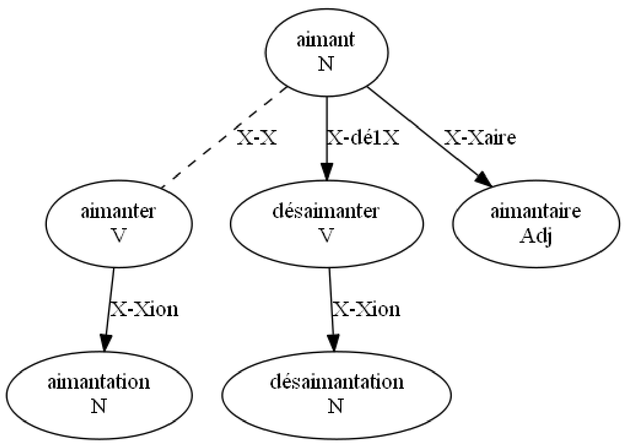
# Organizing and Improving a Database of French Word Formation Using Formal Concept Analysis

**Nyoman Juniarta, Olivier Bonami, Nabil Hathout, Fiammetta Namer, Yannick Toussaint**
***Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France***
**nyoman.juniarta@loria.fr**

## 1. Introduction

*Démonette*: a derivational database that systematically describes the **derivational properties** of a fragment of the French lexicon.

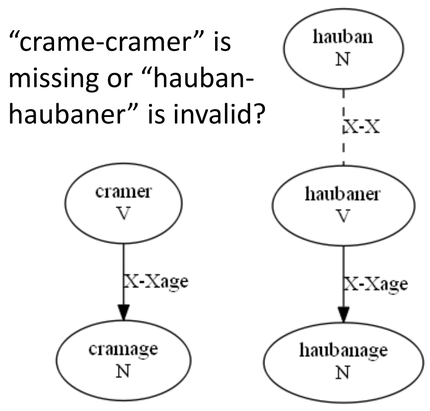



The derivational graph of family *aimant*

Research questions:

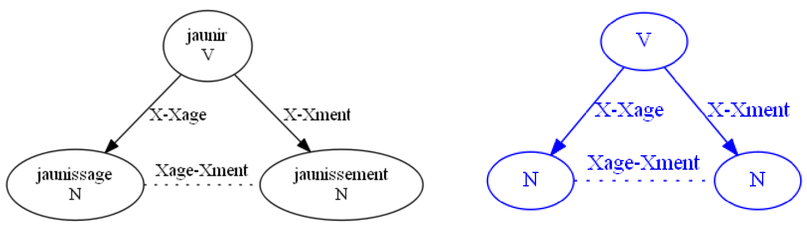
1. How to systematically represent the relations among 13K families
2. How to detect families having **anomalies**, i.e. having either **missing** or **incorrect** derivations.

"crame-cramer" is missing or "hauban-haubaner" is invalid?



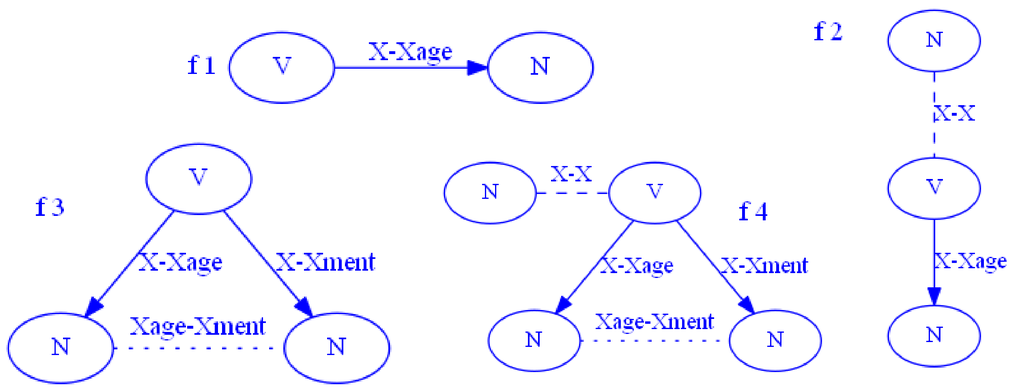

## 2. Fingerprint

Fingerprint of a family is the family's graph without lexemes.



