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

Nyoman Juniarta, Olivier Bonami, Nabil Hathout, Fiammetta Namer, Yannick Toussaint

Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 3969–3976
Marseille, 20-25 June 2022

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

Nyoman Juniarta[1], Olivier Bonami[2], Nabil Hathout[3], Fiammetta Namer[4], Yannick Toussaint[1]
[1] Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France
[2] Université de Paris, CNRS, Laboratoire de linguistique formelle, F-75013, Paris, France
[3] CLLE, CNRS, Université de Toulouse Jean Jaures
[4] ATILF, Université de Lorraine, CNRS
{nyoman.juniarta, yannick.toussaint}@loria.fr
olivier.bonami@linguist.univ-paris-diderot.fr, nabil.hathout@univ-tlse2.fr, fiammetta.namer@univ-lorraine.fr

Abstract

We apply Formal Concept Analysis (FCA) to organize and to improve the quality of Démonette2, a French derivational database, through a detection of both missing and spurious derivations in the database. We represent each derivational family as a graph. Given that the subgraph relation exists among derivational families, FCA can group families and represent them in a partially ordered set (poset). This poset is also useful for improving the database. A family is regarded as a possible anomaly (meaning that it may have missing and/or spurious derivations) if its derivational graph is almost, but not completely identical to a large number of other families.

Keywords: derivation, formal concept analysis, morphology, subgraph matching

1. Introduction

Démonette2 (Hathout and Namer, 2016; Namer et al., 2019; Namer and Hathout, 2020) is a derivational database that systematically describes the derivational properties of a fragment of the French lexicon. In the database, an entry corresponds to a pair of lexemes from the same derivational family.

A derivational family is a set of lexemes connected by derivational relations. Fig. 1 shows examples of five derivational families. Each family is shown with its pair(s) of lexemes and the morphological pattern relating each pair. The family of *cramer*, for example, has only one pair describing the direct derivation from the verb *cramer*$_{\text{V}}$ to the noun *cramage*$_{\text{N}}$ using the morphological pattern X-Xage.

The set of derivations in the family of *cramer* is similar to that of *haubaner*, allowing them to be grouped in the same paradigm. This similarity also poses some questions. Is it possible that a derivational relation (*cramer*$_{\text{V}}$ — *crame*$_{\text{N}}$) is missing in the family of *cramer*? Or conversely, is the derivation *haubaner*$_{\text{V}}$ — *hauban*$_{\text{N}}$ erroneous? It is also possible that these derivations are correct as such, the two families belong then to two different morphological paradigms.

This paper deals with the following questions:

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

The presence of anomalies in Démonette2 is not surprising, given that it is built from heterogeneous resources. This merge generated some mistakes that should be corrected.

The improvement of derivational databases of numerous languages has been studied: Bulgarian (Dimitrova et al., 2014), Croatian (Filko and Šojat, 2017; Filko et al., 2019), Czech (Ševčíková et al., 2018), Hungarian (Trón et al., 2006), and Polish (Dziob and Walentynowicz, 2021). A method based on graph theory has been proposed to improve a German derivational database (Papay et al., 2017). This method focuses on the *fingerprint* of each family, which is the graph describing the family's derivational relations. A fingerprint contains labelled edges with "anonymized" vertices. A family with incorrect derivations is assumed to have a fingerprint that is similar but not identical to the fingerprint of a large number of families.

In this paper, we go further by including part of speech information (verb, noun, adjective, etc.) in each vertex in a fingerprint. Moreover, we also take into account that a fingerprint may be a subgraph of another larger fingerprint, given that a set of derivations may be a subset of another.

Our hypothesis is as follows. Among the families that contain the smaller fingerprint, if a large percentage of them also have d, then the families that do not have d are likely to be incomplete. Conversely, if only a small percentage of them also have d, then this node d is likely to be spurious.

We can consequently use the subgraph relation among fingerprints to define a partial order among derivational families. The systematic representation of families using this partial order can be performed using Formal Concept Analysis (FCA). The partially ordered set (poset) generated by FCA allows us to study the grouping of the families and the overlap among families' set of derivations.

