données, 2^e année INSA Rennes, étudiants de la filière classique, de la filière internationale, et du double diplôme SciencePo-INSA : présentation du modèle relationnel sous ses aspects interrogation (algèbre relationnelle, SQL) et manipulation (création de relations, mises à jour, accès concurrentiels), puis modélisation entités-associations et interfaçage avec Java. **Responsabilité de l'enseignement pour 300 étudiants (14 groupes) et coordination des 11 à 12 intervenants.** 16h Cours-TD, 18h TP (dont 12h de mini-projet encadrées à deux), 24 étudiants par groupe (deux groupes chaque année), depuis 06-07. Polycopié initialement co-rédigé, mis à jour annuellement par mes soins, distribué aux étudiants.
- Bases de données, 3^e année INSA, module de « remédiation » à destination des étudiants entrant directement, depuis l'extérieur, au département INFO de l'INSA, n'ayant pas fait de bases de données. Programme identique au module précédent, sans la partie interfaçage BD-Java. Responsabilité de l'enseignement. 12h Cours-TD et TP, 3 à 15 étudiants, depuis 12-13. Polycopié du cours de 2^e année distribué.
- Bases de données, Master 1 Sciences du Numérique et Sport (SNS) de l'EUR Digisport : l'EUR propose une formation hybride sciences du sport et sciences du numérique à des étudiants ayant suivi des parcours divers. Ce cours vise à doter les étudiants non issus d'une formation initiale en numérique de fondements en modélisation, création et manipulation de bases de données. Création et responsabilité de l'enseignement. 18h Cours-TD et TP, 23 étudiants, depuis 22-23. Polycopié du cours de 2^e année distribué.
- Traitement automatique du langage (TAL) écrit et oral, 3^e année INSA : l'objectif de l'enseignement

est de doter les étudiants des compétences nécessaires pour analyser l'information présente dans des documents écrits ou contenant de la parole. Il présente d'une part les fondements du traitement automatique du texte écrit, potentiellement issu de la transcription de la parole, permettant d'accéder au contenu des documents : représentations des mots et des documents, modélisation du langage et outils fondamentaux du traitement du langage naturel (analyse de séquences). Il se focalise d'autre part sur l'application de ces fondements et outils à certains domaines ou applications du TAL (recherche d'information, analyse syntaxique...). Outre le cours introductif, mes interventions portent principalement sur le 2nd aspect. Responsabilité de l'enseignement et coordination des 3 intervenants, co-création du CM et des TP. 6h CM, 40 étudiants, depuis 15-16. Transparents des CM distribués aux étudiants et disponibles sous Moodle.

- *Traitement automatique des langues et interactions langagières*, 4^e année INSA : le langage, sous sa forme écrite ou orale, est la modalité la plus commune d'interaction et de transmission d'informations. L'objectif de l'enseignement est de doter les étudiants des connaissances nécessaires pour leur permettre d'acquérir des compétences complémentaires en TAL pour accéder à l'information contenue dans de grandes masses de documents, et pour comprendre comment une interface langagière entre un humain et une machine peut être mise en place, à l'écrit ou à l'oral. Il présente les éléments fondamentaux de la compréhension du langage dans un cadre d'interaction dialogique, ainsi que les fondements de la génération de texte écrit. Il aborde brièvement la synthèse de parole à partir du texte. Enfin, il s'intéresse à certaines applications phares du TAL permettant de gérer les volumes conséquents de données, d'y naviguer et d'en extraire des informations précises (résumé automatique, extraction d'informations, etc.). Mes interventions concernent principalement ce 2nd aspect. Responsabilité de l'enseignement et coordination des 3 intervenants, co-création du CM. 4h CM, de 9 à 40 étudiants, depuis 16-17. Transparents des CM distribués aux étudiants et disponibles sous Moodle.
- *Autres* : chaque année, j'assure également le suivi de stages effectués pendant l'été par les étudiants de 3^e ou 4^e années (avec relecture et notation des rapports), et des stages de fin d'études d'étudiants de 5^e année (visites (physique ou en visio), relecture et notation des rapports et soutenances). Par ailleurs, pendant ma période de responsable de la composante INSA de l'IRISA, j'ai fait des interventions de présentation de l'UMR dans le cadre de modules d'information des étudiants de 1^{re} et 2^e années.

2.3. *Direction et animation de formations, dont partenariats internationaux*

Responsabilités de formations ou filières

- **De septembre 16 à septembre 21**, j'ai été **responsable de la 5^e année** du département informatique de l'INSA. Cette tâche implique, entre autres, outre le suivi des étudiants et un certain nombre de réunions, la production et la saisie des emplois du temps, le suivi des notes et l'organisation des jurys. La particularité de cette 5^e année regroupant 70 à 75 étudiants est le nombre très important de situations particulières des élèves : certains font leur 1^{er} semestre à l'étranger (25 à 30), d'autres sont en double-diplôme (à l'étranger, en master 2 recherche, en école de commerce ou de gestion, avec ou sans année de césure...), certains effectuent une scolarité « à cheval » entre la 4^e et la 5^e année... ce qui conduit à un besoin de suivi très individualisé.
- **De septembre 06 à août 11**, j'ai été **co-responsable du Master 2 recherche en informatique** co-habilité par 8 établissements d'enseignement supérieur de Bretagne. De **06-07 à 08-09**, j'assurais **plus de 95% de la responsabilité et du travail**; à partir de 09-10, un changement de co-responsable a fait qu'après quelques mois de mise en place, la co-responsabilité **s'est équilibrée à 50%**. Outre la mise en place de ce diplôme assurée seule (cf. ci-dessus), cette tâche très prenante impliquait la gestion quotidienne du diplôme, regroupant entre **50 et 60 étudiants** présents sur **4 sites** géographiques et engagés dans **54 unités d'enseignement** assurées par **110 intervenants**, le suivi des notes et la tenue des jurys, le recrutement des étudiants pour l'INSA et l'université de Rennes 1, mais également l'initiation d'actions de partenariats pour ce diplôme (obtention de la labellisation par le pôle de compétitivité Images et réseaux, convention INRIA, mise en place de partenariats internationaux...), la prise en compte de la diversité des statuts des établissements impliqués, ou encore la gestion du site web...

- Autres responsabilités liées au Master 2 recherche en informatique : de 06 à 11, également responsable de l'antenne de l'INSA de Rennes pour ce M2R ; de 04 à 06, responsable adjointe de ce M2R (20%) ; de 99 à 08, co-responsable de la spécialité Intelligence Artificielle et Image de ce M2R (90%).
- De février 07 à juin 17, j'ai été **co-responsable (50%) du parcours Traitement et modélisation de masses de données** en 3^e, 4^e et 5^e années du département informatique de l'INSA (cf. ci-dessus). Ce parcours s'est arrêté en 3 ans lors de la mise en place progressive de la nouvelle maquette d'enseignement du département.
- De 90 à 93, j'ai été **chargée des études en 2^e année du département STID** de l'IUT de Vannes. Cette tâche consistait, entre autres, à coordonner les enseignements, à élaborer, chaque semaine, les emplois du temps, à gérer les notes ainsi que l'organisation des jurys et des contrôles.

2.4. *Diffusion de la culture à travers le développement de la culture scientifique, technique, industrielle*

- Deux articles de médiation dans la revue *Interface, La revue INSA ALUMNI*, N° 136, 4^e trimestre 19
 - Entretien avec deux enseignantes autour de l'intelligence artificielle, Peggy Cellier et Pascale Sébillot, pages 30-31 (2 colonnes) ; entretien, relecture et correction de la version initiale proposée
 - Enseignement et pratique de l'intelligence artificielle au département informatique de l'INSA Rennes, Peggy Cellier, Jean-Louis Pazat et Pascale Sébillot, page 113, 2 colonnes ; co-rédaction

2.5. *Rayonnement et activités internationales*

- De 05 à 06, j'ai été co-responsable (50%) de la gestion des aspects internationaux du Master 2 recherche en informatique pour l'INSA.
- De 91 à 01, j'ai **mis en place** et assuré la **responsabilité** pour Rennes 1 de la **coopération Erasmus** entre le Trinity College de Dublin, l'Université de Rennes 2 et l'Université de Rennes 1.

3. Activité scientifique

3.1. *Présentation des thématiques de recherche*

Résumé rapide de mes thématiques

Mes activités de recherche se situent dans le domaine du traitement automatique des langues (TAL). Après une thèse portant sur des aspects syntaxiques de la langue, mes centres d'intérêt se sont orientés vers la sémantique, en particulier sur l'acquisition d'éléments de sémantique lexicale en corpus par des méthodes d'apprentissage automatique. La création de l'équipe TexMex de l'IRISA (techniques d'exploitation des documents multimédias) en 2002, regroupant des spécialistes des différentes modalités (texte, parole, image, vidéo), de bases de données et de statistiques, a été pour moi l'occasion d'orienter mes activités vers l'accès au contenu sémantique de grandes collections de documents multimédias. Dans le cadre des thèses que je co-encadre depuis 2004 avec G. Gravier (au sein de TexMex puis de son prolongement LinkMedia en 2014), nous exploitons, en tirant profit de nos compétences initiales respectives en TAL et en reconnaissance automatique de la parole (RAP), les textes et la transcription automatique de la parole contenue dans les flux télévisuels et radiophoniques pour accéder à la sémantique des documents. Outre nos premiers travaux communs sur l'amélioration de la qualité de la transcription à l'aide de TAL, nos recherches portent d'une part sur la structuration de documents et de collections de documents multimédias et, d'autre part, sur le couplage entre contenus multimédias et bases de connaissance (structurées) pour co-enrichir les deux pans et permettre, dans le cadre d'applications de type journalisme de données (*data journalism*) ou vérification des faits (*fact checking*), de tirer conjointement profit des informations présentes dans les documents non structurés et les bases de connaissance ou ontologies. **Je présente ci-dessous les sujets et résultats les plus marquants de ces recherches au cours des quatre dernières années.**

Structuration de collections de documents multimédias : dans le cadre du projet ANR ARCHIVAL, nous nous intéressons (thèse de D. H. Nguyen, débutée en sept. 20) à l'établissement, par des techniques d'apprentissage, de liens sémantiques intra et inter-documents au sein d'archives multimédias, et, plus

particulièrement, à l'exploration des possibilités offertes par les mécanismes d'attention pour expliquer les relations établies. Si plusieurs études empiriques laissent entendre, à partir de l'analyse d'exemples, que les cartes d'attention offertes par ces mécanismes peuvent fournir une justification des décisions prises, seules quelques-unes s'intéressent à l'évaluation de la *plausibilité* des explications fondées sur ces cartes d'attention, c'est-à-dire l'utilité, pour des humains, des cartes pour leur permettre de comprendre les décisions. Nous avons étudié les facteurs qui influencent la plausibilité d'un mécanisme d'attention entre 2 encodeurs récurrents LSTM dans une tâche d'inférence en langue naturelle, où 2 phrases sont comparées pour décider si elles sont en relation d'implication, de contradiction ou sans relation, tâche proche de notre problématique d'explication de liens établis entre fragments de documents. Nous avons comparé, sur le corpus e-SNLI, les poids d'attention croisée entre les deux encodeurs avec des annotations fournies par des humains et par une heuristique fondée sur la similarité sémantique entre les mots des 2 phrases et se focalisant sur certaines catégories morphosyntaxiques. Nous avons montré que l'heuristique présente une corrélation raisonnable avec les annotations humaines, contrairement au modèle initial qui conduit à une diffusion des poids sur l'ensemble des *tokens* d'entrée. Nous avons proposé des règles simples de filtrage (masques issus d'une analyse morphosyntaxique) et de régularisation par minimisation de l'entropie pour forcer la parcimonie, aboutissant à une amélioration sensible de la plausibilité des poids d'attention. Nous avons étendu cette étude initiale en proposant deux approches qui intègrent dans le critère d'apprentissage la notion de plausibilité. La première, entièrement supervisée, s'appuie sur l'annotation fournie par les experts afin de biaiser les poids d'attention vers des valeurs plausibles par rapport à l'explication attendue ; la seconde, semi-supervisée, génère automatiquement l'annotation utilisée pour superviser l'apprentissage par une heuristique qui considère les noms, verbes et adjectifs comme pertinents. Nous avons comparé ces deux approches supervisées à la régularisation des poids d'attention par un critère d'entropie sur trois corpus : e-SNLI et deux corpus de classification de textes exploités par des bi-LSTM hiérarchiques : HateXplain (détection de messages haineux) et Yelp (classification en polarité). Sur e-SNLI, les résultats expérimentaux montrent que si les trois méthodes permettent d'accroître la plausibilité des poids d'attention sans sacrifier la performance en prédiction, elles favorisent des comportements différents dans la carte d'attention (précision favorisée par la régularisation, rappel par la supervision). Sur les deux autres corpus, les trois techniques permettent aussi d'améliorer la plausibilité dans le cadre d'un classifieur bi-LSTM hiérarchique ou non. Ces gains sont toutefois limités et plafonnés en fonction du nombre de couches de bi-LSTM utilisées pour contextualiser la représentation des *tokens* avant la couche d'attention.

Couplage contenus et bases de connaissance : initiés dans le cadre de l'Inria Project Lab iCODA, certains de nos travaux actuels portent sur l'exploitation conjointe de contenus multimédias et d'ontologies ou bases de connaissance dans des buts d'*analytics*. Notre cas d'étude-cible est celui du journalisme des données. Nous avons débuté cette exploitation conjointe de données et de contenus en proposant une méthode de liage collectif de mentions d'entités présentes dans des contenus textuels à des entités d'une base de connaissance tirant profit de la structure de la base RDF pour désambiguïser les entités candidates (thèse de C. B. El Vaigh, soutenue début 2021). Plus précisément, nous avons proposé une mesure de relation (*relatedness*) entre entités qui prend en compte les chemins de longueur supérieure ou égale à un dans la base de connaissance. Nous avons également mis au point une technique d'alignement des entités présentes dans deux bases de connaissance fondée sur l'apprentissage d'un modèle conjoint d'*embedding* des bases dans un même espace. Notre solution repose sur un critère qui, au lieu de chercher à minimiser explicitement la distance entre les entités des deux bases à aligner, fournies en amorce (*seed*), cherche à maximiser globalement la corrélation entre les dimensions de l'*embedding* des entités alignées de l'amorce, diminuant ainsi indirectement la distance entre elles. Cette prise en compte globale de l'amorce permet une meilleure résistance au bruit qu'elle peut contenir. Nous avons montré que cette robustesse accrue autorise le remplacement de la création manuelle de l'amorce par une technique automatique de génération.

