

**Projet de collaboration avec (FAP, BRESIL)
(2014-2016)**

Responsable scientifique : *Laurent Besacier, LIG – UMR 5217, INSII*

RAPPORT D'ACTIVITÉ 2015

A. MOBILITE TRANSNATIONALE

A.1 - Accueil, dans le laboratoire français, de chercheurs du laboratoire partenaire étranger

Objet de l'accueil, date, nom du chercheur, qualité, laboratoires d'origine et d'accueil, durée du séjour, si le chercheur a donné un séminaire indiquer le titre

a) Séjour de **Aline Villavicencio** et **Marco Idiart** (Professeurs, UFRGS - Brésil) en France en Novembre 2015 : au LIF du 8 au 12 Novembre 2015 ; au LIG - 16 Novembre au 20 Novembre 2015 ; au Lattice les 21-22 Novembre 2015. Objet: « Projet AIM-WEST – Echanges – Stages – Futures collaborations »

b) Séjour de **Pablo Botton Da Costa** (Etudiant en Master) au LIG - 23 Octobre 2015 au 23 janvier 2016. Objet: « Multilingual neural parsing »

A.2 - Séjours, dans le laboratoire partenaire étranger, de chercheurs du laboratoire français

Objet du séjour, date, nom du chercheur, qualité, laboratoires d'origine et d'accueil, durée du séjour, si le chercheur a donné un séminaire indiquer le titre

a) 3 séjours de chercheurs français ont été financés pour une mission à Porto-Alegre (Brésil) du 19 Octobre au 23 Octobre 2015. Les trois participants ont été :

- **Laurent BESACIER** (Professeur, LIG)
- **Carlos ramisch** (Maître de Conférences, LIF)
- **Isabelle TELLIER** (Professeur, LATTICE)

Un compte-rendu de ce séjour et les présentations réalisées sont disponibles en ligne sur :

<http://aim-west.imag.fr/project-meetings/2nd-aim-west-workshop-porto-alegre/>

b) Un élève ingénieur de Polytech' Grenoble (**Rama CODAZZI**) à fait un stage d'été à San Carlos (Brésil) co-encadré par Hélène Caseli (Ufscar) et Laurent Besacier (LIG)

A.3 – Organisation de réunions de travail, séminaires etc. dans le cadre du projet

Indiquer l'objet de la réunion, date, lieu (laboratoire partenaire ou ville), nombre total de participants, identifier les participants français (nom, qualité, laboratoire de rattachement, durée de la mission).

Vous pouvez également donner ces renseignements sous forme de tableau Excel.

Un bilan des réunions de travail sur le projet est donné ci-dessous (en anglais). Il se trouve aussi en ligne sur le site Web du projet AIM-WEST (**<http://aim-west.imag.fr/project-meetings/>**) (bilan en anglais)

- **January 21st, 2015** - Short **Skype** meeting between AIM-WEST partners
- **June 2015** (exact date not found) - Short **Skype** meeting between AIM-WEST partners to prepare the Workshop in Porto-Alegre
- **20-21 October 2015** AIM-WEST meeting in Porto Alegre (**Brasil**)
- **16-20 November 2015** AIM-WEST (LIG-UFRGS) meeting in Grenoble (**France**)

B. TRAVAUX EN COLLABORATION

B.1 – Etat d’avancement du projet scientifique

5 pages maximum pour l’année en cours ou 10 pages maximum pour les projets arrivant à terme. Le nom des chercheurs impliqués sera précisé

(en anglais)

Most of the work done during 2015 is summarized in the project meeting presentations available on <http://aim-west.imag.fr/project-meetings/2nd-aim-west-workshop-porto-alegre/>
Some of the tasks planned in 2014 and achieved in 2015 are summarized below.

Task 1 Release a usable Fr-En-Pt-BR corpus : has been done in 2015 (data extracted from TED corpus) by a summer Internship from Polytech’ Grenoble (Rama Codazzi) co-supervised by Helena Caseli & Laurent Besacier

Task 2 Organize internally a Fr-En MT shared task : has been done in 2015 in order to evaluate the behavior of several MT systems on translating several types of MWEs. The results of this shared task, as well as other works done by LIG during AIM-WEST are summarized online on <http://aim-west.imag.fr/wp-content/uploads/2015/10/AIM-WEST-LIG2015.pdf>

Task 3 extend MWE annotation done in French to English and Portuguese : has been started for English (*no online document yet*)

Task 4 POS tagging or morphological analysis of non contiguous MWEs : an internship at Lattice focused in 2015 on this topic. More precisely, the intern worked on *Labeling Multi Word Expressions with CRFs* – more details on <http://aim-west.imag.fr/wp-content/uploads/2015/10/tellier-AIM-West.pdf>

Task 5 Study of the interplay between complex function words and parsing
see main results summarized on

<http://aim-west.imag.fr/wp-content/uploads/2015/10/presentation-morph-jsl2015.pdf>

Task 6 Speech recognition and MWE expressions

Speech Processing aspects were postponed to 2016. An internship between UFRGS and LIG is planned on this topic. The subject below has been posted to Master (and engineering degree) students in France.

