Abstract

We examine a case of base variation related to property nouns formation: namely, -ité suffixed French nouns expressing the character proper both to those who belong/are related to a place (town, country...) and/or to the place itself (henceforth Ethnic Property Nouns: EPNs). The study is based upon an important web-extracted corpus and shows that, at large scale, speakers coin EPNs either from toponyms (PORTUGAL > PORTUGALITÉEPN ‘portugal-ness’ =
‘portugueseness’), from related ethnic adjectives (AFRIQUE ‘Africa’ > AFRICAIN ‘African’ > AFRICANITÉEPN ‘africanness’) or from both (BELGIQUE ‘Belgium’ > BELGICITÉEPN ‘Belgium-ness’; BELGE ‘Belgian’ > BELGITÉEPN ‘Belgianness’). Several examples testify that these base variations are unrelated to meaning but rather correlated with four formal competing constraints: among them, what we call ‘lexical pressure’ can explain the form of the output. A survey experiment is then described, which corroborates our analysis. Finally, the scope of our conclusions goes beyond French EPNs, as they apply to other word formation rules, in many languages.

1. Introduction

Following the research initiated in Dal & Namer (2005), this paper deals with -ité suffixed French nouns expressing the property both of those who belong or are related to a place (town, region, country, continent), and/or of the place itself. We henceforth call these nouns “ethnic property nouns” (EPNs). Though this study has been performed on French data, the results obtained are transposable to other languages – at least, some Romance languages. The issue we address here follows from two observations. First of all, there are two ways to form a French EPN: either from an ethnic adjective base, or from a toponym, even if they lead to a single semantic output (§ 2). Second (§
3), this variation is recurrently observed. To back this observation up, we use a massive set of data mainly collected from the Internet. Faced with this data, our hypothesis (§ 4) is twofold: first, this variation is a matter of competition between constraints on the output form; second, it is possible to rank these constraints in order to predict what new EPNs should look like. A survey experiment is also reported on, that confirms our assumptions. To conclude (§ 5), we draw theoretical consequences from the observed phenomena and their analysis.

2. Issue

Examples (1a) to (1e) provide some contexts in which the data we are interested in occur. They were collected from the Web in July 2007.

(1) a. L’hystérie de la Belgité : l’hystérie dans la littérature belge de langue française.

Belgianness hysteria: histeria in French language Belgian literature.

b. Le retour de l’Alsace au Reich en 1870 renforça la germanicité des communautés rurales.
As Alsace went back to the Reich in 1870, this reinforced germandness among rural communities.

c. On a reproché à Balzac de s’être trompé sur le sens de l’italianité... Peut-être faudrait-il distinguer entre italianité et rêverie italienne (…) 

*Balzac has been criticized for getting the wrong meaning of italianness... maybe italianness should be distinguished from Italian daydreaming (…)*

d. La francité, c’est d’abord l’esprit français, tel qu’il apparaît encore dans la langue française.

*France-ness [frenchness] is first of all the French spirit, as it still appears in the French language.*

e. En pleine période de trouble au Liban, la banque Byblos qui fait de la “Libanité” le cœur de ses valeurs (…) 

*In the middle of a troubled period in Lebanon, Byblos bank, which makes “Lebanon-ness” [lebaneseness] the heart of its own values, (…)*
f. (...) un projet de recherche dans le but de comprendre ladite
terme particulière de vivre cette "portugalité" silencieuse dans
l’espace familial ou associatif (…)

(...) a research project aiming to understand this particular way
of living this silent "Portugal-ness" [portugueseness] in a family
or community environment, (…)

These examples illustrate the fact that there are two possibilities for a speaker
to coin a new EPN: from a simple adjective (1a: belge ‘belgian’) or a
complex one (1b: GERMANIQUE ‘german’, 1c: ITALIEN: ‘italian’), or directly
from the toponym (1d: FRANCE, 1e: LIBAN, 1f: PORTUGAL). We will examine
these two ways successively.

2.1 Adjective-based -ité EPNs

Examples as (1a-c) are instances of the general French -ité Word Formation
Rule. According to this rule, the input usually is a predicative adjective, and
the output is the corresponding property noun. Table (1) provides some
examples of lexemes formed according to this general rule.

<table>
<thead>
<tr>
<th>Input: predicative adjective</th>
<th>Output: corresponding property noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>BELGE ‘belgian’</td>
<td>BELGÉTÉ</td>
</tr>
<tr>
<td>GERMANIQUE ‘german’</td>
<td>GERMANIQUE-ITÉ</td>
</tr>
<tr>
<td>ITALIEN: ‘italian’</td>
<td>ITALIEN-ITÉ</td>
</tr>
<tr>
<td>FRANCE</td>
<td>FRANÇAÎTÉ</td>
</tr>
<tr>
<td>LIBAN</td>
<td>LIBAN-ITÉ</td>
</tr>
<tr>
<td>PORTUGAL</td>
<td>PORTUGAÎTÉ</td>
</tr>
</tbody>
</table>
Table 1: applying French -ité Word Formation Rule