Grouping families according to their paradigm or fin-

Family of *roder* *roder*$_{\text{V}}$ → *rodage*$_{\text{N}}$: X-Xage 'run in' 'running in'
Family of *cramer* *cramer*$_{\text{V}}$ → *cramage*$_{\text{N}}$: X-Xage 'burn' 'burning'
Family of *haubaner* *haubaner*$_{\text{V}}$ → *haubanage*$_{\text{N}}$: X-Xage 'stabilize by shroud' 'shroud-stabilizing' *haubaner*$_{\text{V}}$ — *hauban*$_{\text{N}}$: X-X 'stabilize by shroud' 'nautical shroud'
Family of *jaunir* *jaunir*$_{\text{V}}$ → *jaunissage*$_{\text{N}}$: X-Xage 'turn yellow' 'yellowing' *jaunir*$_{\text{V}}$ → *jaunissement*$_{\text{N}}$: X-Xment 'turn yellow' 'having yellowed' *jaunissage*$_{\text{N}}$ ⋯ *jaunissement*$_{\text{N}}$: Xage-Xment 'yellowing' 'having yellowed'
Family of *ajouter* *ajouter*$_{\text{V}}$ → *ajoutage*$_{\text{N}}$: X-Xage 'add' 'nozzle' *ajouter*$_{\text{V}}$ → *ajoutement*$_{\text{N}}$: X-Xment 'add' 'text addition' *ajouter*$_{\text{V}}$ — *ajout*$_{\text{N}}$: X-X 'add' 'something added' *ajoutage*$_{\text{N}}$ ⋯ *ajoutement*$_{\text{N}}$: Xage-Xment 'nozzle' 'text addition'

Figure 1: Examples of five small derivational families. Démonette2 has three types of derivation: direct, indirect, and undecidable; here represented by →, ⋯, and — respectively. The family of *jaunir* illustrates the incompleteness of the database, as it is missing *jaune* 'yellow'.

gerprint offers many advantages. One principal advantage is to provide a new organization of lexicon. The poset also allows us to easily obtain the number of families having a particular fingerprint, which is needed in detecting incorrect derivations as previously explained.

2. Formal Concept Analysis

FCA is a mathematical framework based on lattice theory and used for classification, data analysis, and knowledge discovery (Ganter and Wille, 1999). From a formal context, FCA builds all formal concepts, and arranges them in a concept lattice.

A formal context is a binary table of objects and their attributes, usually depicted as a cross-table with set of objects G as rows and set of attributes M as columns. A cross in the table signifies that object $g \in G$ has attribute $m \in M$. An example of a formal context with 7 objects $(g_1, g_2, \cdots g_7)$ and 7 attributes $(m_1, m_2, \cdots m_7)$ is given in Table 1.

A formal **concept** is a pair $(A \subseteq G, B \subseteq M)$, where A and B are called extent and intent, respectively. A formal concept corresponds to the maximal set of objects A sharing a set of attributes B, while B is the maximal set of attributes shared by A. The pair $(\{g_2, g_3\}, \{m_1, m_2, m_3, m_5, m_6\})$ is a concept from Table 1. In the cross-table depiction, a formal concept

	m_1	m_2	m_3	m_4	m_5	m_6	m_7
g_1		×					
g_2	×	×	×		×	×	×
g_3	×	×	×		×	×	
g_4		×	×	×			×
g_5		×			×		
g_6		×		×			×
g_7	×	×	×		×		×

Table 1: A formal context

corresponds to a maximal rectangle of crossed cells.

A concept (A_1, B_1) is **subconcept** of (A_2, B_2), written as $(A_1, B_1) \leq (A_2, B_2)$, iff $A_1 \subseteq A2$. In this case, (A_2, B_2) is superconcept of (A_1, B_1). With the subsumption relation $\leq$, the set of all concepts is partially ordered, and can be represented as a concept lattice.

2.1. Concept Lattice

The concept lattice of Table 1 is illustrated in Fig. 2 left. It shows all concepts from Table 1 and the $\leq$ relation among them. A concept is depicted as a rectangle with three parts, from top to bottom: concept number, intent, and extent. Given two concepts C_a and C_b, if there is a downward path from C_a to C_b, then C_a is superconcept of C_b, and consequently C_b is subconcept of C_a.

The lattice in Fig. 2 left is illustrated in a **full** form. All objects and attributes of the extent and intent, respectively, are listed. A **simplified** but equivalent form consists in listing only objects and attributes the first time they occur in the lattice. **Inheritance** from top to bottom for attributes, and from bottom to top for objects, allows us to calculate the complete extent and intent for each concept. This simplified version is illustrated in Fig. 2 middle.

Thus, an object and an attribute are **introduced** in the lowest and the highest concept, respectively, where they are present. In Fig. 2 middle, m_6 is only shown in C_8, the first time it occurs from top to bottom. The extent of C_8 is g_3 and its inherited object: g_2, while its intent is m_6 and its inherited attributes: m_1, m_2, m_3, and m_5.

2.2. AOC-Poset

A concept lattice may have $2^{min(|G|,|M|)}$ concepts. The large number of concepts that can be found in a formal context could render the lattice too complex to be explored. The use of an Attribute-Object-Concept (AOC) poset (Dolques et al., 2013) is one approach to reduce this complexity.

Instead of retrieving all concepts, the AOC-poset is restricted to object-concepts (which introduce at least one object) and attribute-concepts (which introduce at least one attribute), keeping the $\leq$ relation among selected concepts. In Fig. 2 middle, we can see that C_6 and C_{12} are neither object-concept nor attribute-concept. Consequently, they are not included in the corresponding AOC-poset.

