**Aide à la finalisation de corpus oraux ou multimodaux pour diffusion, valorisation et dépôt pérenne**

Le consortium IRCOM de la TGIR Corpus et l’EquipEx ORTOLANG s’associent pour proposer une aide technique et financière à la finalisation de corpus de données orales ou multimodales à des fins de diffusion et pérennisation par l’intermédiaire de l’EquipEx ORTOLANG. Cet appel ne concerne pas la création de nouveaux corpus mais la finalisation de corpus existants et non-disponibles de manière électronique. Par finalisation, nous entendons le dépôt auprès d’un entrepôt numérique public, et l’entrée dans un circuit d’archivage pérenne. De cette façon, les données de parole qui ont été enrichies par vos recherches vont pouvoir être réutilisées, citées et enrichies à leur tour de manière cumulative pour permettre le développement de nouvelles connaissances, selon les conditions d’utilisation que vous choisirez (sélection de licences d’utilisation correspondant à chacun des corpus déposés).

Cet appel d’offre est soumis à plusieurs conditions (voir ci-dessous) et l’aide financière par projet est limitée à 3000 euros. Les demandes seront traitées dans l’ordre où elles seront reçues par l’ IRCOM. Les demandes émanant d’EA ou de petites équipes ne disposant pas de support technique « corpus » seront traitées prioritairement. Les demandes sont à déposer du 1er septembre 2013 au 31 octobre 2013. La décision de financement relèvera du comité de pilotage d’IRCOM. Les demandes non traitées en 2013 sont susceptibles de l’être en 2014. Si vous avez des doutes quant à l’éligibilité de votre projet, n’hésitez pas à nous contacter pour que nous puissions étudier votre demande et adapter nos offres futures.

Pour palier la grande disparité dans les niveaux de compétences informatiques des personnes et groupes de travail produisant des corpus, L’ IRCOM propose une aide personnalisée à la finalisation de corpus. Celle-ci sera réalisée par un ingénieur IRCOM en fonction des demandes formulées et adaptées aux types de besoin, qu’ils soient techniques ou financiers.

Les conditions nécessaires pour proposer un corpus à finaliser et obtenir une aide d’IRCOM sont :

* Pouvoir prendre toutes décisions concernant l’utilisation et la diffusion du corpus (propriété intellectuelle en particulier).
* Disposer de toutes les informations concernant les sources des corpus et le consentement des personnes enregistrées ou filmées.
* Accorder un droit d’utilisation libre des données ou au minimum un accès libre pour la recherche scientifique.

Les demandes peuvent concerner tout type de traitement : traitements de corpus quasi-finalisés (conversion, anonymisation), alignement de corpus déjà transcrits, conversion depuis des formats « traitement de textes », digitalisation de support ancien. Pour toute demande exigeant une intervention manuelle importante, les demandeurs devront s’investir en moyens humains ou financiers à la hauteur des moyens fournis par IRCOM et ORTOLANG.

IRCOM est conscient du caractère exceptionnel et exploratoire de cette démarche. Il convient également de rappeler que ce financement est réservé aux corpus déjà largement constitués et ne peuvent intervenir sur des créations ex-nihilo. Pour ces raisons de limitation de moyens, les propositions de corpus les plus avancés dans leur réalisation pourront être traitées en priorité, en accord avec le CP d’IRCOM. Il n’y a toutefois pas de limite « théorique » aux demandes pouvant être faites, IRCOM ayant la possibilité de rediriger les demandes qui ne relèvent pas de ses compétences vers d’autres interlocuteurs.

Les propositions de réponse à cet appel d’offre sont à envoyer à ircom.appel.corpus@gmail.com. Les propositions doivent utiliser le formulaire de deux pages figurant ci-dessous. Dans tous les cas, une réponse personnalisée sera renvoyée par IRCOM.

Ces propositions doivent présenter les corpus proposés, les données sur les droits d’utilisation et de propriétés et sur la nature des formats ou support utilisés.

Cet appel est organisé sous la responsabilité d’IRCOM avec la participation financière conjointe de IRCOM et l’EquipEx ORTOLANG.

Pour toute information complémentaire, nous rappelons que le site web de l'Ircom (<http://ircom.corpus-ir.fr>) est ouvert et propose des ressources à la communauté : glossaire, inventaire des unités et des corpus, ressources logicielles (tutoriaux, comparatifs, outils de conversion), activités des groupes de travail, actualités des formations, ...

L'IRCOM invite les unités à inventorier leur corpus oraux et multimodaux - 70 projets déjà recensés - pour avoir une meilleure visibilité des ressources déjà disponibles même si elles ne sont pas toutes finalisées.

Le comité de pilotage IRCOM

Utiliser ce formulaire pour répondre à l’appel : Merci.

Réponse à l’appel à la finalisation de corpus oral ou multimodal

Nom du corpus :

Nom de la personne à contacter :

Adresse email :

Numéro de téléphone :

Nature des données de corpus :

Existe-t’il des enregistrements :

Quel média ? Audio, vidéo, autre…

Quelle est la longueur totale des enregistrements ? Nombre de cassettes, nombre d’heures, etc.

Quel type de support ?

Quel format (si connu) ?

Existe-t’il des transcriptions :

Quel format ? (papier, traitement de texte, logiciel de transcription)

Quelle quantité (en heures, nombre de mots, ou nombre de transcriptions) ?

Disposez vous de métadonnées (présentation des droits d’auteurs et d’usage) ?

Disposez-vous d’une description précise des personnes enregistrées ?

Disposez-vous d’une attestation de consentement éclairé pour les personnes ayant été enregistrées ? En quelle année (environ) les enregistrements ont eu lieu ?

Quelle est la langue des enregistrements ?

Le corpus comprend-il des enregistrements d’enfants ou de personnes ayant un trouble du langage ou une pathologie ?

Si oui, de quelle population s’agit-il ?

 Dans un souci d’efficacité et pour vous conseiller dans les meilleurs délais, il nous faut disposer d’exemples des transcriptions ou des enregistrements en votre possession. Nous vous contacterons à ce sujet, mais vous pouvez d’ores et déjà nous adresser par courrier électronique un exemple des données dont vous disposez (transcriptions, métadonnées, adresse de page web contenant les enregistrements).

Nous vous remercions par avance de l’intérêt que vous porterez à notre proposition. Pour toutes informations complémentaires veuillez contacter Martine Toda martine.toda@ling.cnrs.fr ou à ircom.appel.corpus@gmail.com.