<table>
<thead>
<tr>
<th>Word</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>BANAL</td>
<td>'banal'</td>
</tr>
<tr>
<td>BANALITÉ</td>
<td>'banality'</td>
</tr>
<tr>
<td>BRUTAL</td>
<td>'brutal'</td>
</tr>
<tr>
<td>BRUTALITÉ</td>
<td>'brutality'</td>
</tr>
<tr>
<td>TRANSITIF</td>
<td>'transitive'</td>
</tr>
<tr>
<td>TRANSITIVITÉ</td>
<td>'transitivity'</td>
</tr>
</tbody>
</table>

Table 2 presents additional EPNs which can be regarded as instances of this WFR. Input adjectives are all ethnic adjectives. They are often formed on toponyms (e.g. in (a) AFRICAIN<sub>ADJ</sub> < AFRIQUE<sub>P RN</sub>), though this is not always the case (e.g. in (d) CATALAN<sub>ADJ</sub> is semantically but not morphologically related to CATALOGNE<sub>P RN</sub> ‘Catalonia’; so is MAGYAR<sub>ADJ</sub> to HONGRIE<sub>P RN</sub> ‘Hungary’ in (f)). We will briefly come back to this in sections 4.2.1 and 4.2.3.

@ Insert Table 2 here

<table>
<thead>
<tr>
<th>Input: ethnic adjective</th>
<th>Output: EPN</th>
</tr>
</thead>
<tbody>
<tr>
<td>a  AFRICAIN ‘african’</td>
<td>AFRICANITÉ ‘africanness’</td>
</tr>
<tr>
<td>b  AMÉRICAIN ‘american’</td>
<td>AMÉRICANITÉ ‘americanness’</td>
</tr>
<tr>
<td>c  ASIATIQUE ‘asian’</td>
<td>ASIATICITÉ ‘asianness’</td>
</tr>
<tr>
<td>d  CATALAN ‘catalan’</td>
<td>CATALANITÉ ‘catalanness’</td>
</tr>
<tr>
<td>e  IVOIRIEN ‘ivorian’</td>
<td>IVOIRIANITÉ ‘ivorianness’</td>
</tr>
<tr>
<td>f  MAGYAR ‘hungarian’</td>
<td>MAGYARITÉ ‘hungarianness’</td>
</tr>
<tr>
<td>g  MALIEN ‘malian’</td>
<td>MALIANITÉ ‘malianness’</td>
</tr>
<tr>
<td>h  SYRIEN ‘syrian’</td>
<td>SYRIANITÉ ‘syrianness’</td>
</tr>
</tbody>
</table>

Table 2: examples of adjective-based EPNs
2.2 EPNs directly formed on toponyms

There is however a second way to coin EPNs. Actually, the suffix -ité can be directly applied to the toponym: this is what happens with FRANCITÉ, LIBANITÉ or PORTUGALITÉ in (1d-f). Other examples are given in Table 3.

@ Insert Table 3 here

<table>
<thead>
<tr>
<th>Input: toponym</th>
<th>Output: EPN</th>
</tr>
</thead>
<tbody>
<tr>
<td>a AMÉRIQUE 'America'</td>
<td>AMÉRICITÉ 'America-ness=americanness'</td>
</tr>
<tr>
<td>b BELGIQUE 'Belgium'</td>
<td>BELGICITÉ 'Belgium-ness=belgianness'</td>
</tr>
<tr>
<td>c VIETNAM 'Vietnam'</td>
<td>VIETNAMITÉ 'Vietnam-ness=vietnameseness'</td>
</tr>
</tbody>
</table>

Table 3: examples of toponym-based EPNs

For FRANCITÉ, LIBANITÉ in (1d-e) and, in Table 3, AMÉRICITÉ and VIETNAMITÉ, this claim is possibly disputable. In a theoretical perspective that admits truncation in Word Formation (Aronoff 1976; Corbin 1987), it could be argued that these EPNs are regularly built on adjectives, with the truncation of the adjectival suffix –ais (for FRANÇAIS ‘French’ or LIBANAIS ‘lebanese’), that of -ain (for AMÉRICAIN) or -ien (for VIETNAMIEN ‘vietnamese’). According to this analysis, FRANCITÉ (/frɑ̃siti/) for instance, would result from the
application of the -ité suffixation rule on the ethnic adjective FRANÇAIS (/fʁɑ̃s/), assuming -ais (/ɛ/) truncation.

However, this hypothesis can no longer be retained for examples such as PORTUGALITÉ (/pɔʁtygalit/) in (1f) or BELGICITÉ in Table 3. Each of these examples contains a phonic sequence which is absent from the ethnic adjective (respectively, BELGE - /bɛlʒ/ and PORTUGAIS - /pɔʁtyʒ/ ‘portuguese’), but is present in the toponym². The presence of this sequence indicates that the only possible base is the corresponding toponym.

As we can see in Table 4 with such examples as AMÉRICITÉ/AMÉRICANITÉ, BELGITÉ/BELGICITÉ, it can be the case that both constructions be attested on the Web for what we consider to be a single semantic output (the issue will be debated below, in section 4.1). But, in this case, both adjective-based and toponym-based EPNs are found, even though they often occur with very different frequencies. Table 4 provides such examples and indicates variable ratios among -ité noun pairs. Figures sometimes favour deadjectival formations (e.g. line a, AMÉRICANITÉ is almost 2000 times more frequent than AMÉRICITÉ). Sometimes, on the opposite, they give advantage to denominal ones (compare, for instance, line e, figures for LIBANITÉ and LIBANAISITÉ).

