



HAL
open science

The MoNoPoli database

Mathilde Huguin

► **To cite this version:**

Mathilde Huguin. The MoNoPoli database. Third International Workshop on Resources and Tools for Derivational Morphology (DeriMo 2021), Fiammetta Namer; Nabil Hathout; Stéphanie Lignon; Magda Ševčíková; Zdeněk Žabokrtský, Sep 2021, Nancy, France. pp.76-85. hal-03374883

HAL Id: hal-03374883

<https://hal.science/hal-03374883v1>

Submitted on 12 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

The MoNoPoLI database Or how to catch *Macron-itis*

Mathilde Huguin

ATILF (CNRS-UL) UMR 7118 / Nancy, France
mathilde.huguin@univ-lorraine.fr

Abstract

In this article we present our method to build a derivational database of French deanthroponyms, which we call MoNoPoLI for *Mots construits sur Noms propres de personnalités Politiques*, ‘complex words based on politician proper names’. MoNoPoLI contains 6,545 complex words amounting to a total of 55,030 tokens and includes almost only neologistic forms. The Web is the only conceivable resource for collecting them: it alone gives massive access to discourse genres that contain neologisms. To feed the database, a program automatically generates the set of all possible derived words. Generated forms are then used as queries on the Web. Attested forms are kept with their context. This method provides a potential alternative to collect data that cannot be found elsewhere. Finally, this article describes some of the remarkable results obtained with the analysis of the deanthroponyms of MoNoPoLI.

1 Introduction

We study French deanthroponyms, i.e. words morphologically built on proper names (Schweickard, 1992; Leroy, 2008; Schlücker and Ackermann, 2017) that refer to contemporary political figures, henceforth PPN ‘politician proper name’. These data have the particularity of being absent from most existing French corpora since they are neologisms and often exhibit the properties of nonce-formations. According to Bauer (1983) and Dal and Namer (2018), nonce-formations are words deemed to be new by their creators and used intentionally to meet an immediate need in a given context. In (1a) and (1b), the forms *Macronite* (‘Macron-itis’) and *aubrycratie* (‘aubry-cracry’) are intentionally used by writers to express their aversion to the referents of the PPNs or their political ideas/actions.

- (1) a. *Un nouveau cas de **Macronite** aiguë était signalé en France. (Emmanuel Macron)*
‘A new case of acute Macron-itis was reported in France.’
- b. *Il faut dire à ces militants de ne pas confondre démocratie et **aubrycratie**. (Martine Aubry)*
‘It is necessary to say to these militants not to confuse democracy and aubry-cracry.’

As these derivatives are most often neologistic nonce-formations, the constitution of the corpus required the elaboration of a specific methodology which we describe in this article. This methodology consists of two steps. First, we automatically generate hypothetical deanthroponym *candidates* (e.g. *françoishollandien*, ‘françoisholland-ian’, *lepenphobe*, ‘lepen-phobic’) using the 89 PPNs and 90 suffixes of Huguin (2018). We then look up these candidates in context, using the Web. We also present some of the results we obtained focusing on two types of atypical constructions which characterize deanthroponyms.

Our presentation will be structured as follows. We explain why the Web is the most likely resource to contain deanthroponyms (§2). We discuss two types of possible strategies to collect them and evaluate

Each example of deanthroponym is provided with its context; the base PPN is indicated between brackets and the stem/suffix boundary is marked by a hyphen in the English translation.

Both lists are available online at this address: <https://perso.atilf.fr/mhuguin/accueil/these/documents/>.

their theoretical implications (§3). We present the architecture of the program that allowed their generation (§4) and the results of our collection (§5). Finally, we present some of the results that emerged from the analysis of our data (§6).

2 Where to find data

In order to know where to look for deanthroponyms, one must ask about their characteristics, including their degree of institutionalization (Hohenhaus, 2005) or their discursive function.

The list of PPNs used to generate candidates contains names of politicians who have held a prominent position in French politics (Presidents, leaders of political party, etc.). They have occupied this role since 1981, so they are contemporary referents (e.g. Jacques Chirac, Emmanuel Macron). As we have selected names of current personalities, we expect that the words based on PPNs are recent creations, i.e. neologisms (Štekauer, 2002; Kerremans, 2015).

Given that these politicians make decisions that directly impact the French people, one can assume that deanthroponyms formed on their names will occur in puns, jokes, criticisms or claims. Hence we can expect that the complex words we are going to find will occupy argumentative or humorous functions. Therefore they display the characteristics of nonce-formations (Hohenhaus, 2015) as in (2). In (2a), alongside their morphological creations *royaliste* ('follower of Ségolène Royal') and *montebourgeois* ('follower of Arnaud Montebourg'), the writer inserts a meta-discursive comment: "I don't know if that's how you say it". The deanthroponym *hollandophobe*_{Adj} 'hollando-phobic' (2b) appears in a sequence that contains several terms of the same series (*Xphobe*_{Adj}), which Tanguy (2012, p. 104) calls suffixal outbursts. Comments and outbursts are among the structures that Dal and Namer (2018) have coined (meta)discursive and that often overlap with nonce-formations.