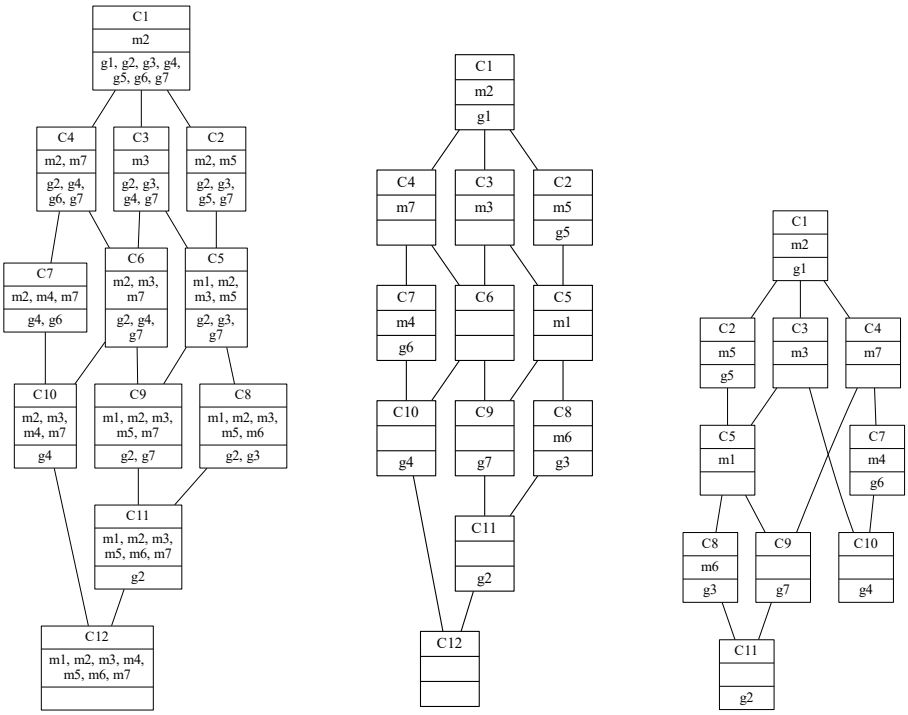

Figure 2: The concept lattice in full form (left), the concept lattice in simplified form (middle), and the AOC-poset in simplified form (right) of the formal context in Table 1.

We define the number of **levels** in an AOC-poset as the longest top-down path. The AOC-poset in Fig. 2 right contains 5 levels ($C_1 \rightarrow C_2 \rightarrow C_5 \rightarrow C_8 \rightarrow C_{11}$).

We also define a concept C_a as a **child** of C_b if $C_a \leq C_b$ and there is no concept between them. The concept C_9 in Fig. 2 right is a lower neighbor of C_4 and C_5. Those two concepts are called **parents** of C_9.

2.3. Implication and Association Rules

Concept lattice and AOC-posets are both useful for extracting implication and association rules, since they provide hierarchical information among objects and attributes. For example, in Fig. 2 middle, m_1 and m_5 are introduced in C_5 and C_2, respectively, with $C_5 \leq C_2$. This means that any object having m_1 also has m_5, written as the implication $m_1 \rightarrow m_5$.

Conversely, we can also find the association rule $m_5 \rightarrow m_1$ with confidence $3/4 = 0.75$, since there are 4 objects in C_5 and 3 objects in C_2. This means that among 4 objects having m_5, 3 of them have also m_1.

The rules present in the lattices can be extracted from the corresponding AOC-poset. Consequently, the AOC-poset, which is simpler, is preferred.

Furthermore, an association rule can also highlight anomalies among objects and attributes, by looking at the confidence value. In the previously mentioned rule $m_5 \rightarrow m_1$, we see that among 4 objects having m_5, only one of them does not have m_1. This may indicate that this one object, g_5, is an anomaly. This anomaly detection is the basis of our detection of incorrect derivations in Démonette2.

3. Methodology

In this section we explain how to use FCA to represent derivational families in Démonette2 and to find possible inaccuracies in it.

Démonette2 is a database whose entries are pairs of derivationally related lexemes. Each entry has 38 columns describing the two lexemes and their relation. In this paper, we are focusing only on: the two lexemes and their part of speech (noun, verb, etc.), the orientation of the relation (direct, indirect, undecidable), and the morphological patterns relating the lexemes (X-Xion, X-Xment, etc.)

3.1. Démonette2 as Graphs

In a graph, we assign the lexemes as the labels of vertices, while the morphological patterns serve as the label of edges. The orientation of the derivation determines the edge type. Direct derivations are represented by solid directed edges, while indirect and undecidable derivations are represented by dashed and dotted undirected edges respectively. The graphs of families of the *cramer* and *haubaner* are depicted in Fig. 3.

