

**Rapport d'examineur sur une demande
de subvention d'un programme de recherche**
CRSNG – janvier 2002

Examineur : Malek Boualem (France Télécom R&D, France)
No NIP du CRSNG : 256425

Candidat : XXXXX
No de la demande : 249630
Nom du programme : RGPIN
No du comité : 331
Durée du programme de recherche : 4 ans

Titre de la proposition :
Amélioration de la traduction probabiliste par un modèle Analyse-Transfert-Génération.

1. Profil du chercheur candidat

Connaissances, expertise et expérience :

Le chercheur a travaillé dans plusieurs établissements universitaires et laboratoires de recherche en France, en Suisse, en Suède et au Canada. Actuellement, il occupe un poste de professeur titulaire. Cette expérience est incontestablement très profitable pour ses activités scientifiques. Par ailleurs, le chercheur est familier aussi bien avec le domaine du traitement automatique de l'écrit qu'avec celui de l'oral (reconnaissance de la parole). Cette expertise multidisciplinaire est sans doute très profitable à d'autres domaines et à d'autres chercheurs. En outre, elle aura des incidences substantielles dans l'évolution des domaines de l'écrit et de l'oral. En effet, ces deux domaines tendent à se rapprocher autour de services offrant une qualité, en constante progression, de la communication Homme-Machine.

Qualité des contributions et leur incidence :

Les contributions du chercheur concernent une multitude de domaines ayant un lien plus ou moins rapproché du domaine faisant l'objet de son programme de recherche (alignement multilingues, traduction assistée par ordinateur, mémoires de traduction, etc.). En outre, les contributions publiées dans des revues ou dans des ouvrages collectifs témoignent de leur qualité. Le chercheur pourra sans doute tirer profit de ces travaux antérieurs dans le cadre de son programme de recherche. Les contributions du chercheur sont également profitables à d'autres domaines de recherche abordés ou non par les membres de son laboratoire, tels que la prosodie dans la reconnaissance de la parole.

Complémentarité des membres du groupe :

Le projet de recherche proposé s'insère de façon complémentaire et homogène avec les travaux menés par les chercheurs du laboratoire de rattachement du candidat (outils d'aide à la traduction, concordances, appariement multilingue de textes, génération automatique de textes, etc.).

2. Mérite de la proposition

Originalité et innovation :

Plusieurs méthodologies ont été proposées dans le domaine de la traduction automatique du langage naturel depuis plus de cinquante ans. Parmi ces méthodologies, l'approche statistique et probabiliste, initiée par IBM, a rapidement connu ses limites. Par ailleurs, l'une des approches les plus suivies est celle dite à transfert et pour laquelle certains systèmes ont vu le jour et ont été largement diffusés (sur la toile notamment). Ces systèmes n'offrent également pas la qualité escomptée.

Le projet de recherche décrit ici, propose de fusionner et tirer profit de ces deux méthodes de traduction. Il s'agit là d'un véritable projet de recherche innovant et d'un challenge scientifique fort intéressant.

D'autre part, à travers ce projet, le candidat anticipe, avec mesure, l'intérêt croissant de la communauté scientifique de son domaine pour le formalisme des grammaires probabilistes. Ceci témoigne de la maîtrise du chercheur du contexte et des portées de son programme de recherche.

Importances des travaux et contributions prévues à la recherche :

Il s'agit d'un programme de recherche pour lequel le chercheur a déjà amorcé des réflexions et des travaux concrets (c.f. communications et publications sur le sujet, en particulier celles qui concernent la traduction probabiliste et les grammaires probabilistes). Ce projet ne se situe donc pas dans un registre complètement exploratoire où les risques peuvent être plus ou moins importants.