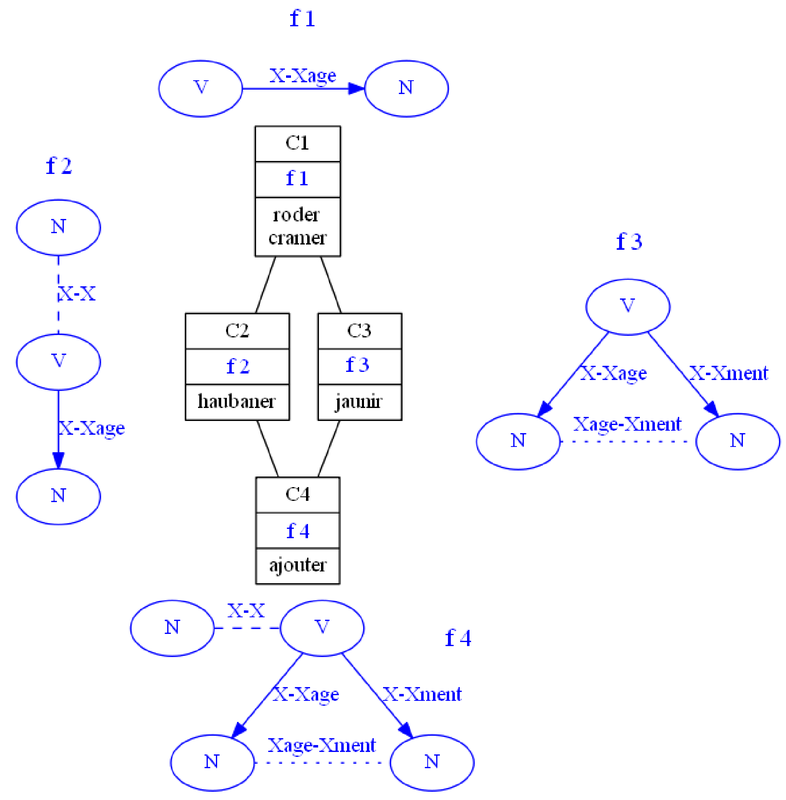

A fingerprint can correspond to multiple families.
Among 13K families, there are 4K unique fingerprints.

## 3. Formal context



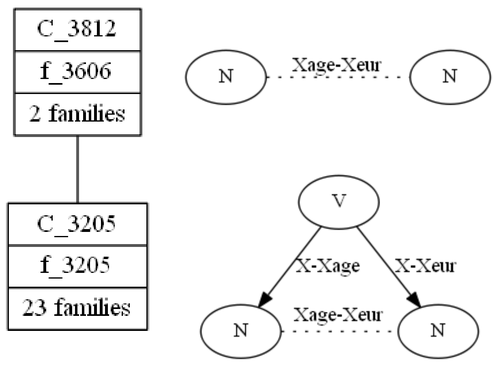

| Family | f 1 | f 2 | f 3 | f 4 |
|---|---|---|---|---|
| *roder* | × | | | |
| *cramer* | × | | | |
| *haubaner* | × | × | | |
| *jaunir* | × | | × | |
| *ajout* | × | × | × | × |

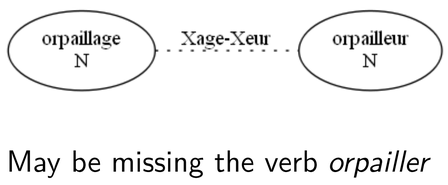
## 4. AOC-poset of derivational families



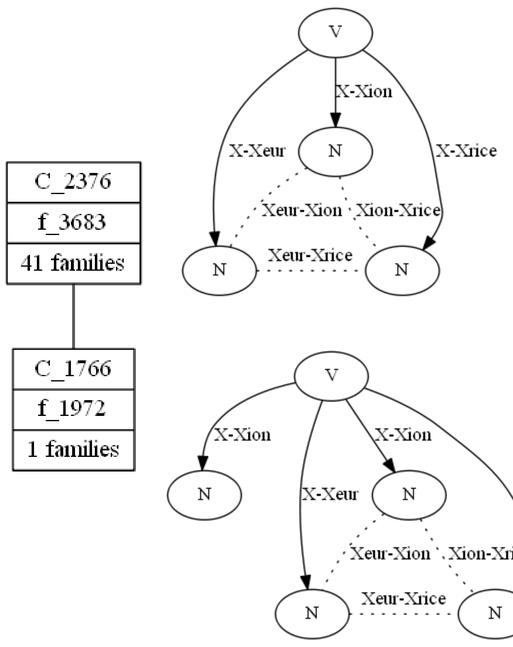

- Shows how a fingerprint can *develop* to other fingerprints
- Shows how a fingerprint is a *combination* of other fingerprints

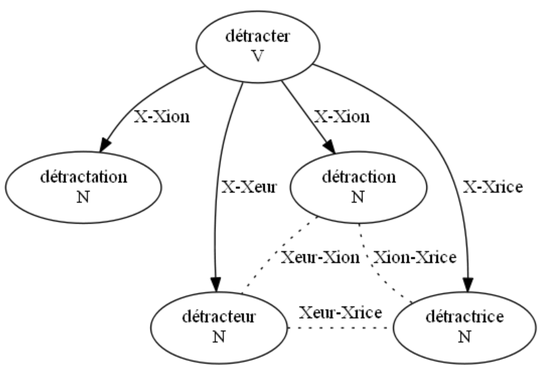
## 5. Anomaly detection using AOC-poset




May be missing the verb *orpailler*




*détracter-détractation* may be incorrect

## 6. References

1. Dolques, X. et al. (2013). AOC-posets: a scalable alternative to concept lattices for relational concept analysis. In *Proceedings of the 10th International Conference on Concept Lattices and Their Applications (CLA)*.

2. Papay, S. et al. (2017). Evaluating and improving a derivational lexicon with graphtheoretical methods. In *Proceedings of the 1st International Workshop on Resources and Tools for Derivational Morphology (DeriMo)*.

3. Namer, F. et al. (2019). Demonette2 - Une base de donnees dérivationnelle du français à grande échelle : premiers résultats. In *Actes de la Conference sur le Traitement Automatique des Langues Naturelles (TALN) PFIA 2019*.

Laboratoire de linguistique formelle
Loria Laboratoire lorrain de recherche en informatique et ses applications
atilf ANALYSE ET TRAITEMENT INFORMATIQUE DE LA LANGUE FRANÇAISE
Clle ERSS UMR 5263 COGNITION, LANGUES, LANGAGE, ERGONOMIE