@@ Insert Table 4 here
<table>
<thead>
<tr>
<th>Toponym/ Ethnic Adj</th>
<th>Adjective-based EPN: # occ</th>
<th>Toponym-based EPN: # occ</th>
<th>Ratio: # Adj-based/# Toponym-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>a AMÉRIQUE/AMÉRICAIN</td>
<td>AMÉRICANITÉ: 38,700</td>
<td>AMÉRICITÉ: 21</td>
<td>1842,85</td>
</tr>
<tr>
<td>b BELGIQUE/BELGE</td>
<td>BELGITÉ: 142</td>
<td>BELGICITÉ: 31</td>
<td>4,58</td>
</tr>
<tr>
<td>c CAMEROUN ('Cameroon')/CAMEROUNAIS ('Cameroonian')</td>
<td>CAMEROUNAISITÉ: 2</td>
<td>CAMEROUNITÉ: 698</td>
<td>0,003</td>
</tr>
<tr>
<td>d IRAK/IRAKIEN ('Iraki')</td>
<td>IRAKIANITÉ: 3</td>
<td>IRAKITÉ: 116</td>
<td>0,036</td>
</tr>
<tr>
<td>e LIBAN/LIBANAIS</td>
<td>LIBANAISETÉ: 1</td>
<td>LIBANITÉ: 1230</td>
<td>0,0008</td>
</tr>
<tr>
<td>f PORTUGAL/PORTUGAIS</td>
<td>PORTUGAISITÉ: 2</td>
<td>PORTUGALITÉ: 170</td>
<td>0,012</td>
</tr>
<tr>
<td>g SÉNÉGAL/SÉNÉGALAIS ('Senegalese')</td>
<td>SÉNÉGALAISITÉ: 3</td>
<td>SÉNÉGALITÉ: 271</td>
<td>0,011</td>
</tr>
</tbody>
</table>

Table 4: Number of occurrences for adjective-based and toponym-based EPNs

(Yahoo™, July 18th 2007)

Confronted with these data, our attempt has thus been to determine whether the choice between these two forms (toponym-based or ethnic adjective-
based) is randomly made, or whether it is possible to predict what newly coined EPNs would look like.

3. Collected data: preliminary observation

To achieve this task, a data collection was performed. As detailed in § 3.1, it has provided us with more than 200 EPNs found on the Internet, with their appearance context and their number of occurrences. Investigation conducted on this data is described in § 3.2 and § 3.3.

3.1 Methodology

Collecting EPNs from the Internet required the following steps to be performed. First, a list of 145 toponyms was set up from the PACTOLS thesaurus (Lequeux 2005), referring to very well known countries, regions or towns. Then, each member of this list was mapped onto its morphologically and/or, if any, semantically corresponding ethnic adjective(s). For instance, ITALIE was linked to ITALIEN; HONGRIE, on the other side, was related to two ethnic adjectives: HONGROIS (‘Hungarian’) and MAGYAR. All in all a list of 411 candidate bases for EPNs was compiled, and each of them served to
automatically generate its -ité ending EPN counterpart. In other words, a program automatically built a list of -ité ending potential forms based either on toponyms (for instance the candidate form hongrité was generated from HONGRIE, and italité was coined after ITALIE) or on ethnic adjectives (for example, HONGROIS, MAGYAR and ITALIEN respectively gave rise to the potential forms hongroisité, magyarité and italienité). Of course, during this generation task, the question whether the resulting forms were attested or not in dictionaries was irrelevant: the aim was to obtain a list of 411 morphologically well-formed potential -ité ending ethnic property nouns. These generated forms were used as Yahoo queries by means of the WaliM robot (Namer 2003), so that only those EPNs actually found on Yahoo indexed Web documents were kept.

At this point, these raw results were cleaned up, in order to end up with 203 out of our 411 originally generated EPNs: elements from wordlists, misspelled words, jokes, and so on have been manually discarded.

3.2 Three EPN base types

As indicated in Table 5, these validated EPNs happen to be equally divided into toponym-based nouns and adjective-based ones (as each type has 75
tokens). For 17 other EPNs, the base is a suppletive form. Among them, /sin/ in SINITÉ corresponds to either CHINEPRE (‘China’) or CHINOISADJ (‘Chinese’), NIPPON (‘Japanese’) (in NIPPONITÉ) acts as JAPONAISADJ (‘Japanese’), and MAGYAR (in MAGYARITÉ) substitutes for HONGрослADJ. For the remaining 36 property nouns, it is impossible to decide whether they are toponym- or adjective-based, inasmuch as affixing -ité produces the same output form. For instance, BULGARITÉ can be formed from either the adjectival base BULGARE (‘Bulgarian’), or from the toponym BULGARIE (‘Bulgaria’), with a haplology between both /i/ (/bul.ga.ри.ite/).