La thèse de C. Mallart (Cifre Ouest-France, soutenue en novembre 2022) s'est, elle, intéressée au problème de l'enrichissement mutuel des contenus et des bases de connaissance, dans le contexte réel d'une entreprise de presse principalement dédiée à l'information locale. Elle explore en particulier l'extraction de relations des textes journalistiques pour nourrir les bases de connaissance. Au vu des phrases des corpus manipulés dans lesquelles la proportion de couples d'entités effectivement liées est faible par rapport au nombre de cooccurrences d'entités, un travail initial a porté sur la mise en place d'une phase de détection dans les phrases de la présence ou non d'une relation entre deux entités, en amont de la classification de la relation.

Cette détection efficace, fondée sur l'analyse par un réseau LSTM du plus court chemin de dépendances entre les entités combinée à leurs types, a aussi montré sa capacité à améliorer la classification des relations. Dans notre contexte journalistique dans lequel, d'une part, le temps, la puissance de calcul et la disponibilité d'annotateurs humains sont limités, et, d'autre part, les relations à extraire sont présentes dans des phrases où de nombreuses entités sont non liées, sont de classes variées et spécifiques, potentiellement très mal équilibrées, et pour lesquelles nous ne disposons que de peu ou d'aucune(s) donnée(s) annotée(s), l'utilisabilité des méthodes d'extraction de relations état de l'art est questionnable. Pour nous adapter à ces contraintes et offrir des solutions utilisables, nous avons proposé un *pipeline* d'extraction de relations fondé sur de l'apprentissage actif, composé du modèle léger précédent de détection des relations réelles et d'un modèle de classification des relations. Nous avons comparé plusieurs choix de modèles de classification dans ce scénario, d'un basique prenant en entrée la moyenne des *embeddings* des mots des phrases considérées à des réseaux de neurones graphiques (C-GCN) ou fondés sur BERT ; et nous avons exploré plusieurs stratégies de choix d'exemples à annoter pour l'apprentissage actif, afin de trouver l'approche la plus rentable et la plus précise. Nous avons montré que les approches légères sont un choix plus sûr dans ce contexte particulier.

Nous continuons à étudier la façon dont le TAL permet d'enrichir une base de connaissance, dans un contexte de frugalité nécessaire des modèles utilisés pour extraire les relations des textes, en nous intéressant (thèse de H. Thomas, débutée en octobre 22) à l'étude de l'apport des techniques d'apprentissage frugales en données (*zero shot et few shot learning*) à la détection et classification de relations. Une étude préliminaire portant sur une analyse approfondie des modélisations de relations de plusieurs architectures de modèles de typage de relations, et visant, par la définition de critères variés, à mieux comprendre l'organisation de l'espace latent des modélisations et ses propriétés, a donné lieu à un article accepté à la conférence francophone TALN 2023.

3.2. Publications

- Cf. Annexes page 21

3.3. Encadrement et animation recherche

Encadrements (voir détails en annexe)

Thèses dirigées et soutenues	Thèses dirigées et en cours	Post-doctorats	M2 recherche
17	2	1	27

Animation de la recherche

- 04-08 : partenaire du **réseau d'excellence européen MUSCLE** (*Multimedia Understanding through Semantics, Computation and Learning*) qui se focalisait sur l'accès au contenu de documents multimédias en regroupant des personnes venant des domaines de l'apprentissage artificiel et de l'accès sémantique aux contenus (Communauté européenne, 203 000 €).
- **Membre du bureau du GDR TAL depuis 19, et co-responsable d'un groupe de travail de 19 à déc. 20** : suite à une phase de pré-GDR, le CNRS a acté en janvier 20 la création d'un GDR dédié au traitement automatique des langues (journée inaugurale du pré-GDR : 14/03/19). Pendant l'année 18, j'ai été membre du comité des animateurs des (8) axes de réflexion du pré-GDR, plus particulièrement en charge, avec B. Favre d'Aix-Marseille Univ., de l'axe Intermodalité et multimodalité. Nous avons contacté une trentaine de personnes pour parvenir à la rédaction d'une synthèse sur cette thématique, présentée au comité scientifique et aux animateurs des autres axes en juin 18. Cette présentation a conduit à la création d'un **groupe de travail du GDR** sur cette thématique **Intermodalité et multimodalité**, groupe dont j'ai assuré la **co-responsabilité de 19 à déc. 20**. Les orientations de ce groupe de travail ont été présentées, illustrées par un exposé invité et discutées au cours d'une journée commune à deux autres groupes en octobre 19 à Grenoble. Nous avons également coorganisé, avec un autre groupe de travail du GDR, une journée « TAL et sciences cognitives » à Marseille en décembre 2020 (distanciel). Je suis **également membre du bureau du GDR depuis 19**.

- **Membre du comité permanent (CPerm) de la conférence TALN de juin 13 à septembre 20** (cooptée de 13 à 17 et nommée de 17 à 20). Le CPerm est en charge du suivi de l'organisation de la conférence annuelle, la sélection des sites, la nomination des présidents des comités, etc. Composition : 3 membres cooptés, les organisateurs de TALN de l'année en cours, de l'année précédente et de l'année suivante, et 2 membres du CA de l'ATALA.
- 03-09 : membre du **conseil d'administration** de l'Association pour le traitement automatique des langues (ATALA), élue en 2003, réélue en 2006.

Organisations de conférences, journées d'études

- *Audience internationale - conférences*
 - JADT'2002 (Journées internationales d'analyse statistique des données textuelles), Saint-Malo, mars 2002 ; 177 participants (co-responsable organisation 50%).
 - LACL'2001 (*Logical Aspects of Computational Linguistics*), Le Croizic, juin 2001 (membre du comité d'organisation).
- *Audience nationale - conférences*
 - **CORIA-TALN-RECITAL 2018** (Conférence en recherche d'information et applications, Traitement automatique des langues naturelles, Rencontre des étudiants chercheurs en informatique pour le traitement automatique des langues), Rennes, 14-18 mai 2018 ; 250 participants ; membre du comité d'organisation de l'ensemble des 3 conférences et **co-président** (avec V. Claveau) **du comité de programme de TALN**
 - **TALN 2015**, Caen, juin 2015 ; 180 participants ; **co-président par intérim du comité de programme** avec Emmanuel Morin (U. Nantes) et Philippe Blache (Aix-Marseille U.) : suite à un problème de santé des deux personnes en charge du comité de programme, nous (en tant que membres cooptés du CPerm de TALN) avons dû prendre en charge le comité de programme
 - TALN'2003 et RECITAL'2003, Batz sur mer, juin 2003 ; 160 + 30 participants ; membre du comité d'organisation.
- *Audience nationale – journées d'études*
 - **Journée du GDR TAL 2022** sur le thème de « l'accès à l'information », Rennes, 19 octobre 2022 ; membre du comité d'organisation (aspects scientifiques et organisationnels), 80 participants
 - ETAPE 2012 (workshop de la campagne d'évaluation des systèmes de reconnaissance automatique de la parole), 18-19 juin 2012 ; 35 participants ; membre du comité d'organisation.

Contrats et relations académiques et industrielles

Pour les contrats les plus récents, les montants arrondis sont fournis.

Responsabilités de contrats

Contrats non industriels

- ARED Région Bretagne (1/2 bourse) et AI4SDA (1/2 bourse du programme doctoral *AI for semantic data analytics* de l'ANR), 1/10/2022 - 30/09/2025 servant à financer la thèse de Hugo Thomas, co-encadrée avec Guillaume Gravier : Détection de relations à partir de peu ou pas d'exemples dans les archives de presse. **Co-porteur à 50%** (avec G. Gravier Irisa/Cnrs)
- Projet Labex CominLabs LIMAH (*Linking media in acceptable hypergraphs*), 03/2014 - 12/2018, impliquant l'équipe TexMex de l'IRISA, PREFics (sociologie) et le CRPCC (psychologie cognitive) de Rennes 2, IODE (droit) de Rennes 1, et l'équipe TALN du LINA (Nantes). Projet, visant l'organisation de collections de documents multimédias sous forme d'hypergraphes (où les liens reflètent une proximité sémantique entre (fragments) de documents), ainsi que l'évaluation de l'impact de ces collections structurées sur les usages et la technologie (542 300 €). **Co-porteur à 50%** (avec G. Gravier Irisa/Cnrs)
- ARED NAVITEL - Structuration sémantique de flux multimédia, 2007-2010, allocation de la région Bretagne ayant financé pour moitié la thèse de C. Guinaudeau (44 750 €). **Co-porteur à 50%** (avec G. Gravier Irisa/Cnrs)

- De 05 à fin 06, porteur d'un accord de coopération entre les universités de Montréal et de Rennes 1, visant à combiner des travaux développés au sein de TexMex et de l'Observatoire linguistique sens-texte pour détecter automatiquement des relations paradigmatiques dans des corpus (12 000 €).
- De mars 99 à mai 01, porteur d'une action de recherche partagée de l'Agence Universitaire de la Francophonie (réseau Francil), intitulée « Acquisition automatique d'éléments du Lexique Génératif pour améliorer les performances de systèmes de recherche d'informations » et regroupant l'ISSCO de Genève, l'Université libre de Bruxelles et l'ERSS de Toulouse.
- De 95 à 97, porteur d'un programme de recherches avancées de l'Association Franco-Chinoise pour la Recherche Scientifique et Technique pour développer avec l'Institut de Mathématiques de l'Académie des Sciences de Beijing une interface de prise en compte de réponses d'élèves éventuellement agrammaticales pour un tuteur intelligent.
- De 94 à 96, porteur d'un contrat CNRS de collaboration avec le laboratoire d'informatique IICP (Institute of Informatics and Control Problems) de l'Académie des Sciences d'Almaty (Kazakhstan), portant sur l'analyse du Kazakh.

Contrats industriels

- Cifre Ouest-France, 3/12/2018 - 2/12/2021, servant à financer la thèse de Cyrielle Mallart co-encadrée avec Guillaume Gravier et Michel Le Nouy (Ouest-France) : Construction incrémentale dynamique de graphes de connaissance par fouille de contenus (45 000 €). **Co-porteur à 50%** (avec G. Gravier Irisa/Cnrs)
- Cifre INA, 18/04/2011 - 17/04/2014, servant à financer la thèse de Ludivine Kuznik co-encadrée avec Guillaume Gravier et Jean Carrive (INA) : Structuration et navigation dans des archives documentaires. Suite à de graves problèmes de santé, Ludivine a dû arrêter sa thèse (15 000 € HT). **Co-porteur à 50%** (avec G. Gravier Irisa/Cnrs)
- Cifre Orange Labs Rennes, 01/10/2009 - 30/09/2012, servant à financer la thèse de K. Elagouni : *Combining neural-based approaches and linguistic knowledge for text recognition in multimedia documents* (24 000 € HT). **Porteur à 100%**
- De février à octobre 02, co-porteur d'un contrat avec Thalès-Communications, intitulé « Analyse des caractéristiques d'un auditoire en vue de la conception d'un logiciel d'argumentation », réalisé avec P. Besnard (IRIT Toulouse) (6 300 €).
- De février à juillet 00, co-porteur d'un contrat avec Thomson-CSF, intitulé « Analyse linguistique pour la conception d'un logiciel d'argumentation », réalisé avec P. Besnard (IRIT Toulouse).
- De 93 à 96, porteur d'un contrat CNET (maintenant Orange Labs) (CTI) portant sur la détermination de règles d'interprétation des composés anglais et français et leur intégration au système d'indexation des services télématiques du CNET. Ce contrat a financé la thèse de C. Fabre.

Autres relations nationales et internationales (dont participations à des contrats)

Partenariat au sein d'actions et de contrats internationaux

- De 2018 à 2021, partenaire du projet READ-IT (*Reading Europe advanced data investigation tools*) JPI Cultural Heritage regroupant des partenaires TIC et SHS de 4 pays européens. Le projet s'intéresse au repérage et à l'analyse des situations de lecture dans des corpus multimédias du 18^e au 21^e siècle (porteur B. Ouvry-Vial (Univ. Le Mans), global : 831 000 € ; pour l'équipe : 129 000 €). Travail sur le développement d'outils de TAL permettant de faciliter le repérage des situations de lecture dans les textes, ainsi que l'annotation de corpus pour le développement de méthodes d'apprentissage ; co-encadrement d'un ingénieur.
- De 08 à 13, partenaire de **QUAERO**, vaste programme de recherche et projet applicatif franco-allemand portant sur le domaine de la description multimédia et des moteurs de recherche. Travail personnel focalisé sur l'acquisition de terminologie et de relations sémantiques à partir de corpus pour l'annotation de matériau textuel, et sur la structuration automatique de flux multimédias (Oseo, 980 000 €).
- En 08, partenaire de l'équipe associée INRIA *Automatic Structuring of TV Streams* associant TexMex à l'Université de Nagoya et au National Institute of Informatics de Tokyo (Japon). Travail sur l'exploitation conjointe de divers indices multimédias issus des flux télévisuels à des fins de structuration, segmentation, classification et suivi de sujets (20 000 €).

