Contes de Bodon : construccion d'un còrpus per d'analisis comparatius de las estructuras temporalas de l'occitan

Louise Garrigou et Myriam Bras

Université de Toulouse, CLLE, UMR5263, CNRS & Université Toulouse Jean Jaurès

L'expression linguistique du temps représente un champ de recherche central en sémantique et pragmatique. Le projet global dans lequel s'inscrit notre proposition vise à produire des descriptions scientifiques destinées à étayer les études dédiées aux temps verbaux et autres marqueurs temporels en occitan. Il s'inscrit à la fois dans une visée descriptive et dans une visée comparative entre les structures temporelles de l'écrit et de l'oral.

Nous présenterons dans cette communication un corpus spécifique construit pour une étude de ce type. Il réunit des textes écrits et des textes oraux transcrits relevant du genre du conte. Pour rendre les données de l'occitan accessibles nous avons choisi d'associer aux textes leurs traductions en français.

Notre corpus est donc un corpus aligné occitan/français, partitionné selon la distinction oral/écrit. La constitution du sous-corpus oral pour l'occitan s'est appuyée sur des enregistrements audio ou video de Joan Bodon en situation de conteur réalisés entre 1964 et 1974. Nous y avons sélectionné quatre contes : *la pèl*, *lo sang del Drac*, *l'enfant polit*, *l'auca*, que nous avons transcrits. Nous avons rassemblé les versions littéraires de ces mêmes contes, issues des *Contes del Drac* de Bodon, figurant déjà dans la base BaTelÒc (base textuelle pour la langue occitane, `http://redac.univ-tlse2.fr/bateloc`), pour former le sous-corpus écrit pour l'occitan. Ainsi construit, le corpus offre la possibilité d'associer la transcription de la performance orale du conteur avec la version littéraire du même conte. En cela, il présente pour nos objectifs de recherche un intérêt majeur.

Le corpus a été complété par les traductions françaises des contes littéraires présentes dans l'édition des contes littéraires, et par nos traductions des transcriptions des performances orales. Il a été intégré dans la base Parcolab (corpus parallèle en français, serbe, espagnol, occitan, anglais, `http://parcolab.univ-tlse2.fr/parcolaf/`) permettant des études comparatives entre les langues.

Le corpus contient actuellement 16 textes (4 contes en modalité écrit/oral

et en occitan/français). Notre objectif est d'étendre le corpus à l'ensemble des contes pour lesquels nous disposons d'une performance orale de Joan Bodon.

Pour les besoins de l'analyse sémantique, nous avons mis en place une méthodologie d'annotation des marqueurs temporels et aspectuels afin de mettre au jour les structures temporelles des textes. Cette annotation concerne trois grands types de marqueurs temporels : les temps verbaux, les adverbiaux de localisation temporelle et les connecteurs temporels, qu'il s'agit de repérer et d'identifier. Nous donnerons quelques exemples de contes annotés et de leurs structures temporelles.