


|   |                                 |   |
|---|---------------------------------|---|
|  | <b>CONTENUS ET INTERACTIONS</b> | Réservé à l'organisme gestionnaire du programme<br>N° de dossier : ANR-08-XXXX-00<br>Date de révision : |
|   | Document de soumission <b>B</b> | <b>Edition 2008</b>   |

|  |  |
|--|--|
| <b>Acronyme/Acronym</b>  | <b>ETAPE</b>                                       |
| <b>Titre du projet/Proposal title</b><br><i>(en français/ inFrench)</i>  | Évaluations en Traitement Automatique de la Parole |
| <b>Titre du projet/Proposal title</b><br><i>(en anglais/ in English)</i> | Evaluations in Automatic Speech Processing         |

## **1. Programme scientifique et technique/Description du projet.** ***Technical and scientific description of the activities***

### **1.1 Problème posé.**

L'accès à des contenus audio ou audiovisuels, professionnels ou personnels, disponibles sur des supports variés, n'est possible que par le développement de technologies efficaces, offrant des possibilités d'accès à la demande dans un but de loisir, d'information ou de culture ou bien pour la gestion du patrimoine numérique. De plus, afin de pouvoir traiter une grande variété de contenus mélangeant productions professionnelles et amateurs, de qualité très diverses et portant sur des sujets extrêmement variés, il est nécessaire de développer des outils très robustes ou encore suffisamment génériques pour s'adapter aux changements de conditions.

Par ailleurs, malgré la diversité des sources (radio, TV ou web) et l'hétérogénéité des contenus (informations, sport, culture, jeux, ...), le traitement de la parole est un élément clé des futures chaînes de l'information numérique en ce sens qu'il permet de comprendre le sens d'un document au-delà de la forme. On se heurte dès lors aux limites des technologies actuelles du traitement de la parole. Si les outils sont maintenant performants sur un type précis de documents, un style de parole (parole préparée), et pour des conditions acoustiques stables, leurs performances diminuent fortement en présence de parole spontanée ou dans des conditions acoustiques nouvelles et l'adaptation des outils existants à de nouvelles conditions d'utilisation est généralement lourde.

C'est autour de ces problèmes que s'articule le projet ETAPE. L'objectif premier est de faire progresser les technologies en traitement automatique de la parole, notamment concernant la parole spontanée et la langue française, par le biais de la production de ressources, par la dynamique de l'évaluation et par la mise en relation des acteurs du domaine et des domaines connexes. De plus, la mise en œuvre d'une campagne d'évaluation permet une mesure objective des performances de ces technologies pour des tâches d'indexation, favorisant ainsi la migration de ces technologies d'un contexte académique vers un contexte industriel.

### **1.2 Contexte et enjeux du projet.**

#### **1.2.1 Contexte scientifique**

Le traitement automatique de la parole, quel que soit le type de parole ou les conditions acoustiques, constitue en enjeu majeur pour l'exploitation des masses de données numériques, notamment pour l'extraction de méta données permettant un accès sémantique aux contenus. Si, comme l'ont montré les précédentes campagnes d'évaluation organisées en France, les technologies de la parole ont atteint une certaine maturité pour des contenus contrôlés, il n'est à l'heure actuelle pas possible d'évaluer de manière objective les performances de ces technologies sur des données représentatives des contenus disponibles au grand public. En particulier, la parole spontanée dans des fonds sonores variés (studio, intérieur, extérieur) est très fortement présente dans ces contenus et pose problème aux technologies actuelles. Bien que les solutions techniques existantes permettent

de traiter des contenus tels que les débats, les interviews, les jeux, voire les podcasts, elles ne sont pas adaptées à ce type de contenu. De plus, il n'existe à l'heure actuelle pas de mesure objective des performances sur de tels contenus. Or ce type de contenus représente une bonne partie des données diffusées la télévision ou sur Internet, données pour lesquels des outils d'indexation et d'enrichissement des contenus sont indispensables. La parole spontanée, dans des conditions acoustiques souvent dégradées, est aussi présente dans la plupart des vidéos personnelles telles que celles présentes sur des sites comme YouTube. À nouveau, disposer d'outils permettant d'accéder à la sémantique de tels documents et mesurer les performances de ces outils présentent un enjeu considérable pour le traitement et la diffusion des contenus numériques.

Les campagnes d'évaluation des technologies vocales, menées aux Etats-Unis depuis les années 90 et en France depuis 1997, ont permis, outre la mesure des performances des technologies pour une tâche à une date donnée, des progrès technologiques considérables dont les raisons sont multiples. La mise à disposition de bases de données annotées, la fédération d'une communauté d'acteurs ou encore la définition de métriques d'évaluation sont parmi les principales. C'est dans cette logique d'évaluation des technologies de la parole que s'inscrit le projet ETAPE.

Nous donnons tout d'abord un aperçu des campagnes d'évaluation des technologies vocales mises en œuvre aux Etats-Unis, en Europe et en France, en mettant en exergue leur impact scientifique et technologique, avant de décrire les enjeux et objectifs de la campagne d'évaluation développée dans le projet ETAPE.

### 1.2.2 État de l'art en évaluation des technologies vocales

Aux États-Unis, le DARPA, relayé par NIST, a organisé une évaluation annuelle des systèmes de transcription des bulletins d'information radios et télés (HUB-4 Broadcast News) de 1996 à 1999. L'intégration d'informations annexes, pour enrichir la transcription, est apparue avec les campagnes annuelles d'extraction automatique de contenu (Automatic Content Extraction, 1999-2008) qui incluent une détection des entités nommées dans les transcriptions automatiques d'émissions et les campagnes annuelles de détection de thème (Topic Detection and Tracking, 1998-2004). Depuis 2002, l'enrichissement des transcriptions avec des informations concernant les locuteurs fait également l'objet d'une campagne d'évaluation (Rich Transcription, 2002-2007 – DARPA Effective Affordable Reusable Speech-To-Text (EARS) Program) qui s'est progressivement élargie au traitement d'autres types de parole : parole conversationnelle téléphonique (à partir d'octobre 2003 après les campagnes Conversational Telephone Recognition 1997-2001) et enregistrement de réunions (Meeting à partir de 2005). Le NIST organise également depuis 1996 des campagnes annuelles portant sur la vérification du locuteur (Speaker Recognition Evaluation) tandis que la segmentation selon le locuteur est présente depuis quelques années dans les évaluations Rich Transcription.

Ces campagnes ont connu un succès marqué, par le nombre et la qualité des participants, et par le nombre de publications ; elles ont également permis de dynamiser les recherches dans le domaine du traitement du langage parlé, en mettant à la disposition des chercheurs des ressources indispensables (corpus, instruments de mesure des performances...) mais également en favorisant les interactions entre les acteurs du domaine. Cependant, les effets de ces campagnes se sont faits sentir essentiellement outre-atlantique, pour des raisons évidentes de proximité et de langue, les ressources proposées portant essentiellement sur l'anglais.

En Europe, le projet THISL (Thematic Indexing of Spoken Language 1997-2000 – EC Research and Technical Development Program) a favorisé le développement de la recherche autour d'applications multimédia basées à la fois sur les techniques de traitement du langage naturel (texte) et la reconnaissance de la parole continue grand vocabulaire (transcription de la parole) dans un but de recherche d'information multimédia utilisant des interfaces vocales. Le programme Cross-Language Evaluation Forum (CLEF) a récemment proposé une tâche de réponse à des questions sur des transcriptions automatiques (QAs) sans attacher d'importance à la transcription (fournie aux participants) ou aux interactions entre transcription et traitement des langues.

En France, une première campagne d'évaluation a été organisée en 1997, dans le cadre des ARC (B1) de l'Agence Universitaire de France (AUF), portant sur la reconnaissance de parole lue. La communauté francophone rassemblée autour de l'étude de la parole et du langage a ensuite tiré de larges bénéfices de la première campagne ESTER (2002-2005), organisée dans le cadre du projet Technolanguage EVALDA et portant sur les bulletins d'informations radiophoniques. De très nombreuses publications ont permis à la communauté d'assurer un rayonnement scientifique national et international : on dénombre plus de 50 publications dans les principales conférences internationales du domaine<sup>1</sup> ! Plusieurs thèses ont également été menées autour des systèmes développés à l'occasion de la campagne ESTER. Outre les progrès scientifiques dans le domaine des STIC et le renforcement de la structuration de cette communauté, une ouverture vers les disciplines des

<sup>1</sup> Une listes de ces publications peut être consultée sur le site ESTER, <http://www.afcp-parole.org/ester/index.html>.

Sciences Humaines et Sociales a été initiée. Par exemple, certains participants à la campagne ESTER se sont fédérés avec des équipes de traitement automatique des langues, pour proposer un projet ANR labellisé en 2006 : le projet EPAC. Celui-ci se focalise sur le traitement de la parole conversationnelle et se base sur les données mises à disposition lors de la première campagne ESTER. L'Action de Recherche Concertée INRIA Rapsodis, alliant sémantique, syntaxe et transcription automatique, bénéficie également des retombées de la campagne ESTER. Dans le même ordre d'idée, des études concernant les alignements phonétiques automatiques de certains des participants de la campagne ESTER et leur utilisation pour des études en phonétique ont été menées par des chercheurs de le Laboratoire de Phonétique et Phonologie (Université de Paris 3). Parmi les retombées de cette même première campagne, et pour montrer combien l'émulation qu'elle a provoquée a été forte, plusieurs équipes ayant pu développer et/ou évaluer leur système de transcription automatique dans le cadre de la campagne ESTER, ont pu ensuite le faire évoluer et le présenter à la campagne européenne TC-STAR 2007 (Speech-to-Speech Translation 2005-2007). Finalement, il convient de souligner que, grâce à l'implication au niveau européen de ELDA dans le domaine de l'évaluation, les données ESTER ainsi que les transcriptions automatiques de certains participants ont servi de base à la tâche « Question Answering on Speech Transcriptions » (QAst) des prochaines évaluations CLEF (Cross-Language Evaluation Forum).

À l'initiative de la Délégation Générale pour l'Armement (DGA) et de l'Association Francophone de la Communication Parlée (AFCP), une deuxième campagne proche de la première campagne ESTER a été organisée avec la participation de ELDA. Démarrée en janvier 2008, la campagne ESTER 2 doit s'achever par une campagne d'évaluation en novembre 2008. Elle porte principalement sur le même type de contenus et de tâches que la première campagne ESTER, à savoir des bulletins d'informations à la radio, avec toutefois l'introduction de voix accentuées, ces dernières posant un véritable défi pour les systèmes de transcription automatique. Quelques émissions autres que les bulletins d'informations sont présentes dans le corpus ESTER 2 mais en quantité très limitée. La campagne ESTER 2 a également été l'occasion d'introduire de nouvelles tâches exploratoires comme la transcription aidée par des données contemporaines. Actuellement dans la phase de développement des systèmes, cette campagne regroupe les acteurs de la première campagne ESTER autour d'un nouvel ensemble de données.

