



INS2I - 2018

APPEL A PROJETS

JCJC INS2I 2018

### Identification

Nom du porteur du projet	Flavien BREUVART
Unité de recherche	LIPN
Employeur du porteur	Université Paris 13
Adresse e-mail du porteur	flavien.breuvart@lipn.univ-paris13.fr
Titre long du projet	Confluence Randomisée Pour des Réductions Aléatoires
Acronyme du projet	CoRPoReAl

### Résumé du projet

Historiquement, formaliser la théorie de la réécriture a permis d'attaquer, et même de systématiser, la résolution de problèmes très complexes dans un cadre uniformisé. Avec l'essor des algorithmes randomisés et des langages associés, chaque communauté a dû généraliser ses outils. Malheureusement, ces extensions ne sont pas canoniques et ne forment pas une théorie uniformisée ; en particulier, de nombreux problèmes classiques, comme la confluence, y sont difficilement exprimables. CoRPoReAl devra formaliser cette théorie en restant le plus près du formalisme de Terese [1]. Il s'agira alors d'étudier la notion résultante de confluence, que nous appelons confluence randomisée, et de valider ou d'invalider les théorèmes et méthodologies fondamentales associés (Newmann, paires critiques, chasse de diagrammes...).

Membres du projet	Qualité / Titre	e-mail	Employeur	Unité de recherche
Flavien BREUVART	Maître de Conférences	flavien.breuvart@lipn.univ-paris13.fr	Université Paris 13	UMR 7030

### Description du projet (5 pages maximum)

Cette partie présentera :

- les objectifs scientifiques et technologiques : décrire l'ambition et la nouveauté du projet mais aussi évaluer le potentiel de rupture/d'avancée dans le domaine,
- la pertinence du projet : le projet est-il en lien avec les sciences de l'information et leurs interactions ? Quelles sont les retombées attendues du projet ?
- le programme de recherche : verrous scientifiques, méthodologie et programme prévisionnel. Donner la liste des livrables du projet (publications, logiciels, etc. . .).
- la cohérence du projet : indiquer les compétences requises pour mener à bien le projet, montrer au besoin la complémentarité et les apports de chaque membre, donner un court curriculum du porteur et lister quelques références pertinentes dans le domaine en lien direct avec la proposition.
- la justification détaillée du budget demandé.

CoRPoReAl facilitera le développement de la formalisation, la vérification et l’optimisation des langages probabilistes, thème devenu critique avec l’utilisation croissante de programmes randomisés, dans le domaine de l’apprentissage notamment.

Il n’y a actuellement aucun formalisme permettant de comparer, à un niveau abstrait, des stratégies d’évaluation et de compilation de langages probabilistes. En particulier, il est très difficile d’évaluer la robustesse de la sémantique de ces langages face aux optimisations de compilation. CoRPoReAl devra remédier à ce manque.

Un formalisme abstrait commun permettra, aussi, d’engager des collaborations avec la communauté des automates probabilistes; collaborations qui déboucheront sur des avancés réciproques en plus d’un rapprochement des communautés.

## 1 Objectifs scientifiques et technologiques

**Background.** Nous nous plaçons ici dans le cadre de la théorie de la réécriture présenté dans le Terese [1]. En particulier, nous avons en tête l’étude des systèmes de réécriture de termes (TRS) et leur version d’ordre supérieur (HRS), mais le projet ayant pour but de poser les bases d’une nouvelle technologie, nous resterons principalement dans le cadre de systèmes abstraits (ARS).

La réécriture est un domaine vaste, mais on peut le réduire à une étude du non-déterminisme résultant d’une séquence de choix arbitraires dans l’évaluation d’un objet dont la sémantique est plus restreinte. Il s’agit, en particulier, de prouver des propriétés de confluence, formalisant l’idée que les choix non-déterministes n’impactent que l’efficacité de la réduction, pas la sémantique (le résultat).

L’utilisation de propriétés fines de théorie de la réécriture, telles que la confluence, la normalisation ou la standardisation, a été crucial pour l’étude et l’engouement porté au  $\lambda$ -calcul et à la programmation fonctionnelle en général (ainsi que dans d’autres domaines).

Même si l’implémentation d’un langage de programmation nécessite de fixer une stratégie de réduction (entendre déterminisation du processus de choix des réductions), la présence de ces théorèmes en background permet une grande flexibilité sémantique ainsi que des optimisations importantes du compilateur et des algorithmes.