A **fingerprint** (Papay et al., 2017) is a graph that illustrates the structure of a family, obtained by keeping the label of edges without labeling any vertice. Here we extend this notion by keeping part of speech information in the vertices. The fingerprints of *cramer* and

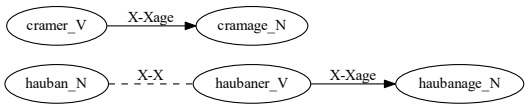

Figure 3: The graphs of *cramer* (top) and *haubaner* (bottom). A solid directed edge and a dashed edge represent a direct derivation and an undecidable derivation respectively.

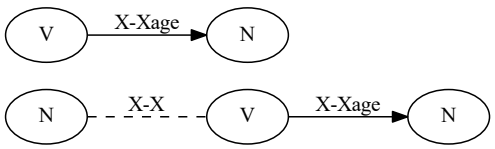

Figure 4: The fingerprint of *cramer* and *roder* (top), which is a subgraph of the fingerprint of *haubaner* (bottom).

haubaner are depicted in Fig. 4. A fingerprint can correspond to several families since a fingerprint does not contain the lexemes. This is the case of the family of *cramer* and *roder*, who share the fingerprint shown in Fig. 4 top.

The graph g_1 is a subgraph of g_2 (written $g_1 \subset g_2$) if g_1 is included in g_2. In Fig. 4, the top fingerprint is a subgraph of the bottom fingerprint, since the latter contains all vertices and edges of the former.

3.2. Creation of AOC-Poset

To build an AOC-poset, first we have to construct a formal context. As explained in Section 2, a formal context is a table of objects and their attributes. In this section, we present how to build formal context using fingerprints.

In the formal context, each object corresponds to a family, and each attribute corresponds to a fingerprint $f \in F$. The set of fingerprints F is partially ordered by subgraph inclusion. A family g possesses its own fingerprint f_g and any fingerprint f_i such that $f_i \in F$ and $f_i \subset f_g$.

The formal context of families in Fig. 1 is shown in Table 2. The family of *roder* and *cramer* share the same fingerprint f_1. The family of *haubaner* has not only its own fingerprint f_2, but also f_1 as $f_1 \subset f_2$. The family of *ajouter* has f_4, and also f_1, f_2, and f_3 since they are subgraphs of f_4.

The corresponding AOC-poset in simplified form is depicted in Fig. 5, together with the fingerprint of each concept.

Any resulting concept is an attribute concept, introducing exactly one fingerprint and one or more families sharing that fingerprint. Consequently, from the resulting AOC-poset, we can see how a family's fingerprint extends to another fingerprint, and how two fingerprints combine to form another fingerprint.

From Fig. 5, we see that f_1 grows by adding an indirect X-X derivation to become f_2 or by adding two derivations to become f_3. We also see that f_2 and f_3 combine to form f_4.

Family	Fingerprints			
	f_1	f_2	f_3	f_4
roder	×			
cramer	×			
haubaner	×	×		
jaunir	×		×	
ajouter	×	×	×	×

Table 2: The formal context of the 5 families in Fig. 1 and the 4 fingerprints shown in Fig. 5.

4. Poset Exploration

Currently, the Démonette2 database contains 51,830 unique pairs of lexemes and their description. From these pairs, we obtain 13,897 families and 4,849 different fingerprints. The number of families associated to each fingerprint is varied: 4,181 fingerprints are only associated to one family, while one small fingerprint (having only two nodes) shared with 1,381 families.

The formal context contains 13,897 objects and 4,849 attributes. The resulting AOC-poset has 8 levels and 4,849 concepts – one for each fingerprint – while 69 concepts are isolated, i.e. having neither subconcept nor superconcept.

There are 209 "top" concepts – having no parent and at least one child – which can be regarded as the beginning of a new paradigm.

The maximum number of parents and children of a given concept is 83 and 124 respectively. Among concepts having at least 1 parent, the average number of parents is 5.06; while among the concepts having at least 1 child, the average number of children is 10.08.

To detect possible incorrect derivations in a family, we focus on pairs of neighboring concepts with their corresponding association rule. Consider the concepts C_1 and C_2 in Fig. 6, with the fingerprint $f_1 \subset f_2$. The association rule $f_1 \rightarrow f_2$ has confidence $14/15$, meaning that among 15 families having f_1, only one of them does not have the larger f_2. Therefore, this one family may be missing some derivations.

On the other hand, consider two concepts C_3 and C_4 in Fig. 6, in simplified form, with the fingerprint $f_3 \subset f_4$. The association rule $f_3 \rightarrow f_4$ has confidence $1/16$, meaning that 15 families have exactly f_3, and only one has f_3 and f_4. Therefore, this one family may have some incorrect additional derivations.

Families with missing derivations. Fig. 7 presents one case of missing derivation found in Démonette2.

