

Le sens de la famille : analyse du vocabulaire de la parenté par les plongements de mots

Ludovic Tanguy¹, Cécile Fabre², Nabil Hathout³, Lydia-Mai Ho-Dac⁴

¹CLLE : CNRS & Université de Toulouse – ludovic.tanguy@univ-tlse2.fr

²CLLE : CNRS & Université de Toulouse – cecile.fabre@univ-tlse2.fr

³CLLE : CNRS & Université de Toulouse – nabil.hathout@univ-tlse2.fr

⁴CLLE : CNRS & Université de Toulouse – lydia-mai.ho-dac@univ-tlse2.fr

Abstract

In this study, we propose a corpus analysis of an area of the French lexicon that is both dense and highly structured: the vocabulary of family relationships. Starting with a lexicon of 25 nouns designating the main relationships (son, cousin, mother, grandfather, sister-in-law etc.), we examine how these terms are positioned in relation to each other through distributional analyses based on the use of these terms in corpora. We show that distributional information can capture certain features that organize this vocabulary (descent, alliance, siblings, genre), in ways that vary according to the different corpora compared.

Keywords: kinship terms, family relation, word embeddings.

Résumé

Dans cette étude nous proposons une analyse en corpus d'une zone du lexique français à la fois dense et très structurée : le vocabulaire des relations familiales. En partant d'un lexique de 25 noms désignant les principales relations (fils, cousine, mère, grand-père, belle-sœur etc.), nous avons examiné la façon dont ces termes se positionnent les uns par rapport aux autres au travers des analyses distributionnelles basées sur l'emploi de ces termes en corpus. Nous montrons que l'information distributionnelle permet de capter certaines caractéristiques qui organisent ce vocabulaire (ascendance, alliance, germanité, genre), selon des modalités qui diffèrent dans les différents corpus que nous analysons.

Mots clés : noms de parenté, relations familiales, plongements de mots