@@ Insert Table 5 here

<table>
<thead>
<tr>
<th>Base type</th>
<th>Number of different EPNs</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toponym</td>
<td>75</td>
<td>AMÉRICITÉ, BELGICITÉ, BRETAGNITÉ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(‘Brittany-ness’), GALICITÉ (‘Galicia-ness’), GUATÉMALITÉ (‘Guatemala-ness’), PORTUGALITÉ</td>
</tr>
<tr>
<td>Ethnic Adjective</td>
<td>75</td>
<td>AUSTRALIANITÉ (‘australianness’), AUVERGNATITÉ (‘from-the-Auvergne-ness’), BRETONNITÉ (‘bretonness’), HISPANICITÉ (‘hispanicness’), MALIANITÉ (‘malianness’), SÉNÉGALAITÉ (‘senegaleseness’)</td>
</tr>
<tr>
<td>Suppletive base</td>
<td>17</td>
<td>MAGYARITÉ, NIPPONITÉ, SINITÉ</td>
</tr>
<tr>
<td>Undecidable</td>
<td>36</td>
<td>BULGARITÉ, PICARDITÉ (‘Picardy-’)</td>
</tr>
</tbody>
</table>
3.3 Heterogeneous quantitative results

We can make two observations from the figures given in Tables 4 and 5:

(1) only half of the automatically generated EPNs are actually found on Web documents. Indeed, we would expect nouns such as AUTRICHITÉ (‘Austrian-ness’), AUTRICHIANITÉ (‘austrianness’), or, better, VIENNITÉ (‘Vienna-ness’), VIENNOISITÉ (‘vienneseness’), but none of them is present on the Web, no more than AUSTRITÉ (despite the existence of AUSTRO-HONGROIS ‘austro-hungarian’). Though this is not the subject of this paper, it would be interesting to investigate the reasons for these gaps, when they are not caused by prosody and morphophonology (§ 4 is devoted to these topics).

Another illustration of lexical gaps is provided below by examples in Table 6. ALGÉRIANITÉ and LIBÉRIANITÉ, in (a1) and (a2), share the same number of syllables and their respective bases ALGÉRIEN (‘algerian’) and LIBÉRIEN (‘liberian’) both end with the same final sequence /ʁjɛ̃/; so, from both prosodic and morpho-phonological points of view, their probability to be created is the same. Yet, this is not the case. The observation of EPNs in (b1) and (b2) offers the same contrast, between their significant difference in
occurrences and the prosodic and rhymic similitudes of their bases (/iɛ.l̃d/ versus /is.l̃d/). The most likely reason that can be assigned to these discrepancies – and that requires further investigation – is bound to extralinguistic grounds: that is the need (or, on the contrary, the lack of need) for identity or nation assertion.

@@ Insert Table 6 here

<table>
<thead>
<tr>
<th>Toponym or ethnic Adjective</th>
<th>EPN</th>
<th>Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>a₁</td>
<td>ALGÉRIEN al.ʒe.ʁjɛ</td>
<td>ALGÉRIANITÉ alʒeʁjanite</td>
</tr>
<tr>
<td>a₂</td>
<td>LIBÉRIEN li.be.ʒe̝</td>
<td>LIBÉRIANITÉ libeʒjanite</td>
</tr>
<tr>
<td>b₁</td>
<td>IRLANDE iʁ.l̃d</td>
<td>IRLANDITÉ iʁl̃dite</td>
</tr>
<tr>
<td>b₂</td>
<td>ISLANDE iʁ.l̃d</td>
<td>ISLANDITÉ iʃl̃dite</td>
</tr>
</tbody>
</table>

Table 6: Unexpected non-occurring EPNs

(2) The second observation has to do with frequency variability for EPN occurrences: as indicated in Table 5, frequencies for the 203 nouns collected on the Web vary a lot. Actually, they range on a scale from 1 to 27,100. Table
7, in which noun sets are ordered according to increasing frequency, shows that the largest noun set (almost half of our corpus) has unfrequent, if not rare, occurrences (less than 10 indexed pages). For instance ALSACITÉ occurs 4 times, ANTILLITÉ only once. On the other hand, more than half of the nouns have occurrences ranging from 10 to 1000. Surprisingly, only 3 of the 15 most frequent nouns (more than 1000 occurrences) are stored in the biggest multi-volume French dictionary of general language: namely the *Trésor de la langue française* (*TLF*). These nouns are GERMANITÉ, FRANCITÉ, and ITALIANITÉ⁴.

@@ Insert Table 7 here

<table>
<thead>
<tr>
<th>Occurrences interval</th>
<th>Number of different EPNs</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-9</td>
<td>91</td>
<td>ALSACITÉ ('Alsace-ness'), ANTILLITÉ ('West-Indies-ness'), AUVERGNATITÉ, BELGICITÉ, PORTUGAISITÉ</td>
</tr>
<tr>
<td>10-99</td>
<td>55</td>
<td>ASIATICITÉ ('asianness'), AUSTRALIANITÉ, BIRMANITÉ ('Burma-ness' or 'burmeseness'), BURUNDITÉ</td>
</tr>
<tr>
<td>100-999</td>
<td>40</td>
<td>BELGITÉ, BRETONNITÉ, CAMEROUNITÉ, IRAKITÉ, PORTUGALITÉ</td>
</tr>
<tr>
<td>1000-10,000</td>
<td>7</td>
<td>AFGHANITÉ, ALGÉRIANITÉ, AMÉRICANITÉ, ARMÉNITÉ ('Armenia-ness'), BOLIVIANITÉ ('bolivianness'), CONGOLITÉ ('Congo-ness'), ITALIANITÉ</td>
</tr>
</tbody>
</table>
In conclusion, when creating EPNs, speakers actually do make a decision as far as base category is concerned: this is what variation in figures illustrates, as shown in Tables 5 and 7. The question arises whether this choice is a free decision, or whether it is based on constraints, and, if so, which constraints. Correlatively, when both the toponym and the adjective are used to produce two output forms with the same meaning, it should be explained why these output forms occur with such differences in frequency. Section 4 addresses these issues, and proposes a tentative answer to the above questions.