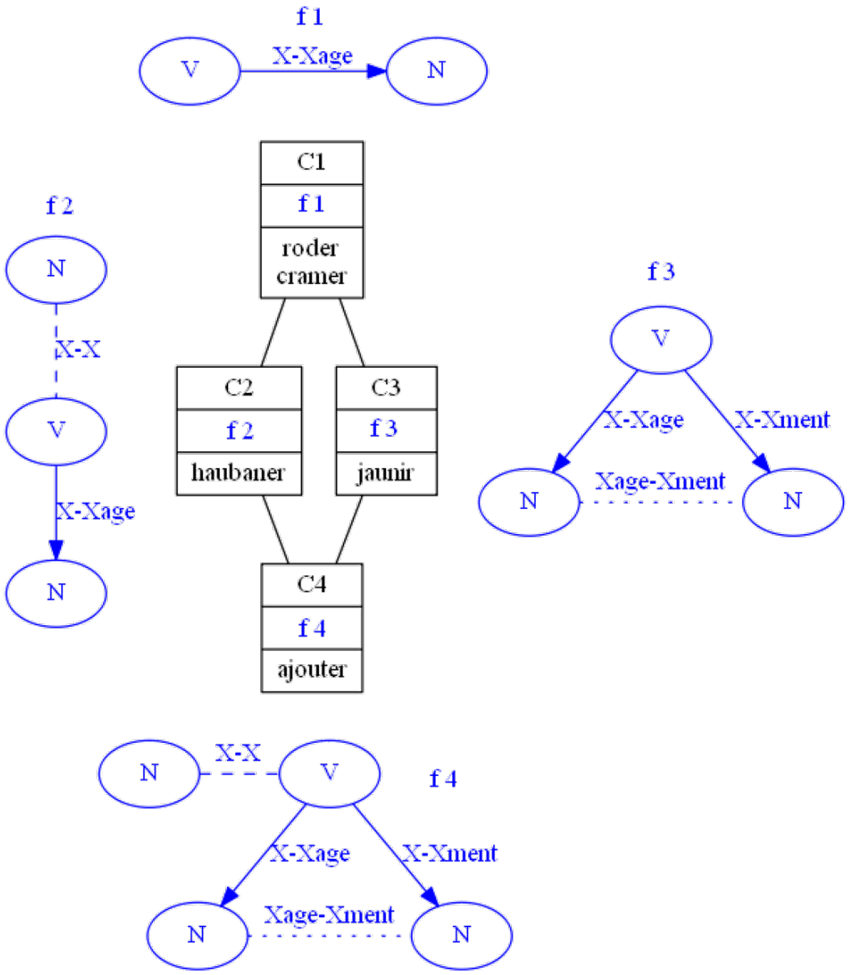

Figure 5: AOC-poset (in black) of Table 2 and to ease understanding we draw in blue the fingerprints involved in the intent of the concepts .

We see concept C_{3812} introduces fingerprint f_{3606}, and one of its subconcepts C_{3205} introduces f_{3842}. These two fingerprints are illustrated beside the concepts, and we see that f_{3606} is a subgraph of f_{3842}, with the latter having two more derivations (direct X-Xage and X-Xeur). Among 708 families having indirect derivation Xage-Xeur, 695 of them also have the direct derivations X-Xeur and X-Xage (e.g. the family *survolter* with lexemes $survolter_{\mathrm{V}}$, $survoltage_{\mathrm{N}}$, and $survolteur_{\mathrm{N}}$).

The remaining 13 families could be missing the X-Xage and X-Xeur derivations and should be checked by hand. An actual example of a missing derivation is that of the family *orpailleur*, which has only two lexemes $orpaillage_{\mathrm{N}}$ and $orpailleur_{\mathrm{N}}$, and an indirect derivation between them, without the lexeme $orpailler_{\mathrm{V}}$. We then propose the addition of the missing lexeme and the two missing derivations.

The family *radeau*, illustrated in Fig. 8, is another relevant example among the 13 families. This family has the indirect Xage-Xeur and direct X-Xage, but is missing the direct derivation X-Xeur from $radeler_{\mathrm{V}}$ to $radeleur_{\mathrm{N}}$. Notice that f_{3606} is a subgraph of *radeau*'s fingerprint. This shows the usefulness of FCA and its capacity to detect an anomaly in a family's subgraph.

Families with incorrect derivations. To find possibly incorrect derivations in Démonette2, we focus on two neighboring concepts where the subconcept introduces a low number of families compared to its superconcept.

This is the case of concepts C_{2376} and C_{1766} shown in Fig. 9. There are 41 families that have the fingerprint f_{3683}, while only one family has f_{1972}, which is f_{3683} with one additional X-Xion derivation. This one family is *détracter*, shown in Fig. 10. This family has the extra direct X-Xion derivation from $détracter_{\mathrm{V}}$ to $détractation_{\mathrm{N}}$, which is a rarely used synonym of $détraction_{\mathrm{N}}$.