Partenariat au sein d'actions et de contrats nationaux

- De 2019 à 2023, partenaire du projet ANR ARCHIVAL regroupant 4 équipes de chercheurs en sciences de l'information et de la communication et en informatique autour de la valorisation des archives et du partage des connaissances pour les arts, la culture et le patrimoine (global : 736 000 € ; pour l'équipe : 200 500 €). Travail sur l'établissement de liens sémantiques entre parties de documents, et explication de ces liens en exploitant des cartes d'attention (co-encadrement de la thèse de D. H. Nguyen).
- De 2022 à 2023, partenaire du programme de recherche Le Télégramme-IRISA/CNRS regroupant des membres de l'équipe de rédaction du quotidien et de l'équipe Linkmedia de l'IRISA autour d'une problématique de journalisme augmenté (84 000 €). Co-encadrement d'un ingénieur chargé de mettre au jour des critères qui font qu'un article est lu, de proposer, suite à ces analyses, des angles de rédaction de sujets pour rendre les articles attractifs, etc.
- De 2017 à 2022, partenaire du projet Inria Project Lab (IPL) iCODA impliquant 4 équipes Inria et des partenaires du monde des médias ; ce projet porte sur l'inférence d'informations et de connaissance par inspection conjointe de sources de données hétérogènes (données structurées et contenus non structurés) (global : 500 000 € ; équipe : 150 000 €). Travail sur l'établissement de liens entre mentions d'entités dans des textes et entités de bases de connaissance, et alignement de bases de connaissance (co-encadrement de la thèse de C. B. El Vaigh).
- De 08 à 09, partenaire de l'action de recherche coopérative INRIA Rapsodis regroupant des membres des équipes TexMex et Metiss de l'IRISA, Parole et Talaris du LORIA et d'une équipe du CEA-LIST/LIC2M, et ayant pour but d'étudier l'introduction d'informations syntaxiques et sémantiques au sein d'un système de reconnaissance de la parole (39 000 €).
- De 04 à 07, partenaire de l'Action concertée incitative Masse de données Demi-ton regroupant des membres des équipes TexMex et Metiss de l'IRISA et de l'INA, et ayant pour objet la description multimodale pour une structuration automatique des flux de télévision (Ministère de la recherche, 79 000 €).
- De 00 à 03, partenaire du projet Caderige (Catégorisation automatique de documents pour l'extraction de réseaux d'interactions géniques), action inter-EPST bio-informatique CNRS, INSERM, INRA, INRIA, Ministère de la Recherche qui regroupait des membres de l'IMAG, du LIPN, du LRI, de l'IRISA et de deux laboratoires INRA : MIG et INRA-ENSAR. Son objectif était de filtrer, dans des bases textuelles de bioinformatique, les textes parlant spécifiquement d'interactions géniques, et d'extraire de ces textes des réseaux de telles interactions (40 000 €).

Autres collaborations internationales

- En plus des collaborations internationales mentionnées dans les contrats cités ci-dessus, **collaborations non contractualisées**
 - avec l'équipe *Language intelligence and information retrieval* de M.-F. Moens à **KU Leuven**. Débutée par l'invitation de M.-F. Moens à participer au jury de la thèse de G. Lecorvé fin 2010, cette collaboration s'est poursuivie avec ma participation aux deux soutenances (privée et publique) de celle de son doctorant G.-J. Poulisse en 2012. Elle s'est intensifiée avec le séjour (avril-juin 2014) d'une de mes doctorantes à Leuven, notre co-participation à la tâche *Search and Hyperlinking* de MediaEval 2014 et la co-rédaction d'un article à SLAM 2015. Nos travaux communs portent principalement sur la segmentation thématique hiérarchique ;
 - avec l'*Insight Centre for Data Analytics* de **Dublin City University**. Invitation d'un de mes doctorants pendant 3 mois à l'automne 2018 par Alan Smeaton pour travailler sur l'évaluation de résumés textuels et multimédias.

3.4. Valorisation de la recherche

Logiciels

Logiciels déposés à l'Agence pour la Protection des Programmes, créés dans le cadre de thèses dirigées

- *Asares* : logiciel d'apprentissage automatique de patrons morpho-syntaxiques et sémantiques d'extraction sur corpus de couples de mots suivant des relations sémantiques données (n°IDDDN.FR.001.0032.000.S.C.2005.000.20900)

- *Faestos (Fully automatic extraction of sets of keywords for topic characterization and spotting)* : logiciel de détection et caractérisation de thèmes n°IDDDN.FR.001.470029.000.S.P.2006.000.40000)
- *2Pac* : logiciel de constitution automatique de classes sémantiques (n°IDDDN.FR.001.470028.000.S.P.2006.000.40000)

Autres logiciels

- IRISA News Topic Segmenter (*irints*) : logiciel dédié à la segmentation thématique non supervisée de textes, particulièrement adapté pour traiter des transcriptions automatiques. Sa licence a été transférée à plusieurs de nos partenaires industriels dont Exalead qui l'a intégré à son démonstrateur Voxlead News. Voxlead a été primé à ACM Multimedia 2010 et par le réseau d'excellence META-NET ("for outstanding contribution to a multilingual Europe") lors du META-FORUM qui s'est tenu à Budapest en juin 2011
- Services Web Otis ! et Hi-toSt, disponibles sur la plate-forme allgo d'Inria, issus respectivement des travaux des thèses de Camille Guinaudeau et de Anca Roxana Simon que j'ai co-encadrées.

Participations à des campagnes d'évaluation internationales

- Participation à la tâche *HIPE (Named Entity Processing on Historical Newspapers)* proposée dans le cadre de CLEF 2020.
- Participation à la tâche *Video Hyperlinking* proposée dans le cadre de TRECVID Challenge 2015 et 2016. Notre système a terminé 1^{er} en 2016.
- Participation à la tâche *Search and Hyperlinking* proposée dans le cadre de MediaEval 2012, 13, 14 et 15. Participation interne à l'équipe en 12 et 15, en partenariat avec le *Heidelberg Institute of Technology* où C. Guinaudeau était en postdoctorat en 13, et en partenariat avec *KU Leuven* en 14. Selon les conditions d'évaluation considérées, notre système a terminé 1^{er} ou 2^e en 2012. Les évaluations par *crowd-sourcing* très partielles les 3 années suivantes rendent tout classement dénué de sens.

3.5. Rayonnement

Jurys de thèses (hors thèses encadrées)

Année	Rapporteur : 21	Expert internat. : 1	Président : 14	Examineur : 10
2023	G. Le Berre (Montréal et Lorraine)			S. Benslimane (Montpellier)
2022	Q. Portes (Toulouse)		B. Fabre (Rennes)	L.G. Moreno Jimenez (Avignon), É. Simon (Sorbonne Univ.)
2021	A. Behre (Paris-Sud)		A. Perquin (Rennes)	
2018			C. Langlet (Telecom ParisTech)	A. Bérard (Grenoble)
2017	A. Ollagnier (Toulouse), D. Valsamou (Paris-Sud)		J. Lark (Nantes)	Y. Dupont (Sorbonne Paris Cité)
2016	J.-P. Fauconnier (Toulouse)		A. Boucekif (Le Mans)	S. Ben Nasr (Rennes)
2015	S. Bannour (Paris-Nord), M. Asadullah (Paris-Sud)		V. Lelli Leitao Dantas (Rennes), B. Qu (Rennes), H. Boukadida (Rennes)	
2014			M. Hatmi (Nantes), G. Ke (Vannes)	
2013	W. Wang (Paris-Sud), P. Shresta (Nantes), F. Lalleman (Toulouse), N. Foucault (Paris-Sud)		A. Boutet (Rennes)	
2012	C. Adam (Toulouse)	G.-J. Poulisse (KU Leuven)		
2010	S. Pena Saldarriaga (Nantes), A.			

	Zidouni (Aix-Marseille)			
2009	K. Séjourné (Orsay)		S. Ayed (Rennes)	
2008	S. Léon (Aix-Marseille)			
2006			O. Martineau (Rennes)	F. Venant (Paris)
2005	M. Ould Abdel Vetah (Orsay)		R. Kulpa (Rennes)	I. Jacquemin (Rennes)
2004	H. Cherfi (Nancy), N. Fourour (Nantes)			M. Roche (Orsay), V. Perlerin (Caen)
2003	L. Audibert (Aix-Marseille)			

Jurys d'HDR

Année	Rapporteur : 6	Président	Examineur : 4
2023			C. Grouin (Paris-Saclay)
2021	M. Filhol (Paris-Sud)		
2020	R. Dufour (Avignon)		P. Cellier (Rennes)
2019	N. Grabar (Paris-Sud)		
2016			O. Ferret (Paris-Sud)
2014	<i>X. Tannier</i> (Paris-Sud)		
2013	P. Beust (Caen)		
2012			L. Tanguy (Toulouse)
2009	J.-P. Chevallet (Grenoble)		

Dans les 2 tableaux ci-dessus, les noms en italiques signalent les jurys où j'ai également tenu le rôle de président en plus de rapporteur (non comptabilisés en présidence).

Comités de rédaction et de lecture de revues

Comités de rédaction

- **Rédacteur en chef de la revue TAL** (traitement automatique des langues) **de juin 2014 à juin 2023** (4 rédacteurs en chef) : malgré son nom français et souvent considérée comme *la* revue francophone du domaine, TAL publie aussi des articles d'auteurs non francophones en anglais. Les rédacteurs en chef se répartissent le suivi des 3 numéros du volume annuel de la revue. J'ai ainsi pour ma part été, jusqu'à présent, en charge des numéros 63-2, 62-1, 60-3, 59-3, 58-2, 58-1 (co-resp.), 56-3, 55-3 et 55-2 (co-resp.). Le comité de rédaction de la revue se réunit en moyenne 6 fois par an (1/2 à 1 journée de réunion), réunions animées par les rédacteurs en chef.
- **Membre du comité de rédaction de la revue TAL depuis juillet 2009**

Comités de lecture

- Au titre de membre du CR de *TAL*, membre du comité de lecture de (en moyenne) 3 articles de la revue par an depuis 2009
- Numéro spécial de *Document numérique* « Le texte : objet d'analyse et vecteur de connaissances », 2010
- Numéro spécial de *Terminology* « Pattern-Based Approaches to Semantic Relation Extraction » 14(1), 2008
- Numéro 8 de *Glottopol* « Traitements automatisés des corpus spécialisés : contextes et sens », 2006
- Numéros spéciaux de *TAL* « Lexiques sémantiques » et « Linguistique de corpus », 2001
- Numéro spécial de *TAL* « Traitement automatique des langues pour la recherche d'information », 1999

Comités de programme

Conférences d'audience internationale

- IJCAI (International Joint Conferences on Artificial Intelligence) : 2020, 2019, 2018, 2017
- AAAI (AAAI Conference on Artificial Intelligence) : 2018
- ACL (Annual Meeting of the Association for Computational Linguistics) : 2019, 2017
- NAACL (Annual Conference of the North American Chapter of the Association for Computational

Linguistics) : 2018

- ECIR (European Conference on Information Retrieval) : 2020, 2019, 2018, 2017, 2016
- CIKM (ACM International Conference on Information and Knowledge Management) : 2016
- MMM (International Conference on MultiMedia Modeling) : 2019, 2017
- LREC (Language Resources and Evaluation Conference) : 2022, 2020, 2018, 2016, 2014, 2012, 2010, 2008
- ICMLA (International Conference on Machine Learning and Applications) : 2021
- JADT (Journées internationales d'analyse statistique des données textuelles) : 2010, 2008, 2006, 2004, 2002
- KES (Knowledge-Based and Intelligent Information & Engineering Systems) : 2008
- NACAP (North American Conference on Computing and Philosophy) : 2009
- GL (Generative Approaches to the Lexicon) : 2007, 2005, 2003, 2001
- LLL (Learning Language in Logic) : 2005

Conférences d'audience nationale

- TALN (Traitement automatique des langues naturelles) : 2022, 2021, 2020, 2019, 2017, 2016, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, 2005, 2003 (rappel : VP du CP de TALN en 2018)
- CNIA (Conférence nationale en intelligence artificielle) : 2023, 2016
- CORIA (Conférence en recherche d'information et applications) : 2011, 2010, 2005, 2004
- RFIA (Reconnaissance des formes et intelligence artificielle) : 2014, 2010, 2008, 2000
- TIA (Terminologie et intelligence artificielle) : 2019, 2015, 2013, 2011, 2009, 2007, 2005
- DEFT (Défi fouille de textes) : 2010, 2009, 2008, 2007, 2006, 2005
- CIFED (Colloque international francophone sur l'écrit et le document) : 2016, 2014

Autres fonctions et comités de programme de workshops internationaux ou nationaux

- CORIA (Conf. en recherche d'information et applications) : **membre du comité de pilotage** 2014
- SLAM (Workshop on Speech, Language and Audio in Multimedia; satellite de InterSpeech) : 2013
- SemDis, SemDis (Workshop sur la sémantique distributionnelle ; satellite de TALN) : 2014, 2013
- Workshop « Acquisition, apprentissage et exploitation de connaissances sémantiques pour l'accès au contenu textuel » de la plateforme AFIA : 2003
- Workshop « Description des adjectifs pour les traitements informatiques » TALN'99 : 1999
- Compound Nouns: Multilingual Aspects of Nominal Composition : 1994

Invitation, séjours dans des universités et laboratoires à l'étranger

- Mai 2008 : deux semaines à l'Université de Nagoya et au NII (National institute of informatics) Tokyo
- Janvier 2007, juillet 2005, 2004 : respectivement sept, dix et dix jours à l'OLST (Observatoire de linguistique Sens-Texte) de l'Université de Montréal
- Été 1996 : un mois à l'Institut de Mathématiques de l'Académie des Sciences de Beijing
- Mai 1994 : neuf jours au laboratoire d'informatique IICP (Institute of Informatics and Control Problems) de l'Académie des Sciences d'Almaty, Kazakhstan

Séminaires, conférences invitées ou tutoriels invités

Internationaux (titres pour les plus récents uniquement)

- « *Information retrieval and natural language processing* », National institute of informatics (NII), Tokyo, Japon, mai 2008
- « *Automatic speech recognition and natural language processing* », Université de Nagoya, Japon, mai 2008
- Séminaire OLST-RALI de l'Université de Montréal, juillet 2005 (conjoint avec S. Huet), juillet 2004, juillet 1987 et juillet 1986
- Journée CNRT TIM « Text Indexing for Multimedia », Rennes, avril 2002
- Institut de Mathématiques de l'Académie des Sciences de Beijing, août 1996

Nationaux (titres pour les plus récents uniquement)

- « Définition informatique de la donnée numérique et perception des contraintes de propriété

intellectuelle », intervention invitée au colloque Propriété intellectuelle et données dans l'environnement numérique, université de Rennes 1 (avec T. Allard, U. Rennes 1), octobre 2017