### 1.2.3 Enjeux du projet

Les progrès accomplis grâce aux campagnes ESTER, couplés avec les besoins applicatifs décrits précédemment, appellent à un élargissement conséquent des contenus à d'autres types de données que les informations et à d'autres supports que la radio. Il nous paraît également essentiel d'élargir la palette des tâches évaluées – d'une part par l'introduction de tâches proches d'un contexte applicatif et, d'autre part, par de nouvelles tâches exploitant la transcription – et de poursuivre la politique de mise en relation des communautés scientifiques impliquées.

C'est à ces besoins que souhaite répondre le projet ETAPE à travers l'organisation d'une campagne d'évaluation. ETAPE s'inscrit donc pour une part dans la continuité des campagnes ESTER – en ce qu'elles ont stimulé les équipes de recherche et permis des progrès considérables dans le domaines des technologies du traitement de la parole (transcription automatique, structuration par le locuteur ou encore détection des entités nommées) grâce notamment à la mise disposition et la diffusion de corpus annotés – tout en définissant de nouveaux enjeux comme la parole spontanée, la diversité des contenus ou encore le traitement des zones de parole superposées.

Les enjeux scientifiques du projet sont clairement de faire progresser les technologies du traitement automatique de la parole par le biais de la production de ressources, de la définition de nouveaux problèmes, de l'évaluation et de la structuration de la communauté scientifique. Un enjeu important est également de faire émerger de nouveaux axes de recherche. L'implication de plusieurs communautés scientifiques (traitement de la parole, traitement du langage naturel mais aussi phonétique, linguistique), l'enrichissement des annotations par des alignements phonétiques et des arbres syntaxiques et l'évaluation de tâches à la frontière entre traitement de la parole et traitement des langues sont autant de facteurs favorisant l'émergence d'axes de recherche transverses.

Au-delà de ces enjeux scientifiques, le projet vise à favoriser et à promouvoir l'utilisation des technologies de la parole, pour la langue française, dans des applications de structuration et d'accès aux contenus numériques. Tant les progrès technologiques et scientifiques induits par la campagne ETAPE que la mesure objective des performances faciliteront le transfert de ces technologies vers des applications effectives, comme en témoignent les partenariats académie/industrie noués à l'issue de la campagne ESTER. De tels partenariats sont de nature à renforcer le tissu économique en développant la compétitivité des acteurs nationaux, notamment sur la langue française.

### 1.3 Objectifs et caractère ambitieux/novateur du projet.

L'organisation d'une campagne d'évaluation des systèmes de traitement automatique de la parole, capables de traiter la parole spontanée de manière efficace, quel que soit le contenu et le support, a pour principaux buts de faire progresser les technologies de la parole, notamment concernant la parole spontanée, technologies clés pour les générations futures de systèmes d'indexation et d'accès au contenu, et de poursuivre la promotion d'une dynamique de l'évaluation en France, centrée autour du traitement de la parole de langue française et initiée dans le cadre de l'action Technolangue (2002-2005).

On attend du projet des progrès technologiques conséquents permettant un développement massif des technologies vocales dans la chaîne de traitement des contenus numériques. Dans ce but, le projet s'attaque à deux verrous forts : le traitement de la parole spontanée et la diversité des contenus. Les ressources qui seront développées dans le cadre du projet ETAPE sont destinées à venir compléter les ressources existantes, issues des précédentes campagnes d'évaluation, offrant ainsi les conditions nécessaires pour lever ces verrous.

Par ailleurs, une des limites fortes de l'indexation sémantique de documents sonores est due à la séparation des communautés du traitement de la parole et du traitement des langues. Les solutions techniques actuelles consistent à considérer le texte issu de la transcription comme un texte standard. De plus, les informations non sémantiques comme la présence de musique ou encore les tours de parole ne sont que rarement prise en compte. Le projet ETAPE, en fédérant l'ensemble des acteurs autour de données et de tâches communes, a pour objectif de lever ces cloisonnements qui constituent un frein technologique.

Il est maintenant communément admis que le relatif plafonnement des performances des systèmes de transcription automatique de la parole est en grande partie dû à l'absence de connaissances de haut niveau dans les algorithmes mis en œuvre. En effet, pour des raisons historiques, les méthodes actuelles n'utilisent que peu de connaissance phonétique et phonologique. De même, elles n'intègrent que peu, et de manière très indirecte, de connaissance concernant la syntaxe. À travers l'enrichissement des annotations du corpus ETAPE par des alignements phonétiques et des arbres syntaxiques, le projet a pour double ambition de faire progresser les connaissances en syntaxe et phonétique et de favoriser l'utilisation de ces connaissances dans les systèmes de transcription automatique de la parole, permettant ainsi un saut technologique.

Le projet présente un caractère ambitieux tant au niveau des ressources produites qu'au niveau des tâches évaluées. Le corpus ETAPE est particulièrement ambitieux et innovant en terme des contenus envisagés. En effet, il n'existe à l'heure actuelle pas de corpus regroupant productions professionnelles et amateurs, en français comme dans les autres langues. En dehors du corpus EPAC, il n'existe pas non plus pour le français de corpus contenant une telle proportion de parole spontanée, le corpus ETAPE offrant par ailleurs une diversité de sources, et donc de conditions acoustiques, absente du corpus EPAC. Finalement, le corpus créé contiendra des données provenant de documents amateurs. Bien que nous envisagions de sélectionner ces contenus de manière à garantir un niveau minimum de performance<sup>2</sup>, le traitement de la parole sur de tels contenus n'a à notre connaissance jamais été envisagé. Finalement, certaines des tâches évaluées dans la campagne ETAPE sont ambitieuses par leur aspect novateur (détection des zones de parole superposée, adaptation non supervisée) ou par leur côté proche du transfert (transcription à délai contraint, segmentation thématique).

Notons finalement que le projet doit permettre de mettre à disposition de la communauté scientifique les ressources produites en favorisant un partage des ressources dérivées (résultats de traitements automatiques, graphes de mots, annotations complémentaires). On s'attachera particulièrement à la mise en œuvre d'un mécanisme centralisé de partage des ressources entre participants pendant, mais aussi après, la campagne afin d'assurer la pérennité et la réutilisabilité des ressources.

En résumé, le projet a pour objectifs :

- d'améliorer les performances des systèmes de traitement automatique de la parole, en particulier sur des données proches du contexte applicatif de l'indexation multimédia ;
- de mettre à disposition les ressources langagières vitales ;

<sup>2</sup> De nombreux contenus amateurs sont particulièrement problématiques pour les systèmes de traitement automatique de la parole, en particulier pour la transcription. Si il est envisageable d'évaluer les technologies vocales pour certains de ces contenus, d'autres restent encore trop problématiques pour les systèmes existants rendant toute évaluation sans objet sur ce dernier type de documents.

- de définir de nouvelles problématiques scientifiques liées au type de parole et à la nature hétérogène des données (vidéo personnelles, podcasts) ;
- d'assurer une dynamique et structurer les communautés scientifiques impliquées ;
- de définir une méthodologie d'évaluation pour des tâches à la frontière entre traitement de la parole et traitement du langage naturel (reponctuation, détection des EN, segmentation thématique, etc.).

## 1.4 Positionnement du projet.

Le projet s'inscrit dans la dynamique d'évaluation instaurée par le projet Technolangue EVALDA/ESTER mené entre 2003 et 2005 et prolongée par la campagne ESTER 2 en 2008. Les campagnes d'évaluation ESTER ont permis de nets progrès concernant la segmentation et la transcription des informations radio, et une extension de ces progrès à la parole accentuée est attendue avec ESTER 2. Le projet ETAPE a pour but d'étendre encore ces progrès à des types de parole très spontanée et à une plus grande diversité de conditions acoustiques. Il vise aussi à renforcer les liens avec le domaine du traitement du langage naturel, sachant que la prise en compte des niveaux syntaxique et sémantique deviennent prépondérant pour corriger les erreurs résiduelles des systèmes de reconnaissance.

Le projet ETAPE est complémentaire au projet ANR EPAC. Celui-ci a pour objet central la parole spontanée et va développer des outils de traitement automatique et des corpus, mais a prévu de s'appuyer sur des campagnes d'évaluation extérieures pour la mesure comparative des progrès. ETAPE répond à ce besoin. Réciproquement, ETAPE va bénéficier des corpus produits dans EPAC.

Le projet ETAPE est aussi complémentaire au programme Quaero. Celui-ci couvre le domaine du traitement de la parole multilingue en se focalisant sur les podcasts. Le type précis de données évoluera en fonction des besoins applicatifs définis par les industriels du consortium, et dans le cas du français pourraient coïncider pour partie avec celles de ETAPE. Il prévoit pour ce type de situation de s'appuyer sur les campagnes extérieures existantes. C'est ainsi la souplesse d'organisation de Quaero qui permet de conserver la complémentarité en fonction des orientations des projets connexes. Le fait que la DGA soit responsable de l'évaluation à la fois dans ETAPE et Quaero contribue aussi à garantir la bonne complémentarité des deux projets.

Plus généralement, les partenaires du projet veilleront à la bonne complémentarité avec toute campagne d'évaluation similaire, notamment dans le cadre des programmes DARPA/NIST ou CLEF (Cross-Language Evaluation Forum).

Le projet ETAPE se positionne clairement dans l'axe transversal "Mesure des performances des technologies et systèmes" de l'appel à projets Contenus et Interactions". Il constitue une plate-forme réunissant l'ensemble des acteurs du domaine, même si seuls les organisateurs sont financés au titre du projet ETAPE. ESTER avait déjà réuni une dizaine de participants et de nouveaux ont déjà manifesté leur intérêt pour les nouvelles campagnes.

Le projet ETAPE se rattache à plusieurs axes de l'appel à projets. Les progrès visés en traitement de la parole se rapportent clairement à l'enrichissement des contenus et à la navigation multimédia (axe 3).

L'utilisation des contenus personnels (podcasts, web) en plus des contenus professionnels (radio, TV) doit permettre de limiter l'écart technologique entre ces deux types de contenus (axes 1 et 2).

L'évaluation de tâches en traitement du langage naturel en plus du traitement de la parole et l'enrichissement des corpus à des fins de recherche en phonétique et en linguistique contribuent à structurer l'ensemble des acteurs et favorise l'émergence de thèmes de recherche transverses et de technologies hybrides (axe 3).

## 1.5 Description des travaux : programme scientifique et technique.

Avant de présenter la structuration du projet en tâches, il convient d'exposer le cycle de vie d'une campagne d'évaluation et quelques règles de base que nous entendons respecter, ainsi qu'une rapide description des éléments d'un système de transcription et d'indexation de documents sonore que nous entendons évaluer.

### 1.5.1 Déroulement de la campagne

La campagne d'évaluation, d'une durée totale de deux ans et demi, se divise en cinq phases que nous décrivons ici succinctement.

Une première phase de **conception de la campagne**, en collaboration avec les participants pressentis, a pour objectif d'aboutir à un consensus concernant la définition exacte des éléments évalués, des métriques pour la

mesure des performances et des règles du jeu. Cette phase de conception doit également permettre de mieux cibler et définir les ressources qui seront nécessaires. Le travail de conception de la campagne est un travail collaboratif qui associe au processus l'ensemble des acteurs du domaine afin de bien cibler leurs intérêts respectifs. Ce travail de consultation de la communauté implique la structuration du dialogue, en particulier à travers l'organisation de réunions, ainsi qu'un travail de synthèse dont l'aboutissement est le plan d'évaluation et la spécification des ressources nécessaires.