- (2) a. Perdre la raison, *un blog militant*. *Longtemps royaliste, maintenant montebourgeois (je ne sais pas si ça se dit comme ça).*
 'Lose his mind, a militant blog. Long time royal-ist, now montebourge-ian (I don't know if that's how you say it).'
- b. *Il y a de quoi venir phobe : hollandephobe, vallsphobe, taubiraphobe, belkacemophobe, gauchophobe, antifaphobe, imamophobe, racaillophobe.* (*François Hollande, Manuel Valls, Christiane Taubira, Najat-Vallaud Belkacem*)
 'There is enough to come phobic: hollande-phobic, valls-phobic, taubira-phobic, belkacemophobic, lefto-phobic, antifa-phobic, imamo-phobic, gangstero-phobic.'

Neologisms are more frequent in opinion genres than in information genres (Gérard, 2018). They indeed tend to be more frequently attested in less formal—or even satirical—contexts. In order to maximize our chances of obtaining them, we should look for resources where speakers/writers will be able to express themselves freely, and where they will be able to reach a wide audience. Social networks, forums and blogs, which are genres specific to the Web, provide such freedom and audience. To build up our corpus, we used the Web as a resource since it alone provides access to these discursive genres in real time.

Lüdeling et al. (2007); Fradin et al. (2008); Dal and Namer (2015) among others, have shown that the Web is useful for collecting contextualized lexical scarcities. Since search engines are constantly performing new indexing, they provide access not only to forms that have been recorded for a long time but also to very recent coinages. To automatically and massively explore the content of the Web, we used a Web scraping program to query the Bing search engine. Our approach can be described as hypothetico-deductive (Tanguy, 2012, p. 101): we first generated a list of deanthroponym candidates and then searched for contexts containing the members of this list on the Web.

The program we use is provided by the company Data-Observer. Data Observer (www.data-observer.com) is a start-up specialized in the collection, processing and analysis of textual data from the Web.

3 Possible strategies

The hypothetical deanthroponyms used as queries are built from PPNs (3a) by means of suffixes (3b). When generating candidates from such inputs, two strategies can be adopted. Each strategy has its theoretical implications.

- (3) a. *Christiane Taubira, Emmanuel Macron, Jacques Chirac, Jean-Marie Le Pen* [...]
 b. *able, erie, ification, isme, isterie, ix, logue, mètre, oïa, thon, us* [...]

The first strategy, which we call *minimal* strategy, consists in generating only morphologically well-formed candidates. They respect a set of wellformedness morphophonological constraints (Roché and Plénat, 2014) as in *Optimality Theory* (Prince and Smolensky, 1993). For example, this strategy leads to build exclusively *taubirie* /tobiri/ ('taubir-land') from the inputs *Taubira* /tobira/ (from *Christiane Taubira*) and *-ie* /i/ so as to: (i) avoid the hiatus /ai/ proscribed by the markedness constraint *HIATUS (/tobira/ + /i/ = */tobirai/), (ii) tend towards the trisyllabic optimal, required according to the size constraints (Plénat, 2009). The objective of this method is to model the repair strategies instinctively implemented by speakers—and assumed by the linguist—to obtain an *optimal* derivative. This first strategy therefore assumes that speakers/writers always (unconsciously) apply the phonotactic constraints and/or that we are only interested in well-formed deanthroponyms.

With the *maximal* strategy, all possible forms are generated, regardless of their adequacy to well-formedness principles. This strategy corresponds to the hypothesis that a speaker/writer may ignore morphophonological constraints of wellformedness, especially in a situation of spontaneous written expression. For example, the sequence /rari/ from /tobirari/, which corresponds to the attested form *Taubirarie* from (4a) entails that the derivative violates the constraints of faithfulness, size as well as the *Obligatory Contour Principle* (Goldsmith, 1976). Faced with such attested examples, we opt for the *maximal* strategy. Moreover, the hierarchy of phonological constraints is not known. We regularly observe several output variants of a construction process, as the derivatives of (4) attest. The variants in the output of a morphological construction process are due to the idiosyncratic ordering of constraints as shown by Roché (2010). In sum, we choose to generate as many forms as possible using PPN stems or variants thereof and a list of suffixes. (5) is an excerpt from the set of graphical forms obtained from *Sarkozy* (from PPN *Nicolas Sarkozy*) and the French suffix *-able*.

- (4) a. *Mais où sommes-nous ? En France ? Ou Taubirarie ?* (*Christiane Taubira*)
 'But where are we? In France? Or Taubirar-land?'
 b. *Vous vous foutez de qui en Taubirie ?* (*Christiane Taubira*)
 'You do not care who in Taubir-land?'
 c. *Il risque très peu en Taubirasie... no problemo.* (*Christiane Taubira*)
 'He risks very little in Taubiras-land ... no problemo.'

- (5) *sarkozyable, sarkozable, sarkozysable, sarkozysable, sarkozytable, sarkozylable, sarkozydable* [...]