4. EPNs Analysis

4.1 Base variations: not a matter of meaning

Sections 3.2 and 3.3 show that EPNs are either deadjectival (ITALIANITÉ) or detoponymic (FRANCITÉ). Moreover, an important amount of data demonstrates that a single toponym (AMÉRIQUE) can be the origin of EPNs,
both directly (AMÉRICITÉ), or through an adjectival stage (AMÉRICANITÉ). The first question addressed by EPN base variation is thus related to meaning: do speakers want to express different realities when they make use of the adjectival base, and when they use the toponym? For us, the answer is no: our claim is that the choice between a toponymic or an adjectival base is not semantically governed. There are three indications which lead us to incline to this assertion.

First, for many EPNs, the input category is formally unidentifiable. Therefore, semantics cannot be involved:

\[(2) \text{PICARDITÉ} (< \text{PICARD or PICARDIE}); \text{RUSSITÉ} (< \text{RUSSE} ‘Russian’}

or \text{RUSSIE ‘Russia’}); \text{YUGOSLAVITÉ} (< \text{YUGOSLAWE ‘Yugoslav’}

or \text{YUGOSLAVIE ‘Yugoslavia’})

Second, it may be the case that context does not enable the detection of semantical differences between two outputs, when the first is based on a toponym, and the second is based on the corresponding ethnic adjective. For example, in (3)\(^5\), both belgicité and belgité are used with the same possessive determiner \(sa\), and both refer to a property pertaining to a human being (namely a writer, in each case). The same is true in (4): algérité in (4a) and algérianitě in (4b) are used with the same possessive marker. They both refer to the property of being Algerian, and occur in the strictly same context \(fier de (‘proud of’):

(3) a. Des écrivains comme Henri Michaux et Samuel Beckett (…) ont abandonné ce qui faisaient leurs spécificités minoritaires. Michaux a essayé d’effacer toutes traces de sa belgicité, Beckett a abandonné sa langue (…)

Writers as Henri Michaux and Samuel Beckett (…) gave up what made their respective minority specificity. Michaux tried to erase all trace of his Belgium-ness, Beckett abandoned his language (…)

b. Il écrit son premier roman (…), avec les accents sincères de sa belgité (…)

He wrote his first novel (…), with the heartfelt accents of his belgianness

(4) a. Je t’invite donc à être fier de ton algérité.

Therefore, I’m encouraging you to be proud of your Algeria-ness

b. Ces jeunes “beurs”, comme on les appelle, nés en France, ont la nationalité française mais sont fiers de leur algérianité
These so-called young “beurs”, born in France, have French nationality, but are proud of their algerianness

The third clue is illustrated with examples (5-8) below. Each of them contains serial EPNs. Some of them are toponym-based (e.g. PAKISTANITÉ in (5), MAGHRÉBITÉ in (6), AMÉRICITÉ in (7), BELGICITÉ in (8)), others are adjective-based (e.g. ALGÉRIANITÉ in (5), AFRICANITÉ in (6), ITALIANITÉ in (8)), for others, the base is undecidable (e.g. ARABITÉ in (5), SERBITÉ in (7)). As in examples (4), it seems impossible to find a semantic difference that explains the choice between these possibilities:

(5) Là il était toujours question de négrité (plutôt que de sénégalité, d’ivoirité), d’arabité (plutôt que d’algérianité, de tunisité), d’indianité (plutôt que de pakistanité).

There, it was always about negro-ness (rather than Senegal-ness, Ivory-Coast-ness), Arabia-ness/arabness (rather than algerianness, Tunisia-ness), indianness (rather than Pakistan-ness)

(6) Une autre question est celle de l’établissement d’indicateurs de francité, d’africanité, de maghrébité ou autres.
Another issue is that of establishing indicators to France-ness, africanness, Maghreb-ness, or others.

(7) La serbité, mon œil ! Ca n’existe pas plus que la francité, l’américité ou la grécité !

Serbia-ness/serbianness, my foot! That does not exist, no more than France-ness, America-ness or greekness

(8) Un albanais qui a l’albanité en lui la vit aussi naturellement qu’un Mario Spaghettini vit son "italianité" ou un Jean-Jacques Vanderfrite vit sa "belgicité" : sans se poser de question.

If an Albanian holds Albania-ness in himself, then he does as naturally as a Mario Spaghettini lives his Italianness, or a Jean-Jacques Vanderfrite lives his Belgium-ness: without wondering about it.