Une fois clairement définies les ressources nécessaires, la phase de **production des ressources** doit permettre de construire un corpus transcrit qui viendra compléter les données existantes (précédentes campagne d'évaluation ESTER, projet EPAC, etc.) et fournir des données d'apprentissage, de développement et de test pour la campagne. On cherchera à obtenir des données comportant une grande proportion de parole spontanée en favorisant la diversité des sources. En particulier, une partie du corpus sera constituée de données professionnelles issues de la télévision, plus difficiles que les données radiophoniques existantes, tandis que des contenus personnels seront considérés pour la deuxième partie du corpus. Ces données seront annotées en terme de conditions acoustiques (*e.g.* parole/musique, présence de bruits, etc.), de tours de parole et de transcription orthographique selon les standards établis lors des précédentes campagnes. On ajoutera des annotations linguistiques comme les entités nommées ou encore la segmentation thématique. Cette phase de production des corpus prévoit également l'enrichissement des corpus créés ou existants par des annotations phonétiques et syntaxiques. Si ces annotations ne sont pas strictement nécessaires au développement d'un système automatique de segmentation, de transcription et de structuration, elles sont nécessaires pour promouvoir de nouveaux axes de recherche transverses comme la phonétique de corpus (associant la phonétique et le traitement automatique de la parole), l'utilisation de la syntaxe pour la transcription automatique (associant linguistique et traitement automatique de la parole), voire même des études linguistiques plus fondamentales sur la parole spontanée. Enfin, cette phase inclut également la mise au point des outils de mesure des performances.

La phase de **développement des systèmes** permet aux participants désireux de prendre part à la campagne d'évaluation de développer leurs algorithmes et de régler leurs systèmes sur les données de développement diffusés dans le cadre de la campagne, de préférence – mais pas nécessairement – représentatives des données de test. Mises à part les éventuelles corrections et améliorations des transcriptions de référence ainsi que la structuration des échanges de ressources et d'algorithmes entre participants, cette phase ne nécessite pas d'implication forte de la part des organisateurs de la campagne. Notons cependant que, bien que non financée par le projet, la phase de développement possède un coût en terme d'investissement des participants. D'après l'expérience des précédentes campagnes, ce coût est évalué entre 2 et 5 hommes an par participants.

La phase d'**évaluation des systèmes** correspond au déroulement de la campagne d'évaluation et aux mesures de performances et autres analyses qui s'ensuivent. Des données non encore connues sont distribuées aux participants qui doivent rendre les résultats de leurs traitements dans un temps limité. Les données de test seront proches des données de développement mais comporteront également des données représentant des conditions acoustiques nouvelles afin de mesurer la robustesse des algorithmes. En raison de possibles imbrications entre les éléments évalués dans la chaîne de transcription/indexation, la diffusion des données et les dates de soumissions des résultats seront échelonnées. Par exemple, la détection des frontières de phrases nécessite de disposer d'une transcription automatique : afin de favoriser des participations à cette seule tâche, le calendrier sera en décalage par rapport à celui de la tâche de transcription. De plus, dans ce cas, nous solliciterons les participants afin de mettre à disposition leurs transcriptions pour la détection de frontière ou pour favoriser des collaborations entre participants. À l'issue de la soumission des résultats, il convient alors de mesurer les performances et d'analyser les résultats. Une première diffusion des annotations de référence du corpus de test permet de valider la mesure des performances en collaboration avec les participants lors d'une phase d'adjudication, les résultats officiels étant ceux établis après adjudication.

Enfin, une phase finale d'**échange, de diagnostic, de consolidation et de valorisation des résultats** comporte plusieurs objectifs complémentaires. D'une part, il s'agit pour les participants d'analyser dans le détail les résultats obtenus lors de la phase de test afin de faire émerger de nouveaux axes de recherche et de consolider les résultats obtenus par des expériences complémentaires ou des expériences croisées entre participants. Le rôle des organisateurs de la campagne est de susciter des expériences complémentaires et de favoriser les diagnostics entre participants en organisant un échange de ressources. Par exemple, la mise en place d'un espace d'échange de ressources entre participants permet de faciliter de telles expériences de diagnostic. D'autre part, il s'agit pour les organisateurs de diffuser les résultats obtenus lors de la campagne dans le respect des règles établies (cf. section 1.9), et de regrouper dans un *package* d'évaluation l'ensemble des données (corpus, outils de mesure des performances, etc.) nécessaires. La diffusion des résultats se fera principalement par les canaux scientifiques habituels (publications dans des revues et conférences internationales) mais aussi à

travers l'organisation d'une journée de conférence destinée à promouvoir l'usage de ces données dans les communautés scientifiques autre que celle du traitement automatique de la parole. La diffusion du *package* d'évaluation sera assurée par ELDA en garantissant un accès à faible coût pour la recherche académique.

### 1.5.2 Éléments évalués et protocoles

Si la campagne d'évaluation se place dans la cadre applicatif d'un système de génération de méta données et d'indexation des contenus audiovisuels, il n'en demeure pas moins que les éléments constituant un tel système seront évalués séparément. En effet, l'objectif du projet n'est pas de développer et d'évaluer un système complet dans un cadre applicatif donné mais plutôt de mesurer de manière objective les performances des différents éléments nécessaires à la mise en œuvre d'un tel système et de faire progresser ces technologies par la mise à disposition des ressources nécessaires (corpus, méthodologie d'évaluation, etc.), la fédération d'une communauté d'acteurs et l'émergence de nouveaux axes de recherche.

Les éléments d'un système qui seront évalués seront définis de manière précise au démarrage du projet, lors de la phase d'élaboration du plan d'évaluation. Cependant, nous pouvons d'ores et déjà classer les éléments d'un système en trois objectifs : la segmentation, la transcription et l'extraction d'informations. Par ailleurs, pour chacun de ces objectifs, nous pouvons définir quelques éléments importants qui seront considérés dans la définition des tâches et catégories de systèmes évalués.

Nous décrivons succinctement chacun des objectifs avant de conclure sur les liens entre les différents éléments évoqués au sein d'un système complet, liens justifiant l'implication d'un grand nombre de participants d'horizons variés et le travail de structuration des acteurs élaboré dans le projet ETAPE.

#### **Segmentation.**

La segmentation a pour but de fournir un découpage d'un flux contenant de la parole (radio, vidéo, podcast, etc.) selon une information donnée comme les classes sonores présentes ou encore le locuteur. La segmentation du flux sonore est une étape préalable indispensable pour son traitement et son indexation, en fournissant des informations sur la nature des données (musique, parole, parole superposée) et sur leur structure (tours de parole, présence de locuteurs particulier, etc.). Nous nous appuyons sur l'expérience des précédentes campagnes pour définir des tâches de suivi d'événements sonores (détection de parole, détection de musique, détection des zones de double parole, détection d'un locuteur donné) et de segmentation en locuteurs. Si nombre de tâches de segmentation constituent des tâches connues et balisées, pour lesquelles on s'appuiera sur les protocoles et les mesures de performances existantes, il convient de mesurer les progrès effectués ces dernières années sur de nouveaux types de contenus. De plus, on apportera un soin particulier aux cas de la double parole qui reste peu étudié à ce jour. Nous souhaitons également favoriser des travaux sur les locuteurs permettant l'exploitation de sources d'informations variées (acoustique, phonétique, sémantique), par exemple en introduisant de nouvelles tâches, comme la segmentation nominative en locuteurs, qui nécessitent, outre la détection des tours de parole, une analyse sémantique de la transcription. Dans la mesure du possible, nous envisagerons des tâches proches d'un contexte applicatif réel telles que le suivi de locuteur à partir d'un court extrait, tâche répondant aux besoins de requêtes du type « retrouver tous les passages où cette personne est en train de parler » associées à un court extrait de parole. Enfin, notons que la présence de tâches de segmentation qui, à l'opposé de la transcription et de l'extraction d'information, nécessitent peu d'investissement en terme de main d'œuvre et de puissance de calcul permet d'associer de nombreux acteurs à la campagne d'évaluation et à la dynamique qui en résulte.

#### **Transcription.**

La transcription a pour objectif de fournir une représentation lexicale (*i.e.* sous forme orthographique) des zones contenant de la parole dans un document sonore, éventuellement accompagnée de mesures de confiance associées à chacun des mots reconnus par le système. Cette représentation lexicale est, dans la plupart des cas, à la base des traitements suivants visant à l'extraction d'informations et à l'indexation des documents sonores justifiant ainsi pleinement une évaluation portant sur la seule transcription. L'évaluation de la transcription bénéficie d'une longue expérience sur laquelle nous nous appuyons, en particulier concernant la mesure des performances de la transcription comme des mesures de confiance. Une tâche de transcription classique, sans contrainte particulière imposée aux systèmes, permettra d'évaluer les performances en transcription de la parole spontanée et la robustesse des systèmes de transcription aux conditions acoustiques. Par ailleurs, nous nous attacherons à définir des tâches proches des contraintes applicatives afin de favoriser le transfert de technologie

à l'issue du projet. En particulier, de nombreux acteurs du domaine ont exprimé leur intérêt pour l'évaluation de systèmes à délai de réponse contraint. Il est intéressant de noter qu'en couplant délai de réponse contraint et temps réel, l'utilisation de systèmes de transcription automatique sur de grandes collections de données devient possible. Nous envisageons par ailleurs de poursuivre la tâche exploratoire d'adaptation non supervisée des systèmes de transcription initiée dans ESTER 2. En effet, les systèmes de transcription sont en général développés pour un domaine particulier et leur adaptation à un nouveau domaine ou à un nouveau thème est coûteuse. Nous souhaitons donc solliciter des contributions permettant d'adapter les systèmes de manière non supervisée afin d'offrir des solutions technologiques génériques et versatiles.

### **Extraction d'informations.**

Étant données les étapes de segmentation et de transcription, il est possible d'extraire de nombreuses informations utiles au processus d'accès aux contenus numériques, que ce soit à des fins de navigation, de structuration, d'indexation ou encore de visualisation. Dans le cadre de cette évaluation, nous chercherons à évaluer des catégories génériques, faiblement dépendantes du contexte applicatif visé, et jugées comme importantes pour de nombreuses applications. La définition précise des différentes catégories sera établie en collaboration avec les participants pour répondre à leurs objectifs de recherche, autour des trois axes suivants : la **détection des entités nommées**, la **segmentation en phrases** et la **segmentation thématique**. Ces trois catégories, à l'interface entre traitement de la parole et traitement des langues, ont reçu peu d'attention dans la communauté scientifique française en ce qui concerne les documents oraux. Bien que largement étudiée sur des données écrites, la détection des entités nommées dans un flux sonore pose de nombreux problèmes non résolus comme l'a mis en évidence la campagne ESTER. La segmentation en phrases reste relativement peu étudiée et, à notre connaissance, pas pour la langue française. Une telle segmentation nous semble indispensable à l'analyse sémantique de la transcription. Enfin, la segmentation thématique a été étudiée soit dans des cadres applicatifs contraints comme les journaux télévisés, soit sur des données écrites. Or il s'agit là d'éléments indispensables à une chaîne d'analyse de contenus numériques contenant de la parole, dont l'évaluation est indispensable. Si la mise en œuvre des tâches d'extraction d'information était difficilement envisageable par le passé en raison du faible nombre d'acteurs de la transcription automatique, elle est maintenant possible notamment grâce aux retombées du projet ESTER. Dans la mise en œuvre de ces tâches, on apportera un soin particulier à l'échange de données entre participants. En particulier, il convient d'organiser la distribution de transcriptions automatiques afin de favoriser la participation d'acteurs du domaine du traitement automatique des langues qui ne possèdent pas de système de transcription propre. On cherchera par ailleurs à stimuler des recherches offrant une véritable interaction entre parole et langue plutôt qu'une simple juxtaposition de techniques.