In terms of costs and benefits, the maximal strategy produces much more noise than the minimal strategy. The higher the number of queries, the higher the noise. Nevertheless, the noise is an inconvenience has a lesser impact than the dearth of results from the minimal strategy. Noisy results can be filtered out, whereas the lack of data cannot be compensated. In addition, this strategy allows us to collect unexpected formal creations, and, consequently, nonce-formations and extravagant formations that the minimal strategy does not allow us to obtain because it obeys morphological standards.

4 Generating derived forms

We run our candidate generation program on all the graphical forms that realize each PPN in our list and all suffixes from our set. PPNs are indeed realized in different forms, at least 3 (the *first name*, the *last name*, the *full name*), and up to 6, which we call sub-names and present in Table 1. The sub-names of a

PPN are coreferential names that are used both autonomously in syntax and as bases in derivation. Unlike what happens with lexeme stems, derivation rules do not impose constraints on sub-names, which is an additional argument for choosing the maximal strategy. We have shown (Huguin, 2018; Lignon et al., 2019) that the selection of a sub-name depends on sociolinguistic or extralinguistic conditionings such as the gender of the referent: e.g. the *firstname* is more used if the referent is a woman (6).

- (6) a. *On a déjà assez à faire pour lutter contre la najatisation de l'enseignement ! (Najat Vallaud-Belkacem)*
 ‘We are already busy enough fighting against the najat-ization of education!’
 b. *Finally le clientélisme et le clémentinisme se rejoignent. (Clémentine Autain)*
 ‘Finally, clientelism and clémentin-ism come together.’

Sub-names	Examples	Derived forms	Gloss
<i>Last name</i>	Strauss-Kahn	<i>strausskahnité</i>	‘strausskahn-ity’
<i>First name</i>	Dominique	<i>dominiqueur</i>	‘dominiqu-er’
<i>Full name</i>	Dominique Strauss-Kahn	<i>dominiquestrausskahnien</i>	‘dominiquestrausskahn-ian’
<i>Last name 1st part</i>	Strauss	<i>straussophile</i>	‘strauss-ophile’
<i>Last name 2nd part</i>	Kahn	<i>kahnisation</i>	‘kahn-ization’
<i>Acronym</i>	DSK	<i>dskie</i>	‘dsk-land’

Table 1: Sub-names from the PPN *Dominique Strauss-Kahn*

The program inputs and outputs are sequences of characters. These graphical forms encode morphophonological phenomena as well as purely orthographic variations. The program generates all possible tuples formed by a stem of sub-name and a suffix. For each tuple, the outputs of the program correspond to a set of possible derived words that we name Candidates. Each Candidate is obtained by concatenating (\oplus) the form of a sub-name Stem_i^n and a suffix Suff_j (7a) ($0 < j \leq 90$). For a given PPN_{*i*} ($0 < i \leq 89$), the symbol Stem_i^n corresponds to the stem of one of its sub-names *n* ($0 < n \leq 6$), or consists of a variation of this stem (7b) produced by one of the 36 \mathfrak{R} rules of the program.

- (7) a. $\text{Candidate} = \text{Stem}_i^n \oplus \text{Suff}_j$
 b. $\begin{cases} \text{Stem}_i^n = \text{Stem}_i^n \\ \text{Stem}_i^n = \mathfrak{R}(\text{Stem}_i^n) \end{cases}$

Each of them selects two arguments: Stem_i^n et Suff_j . Rules are organized in four blocks, cf. Figure 1.

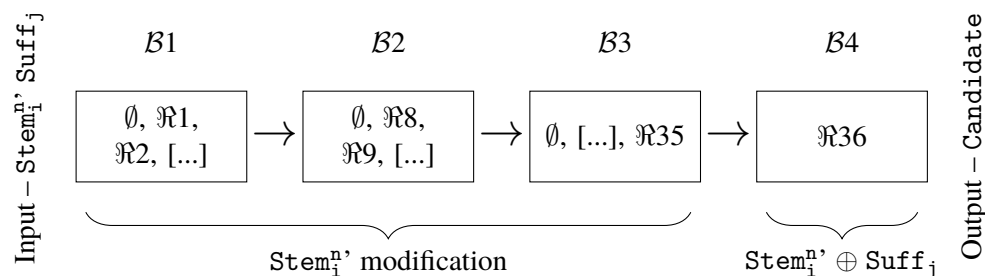


Figure 1: Rule combinations

The rules in the block $\mathcal{B}1$ remove a graphical sequence from Stem_i^n . The rules of $\mathcal{B}2$ add a graphical sequence to Stem_i^n . The rules of $\mathcal{B}3$ operate graphical substitutions. When relevant, the rules are the graphical transcriptions of morphophonological rules: truncation for $\mathcal{B}1$, epenthesis for $\mathcal{B}2$ and allomorphy for $\mathcal{B}3$. Finally, the $\mathfrak{R}36$ rule of $\mathcal{B}4$ concatenates the inputs Stem_i^n and Suff_j . The program

is oriented and acyclic. In each block, the rules are in complementary distribution. The values of Stem_i^a et Suff_j constrain which blocks and which rules can be activated. This organization leads to 136 possible rule combinations. Each stem/suffix input explores the 136 combinations, but a Candidate is only produced when the conditions of application of all the rules of the combination are met. Otherwise, the program tries the next combination.