The conclusion of these investigations is that the role of semantics is irrelevant with respect to speakers’ decisions in terms of EPN bases. Consequently, the choice must be a matter of form.
4.2 Base variations: a matter of form

We formulate the hypothesis that, by default, speakers chose to apply the general -ité suffixation rule to coin new EPNs. That is, adjective-based EPNs are the default case.

However, numerous trends can either favour, or conversely, prevent, the instanciation of this default rule. These trends all apply on the output form in such a way that the base category value would result from the competition of four constraints. These constraints, are briefly presented in (9). They are expressed in terms of avoidance (C₁-C₂) or preference (C₃-C₄):

\[
\text{(9)}
\]

\[
\begin{align*}
\text{[C₁ – dissimiation constraint]} & \text{ Very strong, if not absolute, prevention of final sequences with identical or similar sequences at the base-affix boundary} \\
\text{[C₂ – avoidance constraint]} & \text{ Strong prevention of -aisité (/ezite/ or (/ezite/)) and -oisité (/wazite/) final sequences} \\
\text{[C₃ – lexical pressure]} & \text{ Preference for well represented final sequences in the French attested lexicon} \\
\text{[C₄ – size constraint]} & \text{ Preference for quadrisyllabic outputs}
\end{align*}
\]
4.2.1. Avoidance Strategies

As far as avoidance strategies are concerned, $C_1$ is an example of
dissimilation constraint (cf. Grammont 1895). In the framework of lexeme
formation, dissimilation constraints are meant to prevent two identical or
almost identical phonological sequences from following each other at lexeme
constructional boundaries (cf. Corbin & Plénat 1992; Lignon et al. to appear;
Plag 1998). This explains why, on the Web, there is no occurrence for
YÉMÉNITÉ (‘Yemeniness’), whereas 66 pages (Table 8, line (a)) have been
indexed with YÉMÉNITÉ (‘Yemen-ness’).

More generally, $C_1$ explains the quasi-complete lack of /Njanite/ and /Neanite/
ending EPNs, where N corresponds to the nasal consonants /n/ or /m/ (cf.
Table 8, lines b-d). $C_1$ prevents nearby similar sequences /Nj/ ~ /Ni/ and /Ne/
~ /Ni/.

In particular, at line b, the dissimilation principle leads speakers to apply -ité
directly on the toponym when the adjective is itself obtained by suffixation of
-ien (/jɛ̃/) from this toponym, which in turn ends (1) either with a final nasal
vowel (IRAN /ɪ.ʁaŋ/ > IRANIEN /i.ʁa.njɛ̃/), (2) or with a final syllable starting
with a nasal onset (MAURITANIE /mo.ʁi.ta.ni/ > MAURITANIEN /mo.ʁi.ta.njɛ̃/).
A similar line of reasoning also holds for /neanite/ and /mjanite/ sequences (lines c and d). Moreover, as far as VIETNAMIANITÉ is concerned, we can notice that a further reason for its ungrammaticality is the succession of 3 nasal onsets (/vjɛtnjanite/). As a last concluding remark, we can notice that the frequency difference between ARMÉNIANITÉ (9 occ.) and both IRANIANITÉ and MAURITANIANITÉ (0 occ.) in b can be explained by the value of the vowel occurring in the last but one syllable preceding the -ité suffix: the /anjanite/ sequence in *IRANIANITÉ and *MAURITANIANITÉ leads to the strongly avoided /ani/ segment repetition, whereas the /enjanite/ ending occurring e.g. with ARMÉNIANITÉ is not completely forbidden.

@@ Insert Table 8 here

<table>
<thead>
<tr>
<th>Avoided sequence</th>
<th>Examples</th>
<th>Avoided form: # occ</th>
<th>Observed form: # occ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a /ite/</td>
<td>YÉMÉNITE &gt; YÉMÉNITÉ: 0</td>
<td>YÉMEN &gt; YÉMÉNITÉ: 66</td>
<td></td>
</tr>
<tr>
<td>b /njanite/</td>
<td>ARMÉNIEN &gt; ARMÉNIANITÉ: 9</td>
<td>ARMÉNIE &gt; ARMÉNITÉ: 6390</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IRANIEN &gt; IRANIANITÉ: 0</td>
<td>IRAN &gt; IRANITÉ: 448</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MAURITANIEN &gt; MAURITANIANITÉ: 0</td>
<td>MAURITANIE &gt; MAURITANITÉ: 394</td>
<td></td>
</tr>
<tr>
<td>c /eanite/</td>
<td>MÉDITERRANÉEN &gt;</td>
<td>MÉDITERRANÉE &gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MÉDITERRANÉANITÉ: 0</td>
<td>MÉDITERRANÉITÉ: 659</td>
<td></td>
</tr>
<tr>
<td>d /mjanite/</td>
<td>VIETNAMIEN &gt; VIETNAMIANITÉ: 0</td>
<td>VIETNAM &gt; VIETNAMITÉ: 92</td>
<td></td>
</tr>
</tbody>
</table>