### **Relations entre catégories.**

Afin d'évaluer chacun des éléments d'un système complet, le découpage en catégories présenté ici est nécessaire. Cependant, ces éléments ne sont pas indépendants les uns des autres. Clairement, l'extraction d'informations requiert une transcription, tandis que cette dernière utilise le résultat de la détection des zones de parole. À l'inverse, la segmentation nommée en locuteur ne prend vraiment tout son sens que lorsque la transcription est disponible ; en retour, la segmentation et le regroupement en locuteur bénéficieraient largement d'un apport de la transcription. Dans le même ordre d'idée, l'adaptation non supervisée se base souvent sur une segmentation thématique. Tout en gardant le découpage en catégories pour des raisons pratiques, nous nous efforcerons de promouvoir des travaux de recherche privilégiant les relations entre les catégories, par exemple à travers une évaluation séparée de tels travaux. Soulignons que les catégories « Segmentation nommée en locuteur » et « Transcription avec adaptation non supervisée », non explorées lors des précédentes campagnes, oeuvrent dans ce sens.

Le découpage en catégorie permet également de mobiliser le plus grand nombre d'acteurs possibles, tant dans le domaine du traitement automatique de la parole que dans ceux du traitement des langues et du multimédia, favorisant ainsi l'émergence d'un réseau de compétence et une dynamique d'échange au sein de la communauté scientifique. En effet, d'une part la présence de catégories ne nécessitant pas un investissement matériel et humain lourd, comme la segmentation en événements sonores ou le suivi de locuteurs, offre la possibilité à des petites équipes de s'impliquer dans les travaux d'évaluation. D'autre part, en décalant le calendrier des tâches d'extraction d'informations par rapport à celles de transcription, et en organisant la diffusion des résultats de ces dernières, nous avons pour objectif de permettre l'implication dans la campagne d'évaluation de



participants ne possédant pas de système de transcription automatique (par exemple, des laboratoires dans le domaine du traitement automatique des langues naturelles).

### 1.5.3 Public visé par le projet

Après consultation de la communauté par l'AFCP en prévision du projet ETAPE, on attend clairement la participation des principaux laboratoires du domaine qui ont d'ailleurs massivement participé aux précédentes évaluations AUPELF B1 et ESTER (citons, par ordre alphabétique, LIG, France Télécom R&D, IRISA, IRIT, LIA, LIMSI, LIUM et LORIA), mélangeant partenaires académiques et industriels (France Télécom, Vecsys Research, Telisma). La campagne ESTER 2 a par ailleurs vu l'arrivée de plusieurs nouveaux participants ayant déclaré leur intention de participer à des tâches de segmentation ou d'extraction d'information. Nous sommes en droit d'espérer un maintien de ces nouveaux participants dans la campagne ETAPE. Par ailleurs, la nature plus variée et plus proche d'applications réelles des contenus est susceptible de séduire de nouveaux acteurs pour les tâches de segmentation et d'extraction d'informations. En particulier, on attend du projet une ouverture des participants vers les acteurs du traitement des langues naturelles, notamment sur les tâches d'extraction d'informations, plusieurs projets en cours faisant intervenir des consortiums mélangeant acteurs du traitement automatique de la parole et du traitement automatique des langues (*e.g.* ANR EPAC, ARC INRIA Rapsodis). Un effort particulier sera fait pour attirer ces nouveaux acteurs vers la campagne ETAPE, notamment à travers la mise en place d'un système d'échange permettant à ces laboratoires de disposer de transcriptions automatiques. Si un tel échange n'était pas possible lors des précédentes campagnes du même type en raison du manque de maturité de nombreux systèmes de transcription automatique, il est maintenant tout à fait envisageable, voire nécessaire.

Par ailleurs, l'enrichissement du corpus doit permettre d'élargir considérablement la liste des utilisateurs potentiels du corpus créé dans le cadre du projet ETAPE. En effet, ces enrichissements visent d'une part à développer de nouveaux axes transverses impliquant des acteurs de provenances diverses, tant STIC que SHS (traitement automatique de la parole, traitement automatique des langues naturelles, phonéticien, phonologue, etc.) et, d'autre part, à favoriser l'utilisation de grands corpus annotés dans les domaines relevant des sciences humaines et sociales.

### 1.5.4 Structuration du projet en tâches

Outre la **tâche de coordination**, le projet se décompose en quatre tâches principales qui reflètent les différentes phases du déroulement de la campagne telles que décrites précédemment. Une première **tâche de conception du plan d'évaluation** vise à établir les règles du jeu en donnant une description précise des tâches évaluées, des métriques de mesure des performances et des données. La **tâche de production des ressources** doit permettre la production des données nécessaires et des outils de mesure des performances en suivant les spécifications établies lors de la conception du plan d'évaluation. La **tâche d'évaluation** consiste en la mise en œuvre effective de la phase de test de la campagne. Enfin, la **tâche d'animation scientifique et de valorisation** permet d'une part d'assurer la dynamique scientifique de l'évaluation et de structurer les efforts des participants et, d'autre part, d'assurer la diffusion et la valorisation des résultats et des ressources produites.

Nous décrivons dans la suite le détail des travaux envisagés dans chacune des tâches en listant les partenaires impliqués<sup>3</sup>.

#### **Tâche 0 : coordination.**

Partenaires : AFCP, ELDA, DGA, LPP, LLF

La tâche de coordination regroupe l'ensemble des aspects liés à l'organisation technique et administrative du projet tels que la mise en place des accords de propriété et de diffusion des données, la rédaction des livrables ou encore l'organisation des réunions d'avancement.

#### **Tâche 1 : conception du plan d'évaluation.**

Partenaires : AFCP, DGA, ELDA.

<sup>3</sup> Le partenaire souligné correspond au responsable de la tâche.

La tâche de conception du plan d'évaluation correspond à la première phase de conception de la campagne décrite précédemment. Elle a pour but principal de définir le périmètre exact de l'évaluation en concertation avec les participants potentiels. À partir d'une proposition initiale des organisateurs, articulée autour des catégories que nous venons de définir, une concertation de l'ensemble de la communauté scientifique impliquée doit permettre de définir les protocoles expérimentaux pour chacune des catégories, les données nécessaires pour mener à bien la tâche, les métriques utilisées pour comparer les systèmes ainsi que les règles de participation concernant la publication des résultats. Cette phase de consultation doit permettre d'aboutir, d'une part, à la spécification des données à produire et, en particulier, des annotations nécessaires et, d'autre part, à une première version du plan d'évaluation. À l'issue de la phase de développement des systèmes, une version finale du plan d'évaluation sera établie en fonction de l'expérience acquise par les participants.

## **Tâche 2 : production des ressources.**

Partenaires : ELDA, DGA, LPP, LLF

La tâche de production des ressources regroupe l'ensemble des étapes liées au développement des corpus, à leur enrichissement et au développement des outils d'évaluation (mesure de performances, suivi des modifications, etc). La nature exacte des données (sources, niveau d'annotation) sera spécifiée à l'issue de la première phase de concertation de la tâche 1. Cependant, l'objectif affiché du projet est de produire environ 100 heures de données contenant principalement de la parole spontanée et représentant des conditions acoustiques diverses. Outre l'annotation en conditions acoustiques (parole, musique, etc.), en tours de parole ainsi que l'annotation orthographique, on s'attachera aux annotations nécessaires pour les catégories d'extraction d'informations qui auront été définies. De plus, un sous-ensemble du corpus (environ 50h) sera enrichi par des alignements phonétiques et par des informations syntaxiques.

La tâche de production des ressources, qui constitue un élément clé du projet, se divise naturellement en sous-tâches.

**Sous-tâche 2.1 – production des corpus annotés** : cette première sous-tâche, placée sous la responsabilité de ELDA, regroupe les activités de développement du corpus annoté de l'acquisition des données à la diffusion en passant par l'annotation et le contrôle de qualité. On s'appuiera sur les standards existants en terme de production des données, en exploitant les conventions d'annotation existantes (cf. règles d'annotation du logiciel Transcriber) lorsque cela est possible.

**Sous-tâche 2.2 – enrichissement des corpus** : la sous-tâche d'enrichissement des corpus, placée sous la responsabilité de l'AFCP, vise à sélectionner un sous-ensemble du corpus produit pour l'enrichir par des alignements phonétiques et par des arbres syntaxiques. La sélection du sous-ensemble enrichi se fera en concertation avec l'ensemble des communautés scientifiques impliquées afin d'assurer une représentativité des données par rapport à divers critères comme le sexe des locuteurs, leur accent, le style de parole ou encore l'âge. De tels critères étant classiquement utilisés dans les études en phonétique et en linguistique, cette sélection vise à garantir l'intérêt du corpus enrichi pour ce type d'études, ouvrant ainsi des perspectives de recherche à la frontière entre technologie et science humaines. Tant pour les alignements phonétiques que pour les arbres syntaxiques, l'annotation se fera par correction manuelle d'une ou plusieurs sorties de systèmes automatiques afin d'obtenir des annotations de référence suivant les formats en usage dans les communautés respectives. Dans le cas des alignements phonétiques, les alignements automatiques seront fournis par les participants à la campagne. Dans le cas de l'annotation syntaxique, le système d'analyse syntaxique du Laboratoire de Linguistique Formelle<sup>4</sup> (Université Paris 7) sera utilisé et l'on s'appuiera sur les conventions d'annotation et les formats d'échange définis dans le cadre de la campagne d'évaluation Technolanguage Evalda/Easy<sup>5</sup>

**Sous-tâche 2.3 – infrastructure et outils de l'évaluation** : cette sous-tâche regroupe, sous la responsabilité de ELDA, le développement des outils et infrastructures liés à l'évaluation. Outre le développement de programme de mesure des performances des systèmes, en accord avec les métriques définis à la sous-tâche 2.1, il

<sup>4</sup> B. Crabbé et M. Candito. Expériences d'analyse syntaxique statistique. In Proc. Traitement Automatique des Langues Naturelles (TALN), 2008.

<sup>5</sup> La campagne d'évaluation EASY du projet Technolanguage EVALDA avait pour objectif la création d'une méthodologie d'évaluation des analyseurs syntaxiques en français et son application dans une campagne d'évaluation. Cf. [http://www.technolanguage.net/article.php3?id\\_article=198](http://www.technolanguage.net/article.php3?id_article=198).

conviendra notamment de mettre en œuvre une structure de suivi des modifications des annotations produites (corrections des annotations par les participants, adjudication) et de partage des ressources dérivées (transcriptions automatiques, graphes de mots, alignements phonétiques, etc.). La forme d'une telle structure reste à définir mais on pourra s'appuyer sur le projet d'infrastructure déposé par ELDA dans le présent appel à projets et sur l'expérience acquise avec les espaces web d'échanges de ressources mis en place lors de la campagne ESTER<sup>6</sup>.