Le chercheur possède déjà une expérience substantielle dans le domaine traité par le programme de recherche proposé et faisant l'objet de cette demande de subvention. En effet, il a déjà participé, au sein de son laboratoire actuel, à un projet de recherche sur la traduction interactive, en étendant le pouvoir prédictif du modèle de traduction probabiliste. Ce travail a fait l'objet de plusieurs publications. Le chercheur s'intéresse également au modèle basé sur les mémoires de traduction, dont il étudie la réutilisabilité dans le cadre de ses travaux actuels. D'autre part, le chercheur a travaillé sur la mise en œuvre de systèmes d'alignement de corpus multilingues. Ce travail, qui a été évalué de façon positive dans le cadre d'une action d'évaluation de plusieurs systèmes d'alignement, est également très pertinent pour le programme de recherche proposé. Enfin, au sein de son laboratoire actuel, il est associé à un certain nombre de projets qui traitent des problématiques très liées au programme de recherche qu'il propose. L'un de ces projets a obtenu un financement de la part de la commission européenne.

Par ailleurs, le candidat a déjà clairement identifié les laboratoires avec lesquels il envisage de mener une collaboration, ainsi que les ressources exploitables et réutilisables dans le cadre de son projet.

Clarté des objectifs, pertinence de la méthodologie et faisabilité :

Le candidat situe les objectifs de son programme de recherche de façon claire et justifiée (même si les besoins en traduction de bonne qualité ne nécessitent plus d'être justifiés). Il montre une bonne connaissance des réalisations et des travaux de recherche menés par ailleurs et il explique, de façon réaliste, les avancées technologiques qu'il envisage d'apporter à travers son projet de recherche. La méthode proposée est présentée de manière structurée. On y distingue les fonctionnalités des différents modules du système global, ainsi que le processus de traitement. Cependant, bien que la faisabilité du système global ne soit pas suffisamment justifiée, cela n'altère pas les ambitions de ce programme qui est, avant tout, un programme de recherche.

Points faibles du programme de recherche :

Certains aspects auraient mérité une plus grande attention dans la description du programme de recherche. Cependant, ces critiques sont exprimées dans un but constructif, afin que le chercheur puisse, éventuellement, en tenir compte dans son activité de recherche. A mon sens, elles ne doivent pas remettre en cause la possibilité du financement de ce programme de recherche.

- Le programme de recherche ne spécifie pas clairement s'il se situe dans une approche pour la traduction automatique ou pour la traduction assistée par ordinateur. L'intervention de l'expert ou de l'utilisateur n'est pas clairement identifiée dans le processus de traduction.
- La méthode proposée dans ce programme de recherche ne prévoit pas d'intégrer des mécanismes pour la désambiguïsation (désambiguïsation lexicale par exemple). Le chercheur situe bien son travail au niveau syntaxique et il prévoit de mettre en place un décodeur pour le choix de la meilleure traduction. Il serait sans doute très intéressant de voir les résultats de ce travail.
- Le programme de recherche ne met pas suffisamment en évidence une stratégie pour tirer profit des nombreux échecs dans ce domaine.
- Le programme de recherche ne prévoit pas de transfert technologique à son terme.

3. Contribution à la formation de personnel hautement qualifié (PHQ)

Le candidat a une expérience pédagogique notable dans le milieu universitaire en Europe et au Canada, sur laquelle il semble pouvoir s'appuyer, pour un encadrement efficace, dans ce cadre spécialisé de son programme de recherche. La proposition pour l'encadrement et la formation de personnels hautement qualifiée me paraît judicieuse. En effet, les modules sur lesquels travailleront les différentes personnes sous sa responsabilité, sont suffisamment indépendants (et spécialisés) pour en permettre une bonne maîtrise. En outre, ces modules sont prévus pour être agréés de façon homogène pour la constitution de l'ensemble du système de traduction.

4. Financement demandé

Le budget prévisionnel annoncé par le chercheur, ainsi que les postes identifiés, me semblent être dans les bonnes proportions. Mais il ne prévoit pas de budget pour la propriété intellectuelle.

5. Conclusion

Compte tenu des éléments décrits ci-dessus, je pense que le programme de recherche proposé mérite amplement d'être soutenu.