Limitations of association rules. From Fig. 6 right, the example of f_4 introducing only one family may not mean that this one family has incorrect derivations. This is often the case when a family has extra valid derivations for different spelling, like *essuyement-essuiement*, *debuscage-debusquage*, etc., which should not be regarded as incorrect derivations.

Furthermore, many derivational families are indeed incomplete in the language, and should thus be docu-

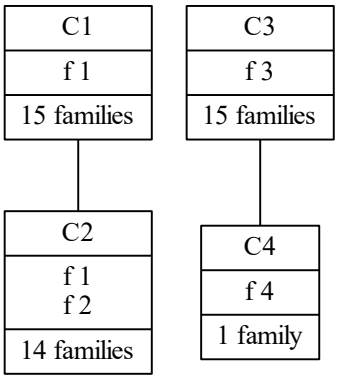

Figure 6: Illustration of two pairs of neighboring concepts. The pair in the left is shown in the full form, while the pair in the right is in the simplified form.

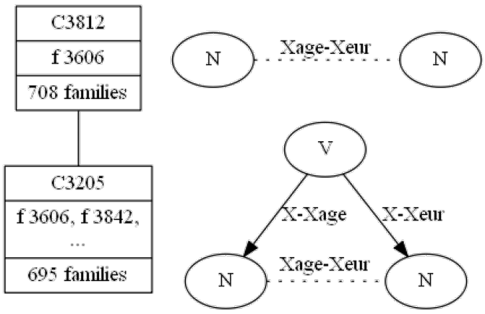

Figure 7: Two neighboring concepts and their fingerprints (f_{3606} and f_{3842}). They differ in X-Xage and X-Xeur.

mented as such in the database. This could lead to a pair of neighboring concepts similar to Fig. 6 right, where instead of incorrect extra derivations in C_4, it is the numerous families in C_3 that are missing some derivations.

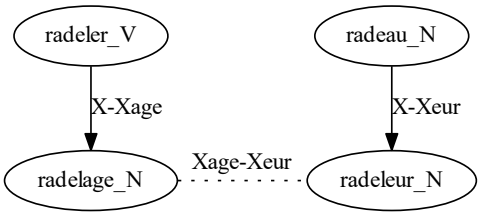

Figure 8: The graph of family *radeau*.

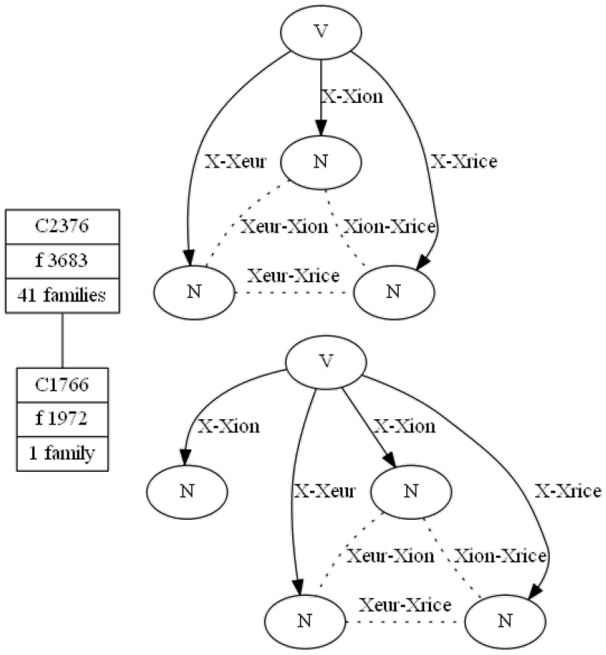

Figure 9: Two neighboring concepts C_{2376} and C_{1766}. C_{2376} introduces the fingerprint f_{3683}, shown in upper right, while C_{1766} introduces f_{1972}, shown in lower right.

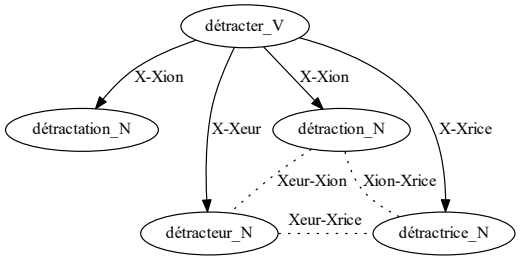

Figure 10: The graph of family *détracter*.

5. Validation

The anomalies found by the method explained in Section 4 can then be presented to linguists to check whether there are actually missing or incorrect derivations. In order to do this validation, we designed a simple web page to visualize our findings.

Missing derivations. A screenshot of the web page for the validation of missing derivation is shown in Fig. 11. In this example, the family *turban* may be missing an X-Xment derivation (in blue). We provide the prediction of the supposedly missing lexeme (*enturbannement*$_{\mathrm{N}}$) by performing a formal analogy (Lavallée and Langlais, 2010; Lepage, 1998; Stroppa and Yvon, 2005) from the two families, e.g.:

emmitoufler$_{\mathrm{V}}$: *emmitouflement*$_{\mathrm{N}}$ =
enturbanner$_{\mathrm{V}}$: ?