- « De la structuration de documents multimédias à la structuration de collections de documents multimédias », intervention invitée de 2h au master Industrie de la langue de l'Université Grenoble Alpes, novembre 2016
- « Introduction aux problématiques et aux solutions du TAL », intervention invitée de 8h à l'INIST (Institut de formation scientifique et technique du CNRS) de Nancy, à destination de personnels de l'INIST non familier du TAL, avril 2016
- « Introduction aux problématiques et aux solutions du TAL », intervention invitée de 6h à l'UFRIST (Unité Régionale de Formation à l'Information Scientifique et Technique) de Rennes à destination d'enseignants-chercheurs, doctorants et professionnels de l'information, Rennes, juin 2015
- « Le traitement automatique des langues face aux données textuelles volumineuses et potentiellement dégradées : qu'est-ce que cela change ? », intervention invitée de 3h dans le séminaire IST INRIA *Big data* : nouvelles partitions de l'information, à destination des professionnels de l'information Saint-Paul-lès-Dax, octobre 2014
- « Exploitation des données multimédias », exposé invité (conjoint avec P. Gros) à la journée *Big Data*, ENSAI de Rennes, novembre 2013
- « Topic analysis of written and oral corpora », Séminaire LORIA, Nancy, avril 2009
- Séminaire France Télécom R&D, Lannion, juillet 2006
- Séminaire du groupe Langues, information et représentations du Limsi, Orsay, avril 2005
- Journées d'étude de l'opération Sémantique et Corpus de l'ERSS La place des méthodes quantitatives dans le travail du linguiste, Toulouse, juin 2004
- Séminaire du département Informatique et télécommunications de l'ENS Cachan - antenne de Bretagne, Bruz, décembre 2003
- Conférence invitée lors de l'atelier « Acquisition, apprentissage et exploitation de connaissances sémantiques pour l'accès au contenu textuel » de la plateforme AFIA, Laval, juillet 2003
- Journées Inria-industrie La recherche d'information dans les réseaux, Rocquencourt, décembre 2001
- Journées Inria Aspects sémantiques du Web, Nancy, janvier 2001
- Séminaire de l'équipe Langue et Dialogue du Loria, Nancy, juillet 2000
- Journée ATALA Acquisition des Langues Assistée par Ordinateur, Paris, janvier 1995
- Journée ATALA Traitements syntaxiques et sémantiques des noms composés, Paris, avril 1993

Comités scientifiques

- De 03 à 07, membre du comité scientifique du programme interdisciplinaire du CNRS « traitement des connaissances, apprentissage et NTIC » (TCAN)

Expertises internationales et nationales

- Natural Sciences and Engineering Research Council of Canada (NSERC) 2015
- Fonds France-Canada pour la recherche 2013
- Information and Communication Technology - Vienna Science and Technology Fund 2012
- Fréquentes évaluations de projets ANR
- Institut Créativité et Innovations d'Aix-Marseille (InCIAM), Aix-Marseille Université : 2022
- Institut Carnot Cognition : 2020
- Labex Digicosme : 2017
- Réseau thématique de recherche avancée Digiteo 2014
- BQR Emploi Paris-Sud 2012
- Conseil régional des Pays de la Loire : 2010
- Évaluation de projets TCAN-CNRS : 2003-2007
- Évaluation de projets ACI-Masse de données : 2006
- Prix de recherche Atlanthèse (bourse de thèse) du conseil général de Loire-Atlantique : 2006

Expertises de projet de recherche

- Projet régional des Pays de la Loire DEPART (LINA, LIUM) : réunion annuelle d'une journée et rédaction d'un rapport d'évaluation annuel transmis par les 3 experts à la région, 2010-2014

4. Responsabilités collectives

4.1. *Présentation générale des responsabilités particulières*

Ayant décrit dans les sections recherche et enseignement mes principales responsabilités liées à ces deux pans du métier d'enseignant-chercheur, je me contente ici de mettre en évidence, dans les rubriques suivantes, mon engagement dans d'autres aspects de responsabilités collectives : responsabilités au sein de l'UMR IRISA (dont direction adjointe), de l'INSA, de l'école doctorale ; implication pendant 8 années au CNU 27 ; participation à des conseils et à de nombreux comités de sélection, etc.

4.2. *Responsabilités administratives*

- **Directrice-adjointe de l'UMR IRISA** depuis janvier 2021.
L'IRISA (Institut de Recherche en Informatique et Systèmes Aléatoires) est l'un des plus grands laboratoires de recherche français dans le domaine de l'informatique et des nouvelles technologies de l'information, implanté à Rennes, Lannion et Vannes, et bénéficiant de neuf établissements tutelles (par ordre alphabétique) : CentraleSupélec, CNRS, ENS Rennes, IMT Atlantique, Inria, INSA Rennes, Inserm, Université Bretagne Sud, Université de Rennes. Structuré en sept départements scientifiques et sept axes transverses, il regroupe actuellement 37 équipes de recherche et plus de 850 personnes. En tant que DU adjointe de l'IRISA, **outre les activités de direction de l'UMR partagées avec mes 2 collègues** (Guillaume Gravier, DU et Sandrine Blazy DU adjointe), j'ai une **mission particulière** consistant à être en charge des **relations avec les composantes et tutelles éducatives de l'UMR**. En sus des concertations pour définir une politique à long terme de recrutements et de profils de postes pertinents pour la stratégie de l'UMR et celle des établissements d'enseignement supérieur, cette mission concerne aussi la définition de moyens permettant d'attirer les étudiants vers la recherche, et d'accroître la porosité entre recherche et enseignement. Ceci implique par exemple de renforcer l'ouverture du laboratoire aux étudiants, sans se limiter au niveau master ou doctorat (visites du laboratoire, d'équipes particulières ; panels explicitant les divers métiers de la recherche ; hackathons...). Réciproquement, cela peut se décliner en invitant les chercheurs à s'investir dans des enseignements, et ce, même au niveau licence. Étudier la façon dont les nouveaux « outils » du PIA3 que sont les EUR (écoles universitaires de recherche) *Cyberschool* et *Digisport*, qui ont ouvert leurs portes à la rentrée 2020-2021, et l'IRISA peuvent s'enrichir mutuellement est également à traiter dans ce cadre. Enfin, on peut aussi citer l'accompagnement des enseignants-chercheurs de l'UMR pour leur donner les moyens de maintenir un équilibre entre enseignement et recherche. L'activité de direction adjointe d'une UMR de la taille de l'IRISA est difficile à quantifier, allant de plusieurs heures à plusieurs jours pleins d'occupation par semaine selon l'actualité.
- **Vice-présidente du comité d'orientation scientifique (COS) de l'IRISA** depuis septembre 2013 (nommée par la direction de l'UMR en 2013).
Le COS regroupe 9 à 10 professeurs ou directeurs de recherche de l'IRISA et est en charge de la prospective scientifique à moyen et long terme (10-20 ans) de l'UMR. Mon rôle consiste à organiser et animer le travail de l'instance, et à prévoir, mettre en place et animer les réunions. Au cours de la période concernée par ce document, j'ai, au cours de l'année 2019, réparti la quarantaine d'équipes de l'IRISA entre les membres du comité, chargés d'interroger plus particulièrement celles-ci sur leur vision à très long terme de leur domaine. Les discussions au sein du COS pour organiser en grandes orientations scientifiques les éléments recensés nous ont permis d'aboutir à un document rédigé, remis à la direction en janvier 2020. Celui-ci a servi pour la partie prospective du rapport d'évaluation HCERES de l'UMR début 2021. Parmi ses autres travaux, le COS a également lancé diverses réflexions et actions concernant le périmètre géographique du laboratoire, sa structuration en départements, les moyens de mise en place de thématiques transverses intra-IRISA ou

inter-laboratoires de notre écosystème, les modalités de politique scientifique en période de pénurie de création de postes, ou encore des réunions avec des directeurs de laboratoires de taille comparable à celle de l'IRISA sur tout ce qui a trait à la prospective. Ces réflexions et actions font ensuite l'objet d'une présentation auprès de la direction scientifique de l'UMR (direction + responsables des 7 départements).

- **Responsable de la composante INSA Rennes de l'IRISA** d'octobre 2016 à août 2021
L'UMR IRISA possède 9 établissements tutelles, dont l'INSA Rennes. J'ai été responsable de la composante regroupant les permanents et non permanents de l'IRISA dont l'INSA est l'employeur. Cette composante de recherche en informatique a la particularité d'être considérée, au sein de l'INSA et dans son fonctionnement par rapport à l'établissement, comme une entité à part entière (c.-à-d. un laboratoire), au même titre que les composantes de recherche d'autres disciplines (chimie, génie civil...) impliquant des membres INSA d'UMR ou non. Par conséquent et malgré une appellation qui peut prêter à confusion, j'ai été responsable à l'INSA pendant 5 ans d'une composante de recherche de 50 personnes (21 enseignants-chercheurs, deux BIATSS, et 25 à 30 doctorants), ayant un budget propre, 10 équipes de recherche (équipes IRISA dont font partie les enseignants-chercheurs de l'INSA), un **conseil que j'ai présidé**. Outre la gestion habituelle d'une telle structure (dialogue de gestion, budget, entretiens annuels, classement de demandes de contrats doctoraux, d'ARED, appels spécifiques à projets financés sur fonds de la composante...), cette tâche impliquait également la participation aux réunions *a minima* mensuelles du conseil réunissant la direction de la recherche et les responsables des diverses composantes de recherche de l'INSA Rennes, ainsi que la gestion des relations avec l'UMR IRISA.
- **Membre du comité des thèses de l'école doctorale MathSTIC** de juin 2015 à avril 2021
Cette tâche consiste à étudier l'ensemble des demandes d'inscription en thèse en informatique ainsi que l'ensemble des demandes de soutenances de thèse en informatique du site rennais. Plus précisément, il s'agit de contrôler la validité des directions et encadrements proposés pour les premières, et celle des rapporteurs et jurys proposés ainsi que la pertinence de la demande en fonction de l'avancée du manuscrit pour les secondes. Nous étions **deux** pour réaliser cette tâche, qui correspond en moyenne à **plus de 150 dossiers à traiter annuellement**.
- **Membre du conseil de l'ED MathSTIC** de 2017 à 2022 (3 réunions annuelles)

4.3. Responsabilités et mandats locaux ou régionaux

Implication locale

- Participations aux conseils (niveau établissement ; enseignement et recherche)
 - De mai 2018 à avril 2022 et de janvier 2011 à juin 2014, membre élu du Conseil scientifique de l'INSA de Rennes.
 - De février 1998 à avril 2002, membre du comité des projets de l'IRISA.
- Participations aux conseils (niveau plus « local » ; enseignement et recherche)
 - Depuis décembre 2021, et de janvier 2006 à novembre 2017, membre élu du Conseil du département informatique de l'INSA de Rennes.
 - De décembre 2011 à août 2021, membre du conseil de la composante INSA Rennes de l'IRISA. La composante (*cf.* ci-dessus) regroupe les enseignants-chercheurs, BIATSS et non permanents de l'UMR IRISA dont l'INSA est l'employeur. J'ai présidé ce conseil depuis ma nomination en tant que responsable de la composante en octobre 2016 jusqu'à mon arrêt de cette fonction en août 2021.
- Présidente du bureau de la section 27 de l'INSA Rennes de 2010 à décembre 2020
Le bureau est constitué de 2 PR et 2 MC. Le président est l'interlocuteur des RH pour tout ce qui a trait aux recrutements d'enseignants-chercheurs et d'ATER de sa section CNU. Concernant les concours de permanents, le rôle du bureau est de faire des propositions de noms pour constituer les comités de sélection pour tous les concours ; le président a la charge de contacter les membres pressentis, de finaliser la constitution des comités et de transmettre aux RH les listes à destination des conseils. Pour ce qui est des ATER, le bureau étudie les candidatures et propose un classement

des candidats. Enfin, pour tous les postes, le bureau participe à la réflexion sur les profils (voire à leur rédaction pour le président), en collaboration avec le directeur du département et le responsable de la composante de recherche. Sur la période de ma présidence, j'ai eu à gérer 3 concours ATER, 3 concours MC et 1 concours PR. Enfin, le bureau ou son seul président a à émettre des avis sur les dossiers de demande de titularisation des MC, ainsi que sur ceux de demande de promotion du directeur de département ou du responsable de la composante de recherche.