Lorsque l’on introduit des opérateurs probabilistes dans un langage [2], par contre, il est habituel de fixer une stratégie de réduction sans aucune flexibilité, quitte à complexifier certaines preuves techniques. La raison principale est en fait très simple : Le manque de formalisme standard et efficace.

**Problématique.** CoRPoReAl se propose de poser un cadre formel et d’y développer une théorie propre inspirée de la théorie classique de la réécriture. Il s’agira d’étudier les théorèmes classiques comme Knuth–Bendix ou Newmann, mais aussi les techniques de preuves comme les paires critiques, la réduction parallèle, ou la chasse de diagrammes.

Dans un premier temps, nous ne considérerons que des système manipulant des distributions dénombrables (non continues). La généralisation aux distributions continues introduit des difficultés supplémentaires qui ne sont pas au coeur de CoRPoReAl et qui sont l’objet de recherches parallèles.

Il faut avant tout se fixer ce qu’est un langage probabiliste, ou, pour abstraire le raisonnement, un

$\beta: \quad (\lambda x.s)t \rightarrow s[t/x]$	$\text{cont}: \quad \text{op} \rightarrow \text{S}(\text{op})$	$r_{\text{tt}}: \quad X \rightarrow \left\langle \frac{1}{2}X + \frac{1}{2}\text{tt} \right\rangle$
$\text{bin}: \quad s \oplus t \rightarrow \left\langle \frac{1}{2}s + \frac{1}{2}t \right\rangle$	$\text{stop}: \quad \text{op} \rightarrow \text{susp}$	$r_{\text{ff}}: \quad X \rightarrow \left\langle \frac{1}{2}X + \frac{1}{2}\text{ff} \right\rangle$
$\text{back}: \quad \text{susp} \rightarrow \left\langle \frac{1}{2}\text{susp} + \frac{1}{2}\text{op} \right\rangle$		
(a) $\lambda$ -calcul avec choix binaire	(b) Comptage avec suspension	(c) Span continu

FIGURE 1 – Exemples de systèmes de réductions probabilistes

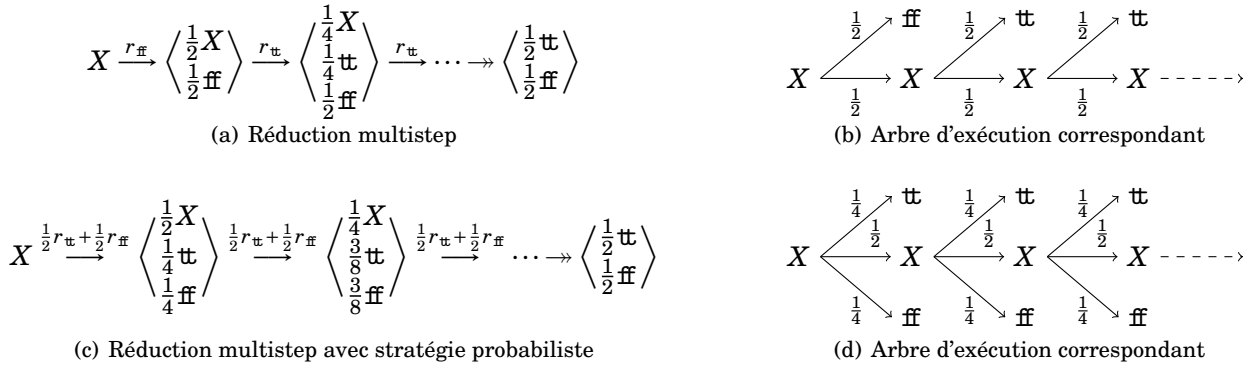


FIGURE 2 – Exemples de réductions multistep et arbres d’exécution associés (Système 1(c))

ARS probabiliste (PARS). On considérera un jeu à deux joueurs dit Arthur-Merlin, avec un joueur non-déterministe (Merlin) qui choisit une règle de réduction (et un redex) et un joueur probabiliste (Arthur) qui applique la réduction. Formellement, on considère qu’une règle est alors de la forme  $t \rightarrow \mu$  où  $\mu$  est une distribution de termes décrivant le tirage d’Arthur. Quelques exemples sont exposés en Figure 1.