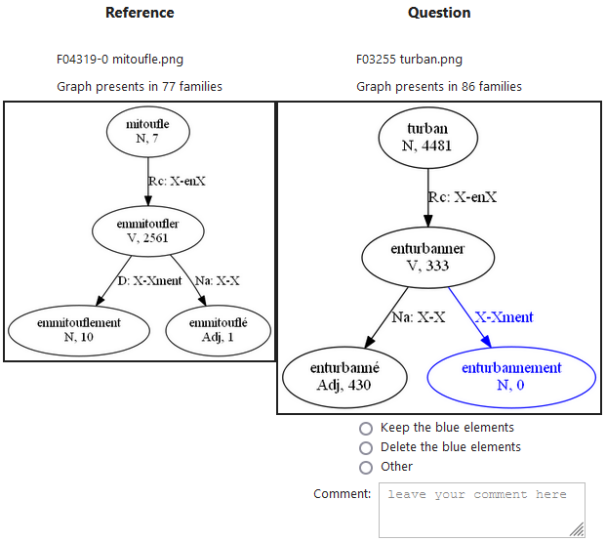

Figure 11: A web-based tool of revision of the possibly missing derivation. The *reference* family is the "normal" family, while the *question* is the family that is missing some derivations in blue. Each lexeme is accompanied by its category and its frequency from FRCOW corpora.

From this analogy, we generate the predicted lexeme *enturbannement* by replacing the affix according to the reference. We then check the frequency of the *enturbannement* in the French web corpus FRCOW (Bildhauer and Schäfer, 2016) to help the linguists decide whether the lexeme exists or not. The linguist/validator can choose one of the three options:

1. keep: the prediction is correct and will be added to the database;
2. delete: the family in question is correct as it is, no change needed in the database;
3. other: for other cases, e.g. when the family in question is indeed missing some derivations, but the given prediction is incorrect.

It should be noted that affix replacement will not always give the correct prediction. This is the case in $hydrocyclique_{\text{Adj}} : hydrocycliser_{\text{V}} = muet_{\text{Adj}} : ?$, where *hydrocyclique* and *muet* do not share any affix. In this case, we put "?" in the prediction, and the linguists can provide the intended lexeme (*mutiser*) in the comment.

Incorrect derivations. A page for the validation of incorrect derivations is shown in Fig. 12. The "delete" option here removes some elements in the family and updates the database, while the "keep" option does not update the database.

Tracing the source. To trace the origin of a derivation, additional information is added for labeling the edges. Beside the morphological pattern, we provide a code that corresponds to a source file. For example, the label of the edge between $palefrenat_{\text{N}}$ to $palefrenier_{\text{N}}$

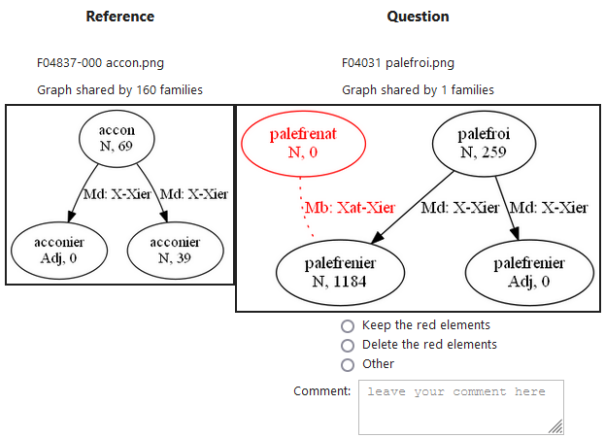

Figure 12: A web-based tool of revision of the possibly incorrect derivation, similar to the web page shown in Fig. 11.

in Fig. 12 is "Mb: Xat-Xier". While "Xat-Xier" is the morphological construction, "Mb" is a code that refers to the source file of the pair.

Update iterations. The result of these validations updates the database, hence updating the AOC-poset. Given a new AOC-poset, we find a new set of anomalies, and a new validation process is needed. The update of Démonette2 is thus iterative, but we can not know beforehand how many iterations it takes to converge.

6. Conclusion

We presented in this paper the potential of FCA in detecting both missing and spurious derivations in Démonette2. This detection is important in improving the quality and completeness of the database. Our methodology can easily be adapted to other word formation databases. Our hypothesis, stated in Section 1, may need some refinements, but it is a good starting point in detecting anomalies in a database.

However, the creation of formal context we presented in this paper is not the only possible way to build a formal context. Instead of calculating the subgraph relation among fingerprints, we may be interested in calculating all frequent subgraphs and study the grouping of families based on which subgraphs they share, leading to a more complex structure. Moreover, in addition to the morphological construction (X-Xier, X-Xion, etc.), the semantic information for each derivation may give us a better grouping of families.