Let us take the example of the input *taubira/ique*. If we follow the first possible combination, we apply the null rule (\emptyset) in each of the blocks \mathcal{B}_1 , \mathcal{B}_2 and \mathcal{B}_3 . Then rule \mathcal{R}_{36} of concatenation in \mathcal{B}_4 gives the Candidate (8). In contrast, the second combination using \mathcal{R}_1 in \mathcal{B}_1 will be discarded by the program. \mathcal{R}_1 corresponds to deleting the final *e* of a stem, hence it cannot be applied to *taubira*. The next rule in \mathcal{B}_1 , i.e. \mathcal{R}_2 , can be applied, since inputs *taubira/ique* satisfy the conditions for application of the \mathcal{R}_2 truncation rule: *taubira* ends with a vowel and *ique* begins with a vowel. \mathcal{R}_2 deletes the vowel *a* at the end of stem, to produce *taubir* (9a). Then, the null rules apply in \mathcal{B}_2 and \mathcal{B}_3 . The output of \mathcal{R}_2 is given as input to \mathcal{R}_{36} in \mathcal{B}_4 to generate the Candidate (9b). Testing all rule combinations exhaustively will eventually produce all other candidates: e.g., with epenthesis (10) and (11).

(8) $\mathcal{R}_{36}(\text{taubira}, \text{ique}) = \text{taubiraique}$

(9) a. $\mathcal{R}_2(\text{taubira}, \text{ique}) = (\text{taubir}, \text{ique})$

b. $\mathcal{R}_{36}(\text{taubir}, \text{ique}) = \text{taubirique}$

(10) $\mathcal{R}_8(\text{taubira}, \text{ique}) = (\text{taubirat}, \text{ique})$

$\mathcal{R}_{36}(\text{taubirat}, \text{ique}) = \text{taubiratique}$

(11) $\mathcal{R}_9(\text{taubira}, \text{ique}) = (\text{taubiras}, \text{ique})$

$\mathcal{R}_{36}(\text{taubiras}, \text{ique}) = \text{taubirasique}$

5 Data collection and annotation

The program produces 110,658 candidate forms, and each is used as a query, i.e. submitted to Bing. The set of attested deanthroponyms, their contexts, the URLs, and the number of pages associated with each query are saved in a tabulated file. A manual post-processing is then applied. It consists, for example, in deleting the entries of candidates homographs to attested lexemes with another meaning, e.g. *hollandais* ‘holland-ese’ is derived from *François Hollande* (12) but more often refers to the inhabitants of Holland.

(12) *Dans le cercle des hollandais, certains émettent l’hypothèse d’une absence du président sortant dans la course présidentielle. (François Hollande)*

‘In the circle of holland-ese, some speculate that the outgoing president will not be in the presidential race.’

The database we obtain contains 6,545 different deanthroponyms, for a total of 55,030 tokens. This corpus contains 3,830 complex words whose formation mode were expected as they were explicitly generated by the program. But Bing’s indexing process accidentally brought back a significant amount of unexpected forms: 40% of the deanthroponyms harvested are not part of our candidate list. For instance, we obtained occurrences of the prefixed noun *anti alliot-marisme*_{Nc} ‘anti-alliot-marism’ (13a) and of the compound adjective *chiraco-raffarinesque*_{Adj} ‘chiraco-raffarinian’ (13b), looking for attestations of the candidates *marisme* and *raffarinesque*.

(13) a. *Est-ce qu’une vague d’anti alliot-marisme peut déferler sur la circonscription [...] ? (Michèle Alliot-Marie)*

‘Can a wave of anti-alliot-marism sweep through the riding?’

b. *La majorité parlementaire chiraco-raffarinesque supprime une bonne partie des moyens financiers permettant à l’INRAP (institut national de recherches archéologiques préventives) d’effectuer des fouilles de sauvetage sur des sites archéologiques menacés par des programmes*

immobiliers. (Jacques Chirac, Jean-Pierre Raffarin)

‘The chiraco-raffarinian parliamentary majority suppresses a good part of the financial means allowing the INRAP (national institute of preventive archaeological researches) to carry out excavations of rescue on archaeological sites threatened by real estate programs.’

Each entry in the database describes an occurrence of one of the 6,545 deanthroponyms collected. This description is decomposed into a hundred or so features which describe, for each deanthroponym: its morphological properties such as its pattern (Xade for *peillonnade* in the Table 2), its category, and its morphophonological characteristics. Other properties do not result directly from the observation of the deanthroponyms but from our own analysis and are absent from Table 2. These include, for example, the semantic category of derivatives. The presentation of all the information contained in MoNoPOLI is beyond the scope of this presentation. However, the reader will be able to find an excerpt of the database and an explanation of each feature online.