Il faudrait alors pouvoir décrire ce que l’on entend par confluence. Par exemple, le système décrit en figure 1(b) décrit un système intuitivement confluent : toute interruption peut être rétablie avec probabilité 1 (mais en temps non borné). Le système figure 1(a), lui, n’est pas confluent puisque  $\langle \lambda x.xx \rangle (s \oplus t)$  peut réduire en  $\langle \frac{1}{2}ss + \frac{1}{2}st \rangle$  ou en  $\langle \frac{1}{4}ss + \frac{1}{4}st + \frac{1}{4}ts + \frac{1}{4}tt \rangle$ , ce qui n’est pas récupérable. Par contre, si l’on restreint l’application de la règle *bin* à des positions linéaires, on récupère bien la confluence.

La généralisation probabiliste des théorèmes classiques, même dans ce cadre est loin d’être triviale. Par exemple, le lemme de Newman est faux lorsque étendu naïvement. En effet, prenons le système de la figure 1(c). Comme décrit figure 2(a) (ou figure 2(b)), quel que soit le premier coup de Merlin il lui est possible de continuer à jouer de sorte à ce qu’un observateur voit ne voit qu’un résultat vrai ou faux à probabilité égale. On a donc une confluence faible. Or, on peut atteindre  $\text{tt}$  et  $\text{ff}$  depuis  $X$  avec probabilité 1... on n’a donc pas confluence.

## 2 Pertinence du projet

**Retombés scientifiques.** En plus de l’apport direct à la formalisation, la vérification et l’optimisation des langages probabilistes ; des retombés importantes sont attendues du côté de la communauté des automates probabilistes.

En effet, les systèmes avec alternance entre choix probabilistes et choix non-déterministes sont très difficiles à étudier et les résultats sont éparpillés. Notre formalisme transversal permettra d’ouvrir de nouvelles perspectives pour des classes spécifiques de tels systèmes.

**Originalité.** Du point de vue des théoriciens des langages de programmation, un langage probabiliste est toujours associé à une stratégie de réduction [3]. Pourquoi ? Simplement du fait de la complexité résultant de l’interaction entre non-déterminisme et probabilités. En réalité, personne ne semble avoir abordé le problème sérieusement.

Du point de vue des automaticiens, si les automates probabilistes sont très étudiés [4], la notion résultante de confluence n’y est pas du tout naturelle. En effet, la confluence induit un changement implicite de comportement du joueur non-déterministe qui a fait certains coups (premier split) puis essaie de se rattraper ; alors que le joueur non-déterministe est généralement considéré comme une entité omnipotente (qui peut être angélique ou démoniaque). Une version de la confluence a néanmoins été étudiée récemment par Timmer, Katoen et van de Pol [5, 6]. Malheureusement, elle est trop spécifique pour transporter aucun résultat traditionnel de la théorie de la réécriture.

**Faisabilité.** Si l’alternance Arthur-Merlin est si complexe, il est raisonnable de se poser la question de la faisabilité de notre approche. Heureusement, un système confluent n’est pas fondamentalement non-déterministe puisque celle-ci formalise l’idée d’une sémantique déterminisme. Cela est d’autant plus vrai que dans un langage de programmation la confluence est structurelle : les règles appliquées sont fondamentalement indépendantes (modulo de légères modifications). Par conséquent, un système confluent randomisé ne devrait pas être beaucoup plus complexe à étudier qu’un système Arthur (avec randomisation mais sans non-déterminisme).

**Dynamique de la recherche.** CoRPOReAl permettra d’initier des collaborations entre automaticiens probabilistes et théoriciens de la preuve. Le porteur, déjà connu de la seconde communauté, devrait acquérir une légitimité suffisante dans le seconde pour, dans un futur proche, monter des projets de rapprochement.

**Retombés pour le porteur.** Le porteur compte ainsi consolider le plus récent de ses trois domaines de recherche, à savoir l’étude de systèmes probabiliste à l’aide des outils de la théorie de la preuve. A l’aboutissement du projet, il devrait y avoir suffisamment de résultats pour demander des fonds plus conséquents.

### 3 Programme de recherche

**WP-0. Formalisation de la réduction multistep (en cours).** Le premier pas est de formaliser la relation de multistep  $\rightarrow$  qui lie un terme  $s$  aux distributions  $v$  accessibles après applications successives de  $\rightarrow$ . Cette relation doit

- être réflexive au sens où  $s \rightarrow \langle s \rangle$  où  $\langle s \rangle$  est un Dirac,
- être transitive par composition point à point,
- contenir  $\rightarrow$ .