The iterative update of Démonette2 should also be investigated, since the AOC-poset structure enables more than one update per iteration. It is also important to filter out the anomalies found in each iteration, since an anomaly that is already treated as valid should not be presented again as an anomaly.

7. Acknowledgment

This work is part of the project DEMONEXT, supported by the ANR-17-CE23-0005 of the French National Research Agency.

8. Bibliographical References

Dimitrova, T., Tarpomanova, E., and Rizov, B. (2014). Coping with derivation in the Bulgarian WordNet. In Heili Orav, et al., editors, *Proceedings of the 7th Global WordNet Conference*, pages 109–117, Tartu, Estonia, January. University of Tartu Press.

Dolques, X., Le Ber, F., and Huchard, M. (2013). AOC-posets: a scalable alternative to concept lattices for relational concept analysis. In Manuel Ojeda-Aciego et al., editors, *Proceedings of the 10th International Conference on Concept Lattices and Their Applications (CLA)*, pages 129–140, La Rochelle, France, October. CEUR-WS.org.

Dziob, A. and Walentynowicz, W. (2021). Enriching plWordNet with morphology. In Piek Vossen et al., editors, *Proceedings of the 11th Global WordNet Conference*, pages 175–181, Pretoria, South Africa, January. Global Wordnet Association.

Filko, M. and Šojat, K. (2017). Expansion of the derivational database for Croatian. In Eleonora Litta et al., editors, *Proceedings of the 1st International Workshop on Resources and Tools for Derivational Morphology (DeriMo)*, pages 27–37, Milan, Italy, October. EDUCatt.

Filko, M., Šojat, K., and Štefanec, V. (2019). Redesign of the Croatian derivational lexicon. In Zdeněk Žabokrtskỳ, et al., editors, *Proceedings of the 2nd International Workshop on Resources and Tools for Derivational Morphology (DeriMo)*, pages 71–80, Prague, Czechia, September. Charles University.

Ganter, B. and Wille, R. (1999). *Formal Concept Analysis: Mathematical Foundations*. Springer, 2nd edition.

Hathout, N. and Namer, F. (2016). Giving lexical resources a second life: Démonette, a multi-sourced morpho-semantic network for French. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the 10th Language Resources and Evaluation Conference (LREC)*, Portorož, Slovenia, May. European Language Resources Association (ELRA).

Lavallée, J.-F. and Langlais, P. (2010). Unsupervised morphological analysis by formal analogy. In Carol Peters, et al., editors, *Multilingual Information Access Evaluation I. Text Retrieval Experiments*, pages 617–624, Corfu, Greece, October. Springer.

Lepage, Y. (1998). Solving analogies on words: an algorithm. In Christian Boitet et al., editors, *Proceedings of COLING-ACL 1998, vol. 1*, pages 728–735, Montreal, Canada, August. Université de Montréal.

Namer, F. and Hathout, N. (2020). ParaDis and Démonette - from theory to resources for derivational paradigms. *Prague Bull. Math. Linguistics*, 114:5–34.

Namer, F., Barque, L., Bonami, O., Haas, P., Hathout, N., and Tribout, D. (2019). Demonette2 - Une base de données dérivationnelle du français à grande échelle : premiers résultats. In Emmanuel Morin, et al., editors, *Actes de la Conférence sur le Traitement Automatique des Langues Naturelles (TALN) PFIA 2019. Volume II : Articles courts*, pages 233–244, Toulouse, France, July. ATALA.

Papay, S., Lapesa, G., and Padó, S. (2017). Evaluating and improving a derivational lexicon with graph-theoretical methods. In Eleonora Litta et al., editors, *Proceedings of the 1st International Workshop on Resources and Tools for Derivational Morphology (DeriMo)*, pages 73–82, Milan, Italy, October. EDUCatt.

Ševčíková, M., Kalužová, A., and Žabokrtskỳ, Z. (2018). A language resource specialized in Czech word-formation: Recent achievements in developing the DeriNet database. In Michal Křen (chair) et al., editors, *Proceedings of the SlaviCorp - Corpora of Slavic Languages*, Prague, Czechia, September. Charles University.

Stroppa, N. and Yvon, F. (2005). Analogical learning and formal proportions: Definitions and methodological issues. *ENST Paris Report*.

Trón, V., Halácsy, P., Rebrus, P., Rung, A., Vajda, P., and Simon, E. (2006). Morphdb. hu: Hungarian lexical database and morphological grammar. In Nicoletta Calzolari (Conference chair), et al., editors, *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*, pages 1670–1673, Genoa, Italy, May. European Language Resources Association (ELRA).

9. Language Resource References

Bildhauer, F. and Schäfer, R. (2016). *COW: Free state-of-the-art web corpora, frequency lists, and link data*. https://corporafromtheweb.org/frcow16/.