PPN	Derivative in context	Pattern	POS	Phonology
<i>Vincent Peillon</i>	<i>une nouvelle peillonnade : la rentrée en août !</i> ‘a new peillon-ade: back to school in August!’	Xade	Nc	/pejɔnad/
<i>Rama Yade</i>	<i>crainte d’une ramayadisation</i> ‘fear of ramayad-ization’	Xisation	Nc	/ramajadizasjɔ/
<i>François Hollande</i>	<i>le chimpanzé à cul rose homo hollandus</i> ‘the pink ass chimpanzee homo holland-us’	homoXus	Nc	/homoolādys/

Table 2: Excerpt of MoNoPOLI

6 Analysis of deanthroponyms: some remarkable results

Unsurprisingly, the PPNs most often used as bases in MoNoPOLI (14a) correspond to the most prominent public figures. They have held a more important position (President vs Member of Parliament) or have been involved in high-profile events (laws, judicial scandals). Jacques Chirac (14b) was President of the French Republic. Dominique Strauss-Kahn (14c) was implicated in scandals (sexual assault and rape). The PPNs of these referents are typically used to create nonce-formations since the referents are subject to controversy. The 13 PPNs in (14a) are the bases of 50% of the deanthroponyms in our corpus.

- (14) a. *Dominique Strauss-Kahn, Marine Le Pen, Emmanuel Macron, Manuel Valls, Jean-Luc Mélenchon, François Mitterrand, Christiane Taubira, Ségolène Royal, François Bayrou, Lionel Jospin, François Hollande, Jacques Chirac, Nicolas Sarkozy*
 b. *On dit "arrête de chiraquer" pour dire arrête de faire des bêtises. (Jacques Chirac)*
 ‘We say "stop chiraqu-ing" to say stop doing stupid things.’
 c. *Enfumages sans feux : après l’éruption mentale de viol-kahnisme sulfureux présumé, retour au volcanisme réel. (Dominique Strauss-Kahn)*
 ‘Smoke and mirrors without fire: after the mental eruption of presumed sulphurous rape-kahnism, back to real volcanism.’

The nonce-formations are identifiable thanks to the meta-discursive signals and the discursive processes (cf. §2) but also sometimes thanks to the morphological processes used. 10% of the deanthroponyms of MoNoPOLI are produced by extragrammatical processes (Dressler and Kilani Schoch, 2005). We provide an overview of the morphological diversity of the content of MoNoPOLI in Table 3. The most frequent types of constructs are formed by suffixation and composition. Derivation by conversion is the least represented, which is certainly an effect of our methodology. Indeed, we only looked for forms corresponding to infinitive verbs.

The excerpt is available at this address: <https://perso.atilf.fr/mhuguin/accueil/these/corpus-monopoli/>.

Process	Frequency	Example	Gloss
Grammatical processes	90%	<i>hollandifier</i> _V	‘holland-ify’
Derivation	58%	<i>jospinerie</i> _{Nc}	‘jospin-ery’
Suffixation	51%	<i>chiraquiste</i> _{Nc}	‘chiraqu-ist’
Prefixation	5%	<i>ex-bovétiste</i> _{Adj}	‘ex-bovét-ist’
Conversion	2%	<i>bayrouer</i> _V	‘to-bayrou’
Compounding	32%	<i>lepénisme-mégretisme</i> _{Nc}	‘lepenism-megretism’
Exagrammatical processes	10%	<i>duflodocus</i> _{Nc}	‘duflo-docus’

Table 3: Overview of the processes of MoNoPOLI

In the following, we show that PPNs are privileged bases for extragrammatical derivation on several levels. On the one hand, they serve as a bases for processes classified as extragrammatical in French such as reduplication, cf. §6.1. On the other hand, they also sometimes lead to the subversion of more classical i.e., more regular processes. We illustrate our remarks by studying the case of compounding in §6.2.

6.1 Extragrammatical process

MoNoPOLI contains 23 deanthroponyms that instantiate the pattern XoX_{suff} as complex words of (15) when X is the stem of a sub-name of PPN. In (15a) *suff* is *-iste* (‘-ist’), in (15b) it is *-ien* (‘-ian’).

- (15) a. *Il s’explique dans un entretien à paraître ce mardi dans les éditions mayennaises d’Ouest-France : sur le plan départemental, c’est la motion la droite forte qui est arrivée largement en tête. Les militants ont choisi la ligne dure, sarkozo-sarkozyste.* (Nicolas Sarkozy)
‘He explains himself in an interview to appear this Tuesday in the Mayenne editions of Ouest-France: on the departmental level, it is the motion of the hard right which arrived largely in head. The militants chose the hard line, sarkozo-sarkozist.’
- b. *En témoigne le remaniement ministériel, qui fait la part plus que belle aux chiraco-chiraquiens : il s’apparente à la formation de la tortue, notent les bons observateurs de la vie politique.* (Jacques Chirac)
‘This is evidenced by the ministerial reshuffle, which gives the lion’s share to the chiraco-chiraquiens: it is similar to the formation of the turtle, remark astute observers of the political sphere.’