Le premier candidat est donc la clôture transitive  $\bigcup^n (\rightarrow^n)$ , mais cette relation a plusieurs défauts :

- La relation  $\rightarrow$  est avant tout une notion d’observation face à un adversaire non-déterministe omnipotent. Cela veut dire que l’opposant a le droit de faire des tirages aléatoires lui aussi. De ce fait, on peut considérer l’application de n’importe quelle combinaison linéaires de règles, comme montré dans la figure 2(c).
- Si on veut pouvoir parler de normalisation, il faut que notre multistep inclue la notion standard d’évaluation, en particulier il faut que  $X \rightarrow \langle true \rangle$  dans le système de la figure 1(c).

Pour répondre à la première condition, on introduit une relation  $\rightsquigarrow$  relevée aux distributions de termes comme suit :  $\mu \rightsquigarrow v$  s’il existe  $(p_i, s_i \rightarrow v_i)_{i \in I}$  et  $(p_j, s_j)_{j \in J}$  pour  $I, J$  dénombrables et disjoints tels que<sup>1</sup>  $\mu = \sum_{i \in I \cup J} p_i \langle s_i \rangle$ ,

et  $v = \sum_{i \in I} p_i v_i + \sum_{j \in J} p_j \langle s_j \rangle$ .

Dans certains cas on peut (et on veut) se passer de la seconde condition, c’est pourquoi on étudie les deux relations :

- $\rightarrow^*$  définit par  $s \rightarrow^* v$  lorsque  $\langle s \rangle \rightsquigarrow^n v$  pour un certain  $n$  ; obtenue après réductions aléatoires suivant une stratégie arbitrairement complexe (potentiellement infinie, non calculable et qui peut elle même introduire des choix aléatoires) mais bornée en temps.
- $\twoheadrightarrow$  qui est la complétion de  $\rightarrow^*$  ; obtenue après réduction aléatoire suivant une stratégie arbitrairement complexe sans borne temporelle..

**WP-1. Formalisation des cellules.** La principale difficulté pour généraliser les théorèmes classiques est de savoir comment et quand remplir les cellules avec de petites cellules (chasse de diagrammes) car il faut bien souvent une infinité convergente de cellules.

En particulier, on voudrait prouver que la propriété du diamant implique la confluence. C’est à dire que si tout  $\rightarrow$ -pic de la forme  $v_1 \leftarrow s \rightarrow v_2$  est clos par une  $\rightarrow$ -vallée  $v_1 \rightarrow \mu \leftarrow v_2$  ; alors tout  $\twoheadrightarrow$ -pic de la forme  $v_1 \leftarrow s \rightarrow v_2$  est clos par une  $\twoheadrightarrow$ -vallée  $v_1 \twoheadrightarrow \mu \leftarrow v_2$ .

---

1. On rappelle que  $\langle s_i \rangle$  représente le Dirac en  $s_i$ .

Il s’agirait aussi de prouver un lemme de Newmann. On a vu que la généralisation intuitive est fautive, au sens où la confluence faible<sup>2</sup> n’implique pas la confluence forte, même dans un système terminant quasi sûrement.<sup>3</sup> On conjecture, néanmoins, des versions plus faibles utilisant des variantes de la confluence faible : la  $\rightarrow^*$ -confluence faible voir même une version mixte qui demande à ce que les pics de la forme  $v_1 \leftarrow s \rightarrow v_2$  soient clos par deux réductions dont l’une est bornée ; c’est à dire que  $v_1 \rightarrow^* \mu \leftarrow v_2$  ou  $v_1 \rightarrow \mu^* \leftarrow v_2$ .

**WP-2. Interprétation coalgébrique.** La relation  $\twoheadrightarrow$  dispose d’une interprétation épurée en théorie des coalgèbres, comme morphisme terminale du système vu comme une coalgèbre de trace [HasuoK07a] dans la catégorie de Kleisli pour la monade probabilité+non-déterminisme (ensembles convexes et fermés de distributions). Cette représentation est utile lorsque l’on travaille à haut niveau d’abstraction car elle induit des théorèmes limites pour passer de  $\rightarrow$  à la réduction multistep  $\twoheadrightarrow$ .

Il s’agit d’une étape importante et difficile, en particulier pour attaquer le WP-3. Nous verrons plus loin qu’un collaboration au Japon est prévu pour travailler sur ce point.