### **Tâche 3 : évaluation.**

Partenaires : DGA, AFCP, ELDA.

Cette tâche correspond à la mise en œuvre de la phase de test de la campagne. Elle regroupe la diffusion des données de test et la récupération des soumissions, en garantissant le respect du calendrier fixé, la mesure effective des performances, la mise en œuvre d'une phase d'adjudication permettant à chaque participant d'analyser les erreurs de son système et de corriger les annotations de références si nécessaire, ainsi que l'analyse détaillée des résultats par les organisateurs.

### **Tâche 4 : animation scientifique et valorisation.**

Partenaires : AFCP, DGA, ELDA.

Cette tâche regroupe l'ensemble des travaux liés à l'animation scientifique de la campagne, à la diffusion des résultats et à la valorisation du projet en général.

Outre l'organisation des discussions avec les participants potentiels concernant l'élaboration du plan d'évaluation, l'animation scientifique de la campagne se fera principalement à travers l'organisation de deux ateliers internes, réservés aux participants, et d'un atelier d'ouverture. Un premier atelier interne à quelques mois de la fin de la phase de développement de systèmes permettra de favoriser les échanges entre participants et d'assurer un retour de ces derniers en vue d'une finalisation des protocoles d'évaluation. Un atelier final interne, sur deux ou trois jours, organisé à l'issue de la phase d'adjudication, permettra aux organisateurs de présenter les résultats finaux et leur analyse. Cet atelier est également l'occasion pour les participants de présenter et de discuter les solutions techniques mise en œuvre, de discuter les perspectives de recherches ouvertes par la campagne d'évaluation et de renforcer les liens au sein de la communauté scientifique. Finalement, un atelier d'ouverture destiné à la promotion du corpus auprès de l'ensemble des communautés scientifiques intéressées (STIC, SHS) permettra de présenter le corpus et les ressources dérivées des participants afin de favoriser l'émergence de thèmes de recherche transverses.

L'action de promotion des ressources produites dans le cadre de ce présent projet ne prend tout son sens qu'en complément d'une politique de diffusion de ces dernières. L'action de valorisation inclut donc le *packaging* des données est leur diffusion par l'agence de distribution ELDA. Afin de permettre un large accès de la communauté scientifique aux résultats du projet, les corpus développés seront diffusés à des fins de recherche à un prix raisonnable.

Finalement, la diffusion des résultats de la campagne se fera principalement par les canaux scientifiques usuels, notamment les publications dans les conférences (LREC, Interspeech, HLT, etc.) et les revues internationales. Un site web dédié à l'évaluation en traitement automatique de la parole, dérivé de l'actuel site Ester, permettra de référencer les publications des organisateurs et des participants ainsi que les ressources mise à disposition par les participants.

## **1.6 Résultats escomptés et retombées attendues.**

On attend en premier lieu de l'organisation d'une campagne d'évaluation la mise à disposition de la communauté scientifique d'un corpus et d'outils de mesure des performances ainsi qu'une mesure objective des performances des systèmes évalués. Dans le cadre de ce projet, le corpus, développé dans un souci de complémentarité avec les corpus existants, doit permettre en premier lieu une amélioration de l'état de l'art dans le domaine du traitement automatique de la parole. On attend particulièrement des résultats concernant la

---

<sup>6</sup> <http://www.afcp-parole.org/ester/ressources.html>

parole spontanée et la robustesse à la diversité des contenus, en plus d'avancées technologiques fondamentales, l'expérience des campagnes d'évaluation précédentes ayant clairement mis en évidence l'impact de telles ressources sur la production en recherche. Cet impact se mesure facilement en utilisant les indicateurs usuels d'évaluation de la recherche (bibliométrie, brevets, *etc.*). Par ailleurs, l'introduction de tâches proches des contraintes applicatives est de nature à favoriser des transferts de technologie de la recherche vers le monde industriel.

Un projet d'évaluation répond également clairement à un objectif de structuration d'une communauté. On attend donc du projet qu'il renforce les synergies et collaborations, non seulement dans le domaine du traitement automatique de la parole, mais aussi, et surtout, dans le domaine de la parole et du langage oral en général. En ce sens, le projet et la campagne sont organisés avec le souci permanent de favoriser les échanges entre l'ensemble des acteurs du domaine, impliqués directement dans l'évaluation ou non. En témoignent l'introduction de tâches alliant traitement automatique de la parole et traitement des langues naturelles, l'enrichissement du corpus ou encore la politique d'accès à faible coût aux données à l'issue de la campagne.

De l'ensemble des facteurs que nous venons d'évoquer, nous sommes en droit d'attendre l'émergence et le renforcement de nouvelles problématiques de recherche impliquant des communautés scientifiques distinctes : indexation de documents sonore contenant de la parole mais aussi phonétique et linguistique de corpus, ou encore utilisation de la syntaxe en traitement automatique de la parole.

## 1.7 Organisation du projet.

Note : le partenaire souligné identifie le responsable scientifique de la tâche.

### Tâche 0 : coordination

- Partenaires : AFCP, ELDA, DGA, LPP, LLF
- Livrables : accord de consortium (T0+6), rapport d'avancement à mi-parcours (T0+15), rapport final (T0+30)
- Durée : T0 – T0+30

Tous les partenaires sont impliqués dans la coordination du projet et dans la rédaction des rapports d'avancement.

### Tâche 1 : conception du plan d'évaluation

- Partenaires : AFCP, ELDA, DGA
- Livrables : proposition de plan d'évaluation (T0+3), spécification des ressources (T0+3), plan d'évaluation final (T0+22)
- Durée : T0 – T0+22

L'AFCP se charge de la consultation de la communauté scientifique. Les trois partenaires se chargent de l'animation scientifique et de la rédaction du plan d'évaluation préliminaire. ELDA se charge de la rédaction de la spécification des ressources. L'AFCP et la DGA se chargent de la rédaction du plan d'évaluation final à l'issue de la phase de développement des systèmes.

La tâche 1 est une étape nécessaire au démarrage du projet et son activité est principalement concentrée sur les trois premiers mois du projet. La tâche reste active jusqu'à la rédaction de la version finale du plan d'évaluation afin de prendre en compte les retours d'expérience des participants.

### Tâche 2 : production des ressources

- Partenaires : ELDA, DGA, LPP, LLF
- Livrables : conventions d'annotations (T0+4), corpus annoté d'apprentissage et de développement (T0+15), corpus de test (T0+24), corpus enrichi (T0+27)
- Durée : T0+3 – T0+27

ELDA se charge de l'acquisition et de l'annotation des données (sujets, tours de parole, transcription orthographique, entités nommées). La DGA assiste dans la phase de validation des données. Le LPP réalise les alignements phonétiques. Le LLF réalise les annotations syntaxiques. Tous participent à la définition des conventions d'annotations.

L'enrichissement des corpus (tâche 2.2) nécessite que la phase d'annotation (tâche 2.1) soit achevée. La mise en place d'une infrastructure et des outils d'évaluation (tâche 2.3) se fait en parallèle de l'annotation.

**Tâche 3 : évaluation**

- Partenaires : DGA, AFCP, ELDA
- Livrables : corpus de test (T0+24), résultats de la campagne (T0+27), version corrigé du corpus (T0+27)
- Durée : T0+24 – T0+27

La DGA est responsable du déroulement de la campagne de test : son rôle inclut la récupération des soumissions, la mesure des performances, l'arbitrage dans la phase d'adjudication. La DGA, en collaboration avec l'AFCP, assure l'analyse détaillée des résultats. ELDA prend en charge la production des données de test, leur diffusion ainsi que la mise à jour du corpus après adjudication.

L'évaluation dépend naturellement de la réalisation des corpus d'apprentissage et de test (tâche 2.1).

**Tâche 4 : animation scientifique et valorisation**

- Partenaires : AFCP, DGA, ELDA
- Livrables : publication de description du corpus (T0+16 et T0+30), publication des résultats (T0+30), actes de l'atelier final (T0+30)
- Durée : T0+12 – T0+30

L'ensemble des partenaires assure la publicité des ressources à travers une participation aux conférences du domaine : publication d'une description du corpus après création du corpus annoté, publication des résultats. L'AFCP est responsable de l'organisation de l'ensemble des ateliers (mi-parcours, ouverture, final) et de la mise à jour du site Internet de la campagne. La DGA assiste l'AFCP dans l'animation des débats scientifiques. ELDA assure le packaging et la diffusion des données de la campagne. L'ensemble des partenaires participe à la valorisation des ressources et des produits dérivés, notamment par la mise en place d'outils de suivi à l'issue de la campagne.

La phase d'animation s'étend sur toute la durée du projet. Ses points clés sont l'organisation des ateliers (mi-parcours à T+22, final à T+28 et ouverture à T+30), les publications de description du corpus avant (T0+16) et après (T0+30) enrichissement et la publication des résultats finaux (T0+30). Cette tâche est fortement dépendante de la création des ressources (tâche 2) et de la réalisation de la campagne de test (tâche 3).

Le chronogramme suivant résume les durées des tâches et les partenaires impliqués :

|   | Partenaires/Partners |   |   |   |   |   |   |   |   |    | Année / Year 1 |   |   |   |    | Année / Year 2 |    |    |    |    | Année / Year 3 |    |    |    |    |    |    |    |  |  |
|---|----------------------|---|---|---|---|---|---|---|---|----|----------------|---|---|---|----|----------------|----|----|----|----|----------------|----|----|----|----|----|----|----|--|--|
|   | 1                    | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 2              | 4 | 6 | 8 | 10 | 12             | 14 | 16 | 18 | 20 | 22             | 24 | 26 | 28 | 30 | 32 | 34 | 36 |  |  |
| Tâche0  |                      |   |   |   |   |   |   |   |   |    |                |   |   |   |    |                |    |    |    |    |                |    |    |    |    |    |    |    |  |  |
| Tâche1  |                      |   |   |   |   |   |   |   |   |    |                |   |   |   |    |                |    |    |    |    |                |    |    |    |    |    |    |    |  |  |
| Tâche2  |                      |   |   |   |   |   |   |   |   |    |                |   |   |   |    |                |    |    |    |    |                |    |    |    |    |    |    |    |  |  |
| Tâche3  |                      |   |   |   |   |   |   |   |   |    |                |   |   |   |    |                |    |    |    |    |                |    |    |    |    |    |    |    |  |  |
| Tâche4  |                      |   |   |   |   |   |   |   |   |    |                |   |   |   |    |                |    |    |    |    |                |    |    |    |    |    |    |    |  |  |
| Livrables / Jalons<br>Deliverables/Milestones                           |                      |   |   |   |   |   |   |   |   |    |                |   |   |   |    |                |    |    |    |    |                |    |    |    |    |    |    |    |  |  |
| Rapports d'avancement / états des dépenses<br>Progress report/expenses  |                      |   |   |   |   |   |   |   |   |    |                |   |   |   |    |                |    |    |    |    |                |    |    |    |    |    |    |    |  |  |
| Accord de consortium/rapport final<br>Consortium agreement/final report |                      |   |   |   |   |   |   |   |   |    |                |   |   |   |    |                |    |    |    |    |                |    |    |    |    |    |    |    |  |  |

- ☺ : Rapport d'avancement semestriel/ 6 month-progress report
- ☺ : Rapport d'avancement semestriel + état des dépenses/ Progress report + expenses
- ★ : Accord de consortium/ Consortium agreement
- ☆ : Rapport de synthèse + récapitulatif des dépenses/ Final report + expenses summary
- : Plan d'évaluation
- : Diffusion du corpus

**Analyse des risques**

Le projet comporte deux point charnières : le premier est lié à la définition du plan d'évaluation et à la spécification des données à annoter (tâche 1); le deuxième est lié à la diffusion du corpus d'apprentissage aux participants (tâche 2).