Table 8: C₁ - Dissimilation constraints for EPNs
Avoidance constraint $C_2$ accounts for the observed property nouns corresponding to ethnic adjectives ending with /e/ or /wa/. Very often, speakers prefer to apply /ite/ directly on the toponym, as illustrated in Table 9. Sometimes this solution conflicts with other constraints. For instance, though violating lexical pressure constraint $C_3$, as will be shown in § 4.2.2, BURUNDITÉ, CAMARGUITÉ and JAPONITÉ (‘Japan-ness’) are more frequent than, respectively, BURUNDAISITÉ (‘burundeseness’), CAMARGUAISITÉ (‘camargueseness’) and JAPONAISITÉ. In other cases, constraint $C_2$ conflicts with size constraint $C_4$ (§ 4.2.2): for instance, RWANDITÉ is produced instead of RWANDAISITÉ (‘rwandanness’). So, as shown in Table 9, the avoidance constraint $C_2$ wins over both lexical pressure ($C_3$) and size ($C_4$) constraints in case of conflict; in other words, $C_2$ seems higher-ranked in the constraint hierarchy. The use of the -ité suffixation rule is not completely forbidden, but rather unlikely (for instance, SÉNÉGALAISITÉ does occur, but only 3 times). On the other hand, it should be noticed that this default rule is actually used to produce rather frequently occurring nouns. But then, the rule does not select the ordinary, standard adjectival base form (according to column 2 in Table 9, the JAPONAIS > JAPONAISITÉ pair is extremely rare, and so are DANOIS (‘Dane’) > DANOISITÉ, HONGROIS > HONGROISITÉ, THAÏLANDAIS (‘Thai’) > THAÏLANDAISITÉ and CHINOIS > CHINOISITÉ): rather, it applies either to the adjective bound suppletive base (/dan/ > DANITÉ, /sin/ > SINITÉ) or to the adjective (possibly learnt) variant: THAI, instead of THAÏLANDAIS, is used to
form THAÏTÉ; MAGYAR and NIPPO respectively replace HONGROIS and JAPONAIS. Of course, this solution requires for speakers to have these suppletive bases stored within their mental lexicons.

@@ Insert Table 9 here

<table>
<thead>
<tr>
<th>/ɛ/ or /wa/ ending ethnic adjectives (suppletive adj.)/ Toponym</th>
<th>Adjective-based EPNs: # occ</th>
<th>Toponym-based EPNs: # occ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjective</td>
<td>Xais/Xois A</td>
<td>Suppletive base</td>
</tr>
<tr>
<td>BURUNDAIS/ BURUNDI</td>
<td>BURUNDAISITÉ: 0</td>
<td>BURUNDITÉ: 20</td>
</tr>
<tr>
<td>CAMARGAIS/ CAMARGUE</td>
<td>CAMARGAISITÉ: 0</td>
<td>CAMARGUITÉ: 7</td>
</tr>
<tr>
<td>DANOIS (DAN)/ DANEMARK</td>
<td>DANOISITÉ: 0</td>
<td>DANITÉ: 25</td>
</tr>
<tr>
<td>HONGROIS (MAGYAR)/ HONGRIE</td>
<td>HONGROISITÉ: 0</td>
<td>MAGYARITÉ: 57</td>
</tr>
<tr>
<td>JAPONAIS (NIPPO)/ JAPON</td>
<td>JAPONAISITÉ: 3</td>
<td>NIPPOITÉ: 202</td>
</tr>
<tr>
<td>RWANDAIS/ RWANDA</td>
<td>RWANDAISITÉ: 0</td>
<td>RWANDITÉ: 46</td>
</tr>
<tr>
<td>SÉNÉGALAIS/ SÉNÉGAL</td>
<td>SENÉGALAITÉ: 3</td>
<td>SENÉGALITÉ: 271</td>
</tr>
<tr>
<td>THAÏLANDAIS/ THAÏ/ THAÏLANDE</td>
<td>THAÏLANDAISITÉ: 0</td>
<td>THAÏTÉ: 2</td>
</tr>
</tbody>
</table>

Table 9: /ɛ/ or /wa/ ending ethnic adjectives and corresponding EPNs
4.2.2. Preference Strategies

The preference expressed in $C_3$, also remarked by Franz Rainer for Spanish (p.c.), is a particular case of lexical pressure. This term describes the effect the attested lexicon can exert on the possible lexicon. Our claim is that, when he/she coins a new EPN, the speaker can be influenced by his/her knowledge of actual French -ité ending nouns, stored in his/her mental lexicon, which we assume to be reflected by dictionaries. Table 10 reports, in increasing frequency order, all phonological sequences $TLF$ /ite/ ending nouns may end with. This investigation voluntarily accounts for all nouns ending with the /ite/ sound, be they simple or complex. The most frequent word endings pertain to lines a to i.

