



FIGURE 4.3: Organisation générale d'une campagne d'annotation.

4.2.1 Mise au point d'une mini-référence

Construire une mini-référence dès le début la campagne présente de nombreux avantages. Cela permet d'abord de tester en conditions réelles le premier jet du guide d'annotation, provenant par exemple de discussions entre le client et le gestionnaire. La construction de la mini-référence permet également d'obtenir une mesure d'évaluation de la fiabilité de l'annotation dès le début de la campagne. Cette mesure pourra par la suite être comparée avec d'autres, effectuées plus tard. En outre, la mini-référence, une fois finalisée, contient les informations nécessaires au calcul des dimensions de complexité de la campagne (voir chapitre 6), qui donneront des indications fines pour sélectionner les outils appropriés, qu'il s'agisse de pré-annotation automatique (voir chapitre 7), d'outils d'annotation (voir section 3.2) ou d'autres solutions (voir section 3.4). Enfin, cette étape permet la sélection d'une mesure d'évaluation de l'accord inter-annotateurs appropriée à la campagne (voir chapitre 8).

Les différentes étapes de la création de la mini-référence sont décrites dans la figure 4.4. Le sous-corpus de référence (ou mini-référence) est un échantillon du corpus d'origine non annoté, sélectionné si possible de manière à être représentatif. Le travail préparatoire (voir section 4.1.2) a normalement permis d'établir une typologie détaillée du corpus, et la création d'un sous-corpus représentatif pour la mini-référence peut donc en principe se faire en sélectionnant des fichiers (ou parties de fichiers) correspondant à chaque type identifié, de manière proportionnée. Le but n'est pas ici d'être parfaitement représentatif, mais de couvrir suffisamment de phénomènes pour mettre au