STAGE : MISE EN PLACE D'UN SYSTÈME DE TRANSCRIPTION AUTOMATIQUE DE LA PAROLE POUR LE PORTUGUAIS BRÉSILIEN

Objectifs du stage:

- 1/ Mise en place d'une système (fondé sur la boîte à outils Kaldi) de transcription automatique de la parole
- 2/ Rassemblement et mise en formes de ressources en portugais brésilien pour l'apprentissage du système (corpus de parole, texte, lexicque)
- 3/ Creation d'une benchmark (corpus de test) pour évaluation des performances de reconnaissance vocale (avec des enregistrements/transcriptions de test) et execution des tests et reporting des resultats
- 4) (si le temps le permet) Adaptation de ce système à la parole de personnes âgées pour un projet local à l'UFRGS

Compétences demandées :

Linux admin, programmation script, comprehension des systemes de reconnaissance vocale et reconnaissance de la parole par machine. Une connaissance du système Kaldi est un plus.

Superviseur Polytech' : Laurent BESACIER (laurent.besacier@imag.fr)

Superviseur Local : Aline Villavicencio (alinev@gmail.com)

Task 7 Word embeddings and multiword expressions

Teams in UFRGS and LIF are jointly conducting experiments on nominal compound interpretation in French, English and Portuguese. Our computational models predict their semantic compositionality using vector representations. Compositionality is important for translation (compositional → word-by-word, idiomatic → full phrase). Some results up to now are :

- 3 new manually annotated datasets containing 180 compounds in each language
- Implementation of tested models in mwetoolkit (<http://mwetoolkit.sf.net>)
- Submission to DIMSUM shared task (part of SEMEVAL 2016)
- Accepted paper at *LREC 2016*, papers being prepared for submission to 2016 conferences

B.2 – Autres activités communes

Activités avec des chercheurs du laboratoire partenaire étranger hors du contexte du projet, projets co-déposés dans le cadre d'appels nationaux ou européens, contrats industriels,...

Objet, cadre, dates, bref descriptif.

a) Un certain nombre de stages co-encadrés France-Brésil ont été proposés et sont en ligne sur : <http://aim-west.imag.fr/research-internships/>

b) Deux cours de type « tutoriels » ont été donnés devant des étudiants de l'UFRGS pendant la réunion AIM-WEST d'Octobre 2015 :

Tutorial I *Distributional semantics: context, vectors, embeddings and similarity* (Carlos ramisch, LIF)
<http://aim-west.imag.fr/wp-content/uploads/2015/10/tuto-distr-aimwest2015.pdf>

Tutorial II *Automatic Speech Recognition* (Laurent Besacier, LIG)
<http://aim-west.imag.fr/wp-content/uploads/2015/10/ASR-TUTORIAL-BESACIER.pdf>

C. PRODUCTION SCIENTIFIQUE CO-SIGNEE AVEC LE PARTENAIRE ETRANGER

a) Liste des publications parues, acceptées ou soumises (préciser) dans des revues avec comité de lecture

b) Liste des publications dans des ouvrages (livres, proceedings, ... préciser)

c) Liste des présentations à des colloques co-signées avec le partenaire étranger
 (indiquer si exposés oraux ou affiches)

- Alexandre Rondon, Helena Caseli, Carlos Ramisch. “Never-Ending Multiword Expressions Learning“, *Proceedings of the 11th Workshop on Multiword Expressions (MWE 2015)*, Denver, USA, p. 45-53, June, 2015. pdf
- Alexis Nasr, Carlos Ramisch, José Deulofeu, André Valli, “Joint Dependency Parsing and Multiword Expression Tokenization“, *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (Volume 1: Long Papers), Beijing, China, July, 2015. pdf
- Zied elloumi, Olivier Kraif, Laurent Besacier. “Integrating Multi Word Expressions in Statistical Machine Translation“, *Proceedings of Workshop on Multi-word Units in Machine Translation and Translation Technology (MUMTTT2015) – part of EUROPHRAS2015 – Malaga, Spain, 1-2 July 2015*. Poster. (*Pas co-signé avec part. étranger mais à la base de la shared task organisée sur AIM-WEST*)
- Agnès Tutin, Emmanuelle Esperança-Rodier, Manolo Iborra, Justine Reverdy “Annotation of multiword expressions in French” *Proceedings of Conference EUROPHRAS2015 – Malaga, Spain, 1-2 July 2015*. (*Pas co-signé avec part. étranger mais à la base de la shared task organisée sur AIM-WEST*)

d) Liste des brevets en co-propriété

e) Autres co-productions (bases de données, plateformes, sites web, portails thématiques... préciser)

site Web du projet AIM-WEST : <http://aim-west.imag.fr>

D. OBSERVATIONS EVENTUELLES

- Nous soulignons que le LIG a financé sur fonds propres des gratifications de stages réalisés sur les thème du projet AIM-WEST (nous avons pensé pouvoir les financer sur le projet lui-même mais ceci n'a pas été possible d'un point de vue administratif)
- Les travaux futurs prévus par chaque partenaire en 2016 sont listés sur <http://aim-west.imag.fr/partners-plans-for-next-year-in-aim-west/>
- Soumission d'un projet CAPES-COFECUB prévue à l'automne 2016
- La prochaine réunion « physique » du projet est prévue à l'automne 2016