- Autres implications
 - Membre du jury de recrutement CRCN/ISFP (*Inria Starting Research Positions*) du Centre de recherche Inria Rennes-Bretagne Atlantique 2022
 - Membre du comité d'admission de la campagne de recrutement ISFP Inria (centre de Rennes Bretagne Atlantique) 2021, 2022 et 2023
 - Membre de la commission d'attribution des bourses de doctorats CORDI-S INRIA 2017, 2014
 - Membre de la commission d'attribution des bourses de post-doctorats INRIA 2014
 - Membre de la commission d'affectation des missions d'enseignement INSA de 2011 à 2013
 - Validation des acquis de l'expérience : de 07 à 12, membre de la commission VAE du département informatique de l'INSA ; de 93 à 00, membre de la commission VAE en Maîtrise d'informatique à l'Université de Rennes 1
 - Médiateur des doctorants de l'IRISA de 96 à mai 00. Cette tâche consiste à chercher à répondre aux interrogations d'un doctorant et à résoudre les diverses difficultés qu'elle ou qu'il peut rencontrer lors de sa thèse, en particulier celles liées à son encadrement
 - Membre du CHS (comité hygiène et sécurité) de l'IRISA de 01 à 11
 - Membre de la commission Communication de l'IRISA/Inria Rennes de mai 00 à avril 02

Implication régionale

- Depuis 2014 : membre du Conseil de la filière informatique nantaise, filière qui fédère sur son axe formation les 3 principaux acteurs de l'enseignement supérieur en informatique de Loire-Atlantique : Nantes Université, Centrale Nantes, IMT Atlantique et Mines Nantes. J'ai été sollicitée pour ce conseil de par mon expérience à fédérer l'informatique au niveau master 2 recherche en Bretagne (3 réunions annuelles).
- Depuis 2017 : membre du conseil de perfectionnement du master informatique de Nantes Université (1 réunion annuelle)

4.4. Responsabilités et mandats nationaux ou régionaux

- **Membre du CNU 27** : nommée de novembre 2011 à novembre 2015, élue de novembre 2015 à novembre 2019
 - à ce titre, évaluation du LIMSI pour l'AERES 10-12 décembre 2013, et évaluation du LIP6 pour l'HCERES 9-11 janvier 2018
- Membre de l'instance d'évaluation de la prime d'excellence scientifique (PES) en 2013

Comité de sélection et commission de spécialistes

- Responsabilités au sein de comités de sélection et commissions de spécialistes
 - Université de Caen : présidente d'un comité de sélection (MC) : 2020.
 - Université de Rennes 1 : présidente d'un comité de sélection (PR) : 2017, d'un comité de sélection (MC) : 2014.
 - INSA Rennes : présidente d'un comité de sélection (PR) : 2022, 2020, 2013 ; de deux comités de sélection (MC) : 2011 et 2010
 - de 94 à 04 : membre du bureau (assesseur) de la commission de spécialistes de la 27^e section de l'Université de Rennes 1

- Participation à des comités de sélection (hors présidences citées ci-dessus)

2023	Le Mans, Nantes, Lannion, INSA Rennes (27° et 32° sections)
2022	Grenoble, Nancy, INSA Rennes
2021	Aix-Marseille, Paris-Sud, ENS Rennes
2020	Le Mans, Paris-Sud, Vannes, INSA Rennes (28° section)
2019	Angers, Avignon, Toulouse, Vannes
2018	Paris-Sud
2017	Brest, Grenoble, Paris-Sud, Toulouse
2016	Lannion, Orléans
2015	Aix-Marseille, Orléans, Toulouse
2013	Lannion
2012	Nantes, Montpellier
2011	Avignon
2010	Avignon, Rennes 2
2009	Caen, INSA de Rennes, Rennes 2

- Participation à des comités pour des postes « repyramidage » (expertises ou auditions)

2023	Rennes 2
2022	Angers, Paris-Saclay, Rennes (section 61)

- Participation à des commissions de spécialistes

06-08	Rennes 2, INSA de Rennes
de 94 à fin 05	Rennes 1
de 95 à fin 05	Bretagne-Sud
de 96 à 97 + de 01 à 04	Nantes

Annexes

Les annexes contiennent successivement les éléments suivants

- liste classée des publications
- encadrements de thèses, de masters 2 recherche et de post-doctorats

Liste classée des publications

La plupart de mes publications sont disponibles sous HAL pour les plus récentes, ainsi qu'à l'URL <http://people.irisa.fr/Pascale.Sebillot/>. Les liens HAL sont fournis ci-dessous pour toutes les publications.

Dans l'ensemble des publications impliquant un doctorant et ses encadrants, notre politique est la suivante : le nom du doctorant est en première position, suivi par les noms des encadrants par ordre alphabétique (d'où ma position systématique en dernière position).

Ouvrages édités ou participations à de tels ouvrages : 9

1. Vincent Claveau, Pascale Sébillot. Actes de la conférence Traitement Automatique de la Langue Naturelle, TALN 2018 : Volume 1 : Articles longs, articles courts de TALN, Rennes, France. 2018.
2. Pascale Sébillot. Le traitement automatique des langues face aux données textuelles volumineuses et potentiellement dégradées : qu'est-ce que cela change ?. In L. Calderan, P. Laurent, H. Lowinger, and J. Millet, editors, *Big data : nouvelles partitions de l'information. Actes du séminaire IST INRIA, octobre 2014*, chapter 3, pages 43-60. De Boeck, 2015.
3. Thomas Gaillat, Pascale Sébillot, and Nicolas Ballier. Automated Classification of Unexpected Uses of This and That in a Learner Corpus of English. In L. Vandelanotte, K. Davidse, C. Gentens, and D. Kimps, editors, *Recent Advances in Corpus Linguistics - Developing and Exploiting Corpora*, pages 309-324. Rodopi, 2014.
4. Vincent Claveau and Pascale Sébillot. Automatic Acquisition of GL Resources, Using an Explanatory, Symbolic Technique. In J. Pustejovsky, P. Bouillon, H. Isahara, K. Kanzaki, and B. Lee, editors, *Advances in Generative Lexicon Theory*, chapter 19, pages 431-454. Springer, 2013.
5. Stéphane Huet, Gwénoél Lecorvé, Guillaume Gravier, and Pascale Sébillot. Toward the Integration of Natural Language Processing and Automatic Speech Recognition: Using Morpho-Syntax and Pragmatics for Transcription. In A. Potamianos, P. Maragos, and P. Gros, editors, *Multimodal Processing and Interaction: Audio, Video, Text*, chapter 9, pages 201-218. Springer Verlag, 2008.
6. Vincent Claveau, Fabienne Moreau, and Pascale Sébillot. Description des textes. In P. Gros, editor, *L'indexation multimédia - Description et recherche automatiques*, chapter 6, pages 163-190. Traité IC2, série Traitement du signal et de l'image, Hermes Lavoisier, 2007.
7. Béatrice Daille, Cécile Fabre, and Pascale Sébillot. Applications of Computational Morphology. In P. Boucher, editor, *Many Morphologies*, pages 210-234. Cascadilla Press, Somerville, 2002.
8. Pascale Sébillot. Traitement automatique des langues et recherche d'information. In *La recherche d'information sur les réseaux II*, chapter 4, pages 137-168. ADBS Éditions, 2002.
9. Patrick Saint-Dizier, Yannick Toussaint, Christophe Delaunay, and Pascale Sébillot. A Natural Language Processing System Based on the Government and Binding Theory. In P. Saint-Dizier and S. Szpakowicz, editors, *Logic and Logic Grammars for Language Processing*, chapter 6, pages 108-140. Ellis Horwood Series in Artificial Intelligence, 1990.

Articles publiés dans des revues d'audience internationale avec comité de lecture : 9

1. Khaoula Elagouni, Christophe Garcia, Franck Mamalet, and Pascale Sébillot. Text Recognition in Multimedia Documents: A Study of two Neural-based OCRs Using and Avoiding Character Segmentation. *IJDAR (International Journal on Document Analysis and Recognition)*, 17(1):19-31, 2 columns, 2014.
2. Camille Guinaudeau, Guillaume Gravier, and Pascale Sébillot. Enhancing Lexical Cohesion Measure with Confidence Measures, Semantic Relations and Language Model Interpolation for Multimedia Spoken Content Topic Segmentation. *CSL (Computer Speech and Language)*, 26(2):90-104, 2012. [CORE rang A]

3. Guillaume Gravier, Camille Guinaudeau, Gwéno   Lecorv  , and Pascale S  billot. Exploiting Speech for Automatic TV Delinearization: From Streams to Cross-Media Semantic Navigation. *EURASIP Journal on Image and Video Processing*, 2011, 17 pages, 2 columns, 2011.
4. St  phane Huet, Guillaume Gravier, and Pascale S  billot. Morpho-Syntactic Post-Processing with N-Best Lists for Improved French Automatic Speech Recognition. *CSL (Computer Speech and Language)*, 24(4):663-684, 2010. [CORE rang A]
5. Pascale S  billot. Symbolic Machine Learning: A Different Answer to the Problem of the Acquisition of Lexical Knowledge from Corpora. *TripleC (Cognition, Communication, Co-operation)*, special issue: selected papers from ECAP 2005 - European Computing and Philosophy Conference 2005, 4(2):277-283, 2006.
6. Mathias Rossignol and Pascale S  billot. Combining Statistical Data Analysis Techniques to Extract Topical Keyword Classes from Corpora. *IDA (Intelligent Data Analysis)*, 9(1):105-127, 2005. [CORE rang B]
7. Vincent Claveau, Pascale S  billot, C  cile Fabre, and Pierrette Bouillon. Learning Semantic Lexicons from a Part-of-Speech and Semantically Tagged Corpus Using Inductive Logic Programming. *JMLR (Journal of Machine Learning Research)*, Special Issue on Inductive Logic Programming, 4:493-525, 2003. [CORE rang A]
8. Fr  d  ric Danna and Pascale S  billot. A Formalization of Student Modeling. *CALL (Computer Assisted Language Learning)*, 10(2):121-147, 1997.
9. Paul Boucher, Fr  d  ric Danna, and Pascale S  billot. Compounds: An Intelligent Tutoring System for Learning to Use Compounds in English. *CALL (Computer Assisted Language Learning)*, 6(3):249-272, 1993.

Articles publi  s dans des revues d'audience nationale avec comit   de lecture : 6

1. Vincent Claveau and Pascale S  billot. Apprentissage semi-supervis   de patrons d'extraction de couples nom-verbe. *TAL (Traitement automatique des langues)*, 45(1):153-182, 2004.
2. Mathias Rossignol and Pascale S  billot. Extraction statistique sur corpus de classes de mots-cl  s th  matiques. *TAL (Traitement automatique des langues)*, 44(3):217-246, 2003.
3. Vincent Claveau, Pascale S  billot, Pierrette Bouillon, and C  cile Fabre. Acqu  rir des   l  ments du lexique g  n  ratif : quels r  sultats et    quels co  ts ? *TAL (Traitement automatique des langues)*, num  ro sp  cial *Lexiques s  mantiques*, 42(3):729-753, 2001.
4. Pierrette Bouillon, C  cile Fabre, Pascale S  billot, and Laurence Jacqmin. Apprentissage de ressources lexicales pour l'extension de requ  tes. *TAL (Traitement automatique des langues)*, num  ro sp  cial *Traitement automatique des langues pour la recherche d'information*, 41(2):367-393, 2000.
5. Jin Zhi, Pascale S  billot, and Fr  d  ric Danna. A Natural Language Interface in Compounds. *Chinese Journal of Advanced Software Research*, 5(4):360-375, 1998.
6. Paul Boucher and Pascale S  billot. Interpr  tation et g  n  ration automatiques de noms compos  s anglais    l'aide de formes logiques. *TAL (Traitement automatique des langues)*, num  ro sp  cial *Traitement automatique de la composition nominale*, 34(2):89-104, 1993.

Communications    des manifestations d'audience internationale avec comit   de s  lection sur article long et actes : 46

1. Duc Hau Nguyen, Cyrielle Mallart, Guillaume Gravier, and Pascale S  billot. Regularization, Semi-supervision, and Supervision for a Plausible Attention-Based Explanation. In *28th International Conference on Natural Language and Information Systems, NLDB*, pages 285-298, Derby, UK, 2023.
2. Cyrielle Mallart, Michel Le Nouy, Guillaume Gravier, and Pascale S  billot. Active learning for interactive relation extraction in a French newspaper's articles. In *International Conference on Recent Advances in Natural Language Processing, RANLP 2021*, pages 886-894, Online, 2021.
3. Duc Hau Nguyen, Guillaume Gravier, and Pascale S  billot. A study of the plausibility of attention between RNN encoders in natural language inference. In *20th IEEE International Conference on Machine Learning and Applications, ICMLA 2021 Special Session Machine Learning for Natural Language Processing*, pages 1-7, Online, 2021.

4. Cheikh Brahim El Vaigh, François Goasdoué, Guillaume Gravier, and Pascale Sébillot. A novel path-based entity relatedness measure for efficient collective entity Linking. In *19th International Semantic Web Conference, ISWC 2020*, Online, 2020.
5. Cheikh Brahim El Vaigh, François Torregrossa, Robin Allesiardo, Guillaume Gravier, and Pascale Sébillot. A correlation-based entity embedding approach for robust entity linking. In *32th International Conference on Tools with Artificial Intelligence, ICTAI, Short Paper*, pages 949-954, Online, 2020.
6. Cheikh Brahim El Vaigh, François Goasdoué, Guillaume Gravier, and Pascale Sébillot. Using Knowledge Base Semantics in Context-Aware Entity Linking. In *19th ACM Symposium on Document Engineering 2019, DocEng'19*, pages 8:1-8:10, Berlin, Germany, 2019. [CORE rang B] (taux d'acceptation 38%)
7. Raheel Qader, Gwénoél Lecorvé, Damien Lolive, and Pascale Sébillot. Disfluency Insertion for Spontaneous TTS: Formalization and Proof of Concept. In *6th International Conference on Statistical Language and Speech Processing, SLSP 2018*, pages 1-12, Mons, Belgium, 2018. (taux d'acceptation 45%)
8. Raheel Qader, Gwénoél Lecorvé, Damien Lolive, Marie Tahon, and Pascale Sébillot. In *Statistical Pronunciation Adaptation for Spontaneous Speech Synthesis. Text, Speech and Dialogue, TSD*, 8 pages, Prague, Czech Republic. 2017.
9. Rémi Bois, Guillaume Gravier, Eric Jamet, Emmanuel Morin, Maxime Robert, and Pascale Sébillot. Linking multimedia content for efficient news browsing. In *ACM International Conference on Multimedia Retrieval, ICMR'17*, pages 301-307, Bucharest, Romania, 2017. [CORE rang B]
10. Rémi Bois, Vedran Vukotic, Vedran, Anca-Roxana Simon, Ronan Sicre, Christian Raymond, Pascale Sébillot, and Guillaume Gravier. Exploiting Multimodality in Video Hyperlinking to Improve Target Diversity. In *23rd international Conference On MultiMedia Modelling, MMM 2017*, 12 pages, Reykjavik, Iceland, 2017. [CORE rang C] (taux d'acceptation 46,3%)
11. Guillaume Gravier, Martin Ragot, Laurent Amsaleg, Rémi Bois, Grégoire Jadi, Éric Jamet, Laura Monceaux, and Pascale Sébillot. Shaping-up Multimedia Analytics: Needs and Expectations of Media Professionals. In *22nd International Conference On MultiMedia Modelling, MMM 2016, special session Perspectives on Multimedia Analytics*, pages 303-314, Miami, USA, 2016. [CORE rang C]
12. Anca-Roxana Simon, Pascale Sébillot, and Guillaume Gravier. Hierarchical Topic Structuring: From Dense Topic Segmentation to Topically Focused Fragments via Burst Analysis. In *10th Recent Advances in Natural Language Processing conference, RANLP 2015*, Short paper, pages 588-595, Hissar, Bulgaria, 2015. [CORE rang C] (taux d'acceptation 38%)
13. Raheel Qader, Gwénoél Lecorvé, Damien Lolive, and Pascale Sébillot. Probabilistic Speaker Pronunciation Adaptation for Spontaneous Speech Synthesis Using Linguistic Features. In *3rd International Conference on Statistical Language and Speech Processing, SLSP 2015*, pages 229-241, Budapest, Hungary, 2015 (taux d'acceptation 37%)
14. Anca-Roxana Simon, Guillaume Gravier, and Pascale Sébillot. Leveraging Lexical Cohesion and Disruption for Topic Segmentation. In *International Conference on Empirical Methods in Natural Language Processing, EMNLP 2013*, pages 1314-1324, 2 columns, Seattle, Washington, USA, 2013. [CORE rang A]
15. Maria Eskevich, Gareth J.F. Jones, Robin Aly, Roeland J.F. Ordelman, Shu Chen, Danish Nadeem, Camille Guinaudeau, Guillaume Gravier, Pascale Sébillot, Tom de Nies, Pedro Debevere, Rik Van de Walle, Petra Galuscakova, Pavel Pecina, and Martha Larson. Multimedia Information Seeking through Search and Hyperlinking. In *ACM International Conference on Multimedia Retrieval, ICMR 2013*, 8 pages, 2 columns, Dallas, Texas, USA, 2013. [CORE rang B]
16. Khaoula Elagouni, Christophe Garcia, Franck Mamalet, and Pascale Sébillot. Text Recognition in Videos Using a Recurrent Connectionist Approach. In *22th International Conference on Artificial Neural Networks, ICANN 2012*, pages 172-179, Lecture Notes in Computer Science, Springer Berlin Heidelberg, Number 7553, Lausanne, Switzerland, 2012. [CORE rang B]
17. Khaoula Elagouni, Christophe Garcia, Franck Mamalet, and Pascale Sébillot. Combining Multi-Scale Character Recognition and Linguistic Knowledge for Natural Scene Text OCR. In *10th IAPR International Workshop on Document Analysis Systems, DAS 2012*, pages 120-124, 2 columns,