La définition du plan d'évaluation et la spécification des contenus à annoter conditionnent le démarrage de la phase de création des corpus. Il y a cependant peu de risques de blocage à ce niveau du projet. En effet, la consultation des participants potentiels organisée par l'AFCP en prévision de la soumission du projet ETAPE fait apparaître un réel consensus sur des contenus TV possédant une forte proportion de parole spontanée (jeux TV, émissions de plateau, débats, etc.). Il existe également un intérêt affiché de nombreux acteurs du domaine pour des contenus réputés plus difficiles comme les vidéos personnelles et les podcasts, la difficulté étant alors de trouver de tels contenus pour lesquels les tâches envisagés sont faisables avec des taux d'erreur raisonnables (en d'autres termes, certains contenus restent encore inenvisageables avec les technologies actuelles) et pour lesquels les droits d'auteurs ne rendent pas leur utilisation impossible. Le risque concernant les vidéos personnelles et podcasts est plus de ne pas trouver une masse suffisante de contenus, ce qui n'empêche pas le déroulement d'une campagne de test.

L'annotation d'un corpus d'apprentissage (et de développement) constitue un deuxième point potentiellement bloquant pour le projet. En effet, la mise à disposition de données d'apprentissage est primordiale au travail de développement des participants. De plus, il est nécessaire de donner un temps de développement suffisant à ces derniers. Nous pensons que le temps imparti à la phase d'annotation des corpus (12 mois) est suffisant pour permettre une diffusion des données à T0+15, la phase de test étant prévue à T0+24. La période de développement de 9 mois peut-être réduite si un retard dans la production des données devait intervenir. Dans le pire des cas, une diffusion des données d'apprentissage en deux temps pourrait être envisagée, une première livraison concernant les contenus TV et la seconde concernant les vidéos personnelles et podcasts.

Un dernier risque concerne le taux de participation à la campagne d'évaluation : le travail de consultation entrepris par l'AFCP avant la rédaction du projet montre cependant un fort intérêt de l'ensemble des acteurs des précédentes évaluations ESTER pour cette nouvelle campagne d'évaluation qui devrait au minimum regrouper les principaux laboratoires académiques du domaine (par ordre alphabétique, LIG, FT R&D, IRISA, IRIT, LIA, LIMSI, LIUM, LORIA).

| <b>TABLEAU des LIVRABLES et des JALONS (le cas échéant)/ Deliverables and milestones</b> |   |  |   |
|--|---|--|---|
| Tâche/<br>Task   | Intitulé et nature des livrables et des jalons/<br>Title and substance of the deliverables and milestones | Date de fourniture<br>nombre de mois à compter de T0/<br>Delivery date, in months starting from T0 | Partenaire responsable du livrable/jalon/<br>Partner in charge of the deliverable/milestone |
| <b>0. Coordination</b>   |   |  |   |
|  | Accord de consortium  | T0+6   | AFCP  |
|  | Rapport à mi parcours   | T0+15  | AFCP  |
|  | Rapport final   | T0+30  | AFCP  |
| <b>1. Conception du plan d'évaluation</b>  |   |  |   |
|  | Plan d'évaluation préliminaire  | T0+3   | AFCP  |
|  | Spécification des données   | T0+3   | ELDA  |
|  | Plan d'évaluation final   | T0+22  | AFCP  |
| <b>2. Production des ressources</b>  |   |  |   |
|  | Conventions d'annotation  | T0+4   | ELDA  |
|  | Corpus d'apprentissage  | T0+15  | ELDA  |
|  | Alignements phonétiques   | T0+27  | LPP   |
|  | Arbres syntaxiques  | T0+27  | LLF   |
| <b>3. Évaluation</b>   |   |  |   |
|  | Corpus de test  | T0+24  | ELDA  |
|  | Résultats de la campagne  | T0+27  | DGA   |
|  | Mise à jour des annotations   | T0+27  | ELDA  |
| <b>4. Animation scientifique et valorisation</b>   |   |  |   |
|  | Publication d'une description des corpus (soumission)   | T0+16  | ELDA  |
|  | Atelier à mi-parcours   | T0+22  | AFCP  |
|  | Publication d'une description des corpus (soumission)   | T0+30  | AFCP  |
|  | Publication des résultats (soumission)  | T0+30  | AFCP  |
|  | Atelier final   | T0+28  | AFCP  |
|  | Atelier d'ouverture   | T0+30  | AFCP  |

## 1.8 Organisation du partenariat. *Description of the Consortium.*

### 1.8.1 Pertinence des partenaires.

L'organisation de la campagne d'évaluation repose sur trois partenaires : l'Association Francophone de la

Communication Parlée (AFCP), le Centre d'Étude Parisien de la Délégation Générale pour l'Armement (DGA/CEP) et ELDA. Ces trois partenaires possèdent une solide expérience, tant individuelle que commune, dans l'organisation de campagnes d'évaluation. En particulier, la collaboration entre ces trois organismes a été à l'origine du succès des campagnes ESTER.

L'AFCP, en tant que société savante chargée d'animer la communauté scientifique dans le domaine de la communication parlée, représente la communauté scientifique impliquée dans la campagne d'évaluation. Son rôle principal dans le projet est donc d'établir le lien avec les participants de la campagne d'évaluation. Il convient d'ailleurs de noter que les orientations concernant le type des données étudiées dans le projet et les tâches envisagées résultent d'une concertation des acteurs du domaine organisée par l'AFCP. Le rôle de l'AFCP dans le projet, à savoir l'animation scientifique, la promotion de la campagne et la valorisation des résultats entre donc pleinement dans les missions de l'association.

ELDA, en tant qu'agence d'évaluation et de diffusion des ressources linguistiques, a pour rôle principal de créer les ressources nécessaires, de diffuser l'ensemble des ressources – passées et présentes – liées à l'évaluation, d'assurer la pérennité de ces dernières et de faciliter la préparation d'évaluations similaires ultérieures. ELDA apporte au projet son expérience en matière de production, de diffusion et de gestion de ressources linguistiques ainsi qu'en matière de valorisation des résultats. Par ailleurs, ELDA fait partie d'un autre consortium qui propose un projet de plateforme dont le but est de fédérer les différentes communautés de la langue et de l'image. Cette plateforme sera un lieu de partage des réflexions méthodologiques mais aussi des ressources et des outils. Le projet ETAPE pourra bénéficier des services web qui seront développés dans le cadre de ce consortium. ELDA a participé à de nombreux projets d'évaluation tels que les projets européens CLEF, CHIL ou TC-STAR, et à de nombreux projets nationaux au sein du programme Technolange. L'implication de ELDA dans de nombreuses campagnes d'évaluation permet de garantir une mutualisation des ressources entre les différentes campagnes.

La participation de la DGA, qui ne représente aucun des laboratoires participants, permet de garantir l'impartialité des évaluations. La DGA a donc pour rôle d'assurer la partie technique de l'évaluation (centralisation des soumissions, mesure des performances, adjudication, etc.). Elle apporte en outre sa compétence scientifique et technique dans la définition des aspects scientifiques de l'évaluation. Tout comme les autres acteurs, la DGA possède une solide expérience en terme d'évaluation à travers son activité de veille technologique. Le Centre d'Expertise Parisien de la DGA est impliqué dans de nombreux projets d'évaluation parmi lesquels le projet Quaero ou encore les actions nationales Technolange et Technovision.

Le Laboratoire de Phonétique et Phonologie (LPP) de l'Université de Paris 3 a pour mission l'enrichissement du corpus par des alignements phonétiques. Le LPP possède une solide expérience dans le domaine à travers sa participation à la création de différents corpus alignés. Les membres impliqués dans le projet ont par ailleurs tous une solide expérience de la phonétique de corpus, domaine sur lequel portent leurs activités de recherche comme en atteste leurs nombreuses publications.

Le Laboratoire de Linguistique Formelle (LLF) de l'Université de Paris 7 possède une longue expérience en matière de création et d'annotation de corpus pour les études en linguistiques. En particulier, Anne Abeillé est l'auteur du treebank de Paris 7, utilisé par une soixantaine de laboratoires dans le monde, et dont les choix d'annotation ont servi de point de départ pour la campagne Technolange Evalda/Easy pour laquelle le LLF a fourni des corpus d'apprentissage. Le LLF est impliqué dans de nombreuses études de corpus oraux ((Corpaix, CRFP, Ester) dans le cadre du projet ANR Program et pour la Grande Grammaire du Français.

Tant le LPP que le LLF portent un intérêt particulier au développement de corpus oraux annotés pour des études en phonétique ou en linguistique, ces corpus étant à la base de leurs travaux de recherche. Par ailleurs, il est important de noter que ces deux laboratoires ne développent pas d'outils automatiques de traitement. Si ils ont un intérêt concernant l'utilisation du corpus développé à des fins de recherche, ils ne participent à aucune activité d'évaluation d'outils de traitement automatique de la parole et de la langue. Outre un avantage lié à l'impartialité des acteurs, le recours à des laboratoires du domaine des Sciences Humaines et Sociales garantit une ouverture des résultats du projet vers la recherche fondamentale en plus du développement technologique.

### **1.8.2 Complémentarité des partenaires.**

Les acteurs du projet sont très clairement complémentaires. Si une campagne d'évaluation ne peut se faire sans une implication de la communauté scientifique concernée, ici par le biais de l'AFCP, elle ne peut se dérouler sans une activité de production et de diffusion de ressources, pas plus que sans un organisme d'évaluation neutre. L'AFCP, ELDA et la DGA assure chacun un des rôles que nous venons d'énoncer, ces rôles bien définis étant distincts bien que complémentaires. Enfin, la valeur ajoutée aux ressources produites par des

annotations phonétique et syntaxique ne peut provenir que d'acteurs de ces deux domaines tels que le LPP et le LLF. Bien que le LPP possède une compétence dans le domaine de la syntaxe, son implication dans le projet se limite à l'alignement phonétique comme en témoignent la liste des personnes impliquées.

Pour conclure sur la complémentarité des acteurs, nous souhaitons souligner que le consortium associe des membres provenant de différentes communautés (technologie, sciences humaines, producteurs de ressources), offrant ainsi une vision très large sur les problèmes traités, à savoir l'évaluation en traitement automatique de la parole et, à plus long terme, l'émergence de nouveaux axes transverses de recherches (phonétique et linguistique de corpus, information syntaxique, phonétique ou phonologique en traitement automatique de la parole, etc.).

