

## **Progetto di Ricerca**

Costruzione di un corpus linguistico multilingue bilanciato e diacronico come base empirica per l'estrazione automatica di informazioni rilevanti per il patrimonio musicale e il suo studio attraverso l'analisi del discorso. Il ricercatore contribuirà alla costruzione dell'intero corpus ma avrà la responsabilità completa della componente in lingua francese. Sviluppo di metodologie per l'annotazione dei testi in relazione all'analisi del discorso e all'individuazione di riferimenti rilevanti per il discorso sul patrimonio musicale, in sinergia con la modellazione ontologica del dominio. Queste attività contribuiscono al Work Package 4 e al Work Package 1 del progetto Polifonia.

## **Piano di attività**

- 1) progettazione e creazione di un corpus multilingue bilanciato e rappresentativo secondo i requisiti necessari alla modellazione dell'ontologia e responsabilità completa per la componente in lingua francese
- 2) analisi e sviluppo di tecniche di NLP adeguate all'estrazione della conoscenza relativamente ai requisiti necessari per la modellazione dell'ontologia. In particolare sarà necessario focalizzarsi sull'analisi sulle strutture morfosintattiche (Universal Dependencies) sul reperimento di tecniche discorsive (Topic Modeling, Sentiment Analysis, Opinion mining) da integrare all'individuazione dei principali attori (Named Entity Recognition) in modo da comporre un quadro di corrispondenze con le categorie dell'analisi del discorso;
- 3) integrazione di una banca dati terminologica multilingue che permetta di allineare le forme lessicali sia in termini di varietà linguistica, che di diversità geografica e/o culturale;
- 4) contribuire allo sviluppo di strumenti di estrazione delle informazioni ad hoc per Polifonia in stretta collaborazione con i partner di Computer Science
- 5) validazione dei risultati
- 6) scrittura di report di progetto e di articoli scientifici (in lingua inglese) per la divulgazione dei risultati di ricerca
- 7) partecipazione a riunioni di progetto