Gold Coast, Queensland, Australia, 2012. [CORE rang B]

Remarque : bien qu'appelé workshop, DAS est une véritable conférence indépendante, sur 3 ou 4 jours, d'où son classement dans cette rubrique.

18. Ali-Reza Ebadat, Vincent Claveau, and Pascale Sébillot. Proper Noun Semantic Clustering Using Bag-of-Vectors. In *Applied Natural Language Conference, ANLP 2012, Special track of the 25th Florida Artificial Intelligence Research Society Conference, FLAIRS 2012*, 6 pages, 2 columns, Marco Island, Florida, USA, 2012.
19. Khaoula Elagouni, Christophe Garcia, and Pascale Sébillot. A Comprehensive Neural-Based Approach for Text Recognition in Videos Using Natural Language Processing. In *ACM International Conference on Multimedia Retrieval, ICMR 2011*, 8 pages, 2 columns, Trento, Italy, 2011. [CORE rang B]
20. Gwénolé Lecorvé, Guillaume Gravier, and Pascale Sébillot. Automatically Finding Semantically Consistent N-Grams to Add New Words in LVCSR Systems. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'11*, 4 pages, 2 columns, Prague, Czech Republic, 2011. [CORE rang B]
21. Julien LawTo, Jean-Luc Gauvain, Lori Lamel, Gregory Grefenstette, Guillaume Gravier, Julien Despres, Camille Guinaudeau, and Pascale Sébillot. A Scalable Video Search Engine Based on Audio Content Indexing and Topic Segmentation. In *2011 Networked and Electronic Media (NEM) Summit*, 6 pages, 2 columns, Torino, Italy, 2011.
22. Camille Guinaudeau, Guillaume Gravier, and Pascale Sébillot. Improving ASR-Based Topic Segmentation of TV Programs with Confidence Measures and Semantic Relations. In *11th Annual Conference of the International Speech Communication Association, Interspeech 2010*, pages 1365-1368, 2 columns, Makuhari, Chiba, Japan, 2010. [CORE rang A] **Sélectionné dans le top 3% de la conférence**
23. Gwénolé Lecorvé, Guillaume Gravier, and Pascale Sébillot. Constraint Selection for Topic-Based MDI Adaptation of Language Models. In *International Conference on Speech and Language Technology, Interspeech'09*, pages 368-371, 2 columns, Brighton, UK, 2009. [CORE rang A]
24. Gwénolé Lecorvé, Guillaume Gravier, and Pascale Sébillot. An Unsupervised Web-Based Topic Language Model Adaptation Method. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2008*, pages 5081-5084, Las Vegas, Nevada, Usa, 2008. [CORE rang B]
25. Stéphane Huet, Guillaume Gravier, and Pascale Sébillot. Morphosyntactic Processing of N-Best Lists for Improved Recognition and Confidence Measure Computation. In *8th Interspeech Conference, Interspeech 2007*, pages 1741-1744, Antwerp, Belgium, 2007. [CORE rang A]
26. Fabienne Moreau, Vincent Claveau, and Pascale Sébillot. Combining Linguistic Indexes to Improve the Performances of Information Retrieval Systems: A Machine Learning Based Solution. In *8e conférence internationale Recherche d'information assistée par ordinateur, RIAO 2007*, 19 pages, Pittsburgh, PA, Usa, 2007.
27. Fabienne Moreau, Vincent Claveau, and Pascale Sébillot. Automatic Morphological Query Expansion Using Analogy-Based Machine Learning. In *29th European Conference on Information Retrieval, ECIR 2007*, pages 222-233, Rome, Italy, 2007. [CORE rang A]
28. Stéphane Huet, Guillaume Gravier, and Pascale Sébillot. Are Morphosyntactic Taggers Suitable to Improve Automatic Transcription? In *Text, Speech and Dialogue, TSD*, volume 4188/2006 of Lecture Notes in Computer Science, pages 391-398. Brno, Czech Republic, 2006.
29. Mathias Rossignol and Pascale Sébillot. Acquisition sur corpus non spécialisés de classes sémantiques thématisées. In *8es journées internationales d'analyse statistique des données textuelles, JADT'2006*, pages 827-837, Besançon, France, 2006. [CORE rang C]
30. Pascale Sébillot. Symbolic Machine Learning: A Different Answer to the Problem of the Acquisition of Lexical Knowledge from Corpora. In *European conference on Computing and Philosophy, E-CAP2005*, 3 pages, Västerås, Sweden, 2005
31. Vincent Claveau and Pascale Sébillot. From Efficiency to Portability: Acquisition of Semantic Relations by Semi-Supervised Machine Learning. In *20th International Conference on Computational Linguistics, COLING'04*, pages 261-267, Geneva, Switzerland, 2004. [CORE rang A]
32. Pierrette Bouillon, Vincent Claveau, Cécile Fabre, and Pascale Sébillot. Acquisition of Qualia Elements from Corpora - Evaluation of a Symbolic Learning Method. In *3rd International*

- Conference on Language Resources and Evaluation, LREC'2002*, pages 208-215, Las Palmas, Canary Islands, Spain, 2002.
33. Mathias Rossignol and Pascale Sébillot. Automatic Generation of Sets of Keywords for Theme Characterization and Detection. In A. Morin and P. Sébillot, editors, *6es journées internationales d'analyse statistique des données textuelles, JADT'2002*, pages 653-664, Saint-Malo, France, 2002.
 34. Pascale Sébillot, Pierrette Bouillon, and Cécile Fabre. Inductive Logic Programming for Corpus-Based Acquisition of Semantic Lexicons. In *Learning Language in Logic, LLL-2000*, pages 199-208, Lisbon, Portugal, 2000.
 35. Pascale Sébillot, Pierrette Bouillon, Vincent Claveau, Cécile Fabre, Laurence Jacqmin, and Jacques Nicolas. Apprentissage en corpus de couples nom-verbe pour la construction d'un lexique génératif. In M. Rajman and J.-C. Chappelier, editors, *5es journées internationales d'analyse statistique des données textuelles, JADT'2000*, audience francophone, pages 205-212, Lausanne, Suisse, 2000.
 36. Ronan Pichon and Pascale Sébillot. From Corpus to Lexicon: From Contexts to Semantic Features. In B. Lewandowska-Tomaszczyk and P.J. Melia, editors, *PALC'99: Practical Applications in Language Corpora*, volume 1 of Lodz studies in Language, pages 375-389. Peter Lang, 2000.
 37. Cécile Fabre and Pascale Sébillot. Semantic Interpretation of Binominal Sequences and Information Retrieval. In *International ICSC Congress on Computational Intelligence: Methods and Applications, CIMA'99, Symposium on Advances in Intelligent Data Analysis, AIDA'99*, 7 pages, Rochester, N.Y., Usa, 1999.
 38. Ronan Pichon and Pascale Sébillot. Automatic Acquisition of Meaning Elements for the Creation of Semantic Lexicons. In *Information Society, IS'98*, pages 69-72, Ljubljana, Slovenia, 1998.
 39. Frédéric Danna, Pascale Sébillot, and Jin Zhi. A Knowledge-Based Approach for Analysing and Understanding the Students' Answers in Compounds. In *Natural Language Processing Pacific Rim Symposium, NLPRS'97*, 7 pages, Phuket, Thaïland, 1997.
 40. Frédéric Danna and Pascale Sébillot. Evaluation of a Student Modeling System. In *Artificial Intelligence and Education, AIED'97*, pages 506-513, Kobe, Japan, 1997.
 41. Pascale Sébillot. A Corpus-Based Approach to Refining a Domain-Independent Compound Interpretation System. In K. Oflazer and H. Somers, editors, *New Methods in Language Processing, NeMLap-2*, pages 133-140, Ankara, Turkey, 1996.
 42. Frédéric Danna and Pascale Sébillot. Modeling Student's Cognitive Temporal Evolution. In *3rd International Conference on Computer Aided Engineering Education, CAEE'95*, pages 317-322, Bratislava, Slovaquia, 1995.
 43. Pascale Sébillot, Frédéric Danna, and Paul Boucher. Expert Model in Compounds, an ITS for Second Language Learning. In *2nd Australian Cognitive Science Conference, CogSci-93*, pages 148-150, Melbourne, Australia, 1993.
 44. Pascale Sébillot. A Logical-Based Language for Feature Specification and Transmission Control. In A. Voronkov, editor, *Logic Programming*, number 592 in Lecture Notes in Artificial Intelligence, pages 452-461. Springer-Verlag, 1992.
 45. Pascale Sébillot. Introduction of Feature Filtering Mechanisms into a Government-and-Binding Parser. In *Al-Azhar Engineering 2nd International Conference, AEC'91*, pages 227-238, Cairo, Egypt, 1991.
 46. Pascale Sébillot. Logical Modelling of Constraints on Acceptable Syntactic Tree Forms for Natural Language Analysis. In Y.A. Feldman and A. Bruckstein, editors, *Artificial Intelligence and Computer Vision*, pages 111-123. Elsevier Science Publishers, North Holland, 1991.

Communications à des manifestations d'audience internationale avec comité de sélection sur article court et actes : 3

1. Thomas Gaillat, Pascale Sébillot, and Nicolas Ballier. Probabilistic Comparing Corpora to Identify Learner-Specific Features of English: The Case of this, that and it. In 3rd Learner Corpus Research Conference, LCR2015, 2 pages, Cuijk/Nijmegen, the Netherlands, 2015.
2. Stéphane Huet, Guillaume Gravier, and Pascale Sébillot. Morphosyntactic Resources for Automatic Speech Recognition. In *6th International Conference on Language Resources and Evaluation, LREC'2008*, 7 pages, Marrakech, Morocco, 2008.
3. Gwénolé Lecorvé, Guillaume Gravier, and Pascale Sébillot. On the Use of Web Resources and

Natural Language Processing Techniques to Improve Automatic Speech Recognition Systems. In *6th International Conference on Language Resources and Evaluation, LREC'2008*, 8 pages, Marrakech, Morocco, 2008.

Articles de workshops internationaux avec comité de lecture : 8

1. Cyrielle Mallart, Michel Le Nouy, Guillaume Gravier, and **Pascale Sébillot**. Confronting active learning for relation extraction to a real-life scenario on French newspaper data. In *2nd Workshop on Interactive Learning for Natural Language Processing, InterNLP 2022@NeurIPS 2022*, 7 pages, New Orleans, USA, 2022.
2. Cheikh Brahim El Vaigh, Cheikh Brahim, Guillaume Le Noé Bienvenu, Guillaume Gravier, Pascale Sébillot. IRISA System for Entity Detection and Linking at CLEF HIPE 2020. In *CLEF HIPE 2020 Online Workshop (CEUR Workshop Proceedings)*, 7 pages, Online, 2020.
3. Rémi Bois, Guillaume Gravier, Éric Jamet, Maxime Robert, Emmanuel Morin, and Pascale Sébillot. Language-based Construction of Explorable News Graphs for Journalists. In *Empirical Methods in Natural Language Processing - Workshop on Natural Language Processing meets Journalism*, pages 31-36, Copenhagen, Denmark, 2017.
4. Anca-Roxana Simon, Rémi Bois, Guillaume Gravier, Pascale Sébillot, Emmanuel Morin, and Sien Moens. Hierarchical Topic Models for Language-based Video Hyperlinking. In *ISCA/IEEE Workshop on Speech, Language and Audio in Multimedia, SLAM 2015*, 4 pages, 2 columns, Brisbane, Australia, 2015.
5. Anca-Roxana Simon, Camille Guinaudeau, Pascale Sébillot, and Guillaume Gravier. Investigating Domain-Independent NLP Techniques for Precise Target Selection in Video Hyperlinking. In *ISCA/IEEE Workshop on Speech, Language and Audio in Multimedia, SLAM 2014*, 3 pages, 2 columns, Penang, Malaysia, 2014.
6. Camille Guinaudeau, Guillaume Gravier, and Pascale Sébillot. Can Automatic Speech Transcripts be Used for Large Scale TV Stream Description and Structuring?. In *1st International Workshop on Content-Based Audio/Video Analysis for Novel TV Services, CBTv 2009*, pages 489-494, 2 columns, San Diego, USA, 2009.
7. Mathias Rossignol and Pascale Sébillot. Automatic Acquisition of Lexical Semantic Information Using Medium to Small Corpora. In *International Workshop on Spoken Languages Technologies for Under-resourced languages, SLTU'08*, pages 92-97, Hanoi, Vietnam, 2008.
8. Pierrette Bouillon, Vincent Claveau, Cécile Fabre, and Pascale Sébillot. Using Part-of-Speech and Semantic Tagging for the Corpus-Based Learning of Qualia Structure Elements. In P. Bouillon and K. Kanzaki, editors, *1st International Workshop on Generative Approaches to the Lexicon, GL'2001*, 10 pages, Geneva, Switzerland, 2001.