Une interprétation coalgébrique permettrait de généraliser nos résultats au delà du cadre de la réécriture probabiliste et de traiter des cas similaires, comme la réécriture avec probabilités continues, la réécriture à plusieurs joueurs (on remplace Arthur par un joueur angélique/may ou démoniaque/must), la réécriture quantique (Arthur peut faire des opérations quantiques) ou des PARS avec des traces plus complexes.

**WP-3. Transfert de technologies.** Une fois ces trois points traités, nous pourrions nous atteler au transfert de nombreux théorèmes et méthodes depuis le cadre classique de la réécriture (ou le cadre de réécriture infinitaire lorsque le lemme de Newmann fait défaut).

En particulier, nous aborderons les problèmes suivants :

- Orthogonalité : Dans le cas particuliers des systèmes orthogonaux, la confluence sera un corollaire direct des résultats obtenus dans le WP-2. La notion d’orthogonalité est néanmoins très restrictive, et il s’agirait d’essayer de l’affaiblir. En particulier, on voudrait prouver la confluence du système de la figure 1(b) à l’aide de telles méthodes.
- Confluence du  $\lambda$ -calcul probabiliste (Figure 1(a)) : Fausse dans le cas général, la confluence est conjecturée avec les restrictions CbN (interdiction de faire des  $\beta$ -réductions en positions non linéaires) ou CbV (interdiction de faire des  $\beta$ -réductions si le membre droit peut être probabiliste).
- Définitions et techniques de standardisation. On parle de standardisation lorsqu’il est possible de décrire, pour chaque paire  $s, v$  tel que  $s \twoheadrightarrow v$ , une unique réduction “de gauche à droite”. *A priori*, cette définition fait sens dans le cas probabiliste ; peut-on y relever les preuves standards ?

**Délivrables.** Nous pouvons attendre un petit article sur la formalisation et ses propriétés, ainsi que un ou deux plus conséquents sur le transfert de technologies.

## 4 Cohérence du projet

**Le coordinateur.** Le coordinateur est maître de Conférences au LIPN (Université Paris 13) depuis septembre 2016, 10 mois, seulement, après sa soutenance de thèse à l’université Paris Diderot.

Sa production scientifique est très dynamique et constituée d’articles dans des conférences/journaux internationaux (LICS, FoSSaCS, ICFP, LMCS). Elle est de plus bien équilibrée entre collaborations (8 collaborateurs différents), travaux indépendants (prix Kleene pour le meilleur papier étudiant dans la conférence LICS), et diversité thématique (4 thématiques différentes).

Au travers de ses publications [7–10], il a montré son expertise sur les thèmes principaux du projet : HRSs, confluence, systèmes de réécriture probabilistes ainsi que la théorie des catégories.

---

2. Les  $\rightarrow$ -pic clos par des  $\twoheadrightarrow$ -vallées.

3. Un système termine presque sûrement lorsque  $s \twoheadrightarrow \mu$  pour  $\mu$  composé uniquement de formes normales.

**Ichiro Hasuo.** Ichiro Hasuo est un expert des structures coalgébriques [11]. Les résultats préliminaires sur le WP.2, en particulier, sont dérivés de ses résultats de thèse sur l'interprétation coalgébriques des sémantiques de traces de chaînes de Markov [12].

Ichiro Hasuo est aussi intéressé par les thématiques de la théorie de la programmation [13] et a accepté de collaborer avec le porteur.

**Stagiaire.** Le porteur recrutera un.e stagiaire de M2 cet été (payé.e par l'université Paris 13) pour travailler sur la thématique. Si le stage se passe bien, une thèse pourrait faire suite (sur fonds de l'école doctorale cette fois).

## 5 Budget (6k€)

**Collaboration avec le Japon (3k€).** Le porteur prévoit un séjour de deux semaines au Japon pour collaborer avec Ichiro Hasuo. Sont budgétisés : 700€ de frais de transport et 2300€ pour 12 nuits au tarif journalier.

**Conférence internationale (2k€).** Le porteur compte assister à la conférence LICS (600€) à Oxford comme partie de FLOC, ainsi qu'à deux jours de workshop (300€). Sont aussi budgétisés 200€ de transport et 6 à 7 nuits sur place (900€ au tarif journalier CNRS).