### 1.8.3 Qualification du coordinateur du projet.

Le coordinateur du projet, chargé de recherche au CNRS depuis 2001, travaille sur le traitement automatique de la parole et sur l'indexation de documents multimédia à l'Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA, UMR 6074). Titulaire d'un diplôme d'ingénieur en mathématiques appliqués de l'Institut National des Sciences Appliquées (INSA) de Rouen et d'une thèse de l'École Nationale Supérieure des Télécommunications, soutenue en janvier 2000, il possède également une expérience industrielle au sein des entreprises ELAN Informatique (1995-1996) et IBM Research (2001-2002).

Les travaux de recherche de Guillaume Gravier portent d'une part sur la reconnaissance de la parole et du locuteur, notamment sur le couplage entre traitement de la parole et traitement des langues naturelles, ainsi que sur les modèles multimodaux pour l'analyse des documents multimédia. Il est l'auteur de nombreuses publications dans des revues et conférences internationales. Au cours des dernières années, il s'est impliqué dans plusieurs projets nationaux (ANR Domus Videum, ACI Demi-Ton, ANR Semim@ges, EVALDA/Ester, Technolanguage.net) et internationaux (Quaero, NoE MUSCLE). Guillaume Gravier est l'un des principaux organisateurs des campagnes d'évaluation ESTER. Dans ce cadre, il a organisé de nombreux séminaires et a présidé le comité scientifique de l'atelier final ESTER en mars 2005. Il a organisé les Journées d'Étude sur la Parole en 2006.

Ancien membre élu du conseil scientifique du Groupe Francophone de la Communication Parlée de la Société Française d'Acoustique, il est membre élu du conseil d'administration de l'AFCP depuis sa création en 2001. Il est membre de l'International Speech Communication Association depuis 1997.

## 1.9 Stratégie de valorisation et de protection des résultats.

La valorisation des résultats se fait selon trois axes principaux :

- publication des résultats par les canaux scientifiques habituels (conférences internationales, revues scientifiques, etc.)
- diffusion et promotion des *kits* d'évaluation par ELDA
- promotion du corpus par l'AFCP, notamment à travers l'organisation d'un séminaire d'ouverture ayant pour vocation de promouvoir l'usage du corpus développé auprès des communautés scientifiques qui ne sont pas concernées par la campagne d'évaluation (principalement la communauté des sciences humaines et sociales) : présentation du corpus développé et des ressources dérivées par les participants.

De manière indirecte, la publication des travaux de recherche menés par les participants sur le corpus pendant mais aussi après la campagne d'évaluation participe à la valorisation des résultats du projet. Notons enfin que l'implication de ELDA et de la DGA dans d'autres activités d'évaluation offre la possibilité de valoriser les résultats du projet ETAPE à travers l'usage de ce corpus dans d'autres cadres que celui prévu dans le présent projet. En particulier, ELDA offrira des services d'évaluation basés sur les réalisations du projet ETAPE, pour les développeurs de technologies ou pour les utilisateurs finaux.

Un accord de consortium concernant le partage et la diffusion des données sera rédigé au démarrage du projet. Cependant, le projet affichant l'objectif d'assurer la diffusion la plus large possible au sein de la communauté scientifique des ressources développées, l'accord de consortium se fondera sur un principe de diffusion de l'ensemble de ces ressources, la propriété des ressources restant acquise au producteur. S'agissant d'annotation de données réalisées par des tiers, ELDA se charge de négocier, le cas échéant, un accord permettant la diffusion des ressources à des fins de recherche. Les membres du projet, ainsi que les participants à la campagne d'évaluation, conservent le droit d'utilisation des données à des fins de recherche. Pour la campagne



d'évaluation, l'accès aux données sera soumis à la signature d'un engagement de participation à la campagne spécifiant les règles d'utilisation des ressources.

Les performances mesurées lors de la campagne d'évaluation seront publiées. Les règles de publication, en particulier concernant l'anonymat des résultats, seront définies en concertation avec les intéressés. Nous souhaitons cependant promouvoir l'idée d'une publication nominative des résultats, par exemple dans un objectif de favoriser le transfert de technologie. Au sein des participants à l'évaluation, la diffusion des résultats et des descriptifs de système à la totalité des participants est obligatoire et ne pourra se faire sous forme anonyme. En particulier, les laboratoires participants s'engagent à diffuser dans ce cadre une description complète du système mis en œuvre pour la campagne d'évaluation. Les organisateurs s'engagent à ne pas publier ces informations. Chacun des participants reste libre de publier ses propres résultats comme il le souhaite. En revanche, il ne pourra publier les résultats des autres sites que sous forme anonyme ou après avoir obtenu l'accord explicite des organisateurs et des laboratoires concernés. Comme pour l'usage des ressources, ces obligations seront contractualisées dans l'engagement de participation à la campagne.

## 2. Justification scientifique des moyens demandés.

### 2.1 Association Francophone de la Communication Parlée

#### 2.1.1 Equipement.

Néant.

#### 2.1.2 Personnel.

Le personnel permanent impliqué dans le projet œuvre pour l'Association Francophone de la Communication Parlée sous le régime du bénévolat valorisé. Les membres du conseil d'administration de l'association impliqué dans le projet ont pour rôle l'animation scientifique de la campagne, à savoir l'organisation de la consultation de la communauté (organisation des ateliers, réunion de préparation), la rédaction du plan d'évaluation ainsi que la diffusion et la valorisation des résultats (atelier d'ouverture, rédaction des articles, maintenance du site web). L'AFCP joue le rôle de coordinateur du projet (organisation des réunions, rapports d'avancement, états financiers).

Les personnes impliquées sont Guillaume Gravier (CR CNRS, IRISA, 15%), Isabelle Ferrané (MdC Univ. P. Sabatier, IRIT, 10%), Georges Linarès (MdC Univ. des Pays d'Avignon et du Vaucluse, LIA, 5%), Sylvain Meignier (MdC Univ. Du Maine, LIUM, 5%) et Laurent Besacier (MdC IMAG, LIG, 5%) soit un effort total de 12,5 hm sur la durée du projet<sup>7</sup>.

#### 2.1.3 Prestation de service externe.

Néant.

#### 2.1.4 Missions.

Le montant demandé pour les missions vise à couvrir les réunions du projet ainsi que la participation à des conférences internationales dans un objectif de diffusion et de valorisation des résultats obtenus. Entre les réunions internes du projet, la participation aux différentes réunions d'élaboration du plan d'évaluation et aux ateliers organisés par l'AFCP dans le cadre du projet, nous prévoyons un total de 15 déplacements impliquant chacun deux représentants de l'association ( $15 * 2 * 300 = 9\ 000$  €).

La diffusion et la valorisation des résultats du projet se fait en grande partie à travers des publications dans les conférences internationales du domaine (LREC, HLT, Interspeech). Nous prévoyons trois missions de participation à de telles conférences ( $3 * 3\ 000 = 9\ 000$  €).

#### 2.1.5 Dépenses justifiées sur une procédure de facturation interne.

Néant.

#### 2.1.6 Autres dépenses de fonctionnement.

L'animation scientifique de la campagne se fera principalement à travers l'organisation de deux ateliers internes, réservés aux participants, et d'un atelier d'ouverture. Un premier atelier interne à quelques mois de la fin de la phase de développement de systèmes permettra de favoriser les échanges entre participants et d'assurer un retour de ces derniers en vue d'une finalisation des protocoles d'évaluation. Un atelier final interne, sur deux ou trois jours, organisé à l'issue de la phase d'adjudication, permettra aux organisateurs de présenter les résultats finaux et leur analyse.

Nous demandons un budget de 30 000 € couvrant les frais de séjour (repas, pauses, nuitées pour l'atelier final), les frais d'impression des actes (CD-ROM), une éventuelle location de salle et la mise en place d'un mécanisme de bourses de voyage permettant une plus grande participation des participants aux ateliers.

<sup>7</sup> Les pourcentages pour les Maîtres de Conférences sont données par rapport au temps de recherche et d'enseignement.

## 2.2 ELDA

### 2.2.1 Equipement.

Coût estimé pour les équipements nécessaires à l'annotation des données :

3 postes d'annotation supplémentaires (PCs + casques d'écoute) = 3 000,00 euros

### 2.2.2 Personnel.

Le budget de ELDA en termes de personnel est estimé pour chaque tâche du projet de la façon suivante :

|   |                    |
|---|--------------------|
| Tâche T0 - Coordination                         | = 24 300,00 euros  |
| Tâche T1 - Plan d'évaluation                    | = 13 500,00 euros  |
| Tâche T2 - Production des ressources            | = 243 737,50 euros |
| Tâche T3 - Evaluations (adjudication)           | = 5 000 euros      |
| Tâche T4 – Animation scientifique, valorisation | = 4 500 euros      |

**Total = 291 037,50 euros**

Personnel non permanent : la plus grande partie des ressources humaines affectées à la tâche T2 (Production des ressources) concernera des postes d'annotateurs / transcripateurs qui feront l'objet de CDDs.

### 2.2.3 Prestation de service externe.

Néant.

### 2.2.4 Missions.

En matière de missions et de déplacements liés au projet, les besoins d'ELDA sont estimées à :

|                             |   |                   |
|-----------------------------|---|-------------------|
| Réunions de projet          | (10 réunions * 2 personnes * 300 euros)   | = 6 000,00 euros  |
| Conférences internationales | (4 conférences * 1 personne * 2500 euros) | = 10 000,00 euros |

**Total = 16 000,00 euros**

### 2.2.5 Dépenses justifiées sur une procédure de facturation interne.

Néant.

### 2.2.6 Autres dépenses de fonctionnement.

Néant.

## 2.3 Délégation Générale à l'Armement / Centre d'Expertise Parisien

### 2.3.1 Equipement.

Néant.

### 2.3.2 Personnel.

Le coût horaire du personnel est de 57 euro/h sur une base de 1350h/an, soit un coût par personne mois (PM) de  $57 * 1350 / 12 = 6412,50$ , arrondi à 6 400 dans le formulaire en ligne.

Nous avons prévu de consacrer

- 4 PM pour la tâche 3 d'évaluation, dont nous sommes responsables
- 2 PM pour la tâche de production de ressources, pour valider les corpus et en particulier les parties de développement et de test,
- 3 PM pour l'ensemble des autres tâches, pour participer à la conception du plan d'évaluation ainsi qu'à la coordination et l'animation scientifique.

### 2.3.3 Prestation de service externe.

Néant.

### 2.3.4 Missions.

En matière de missions et de déplacements liés au projet, les besoins d'ELDA sont estimées à :

|                             |  |                  |
|-----------------------------|--|------------------|
| Réunions de projet          | (15 réunions * 1 personnes * 300 euros)  | = 4 500,00 euros |
| Conférences internationales | (1 conférence * 1 personne * 3000 euros) | = 3 000,00 euros |

|              |  |                         |
|--------------|--|-------------------------|
| <b>Total</b> |  | <b>= 7 500,00 euros</b> |
|--------------|--|-------------------------|

### 2.3.5 Dépenses justifiées sur une procédure de facturation interne.

Néant.

### 2.3.6 Autres dépenses de fonctionnement.

Néant.

## 2.4 Université de Paris 3 / Laboratoire de Phonétique et Phonologie

### 2.4.1 Equipement.

Le coût d'équipement correspond à un poste de travail pour l'annotation.