At first sight, the process used to obtain these derivatives is compounding. One could see in *sarkozo-sarkozyste*_{Adj} and *chiraco-chiraquien*_{Nc} a particular case of the compound XoY_{suff} . For example, the demonym *franco-canadien*_{Adj/Nc} ‘French and Canadian’ is compounded from the bases X *français*_{Adj/Nc}, which appears truncated, et Y *canadien*_{Adj/Nc} (Dal and Amiot, 2008). As in *franco-canadien*_{Adj/Nc}, we find in *sarkozo-sarkyste* the intercalary vowel /o/ typically associated with compounds, and more specifically *learned* compounds, i.e. including the stem of a lexeme inherited from Greek or Latin. In *franco-canadien*_{Adj/Nc}, the vowel can be thought of as subverted, in that it does not mark the learned character of a stem but constitutes an iconic marker of compounding (Dal and Amiot, 2008). Contrary to the compounds which are constituted of two distinct bases, in the deanthroponyms of (15), it is always the same (truncated) stem of X which is used in each element. This observation constitutes a first obstacle, of a formal nature, to the analysis of *sarkozo-sarkozyste*_{Adj} and *chiraco-chiraquien*_{Nc} as compounds. The second obstacle is semantic. The meanings of deanthroponyms of (15) do not correspond to coordination, subordination or apposition if we follow the tripartite classification of compounds of Scalise et al. (2005), for example. Semantically, *sarkozo-sarkozyste* (15a) means ‘very/typically/exclusively sarkozist’ and *chiraco-chiraquiens* (15b) are ‘very/typically/exclusively chiraquian people’. This meaning consists of an amplification or an exaggeration of the property denoted by the base, and this semantic value is

attested in cases of reduplication. The deanthronyms of (15) are therefore not compounds but exhibit characteristics of reduplication.

This intensifying and restrictive polarization of reduplication is notably attested in syntax. In some syntactic reduplications, called *Identical Constituent Compounding* (Hohenhaus, 2004) or *Contrastive Reduplication* (Ghomeshi et al., 2004), the reduplicant allows a meta-discursive comment on the other term. In (16) the reduplication *mad mad* means ‘very/completely mad’. When these are reduplicated nouns the associated paraphrase may be ‘exclusively NOUN’: e.g. *I want salad salad* means ‘only/exclusively salad’ (not *tuna salad* or *compound salad*). These paraphrases have semantic values equivalent to those we have determined for the complex words of (15). In syntax, as in morphology, it is a matter of either intensifying a property or restricting the designated referential class to referents which possess prototypical properties of the class (Kleiber, 1990). Intensification and restriction are well known semantic values for reduplication (Moravcsik, 1978) and attested in many languages (e.g. English, Italian, French, Turkish). In French, reduplication is used exclusively for evaluative purposes. We therefore analyze the deanthronyms *XoXsuff* as resulting from reduplication; more specifically, this is partial pre-reduplication since the reduplicant *Xo* is on the left and does not use the entire phonological material of the base. Finally, we should add that the identified process is not specific to anthroponymic bases since it also applies to ethnic adjectives such as *français*_{Adj} which gives *franco-français*_{Adj} ‘very/typically/exclusively French’.

(16) She’s mad [. . .] Not **mad mad**, but, you know. Out of control. (Hohenhaus, 2007, p. 26)

6.2 Subverted grammatical process

MoNoPoLI contains 1,925 deanthronyms instantiating the general pattern *Xo-(X’(o)-)*Ysuff* where the brackets indicate the optionality of an element and the asterisk notes that the number of components at that position is 0, 1 or more, cf. (17).

- (17) a. *Ainsi, selon une méthode éprouvée, le « camp du bien », pensant pouvoir l’achever, se livre en vain à une exégèse sémantique de sa critique du totem **aubryo-strausskhanien**.* (Martine Aubry, Dominique Strauss-Kahn)
 ‘Thus, according to a tried and fruitlessly tested method, the "camp of the good", thinking to be able to finish it, engages in vain in a semantic exegesis of its criticism of the aubryo-strausskhanian totem.’
- b. *Voilà le fruit de quinze années de **pasquaïo-sarkozo-bessonisme**.* (Charles Pasqua, Nicolas Sarkozy, Éric Besson)
 ‘Here is the fruit of fifteen years of pasquaïo-sarkozo-bessonism.’
- c. *C’est triste que le seul pendant à votre soi-disant pensée unique **bobo-marxo-stalino-taubiro-hollando-demissiono-comploto-lgbt-communiste**, soit juste un propos « anti-système » d’extrême droite.* (Christiane Taubira, François Hollande)
 ‘It’s sad that the only counterpart to your so-called boho-marxo-stalino-taubiro-hollando-resigno-conspiratoro-lgbt-communist unique thought, is just an extreme right-wing "anti-establishment" statement.’