Articles de workshops internationaux sans comité de lecture : 6

1. Rémi Bois, Vedran Vukotić, Ronan Sicre, Christian Raymond, Guillaume Gravier, and Pascale Sébillot. IRISA at TRECVID2016: Crossmodality, Multimodality and Monomodality for Video Hyperlinking. In *Working Notes of the TRECVID 2016 Workshop*, 9 pages, Gaithersburg, United States, 2016.
2. Anca-Roxana Simon, Ronan Sicre, Rémi Bois, Guillaume Gravier, and Pascale Sébillot. IRISA at TrecVid2015: Leveraging Multimodal LDA for Video Hyperlinking. In *Working Notes of the TRECVID 2015 Workshop*, 9 pages, Gaithersburg, United States, 2015.
3. Anca-Roxana Simon, Guillaume Gravier, and Pascale Sébillot. IRISA at MediaEval 2014: Search and Anchoring in Video Archives Task. In *Working Notes Proceedings of the MediaEval 2015 Workshop*, 3 pages, 2 columns, Wurzen, Germany, 2015.
4. Anca-Roxana Simon, Guillaume Gravier, Pascale Sébillot, and Marie-Francine Moens. IRISA and KUL at MediaEval 2014: Search and Hyperlinking Task. In *Working Notes Proceedings of the MediaEval 2014 Workshop*, 2 pages, 2 columns, Barcelona, Spain, 2014.
5. Camille Guinaudeau, Anca-Roxana Simon, Guillaume Gravier, and Pascale Sébillot. HITS and IRISA at MediaEval 2013: Search and Hyperlinking Task. In *Working Notes Proceedings of the MediaEval 2013 Workshop*, 2 pages, 2 columns, Barcelona, Spain, 2013.

6. Camille Guinaudeau, Guillaume Gravier, and Pascale Sébillot. IRISA at MediaEval 2012: Search and Hyperlinking Task. In *Working Notes Proceedings of the MediaEval 2012 Workshop*, 2 pages, 2 columns, Pisa, Italy, 2012.

Communications à des manifestations d'audience francophone avec comité de sélection sur article long et actes : 27

1. Hugo Thomas, Guillaume Gravier, and Pascale Sébillot. Derrière les plongements de relations. In *30^e conférence sur le Traitement Automatique des Langues Naturelles, TALN 2023*, pages 311-322, Paris, France, 2023.
2. Loïc Fosse, Duc Hau Nguyen, Pascale Sébillot, and Guillaume Gravier. Géométrie de l'auto-attention en classification : quand la géométrie remplace l'attention. In *30^e conférence sur le Traitement Automatique des Langues Naturelles, TALN 2023*, pages 137-150, Paris, France, 2023.
3. Loïc Fosse, Duc-Hau Nguyen, Pascale Sébillot, and Guillaume Gravier. Une étude statistique des plongements dans les modèles *transformers* pour le français. In *29^e conférence sur le Traitement Automatique des Langues Naturelles, TALN 2022*, pages 247-256, Avignon, France, 2022.
4. Duc Hau Nguyen, Guillaume Gravier, and Pascale Sébillot. Filtrage et régularisation pour améliorer la plausibilité des poids d'attention dans la tâche d'inférence en langue naturelle. In *29^e conférence sur le Traitement Automatique des Langues Naturelles, TALN 2022*, pages 95-103, Avignon, France, 2022.
5. Cyrielle Mallart, Michel Le Nouy, Guillaume Gravier, and Pascale Sébillot. Relation, es-tu là ? Détection de relations par LSTM pour améliorer l'extraction de relations. In *27^e conférence sur le Traitement Automatique des Langues Naturelles, TALN 2020*, pages 279-287, Online, 2020.
6. Raheel Qader, Gwénolé Lecorvé, Damien Lolive, and Pascale Sébillot. Ajout automatique de disfluences pour la synthèse de la parole spontanée : formalisation et preuve de concept. In *24^e conférence sur le Traitement automatique des langues naturelles, TALN 2017*, pages 1-15, Orléans, France, 2017 (**prix du meilleur article**).
7. Raheel Qader, Gwénolé Lecorvé, Damien Lolive, and Pascale Sébillot. Adaptation de la prononciation pour la synthèse de la parole spontanée en utilisant des informations linguistiques. In *31^e Journées d'Études sur la Parole, JEP*, pages 37-45, Paris, France, 2016.
8. Anca Simon, Guillaume Gravier, and Pascale Sébillot. Évaluation d'une nouvelle structuration thématique hiérarchique des textes dans un cadre de résumé automatique et de détection d'ancres au sein de vidéos. In *23^e conférence sur le Traitement Automatique des Langues Naturelles, TALN 2016*, pages 139-152, Paris, France, 2016.
9. Anca-Roxana Simon, Guillaume Gravier, and Pascale Sébillot. Un modèle segmental probabiliste combinant cohésion lexicale et rupture lexicale pour la segmentation thématique. In *20^e conférence sur le Traitement Automatique des Langues Naturelles, TALN 2013*, pages 202-214, Les Sables d'Olonne, France, 2013.
10. Ali-Reza Ebadat, Vincent Claveau, and Pascale Sébillot. Semantic Clustering Using Bag-of-Bag-of-Features. In *9^e conférence en recherche d'information et applications, CORIA 2012*, pages 229-244, Bordeaux, France, 2012.
11. Khaoula Elagouni, Christophe Garcia, and Pascale Sébillot. Reconnaissance automatique de texte dans des vidéos à l'aide d'un OCR et de connaissances linguistiques. In *23^e colloque GRETSI*, 4 pages, 2 colonnes, Bordeaux, France, 2011.
12. Camille Guinaudeau, Guillaume Gravier, and Pascale Sébillot. Utilisation de relations sémantiques pour améliorer la segmentation thématique de documents télévisuels". In *17^e conférence sur le Traitement Automatique des Langues Naturelles, TALN 2010*, 10 pages, Montréal, Québec, Canada, 2010.
13. Gwénolé Lecorvé, Guillaume Gravier, and Pascale Sébillot. L'adaptation thématique d'un modèle de langue fait-elle apparaître des mots thématiques ?. In *28^{es} Journées d'étude sur la parole, JEP'10*, 4 pages, 2 colonnes, Mons, Belgique, 2010.
14. Camille Guinaudeau, Guillaume Gravier, and Pascale Sébillot. Indices utiles à la cohésion lexicale pour la segmentation thématique de documents oraux. In *28^{es} Journées d'étude sur la parole, JEP'10*, 4 pages, 2 colonnes, Mons, Belgique, 2010.

15. Stéphane Huet, Guillaume Gravier, and Pascale Sébillot. Un modèle multi-sources pour la segmentation en sujets de journaux radiophoniques. In *15e conférence Traitement automatique des langues naturelles, TALN'08*, pages 49-58, Avignon, France, 2008.
16. Gwénolé Lecorvé, Guillaume Gravier, and Pascale Sébillot. Vers une adaptation thématique non supervisée de modèles de langage : utilisation d'internet comme un corpus ouvert. In *27es journées d'étude sur la parole, JEP'2008*, pages 149-152, Avignon, France, 2008.
17. Fabienne Moreau, Vincent Claveau, and Pascale Sébillot. Intégrer plus de connaissances linguistiques en recherche d'information peut-il augmenter les performances des systèmes ? In *4e conférence francophone en Recherche d'information et applications, CORIA 2007*, pages 223-238, Saint-Étienne, France, 2007.
18. Stéphane Huet, Guillaume Gravier, and Pascale Sébillot. Peut-on utiliser les étiqueteurs morphosyntaxiques pour améliorer la transcription automatique ? In *26es journées d'étude sur la parole, JEP'2006*, pages 317-320, Dinard, France, 2006.
19. Mathias Rossignol and Pascale Sébillot. Mise au jour semi-automatique de nuances sémantiques entre mots de sens proches. In *Traitement automatique des langues naturelles, TALN'06*, pages 266-275, Leuven, Belgique, 2006.
20. Vincent Claveau and Pascale Sébillot. Extension de requêtes par lien sémantique nom-verbe acquis sur corpus. In *Traitement automatique des langues naturelles, TALN'04*, 10 pages, Fes, Morocco, 2004.
21. Mathias Rossignol and Pascale Sébillot. Extraction non supervisée sur corpus de classes de mots-clefs. In *14e congrès francophone AFRIF-AFIA de reconnaissance des formes et intelligence artificielle, RFIA 2004*, 10 pages, Toulouse, France, 2004.
22. Vincent Claveau and Pascale Sébillot. Apprentissage symbolique pour l'acquisition de ressources linguistiques. In *Atelier Acquisition, apprentissage et exploitation de connaissances sémantiques pour l'accès au contenu textuel de la plateforme AFIA*, pages 51-61, Laval, France, 2003.
23. Ronan Pichon and Pascale Sébillot. Différencier les sens des mots à l'aide du thème et du contexte de leurs occurrences : une expérience. In *Traitement automatique des langues naturelles, TALN'99*, pages 279-288, Cargèse, France, 1999.
24. Cécile Fabre and Pascale Sébillot. Interprétation automatique des composés nominaux anglais hors domaine : quelles solutions ? In *10e congrès reconnaissance des formes et intelligence artificielle, RFIA'96*, pages 71-79, Rennes, France, 1996.
25. Cécile Fabre and Pascale Sébillot. Interprétation sémantique des composés nominaux anglais et français sans constituant déverbal. In P. Bouillon and D. Estival, editors, *Compound Nouns: Multilingual Aspects of Nominal Composition*, pages 108-124, Geneva, Switzerland, 1994.
26. Patrick Saint-Dizier and Pascale Sébillot. Un processus de filtrage par arbres syntaxiques dans les grammaires discontinues. In *Programmation en logique*, pages 457-472, Trégastel, France, 1986.
27. Patrick Bosc, Patrick Saint-Dizier, and Pascale Sébillot. Compréhension du langage naturel et programmation en logique : application à l'interrogation de bases de données. In *Applications informatiques conversationnelles et le langage naturel*, pages 115-130, Nantes, France, 1984.

Communications à des manifestations d'audience francophone avec comité de sélection sur article court ou résumé et actes : 4

1. François Vignale, Guillaume Le Noé Bienvenu, Guillaume Gravier, and Pascale Sébillot. Utilisation d'approches automatiques pour la reconnaissance des expériences de lecture. In *Colloque Humanistica 2021*, pages 81-83, Rennes, France, 2021
2. François Vignale, Guillaume Le Noé Bienvenu, Guillaume Gravier, and Pascale Sébillot. Je pense que ça traite d'expérience de lecture, à voir... : retour sur une expérience d'annotation collaborative. In *Colloque Humanistica 2021*, pages 84-85, Rennes, France, 2021
3. Rémi Bois, Guillaume Gravier, Pascale Sébillot, and Emmanuel Morin. Vers une typologie de liens entre contenus journalistiques. In *22e conférence sur le Traitement Automatique des Langues Naturelles, TALN 2015*, article court, pages 515—521, Caen, France, 2015.
4. Ali-Reza Ebadat, Vincent Claveau, and Pascale Sébillot. Using Shallow Linguistic Features for Relation Extraction in Bio-Medical Texts. In *18e conférence sur le Traitement Automatique des Langues Naturelles, TALN 2011*, Vol. 2, pages 125-130, papier court, Montpellier, France, 2011.

Autres

Communications à des manifestations d'audience internationale sur invitation et avec actes : 3

1. Pascale Sébillot. Acquiring Elements of Pustejovsky's Generative Lexicon using Inductive Logic Programming. In C. Rétoré and A. Dikovsky, editors, *Programming, Grammar and Logic, GRAL'2000*, Rennes, France, édition Bulzoni, 15 pages. édition Bulzoni, Roma.
2. Pascale Sébillot. Acquérir des informations sémantiques à partir de corpus. In P. Boucher, editor, *Workshop des ateliers en morphologie, Colex, La structure du lexique*, audience francophone, pages 173-181, Nantes, France, 1999. Presse de l'université de Nantes.
3. Pascale Sébillot. Sémantique des composés anglais : approche générative, limites et applications. In *Informatique et langue naturelle, ILN'93*, audience francophone, pages 59-78, Nantes, France, 1993.

Habilitation à diriger des recherches et thèse de doctorat

1. Pascale Sébillot. *Apprentissage sur corpus de relations lexicales sémantiques - La linguistique et l'apprentissage au service d'applications du traitement automatique des langues*. Habilitation à diriger des recherches, Université de Rennes 1, France, 107 pages, décembre 2002.
2. Pascale Sébillot. *Modélisation de filtrage par arbres et par traits dans les grammaires logiques pour l'analyse automatique du langage naturel*. Thèse de doctorat, INSA de Rennes, France, 158 pages, juin 1989.