### 2.4.2 Personnel.

Le personnel permanent correspond aux participations de Cédric Gendrot (MdC, 25%), Cécile Fougeron (CR CNRS, 25%) et Jacqueline Vaissière (PR, 10%).

|              |                            |
|--------------|----------------------------|
| C. Fougeron  | 4,5 PM * 5 472 €= 24 625 € |
| C. Gendrot   | 4,5 PM * 2 154 €= 9 693 €  |
| J. Vaissière | 1,8 PM * 5 182 €= 9 328 €  |

Le coût de personnel non permanent correspond à 12 personnes mois de vacances pour les corrections d'alignements phonétiques (soit environ 50h d'annotations).

### 2.4.3 Prestation de service externe.

Néant.

### 2.4.4 Missions.

Conférences internationales (2 conférence \* 1 personne \* 3000 euros) = 6 000,00 euros

### 2.4.5 Dépenses justifiées sur une procédure de facturation interne.

Néant.

### 2.4.6 Autres dépenses de fonctionnement.

Néant.

## **2.5 Université de Paris 7 / Laboratoire de Linguistique Formelle**

### **2.5.1 Equipement.**

Le coût d'équipement correspond à un poste de travail pour l'annotation.

### **2.5.2 Personnel.**

Le personnel permanent correspond aux participations de Anne Abeillé (PR1, 10%), Elisabeth Delais (CR, 10%), Maria Candito (MdC, 10%), Benoit Crabbé (MdC, 10%) et M. Plancq (IE, 5%).

Le coût de personnel non permanent correspond à 12 personnes mois de vacances pour les annotations syntaxiques (soit environ 50h d'annotation).

### **2.5.3 Prestation de service externe.**

Néant.

### **2.5.4 Missions.**

|                             |  |                  |
|-----------------------------|--|------------------|
| Réunions de projet          | (15 réunions * 1 personnes * 300 euros)  | = 4 500,00 euros |
| Conférences internationales | (2 conférence * 1 personne * 2250 euros) | = 4 500,00 euros |

### **2.5.5 Dépenses justifiées sur une procédure de facturation interne.**

Néant.

### **2.5.6 Autres dépenses de fonctionnement.**

Néant.

## Annexes

Description des partenaires/ *Partners informations* (cf. § 1.8.1) (1 page maximum par partenaire)

### **Association Francophone de la Communication Parlée.**

Site Internet : <http://www.afcp-parole.org>

L'Association Francophone de la Communication Parlée est une association à but non-lucratif (loi du 1<sup>er</sup> juillet 1901) de type société savante, consacrée au soutien, au développement, à la diffusion et à la promotion des différentes spécialités des sciences de la communication parlée, dans la communauté francophone. Fondée en 2001, l'association est affiliée à la International Speech Communication Association (ISCA). Ses membres représentent l'ensemble des acteurs français du domaine de la communication parlée, le conseil d'administration regroupant des membres de la grande majorité des laboratoires français du domaine.

Bien que les sujets semblent très divers, peu d'entre eux peuvent être étudiés sans faire appel simultanément à plusieurs disciplines : traitement du signal, probabilités, informatique, phonétique, linguistique.... L'association s'emploie à rassembler les chercheurs de ces disciplines, à favoriser leurs échanges et à aider à la formation de jeunes chercheurs, par l'organisation de séminaires et par la publication d'ouvrages spécialisés. Une rencontre biennale, "Les Journées d'Etudes sur la Parole" (JEP), permet une synthèse de l'ensemble des travaux de la communauté.

Plus généralement, pour atteindre ses objectifs, l'AFCP se donne, dans son domaine de compétences, les moyens suivants : stimulation et facilitation des échanges scientifiques ; organisation de conférences, de journées scientifiques et d'expositions ; organisation de cours, de séminaires et d'écoles thématiques ; publication et aide à la publication de revues et d'ouvrages ; diffusion de l'information à l'intention des acteurs institutionnels et du grand public ; promotion des relations entre recherche et industrie ; coopération avec d'autres organisations scientifiques ; ainsi que toute action conforme à ses objectifs.

Au cours des dernières années, l'association a été impliquée dans l'organisation des campagnes d'évaluation ESTER 1 et 2 ainsi que dans la réalisation du portail des technologies de la langue, Technolangue.net.

### **ELDA**

Site Internet : <http://www.elda.org>

ELDA, l'Agence pour la Distribution des Ressources Linguistiques et l'Évaluation (en anglais : Evaluations and Language Resources Distribution Agency) a été créée en février 1995 dans le but d'identifier, classer, collecter, valider et produire les ressources linguistiques utiles à la communauté scientifique dans le domaine des HLT (Human Language Technologies). Au fil des ans et des évolutions du domaine, ELDA a élargi ses activités à l'évaluation des technologies de la langue et à la distribution des ressources linguistiques nécessaires pour mener ces travaux d'évaluations. Aujourd'hui, la société est structurée autour de deux départements : *Identification et Distribution de Ressources Linguistiques* et *Production & Évaluation de Ressources Linguistiques*.

Les activités d'ELDA sont supervisées par le président directeur général et fondateur de la société, Khalid Choukri. A l'heure actuelle, douze personnes travaillent à temps complet au sein de la société. Ponctuellement, d'autres personnes viennent compléter l'équipe pour travailler de façon temporaire sur des projets spécifiques.

Au cours de ces dernières années, ELDA s'est impliquée dans de nombreux projets et campagnes d'évaluation, tant au niveau national qu'Européen. ELDA a notamment participé aux projets internationaux TC-STAR et CHIL. L'agence est également fortement impliquée dans le projet CLEF (Cross Language Evaluation Forum). Au niveau national, ELDA a pris part à de nombreux projets du programme Technolangue, notamment au projet ESTER 1.

### **DGA/CEP**

La Délégation Générale pour l'Armement (DGA) est, au sein du Ministère de la Défense, en charge de l'équipement des forces armées et de la préparation de l'avenir. Dans le cadre de cette deuxième mission, elle prépare les capacités technologiques et industrielles futures en soutenant leur développement. Elle finance chaque année plus de 700 millions d'euros de contrats d'études soutenant l'industrie et la recherche. C'est le premier acteur de la recherche de défense en Europe avec 35% de l'effort européen de recherche et technologie de défense.

Au sein de la DGA, le Centre d'expertise parisien (CEP) de la Direction de l'expertise technique (DET) regroupe les capacités d'expertise en région parisienne. Le domaine du traitement de la parole et du langage fait partie des domaines d'expertise du CEP. Le CEP dispose d'une expertise particulière dans l'évaluation des technologies dans ce domaine. Il a notamment été impliqué dans l'organisation des programmes Technolanguage et Techno-Vision, ainsi que du programme Quaero qui est largement fondé sur la méthodologie d'évaluation.

### **Laboratoire de Phonétique et Phonologie, Univ. Paris 3**

Site Internet : <http://lpp.univ-paris3.fr>

Le LPP est une Unité Mixte de Recherche (UMR 7018) spécialisée dans la recherche et l'enseignement en phonétique expérimentale et en phonologie. Le LPP rassemble plus d'une quarantaine de personnes : 7 chercheurs (dont 4 directeurs de recherche), 11 enseignants-chercheurs (dont 3 professeurs), 1 ingénieur, 1 secrétaire-gestionnaire et plus de 25 doctorants et post-doctorants.

Au terme des évolutions, un thème de recherche central s'est imposé : l'intégration de la phonétique et la phonologie qui correspond à une problématique d'actualité, vu les développements importants mais trop indépendants de la phonétique et la phonologie depuis de longues années. Deux thèmes de recherches développés au laboratoire s'avèrent particulièrement pertinents pour le présent projet :

1 - L'étude des variantes de parole est un domaine de recherche qui se situe au cœur des problématiques communes à la phonétique et la phonologie. Nos recherches sur ce thème s'appuient sur une longue expérience dans l'analyse et le traitement des données phonétiques (acoustiques, physiologiques, multilingues...) et leur problématiques s'inscrivent dans une perspective de 'Phonologie de Laboratoire' dont l'objectif à long terme est une modélisation intégrée des différents niveaux de représentations et de traitements (phonétiques, phonologiques et psycholinguistiques).

2 – Ce précédent thème s'accompagne également d'un intérêt à savoir comment les analyses phonétiques utilisées en linguistique peuvent contribuer à l'amélioration des systèmes de reconnaissance et/ou d'alignement automatique de la parole. Des travaux allant dans ce sens ont déjà été publiés en et visent à compléter les besoins de deux communautés complémentaires que sont STIC et SHS.

Les membres du LPP impliqués dans ce projet sont particulièrement impliqués dans ces thèmes de recherche, où l'interaction entre phonétique acoustique, prosodie et phonologie est omniprésente. Cécile Fougeron s'est notamment consacrée à l'analyse acoustique et articulatoire des phonèmes en fonction de leur position prosodique. Cédric Gendrot s'est spécialisé depuis quelques années dans l'analyse acoustique de très grands corpus, comparant ainsi des résultats précédemment analysés sur des corpus plus restreints ou sur des données physiologiques. Son travail a pour but de dégager dans un premier temps les grandes tendances suivies par la majorité des locuteurs, avant de s'intéresser aux stratégies individuelles, dans une optique multilingue (allemand, anglais, arabe, chinois, espagnol, français, italien, portugais, etc.). Jacqueline Vaissière fut ingénieur à France Télécom avant d'être professeur de phonétique à l'université de la Sorbonne Nouvelle. Elle a notamment travaillé sur la reconnaissance automatique de la parole, et sur le problème de la variabilité inter et intra locuteurs, et sur l'introduction des paramètres prosodiques comme aide à la segmentation automatique du continuum.

### **Laboratoire de Linguistique Formelle, Université de Paris 7**

Le Laboratoire de Linguistique Formelle est une unité mixte de recherche du CNRS et de l'Université de Paris 7. La spécificité du Laboratoire de Linguistique Formelle est la combinaison de deux diversités complémentaires : diversité des approches théoriques et diversité des langues étudiées. L'unité des chercheurs du Laboratoire repose sur un ensemble de convictions partagées:



- La linguistique se construit par une interaction constante entre description, analyse, formalisation et théorie.
- Une analyse relativement indépendante des modèles théoriques est possible.
- L'analyse se nourrit de la comparaison approfondie des langues.

Les travaux du Laboratoire de Linguistique Formelle s'inscrivent dans les paradigmes génératifs et énonciatifs. Le paradigme génératif comprend les grammaires minimalistes (MP) et les grammaires de propriétés (HPSG et TAG). Au cadre génératif se rattachent également l'étude du discours dans le cadre *Discourse Representation Theory* ((S)DRT) ; en sémantique et pragmatique formelle, la théorie des quantificateurs généralisés, la sémantique formelle du dialogue, les approches sous-spécifiées (UDRT, MRS) ; en morphologie, les approches Mot et paradigme, la morphologie distribuée (DM) et la morphologie lexématique; en phonologie, la Phonologie du gouvernement, la phonologie basée sur les contraintes, la phonologie autosegmentale et métrique ; les travaux portant sur l'interface prosodie-syntaxe. Le paradigme énonciatif s'articule autour de la théorie des opérations prédicatives et énonciatives.