- The compound *aubryo-strausskhanien*_{Adj} from (17a) has the minimal format of the pattern: *Xo-Ysuff*. It includes the adjective *strausskahnien*_{Adj} (‘strausskahn-ian’) and the sub-names *Aubry* suffixed by /o/. It is the same /o/ that we have already observed in §6.1. It is the typical vowel of learned compounding, subverted from its usual function since, here again, the stem is not inherited. This compound adjective is interpreted as a coordination: ‘aubry-ist and strausskahn-ian’. We can see that the first component is a suffixed truncated adjectival form. The meaning of the compound guides our analysis: since the compound is coordinative and the coordinated elements are, by definition, of the same morpho-semantic type, we conclude that if *Ysuff* is a denominative adjective, X is a denominative adjective like *Ysuff*. So X is probably the truncated form of the relational adjective *aubryiste*_{Adj} (‘aubry-ist’) of the sub-name *Aubry* (*aubryien*_{Adj} ‘aubry-ian’ is attested with a lower frequency).

- In (17b), we see that the minimal pattern can be extended to *Xo-X' o-Ysuff*. We can thus accumulate constituents in /o/. The compound is coordinative as in (17a). So, *Xo* constituents are truncated forms of deanthroponyms of the same nature as *Ysuff* *bessonisme*_{Nc} ('besson-ism') which refers to the ideology of Eric Besson. They are truncated forms of the common nouns *pasquaïsme*_{Nc} ('pasqua-ism') and *sarkozysme*_{Nc} ('sarkoz-ism').
- The examination of (17c), finally shows that what counts for the writer is above all the rhyme in /o/, since *bobo* 'boho' is not suffixed. Moreover, in *bobo-marxo-stalino-taubiro-hollando-demissiono-comploto-lgbt-communiste*_{Adj}, the accumulation of components to the left of the final suffixed component *Ysuff* (*communiste*_{Adj} 'communist') is not limited to a suffixed form in /o/. Indeed, one of the constituents is the acronym LGBT. In any case, all these forms always refer to adjectival properties, as does the last constituent *communiste*_{Adj}. The writer's goal is to accumulate as many constituents as possible, like a outbursts, to distinguish himself (cf. §2).

These compounds are exclusively coordinative. Moreover, the more constituents the writer adds, the more the effect of meaning obtained is that of a cacophony. The longer the deanthroponym, the more original and remarkable it is. In conclusion, even if the compounding process is grammatical, compound deanthroponyms are not always grammatical (especially when they involve more than two bases). In our corpus, compounding is sometimes subverted in the benefit of the writer's argumentation or humor.

7 Conclusion

The method used to create the MoNoPOLI database is reproducible and adaptable to other languages or other inputs (bases or affixes). The database created will eventually be accessible online. It provides a large corpus of contextualized deanthroponyms, which to our knowledge does not exist in French. Moreover, MoNoPOLI presents a real morphological diversity, i.e. a large panel of different processes. It is a possible resource for further research on both words based on anthroponyms and French nonce-formations. Analysing it reveals that deanthroponyms are often nonce-formations constructed by extragrammatical processes. We have also shown that grammatical processes can be subverted to satisfy the enunciative needs of the writer. The latter demonstrates, at the same time, their epilinguistic capacity to play with language.

References

- Laurie Bauer. 1983. *English Word-Formation*. Cambridge Textbooks in Linguistics. Cambridge University Press. <https://doi.org/10.1017/CBO9781139165846>.
- Georgette Dal and Dany Amiot. 2008. La composition néoclassique en français et l'ordre des constituants. In Dany Amiot, editor, *La composition dans une perspective typologique*, Artois Presses Université, Arras, pages 89–113.
- Georgette Dal and Fiammetta Namer. 2015. *Internet*. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen, and Franz Rainer, editors, *Word-Formation: An International Handbook of the Languages of Europe*, De Gruyter Mouton, Berlin, New York, volume 3, pages 2372–2386. <https://doi.org/10.1515/9783110375732-044>.
- Georgette Dal and Fiammetta Namer. 2018. *Playful nonce-formations in French: Creativity and productivity*. In Sabine Arndt-Lappe, Angelika Braun, Claudine Moulin, and Esme Winter-Froemel, editors, *Expanding the Lexicon Linguistic Innovation, Morphological Productivity, and Ludicity*, De Gruyter, Berlin, Boston, 5, pages 203–228. <https://doi.org/10.1515/9783110501933-205>.
- Wolfgang U. Dressler and M. Kilani Schoch. 2005. *Morphologie naturelle et flexion du verbe français*. Gunter Narr Verlag, Tübingen.
- Bernard Fradin, Georgette Dal, Natalia Grabar, Stéphanie Lignon, Fiammetta Namer, Delphine Tribout, and Pierre Zweigenbaum. 2008. *Remarques sur l'usage des corpus en morphologie*. *Langages* 171(3):34–59. <https://doi.org/10.3917/lang.171.0034>.
- Jila Ghomeshi, Ray Jackendoff, Nicole Rosen, and Kevin Russell. 2004. Contrastive Focus Reduplication in English (The Salad-Salad Paper). *Natural Language & Linguistic Theory* 22(2):307–357.