Rapports internes et techniques : 8

1. Raheel Qader, Gwénoél Lecorvé, Damien Lolive, and Pascale Sébillot. Phonology Modelling for Expressive Speech Synthesis: a Review. Rapport de recherche 2020, Irisa, 18 pages, 2014.
2. Stéphane Huet, Pascale Sébillot, and Guillaume Gravier. Utilisation de la linguistique en reconnaissance de la parole : un état de l'art. Rapport de recherche 1804, Irisa, 72 pages, 2006.
3. Fabienne Moreau and Pascale Sébillot. Contributions des techniques du traitement automatique des langues à la recherche d'information. Rapport de recherche 5484, Inria, 34 pages, 2005.
4. Ronan Pichon and Pascale Sébillot. Acquisition automatique d'informations lexicales à partir de corpus : un bilan. Rapport de recherche 3321, Inria, 62 pages, 1997.
5. Cécile Fabre and Pascale Sébillot. Calculability of the Semantics of English Nominal Compounds: Combining General Linguistic Rules and Corpus-Based Semantic Information. Publication interne 974, Irisa, 40 pages, 1995.
6. Frédéric Danna and Pascale Sébillot. Modélisation de l'apprenant dans un tuteur intelligent des composés en anglais. Publication interne 928, Irisa, 33 pages, 1995.
7. Paul Boucher, Frédéric Danna, and Pascale Sébillot. Compounds: An ITS for Learning to Use Compounds in English. Publication interne 718, Irisa, 21 pages, 1993.
8. Cécile Fabre and Pascale Sébillot. Calculabilité de la sémantique des composés anglais et français ; adaptation à l'indexation automatique, Journée CTI thème 6 du CNET, 9 pages, Lannion, France, 1996.

Direction de thèses

J'ai, depuis 93, encadré dix-sept thèses soutenues (16 en informatique, une en linguistique anglaise). Pour celles-ci, j'indique ci-dessous, outre le doctorant, le titre de la thèse, les dates de début et de fin, le financement, le pourcentage d'encadrement personnel, les co-encadrants éventuels, et le devenir du doctorant. J'assure actuellement l'encadrement des travaux de thèse de Duc Hau Nguyen (50%) et de Hugo Thomas (50%). Il est à noter que la politique des équipes TexMex puis LinkMedia, mêlant des spécialistes de différentes modalités, de bases de données, et d'apprentissage, est de co-encadrer *quasi* systématiquement chaque doctorant afin de faciliter le travail en commun. Par ailleurs, j'ai régulièrement été responsable de stages de Master 2 recherche en informatique. En l'absence de mention explicite autre, mon pourcentage de responsabilité des encadrements (thèses et M2 recherche) listés ci-dessous est de 100%.

Encadrements de thèses

Thèses déjà soutenues : 17

1. *Construction dynamique incrémentale de graphes de connaissance par fouille de contenus*, Cyrielle Mallart, début 1/12/18 – soutenance le 23 novembre 2022 (CIFRE Ouest-France). Encadrement : 34% (avec Guillaume Gravier 33% et Michel Le Nouy Ouest-France 33%). Cyrielle Mallart est actuellement ingénieur de recherche dans le cadre d'une ANR JCJC à l'Université de Rennes 2.
2. *Incremental content to data linking leveraging ontological knowledge in data journalism*, Cheikh Brahim El Vaigh, début 1/10/17 – soutenance le 7 janvier 2021 (projet IPL Inria Icodia). Encadrement : 33% (avec Guillaume Gravier 33% et François Goasdoué Enssat Lannion/Université de Rennes 1 33%). Cheikh Brahim El Vaigh est maître de conférences à l'Université de Bourgogne (Dijon).
3. *Hypergraphes orientés multimédias navigables : construction et exploitation*, Rémi Bois, début 1/10/14 – soutenance le 21 décembre 2017 (projet Labex CominLabs Limah). Encadrement : 40% (avec Guillaume Gravier 40% et Emmanuel Morin Université de Nantes 20%). Rémi Bois est ingénieur chez Onfido à Londres, Grande-Bretagne.
4. *Phonology Modeling for Emotional Speech Synthesis*, Raheel Kareem Qader, début 1/01/14 – soutenance le 31 mars 2017 (allocation MESR). Encadrement : 30% (avec Gwénolé Lecorvé Enssat Lannion/Université de Rennes 1 70%). Raheel Kareem Qader est ingénieur R&D (head of R&D) chez Lingua Custodia à Montigny-le-Bretonneux.
5. *Reference in Interlanguage: The Case of "this" and "that". From linguistic annotation to corpus interoperability*, Thomas Gaillat, début 1/10/10 – soutenance le 16 juin 2016 (*thèse de linguistique anglaise* de l'Université Paris Diderot financée sur fonds propres en tant qu'enseignant d'anglais au Scelva, Université de Rennes 1). Encadrement : 50% (avec Nicolas Ballier Université Paris-Diderot 50%). Thomas Gaillat est maître de conférences en linguistique à l'Université de Rennes 2.
6. *Structuration sémantique hiérarchique de collections vidéos*, Anca-Roxana Simon, début 1/10/12 – soutenance le 2 décembre 2015 (allocation MESR). Encadrement : 50% (avec Guillaume Gravier 50%). Anca-Roxana Simon est ingénieur R&D chez ProSiebenSat.1 à Munich, Allemagne.
7. *Combining Neural-Based Approaches and Linguistic Knowledge for Text Recognition in Multimedia Documents*, Khaoula Elagouni, début 1/10/09 – soutenance 28/05/13 (Cifre Orange Labs). Encadrement : 35% (avec Christophe Garcia Liris/Insa Lyon 35% et Franck Mamalet Orange Labs Rennes 30%). Khaoula Elagouni est ingénieur R&D chez MyScript à Nantes.
8. *Toward Robust Information Extraction Models for Multimedia Documents*, Ali-Reza Ebadat, début 1/10/09 – soutenance 17/10/12 (projet QUAERO). Encadrement : 30% (avec Vincent Claveau 70%). Ali-Reza Ebadat est ingénieur de recherche chez MyScript à Nantes.
9. *Structuration automatique de flux télévisuels*, Camille Guinaudeau, début 1/10/08 – soutenance 7/12/11 (projet QUAERO et bourse région). Encadrement : 50% (avec Gravier Guillaume 50%). Camille Guinaudeau est maître de conférences à l'IUT d'Orsay/Université Paris Sud.
10. *Adaptation thématique non supervisée d'un système de reconnaissance automatique de la parole*, Gwénolé Lecorvé, début 1/10/07 – soutenance 24/11/10 (allocation MESR). Encadrement : 50% (avec Guillaume Gravier 50%). Gwénolé Lecorvé est maître de conférences à l'ENSSAT/Université de Rennes 1 à Lannion, actuellement en détachement chez Orange. **Prix de l'AFCP** (association francophone de la communication parlée) en novembre 2010

11. *Apprentissage incrémental pour la construction de bases lexicales évolutives : application en désambiguïsation d'entités nommées*, Thomas Girault, début 1/11/05 – soutenance 18/06/10 (Cifre Orange Labs). Encadrement : 30% (avec Olivier Collin Orange Labs Lannion 70%). Thomas Girault est ingénieur chez Pages jaunes à Rennes.
12. *Informations morpho-syntaxiques et adaptation thématique pour améliorer la reconnaissance de la parole*, Stéphane Huet, début 1/10/04 – soutenance 11/12/07 (allocation MESR). Encadrement : 50% (avec Guillaume Gravier 50%). Stéphane Huet est maître de conférences à l'Université d'Avignon.
13. *Revisiter le couplage traitement automatique des langues et recherche d'information*, Fabienne Moreau, début 1/10/03 – soutenance 7/12/06 (financement Inria). Fabienne Moreau est maître de conférences à l'Université de Rennes 2.
14. *Acquisition sur corpus d'informations lexicales fondées sur la sémantique différentielle*, Mathias Rossignol (allocation MESR), début 1/10/01 – soutenance 26/10/05. Mathias Rossignol est actuellement ingénieur (AI & Analytics expert) chez FPT Smart Cloud à Hanoï, Vietnam.
15. *Acquisition automatique de lexiques sémantiques pour la recherche d'information*, Vincent Claveau, début 1/10/00 – soutenance 17/12/03 (allocation MESR). Vincent Claveau est CR1 CNRS dans notre équipe, détaché auprès de la DGA à Bruz depuis 23.
16. *Modélisation de l'apprenant dans un logiciel d'enseignement intelligemment assisté par ordinateur. Application à un tutoriel intelligent dédié aux composés anglais*, Frédéric Danna, début 1/10/93 – soutenance 8/01/97 (financement Inria). Frédéric Danna est informaticien dans une société privée à Toulouse.
17. *Interprétation automatique des séquences binominales en anglais et en français. Application à la recherche documentaire*, Cécile Fabre, début 1/10/93 – soutenance 18/12/96 (contrat CTI-CNET). Cécile Fabre est professeur à l'Université de Toulouse-Le Mirail.

Thèses en cours : 2

1. *Détection de relations à partir de peu ou pas d'exemples dans les archives de presse*, Hugo Thomas (½ ARED + ½ AI4SDA), depuis 1/10/22. Encadrement : 50% (avec Guillaume Gravier 50%).
2. *Espace multimodal pour la génération et la justification de liens sémantiques entre documents*, Duc-Hau Nguyen (ANR Archival), depuis 1/09/20. Encadrement : 50% (avec Guillaume Gravier 50%).

Encadrements de Master 2 recherche en informatique : 27

1. *Link prediction in knowledge graph embeddings for few-shot news classification*, Hugo Thomas (M2 SIF (M2 recherche en informatique) INSA Rennes, 2021-2022, encadrement 25% (co-encadrants Thomas Girault, Michel Le Nouy (SIPA Ouest-France) et G. Gravier)
2. *Classification automatique de liens au sein d'un corpus journalistique*, Lyes Chelfi (M2 recherche en informatique université de Rennes 1), 2015-2016, encadrement à 33% (co-encadrants Rémi Bois et Guillaume Gravier)
3. *Résumé automatique de texte appliqué au résumé multimodal de vidéos par programmation par contraintes*, Caio Corro (M2 recherche en informatique ENS Rennes), 2013-2014, encadrement à 50% (co-encadrant Patrick Gros)
4. *Segmentation thématique hiérarchique d'émissions télévisuelles transcrites automatiquement*, Anca-Roxana Simon (M2 recherche en informatique INSA de Rennes), 2011-2012, encadrement à 50% (co-encadrant Guillaume Gravier)
5. *Adapter le vocabulaire d'un système de transcription automatique de la parole aux thèmes abordés*, Tissier Florent (M2 recherche en informatique ESIR Université de Rennes 1), 2010-2011. encadrement à 50% (co-encadrant Guillaume Gravier)
6. *Fouille d'opinions*, Sébastien Gillot (M2 recherche en informatique IFSIC), 2009-2010, encadrement à 50% (co-encadrant Vincent Claveau)
7. *Segmentation thématique, suivi de sujets et construction de fils d'actualité dans des journaux TV français*, Julien Fayolle (M2 recherche STI université de Rennes 1), 2008-2009, encadrement à 25% (co-encadrants Guillaume Gravier, Ichiro Ide université de Nagoya, Shin'ichi Satoh NII Tokyo)
8. *Suivi de sujets d'actualité dans les flux télévisuels*, Helmi Ben Hmida (M2 recherche en informatique IFSIC), 2008-2009, encadrement à 50% (co-encadrant Guillaume Gravier)

9. *Contrôle automatisé des contenus de programmes télévisuels*, Camille Guinaudeau (M2 recherche Algorithmique et méthodes de l'information, Université de Caen), 2007-2008, encadrement à 50% (co-encadrant Guillaume Gravier)
10. *Adaptation thématique d'un système de transcription automatique de la parole*, Gwénolé Lecorvé (M2 recherche en informatique INSA), 2006-2007, encadrement à 50% (co-encadrant Guillaume Gravier)
11. *Vers une représentation riche des images*, Sébastien Rius (M2 recherche en informatique IFSIC), 2005-2006, encadrement à 50% (co-encadrant Vincent Claveau)
12. *Structuration automatique de classes sémantiques*, Mathieu Coutaud (M2 recherche en informatique IFSIC), 2004-2005
13. *Intégration de connaissances linguistiques pour la reconnaissance de phrases manuscrites en-ligne*, Solen Quiniou (DEA IFSIC), 2003-2004, encadrement à 33% (co-encadrants Sabine Carbonnel et Éric Anquetil)
14. *Sélection automatique d'unités de textes décrivant des interactions entre gènes*, Pascaline Tchienehom (DEA IFSIC), 2001-2002, encadrement à 50% (co-encadrant Jacques Nicolas)
15. *Apprentissage automatique de catégories de mots par co-training*, Emmanuelle Catz (DEA IFSIC), 2001-2002, encadrement à 50% (co-encadrant Vincent Claveau)
16. *Acquisition sur corpus d'informations lexicales basées sur la sémantique différentielle*, Mathias Rossignol (DEA IFSIC), 2000-2001
17. *Acquisition automatique de lexiques sémantiques basés sur la sémantique différentielle*, Olivier Tardif (DEA IST Génie linguistique, Université de Marne-La Vallée), 1999-2000
18. *Apprentissage d'un lexique génératif à partir d'un corpus étiqueté sémantiquement*, Vincent Claveau (DEA IFSIC), 1999-2000
19. *Acquisition automatique de lexiques sémantiques à partir de corpus*, Antoine Legrand (DEA IFSIC), 1998-1999
20. *Interface en langage naturel pour un Tuteur Intelligent des composés en anglais*, Guillaume Maingourd (DEA IFSIC), 1997-1998
21. *Module pédagogique pour un tuteur intelligent des composés en anglais*, David Galic (DEA IFSIC), 1996-1997
22. *Élaboration d'un module pédagogique pour un tuteur intelligent de mots composés en anglais*, Christine Largouët (DEA IFSIC), 1995-1996
23. *Typage des associations préférentielles dans les composés extraits de corpus*, Ronan Pichon (DEA IFSIC), 1995-1996
24. *De l'extraction automatique au résumé automatique*, Hervé de Beugny d'Hagerue (DEA IFSIC), 1994-1995
25. *Tools for paradigmatic classification in agglutinative languages*, Alain Marie (DEA IFSIC), 1992-1993, encadrement à 30% (co-encadrants de l'Université technique de Budapest : Ivan Bach et Matyas Naszodi)
26. *Modélisation de l'élève dans un tuteur intelligent des composés en Anglais*, Hervé Jehan (DEA IFSIC), 1992-1993
27. *Réalisation d'un logiciel d'EIAO du processus de composition anglais*, Frédéric Danna (DEA IFSIC), 1991-1992

Encadrements de post-doctorat : 1

1. *Bases de données textuelles et recherche d'information exploitant des techniques de traitement automatique des langues : un pas vers une fusion*, Antoine Doucet (bourse Inria), 2006-2007, encadrement à 50% (co-encadrant Laurent Amsaleg)