**Matériel (1k€).** Le porteur achètera un ordinateur portable moyenne gamme pour ses déplacements.

## 6 Bibliographie

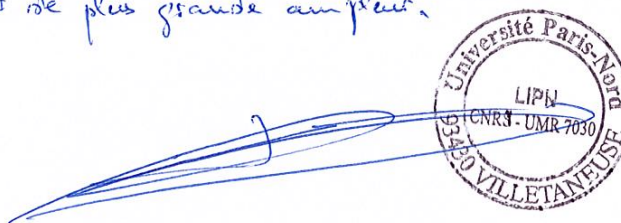
- [1] M. BEZEM, J. W. KLOP et R. de VRIJER, *Term rewriting systems*. Cambridge University Press, 2003.
- [2] N. SAHEB-DJAHROMI, "Probabilistic LCF", dans *MFCS*, 1978, p. 442–451.
- [3] U. DAL LAGO et M. ZORZI, "Probabilistic operational semantics for the lambda calculus", *RAIRO - Theor. Inf. and Applic.*, n° 3, p. 413–450, 2012.
- [4] M. O. RABIN, "Probabilistic automata", *Information and control*, n° 3, p. 230–245, 1963.
- [5] M. TIMMER, M. STOELINGA et J. van de POL, "Confluence reduction for probabilistic systems", dans *TACAS*, P. A. ABDULLA et K. R. M. LEINO, édés., Springer, 2011, p. 311–325, ISBN : 978-3-642-19834-2. adresse : [https://doi.org/10.1007/978-3-642-19835-9\\_29](https://doi.org/10.1007/978-3-642-19835-9_29).
- [6] M. TIMMER, J. KATOEN, J. van de POL et M. STOELINGA, "Confluence reduction for markov automata", *Theor. Comput. Sci.*, p. 193–219, 2016. adresse : <https://doi.org/10.1016/j.tcs.2016.01.017>.
- [7] F. BREUVART, U. D. LAGO et A. HERROU, "On higher-order probabilistic subrecursion", dans *FOSSACS*, J. ESPARZA et A. S. MURAWSKI, édés., 2017, p. 370–386, ISBN : 978-3-662-54457-0. adresse : [https://doi.org/10.1007/978-3-662-54458-7\\_22](https://doi.org/10.1007/978-3-662-54458-7_22).
- [8] F. BREUVART et U. D. LAGO, "On intersection types and probabilistic lambda calculi", submitted, adresse : <https://lipn.univ-paris13.fr/~brevuart/articles/ProbIntTyp.pdf>.
- [9] F. BREUVART, "On the characterization of models of  $H^*$ ", dans *CSL-LICS*, T. A. HENZINGER et D. MILLER, édés., Kleene award, ACM, 2014, 24 :1–24 :10, ISBN : 978-1-4503-2886-9. adresse : <http://doi.acm.org/10.1145/2603088.2603111>.
- [10] M. GABOARDI, S. KATSUMATA, D. A. ORCHARD, F. BREUVART et T. UUSTALU, "Combining effects and coefficients via grading", dans *ICFP*, J. GARRIGUE, G. KELLER et E. SUMII, édés., ACM, 2016, p. 476–489, ISBN : 978-1-4503-4219-3. adresse : <http://doi.acm.org/10.1145/2951913.2951939>.
- [11] C. CÎRSTEA, S. SHIMIZU et I. HASUO, "Lattice-theoretic progress measures and coalgebraic model checking", dans *POPL*, R. BODÍK et R. MAJUMDAR, édés., ACM, 2016, p. 718–732, ISBN : 978-1-4503-3549-2. adresse : <http://doi.acm.org/10.1145/2837614.2837673>.
- [12] I. HASUO et Y. KAWABE, "Probabilistic anonymity via coalgebraic simulations", dans *ESOP*, Springer, Berlin, 2007, p. 379–394.
- [13] K. MUROYA, N. HOSHINO et I. HASUO, "Memoryful geometry of interaction II : recursion and adequacy", dans *POPL*, R. BODÍK et R. MAJUMDAR, édés., ACM, 2016, p. 748–760, ISBN : 978-1-4503-3549-2. adresse : <http://doi.acm.org/10.1145/2837614.2837672>.

### Les financements complémentaires au projet

Financier	Montant	Financement acquis ou demandé	Nature de la dépense financée

### Avis et visa du directeur d'unité

Avis très favorable . Le projet, porté par un jeune HCF, vise à généraliser la réécriture à des langages probabilistes afin de développer leurs formalisation, vérification et optimisation. Il devra déboucher sur la soumission d'un projet de plus grande ampleur.



A handwritten signature in blue ink is written over a circular stamp. The stamp contains the text: Université Paris-Nord, LIPN, CNRS - UMR 7030, and 93430 VILLETANEUSE.