

## Francesca Frontini – CV 2006-2016

Ma thèse doctorale (**Università degli Studi di Pavia**, Italie, 2006-2010) se situe dans le domaine de la **linguistique des corpus**, et du **traitement automatique du langage naturel** (TALN). Mon travail de recherche avait pour but de développer un système destiné à annoter automatiquement l'information syntaxique basique (lemmatisation et PoS tagging) dans des corpus textuels produits par apprenants d'italien langue seconde.

Dans le cadre de mon travail de thèse, je me suis également intéressée aux **techniques de classification et de profilage automatique**, en analysant les corpus ainsi annotés avec des techniques de regroupement automatique (MDS et PCA) permettant de mesurer la distance syntaxique entre des apprenants de langues maternelles différentes.

Après mon doctorat, j'ai poursuivi ma carrière pendant quatre ans comme post-doctorante au **Consiglio Nazionale delle Ricerche (CNR)**, à l'**Institut de Linguistique Computationnelle (ILC) de Pise**. Mon activité à ILC-CNR a été financée par différents projets de recherche européens (EU FP7), nationaux et régionaux dédiés à la construction d'outils pour le **traitement automatique du langage naturel** (TALN) et à la **construction automatique de ressources lexicales**.

Les principales directions de recherches ont été:

- construction (semi) automatique, analyse et évaluation des **ressources lexicales numériques** pour le TALN (construction de ressources de type *wordnet* pour le domaine géographique ; extraction expressions multi-mots, acquisition de schémas de sous-catégorisation verbale, étude de la polysémie régulière)
- développement d'outils pour l'analyse automatique des corpus (en particulier pour l'analyse morphosyntaxique de l'italien, l'analyse du sentiment et la désambiguïsation automatique du sens)
- définition de standards de représentation pour les ressources linguistiques numériques (en particulier modèles pour la représentation en **rdf/owl des ressources linguistiques**);

En 2014-2015 j'ai obtenu une Bourse "Fernand Braudel" de la FMSH pour un séjour de 9 mois à Paris, au sein du **Labex OBVIL et du LIP6 (UPMC)** dans le cadre de projets d'**humanités numériques**.

Au sein du Labex OBVIL, j'ai approché de deux thématiques:

- **stylistique computationnelle**, en particulier le développement d'une méthodologie permettant de détecter les patrons syntaxiques distinctifs dans des textes littéraires
- **annotation et référencement automatique d'entités nommées (NERC et NEL)**

De retour à Pise en tant que chercheur CDD auprès de ILC CNR (2015-2016), j'ai continué mes recherches de linguistique numérique dans le cadre de l'infrastructure européenne pour les ressources linguistiques CLARIN (Common Language Infrastructure for Social Sciences and Humanities), dont ILC est le coordinateur national.

A présent je continue les recherches sur NERC et NEL ainsi que sur la stylométrie en collaboration le LIP6 et le EHESS. Je collabore aussi avec la communauté w3c « OntoLex » pour le développement de standards pour les ressources linguistiques et lexicales dans le web sémantique (*linguistic linked data* et *lexical linked data*). Enfin je maintiens le contact avec le réseau des infrastructures de recherche pour la linguistique et les sciences humaines, en particulier avec l'infrastructure CLARIN.

Quant aux champs d'application, mes travaux en TAL ont trouvé application soit dans le domaine des technologies de la fouille de données pour l'industrie soit dans la construction de méthodologies pour l'analyse textuelle outillée dans les sciences humaines. En particulier mes contacts avec la communauté des humanités numériques se sont récemment très intensifiés avec la publication de

papiers dans des journaux du secteur, la participation à des conférences DH et l'organisation de colloques dans ce domaine.