- John Goldsmith. 1976. *Autosegmental phonology*. Ph. Dissertation, MIT.
- Christophe Gérard. 2018. Variabilité du langage et productivité lexicale: Problèmes et propositions méthodologiques. *Neologica* 12:23–45.
- Peter Hohenhaus. 2004. Identical Constituent Compounding – a Corpus-based Study. *Folia Linguistica* 38(3-4):297–332. <https://doi.org/10.1515/flin.2004.38.3-4.297>.
- Peter Hohenhaus. 2005. Lexicalization and institutionalization. In Pavol Štekauer and Rochelle Lieber, editors, *Handbook of Word-Formation*, Springer, Dordrecht, pages 353–370. https://doi.org/10.1007/1-4020-3596-9_15.
- Peter Hohenhaus. 2007. How to do (even more) things with nonce words (other than naming). In Judith Munat, editor, *Lexical Creativity, Texts and Contexts*, John Benjamins Publishing Company, Amsterdam, Philadelphia, Studies in Functional and Structural Linguistics 58, pages 15–38.
- Peter Hohenhaus. 2015. Anti-naming through non-word-formation. *SKASE Journal of Theoretical Linguistics* 12(3):272–291.
- Mathilde Huguin. 2018. Anthroponyms and paradigmatic derivation in French. *Lingue e Linguaggio* XVII(2):217–232. <https://doi.org/10.1418/91866>.
- Daphné Kerremans. 2015. *A Web of New Words: A Corpus-Based Study of the Conventionalization Process of English Neologisms*. Peter Lang GmbH, Internationaler Verlag der Wissenschaften. <https://doi.org/10.1515/ang-2016-0076>.
- Georges Kleiber. 1990. *La Sémantique du prototype. Catégories et sens lexical*. PUF, Paris.
- Anke Lüdeling, Stefan Evert, and Marco Baroni. 2007. Using web data for linguistic purposes. In Marianne Hundt, Nadja Nesselhauf, and Carolin Biewer, editors, *Corpus Linguistics and the Web*, Rodopi, Amsterdam, New York, 59, pages 7–24. https://doi.org/10.1163/9789401203791_003.
- Sarah Leroy. 2008. Les noms propres et la dérivation suffixale. *Neuphilologische Mitteilungen* 109:55–71.
- Stéphanie Lignon, Fiammetta Namer, Nabil Hathout, and Mathilde Huguin. 2019. When sarkozysation leads to the hollandade, or the rejection of phonological well-formedness constraints by anthroponym-based derived words. In *International Symposium of Morphology (ISM) 2019*. Paris, France.
- Edith Moravcsik. 1978. Reduplicative Constructions. In Joseph H. Greenberg, editor, *Universals of Human Language*, Stanford University Press, Standford, number 3 in Word Structure, pages 297–334.
- Marc Plénat. 2009. Les contraintes de taille. In Marc Plénat, Bernard Fradin, and Françoise Kerleroux, editors, *Aperçus de morphologie du français*, Presses Universitaires de Vincennes, Saint-Denis, pages 47–64.
- Alan Prince and Paul Smolensky. 1993. Optimality theory: constraint interaction in generative grammar. *Technical Report, Rutgers University Center for Cognitive Science and Computer Science Department*.
- Michel Roché. 2010. Base, thème, radical. *Recherches linguistiques de Vincennes* 39:95–134. <https://doi.org/10.4000/rlv.1850>.
- Michel Roché and Marc Plénat. 2014. Le jeu des contraintes dans la sélection du thème présuffixal. In *SHS Web Conferences* 8. pages 1863–1878. <https://doi.org/10.1051/shsconf/20140801143>.
- Sergio Scalise, Antonietta Bisetto, and Emiliano Guevara. 2005. Selection in Compounding and Derivation. In Wolfgang Ulrich Dressler, Dieter Kastovsky, Oskar E. Pfeiffer, and Franz Rainer, editors, *Morphology and its demarcations: Selected papers from the 11th Morphology meeting, Vienna, February 2004*, John Benjamins Publishing Company, Amsterdam, Philadelphia, pages 133–150. <https://doi.org/https://doi.org/10.1075/cilt.264.09sca>.
- Barbara Schlücker and Tanja Ackermann. 2017. The morphosyntax of proper names: An overview. *Folia Linguistica* 51(2):309–339. <https://doi.org/10.1515/flin-2017-0011>.
- Wolfgang Schweickard. 1992. *Deonomastik. Ableitungen auf der Basis von Eigennamen im Französischen*. De Gruyter, Tübingen. <https://doi.org/10.1515/9783110933901>.
- Ludovic Tanguy. 2012. *Complexification des données et des techniques en linguistique : contribution du TAL aux solutions et aux problèmes*. Habilitation à diriger des recherches, Université Toulouse-Le Mirail, Toulouse.
- Pavol Štekauer. 2002. On the Theory of Neologisms and Nonce-formations. *Australian Journal of Linguistics* 22(1):97–112. <https://doi.org/10.1080/07268600120122571>.