@@ Insert Table 10 here

<table>
<thead>
<tr>
<th>/ite/ nouns final sequence in $TLF$</th>
<th>Number of lexemes</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>a /ilite/</td>
<td>454</td>
<td>FUTILITÉ ('futility')</td>
</tr>
<tr>
<td>b /alite/</td>
<td>278</td>
<td>BANALITÉ ('banality')</td>
</tr>
<tr>
<td>c /isite/</td>
<td>123</td>
<td>ATOMICITÉ ('atomicity')</td>
</tr>
<tr>
<td>d /asite/</td>
<td>82</td>
<td>FAMILIARITÉ ('informality')</td>
</tr>
<tr>
<td>g /inite/</td>
<td>26</td>
<td>AFFINITÉ ('affinity')</td>
</tr>
<tr>
<td>h /anite/ (different from /janite/)</td>
<td>24</td>
<td>HUMANITÉ ('humanity')</td>
</tr>
<tr>
<td>i /esite/ or /esite/</td>
<td>23</td>
<td>VÉRITÉ ('truth')</td>
</tr>
</tbody>
</table>
Apart from type frequency, another factor favouring lexical pressure probably is use (or token) frequency. For instance, the small number of /ezite/ ending nouns in dictionaries can be offset by the high use frequency of these nouns (e.g. over 3.2 millions occurrences of “obésité” on the Internet).

The next step in our experiment was to perform the same classification task to our 213 EPNs. As Table 11 shows, this second result is consistent with the previous one:

1) Each of the 16 sequences in Table 10 occur in EPNs,
2) The most frequently occurring EPNs in Table 11 end with one of the 9 most frequent ending sequences of Table 10,
3) Those final sequences which are the most frequently represented among the 213 EPNs are also the most frequent sequences in the /ite/ ending general lexicon, according to Table 10.
This is how it can be explained that AMÉRICITÉ goes up to 10 occurrences (Table 11, line b), beside AMÉRICANITÉ (Table 11, line h), which is well-formed according to the -ité suffixation rule, and very frequent on the Web (8660 occ.): our assumption is that the existence of AMÉRICITÉ is eased by the /isite/ sequence, at third place among /ite/ ending dictionary attested nouns (Table 10, line c). A similar explanation can be given for BELGICITÉ, which has 33 occurrences on the Web, and which coexists with BELGITÉ (112 occ.), that instanciates the -ité rule. Besides, as we shall see below, BELGICITÉ has
the advantage of satisfying $C_4$. Moreover, as \textsc{belge} $>$ \textsc{belgité} is concerned, notice that lexical pressure (given the high rank of the /eite/ final sequence in the attested vocabulary in Table 10) may also be the cause of the existence of the \textsc{belge} to \textsc{belgéité} variant -ité rule application (Table 11, line $e$).

It is interesting to notice that the attempt to model a new EPN on a well-represented ending in the lexicon may lead to form nouns on ethnic adjective bases unattested in French. As we can see in example (10), this is the case for \textsc{anglicité} (514 occ.), whereas the ethnic adjective is \textsc{anglais}, and \textit{anglique} is not an attested alternative; this is also the case for \textsc{antillanité} (942 occ.), formed on \textit{antillan} (the French ethnic attested adjective is \textsc{antillaïs}).

<table>
<thead>
<tr>
<th>EPN</th>
<th>Attested ethnic adjective</th>
<th>Toponym</th>
<th>EPN formal base</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textsc{anglicité}</td>
<td>\textsc{anglais}</td>
<td>\textsc{angleterre} (‘England’)</td>
<td>\textit{anglique}</td>
</tr>
<tr>
<td>\textsc{antillanité}</td>
<td>\textsc{antillaïs}</td>
<td>\textsc{antilles}</td>
<td>\textit{antillan}</td>
</tr>
</tbody>
</table>

The second preference we noticed is a tendency for quadrisyllabic outputs. When speakers create a new EPN, their decision is also guided by prosodic matter, that is, obtaining the optimal output size. The size constraint expressed in $C_4$, actually follows Plénat’s (to appear) hypothesis. It states that, ideally,
French roots in constructed lexemes tend to be disyllabic. /i.te/ consisting itself of two syllables, EPNs are thus expected to be quadrisyllabic. Data in example (11) follow that direction. Though they instanciate the general -ité rule, nouns in the left column are less frequent on the Web than corresponding four-syllable nouns in the right column, directly formed on the toponym:

(11)

<table>
<thead>
<tr>
<th>ALSACIANITÉ: 134</th>
<th>ALSACITÉ: 176</th>
</tr>
</thead>
<tbody>
<tr>
<td>/al.za.sja.ni.te/</td>
<td>/al.za.si.te/</td>
</tr>
<tr>
<td>BELGITÉ: 33</td>
<td>BELGICITÉ: 112</td>
</tr>
<tr>
<td>/bel.ʒi.te/</td>
<td>/bel.ʒi.si.te/</td>
</tr>
<tr>
<td>SOMALIANITÉ: 0</td>
<td>SOMALITÉ: 6</td>
</tr>
<tr>
<td>/so.ma.lja.ni.te/</td>
<td>/so.ma.li.te/</td>
</tr>
</tbody>
</table>

4.2.3. Combining strategies

These major tendencies still require refining; however, they allow us to draw up some rules in order to predict the most likely form for an EPN in French. These rules combine avoidance and preference techniques in a three-way strategy: (1) preference for an adjectival base, (2) choice for a replacing form when the adjective leads to a sequence to be avoided, (3) coexistence of
several forms, when preference constraints are met. Details on the manner
tactics (2) and (3) work are given in what follows.

When deadjectival formation is strongly prevented by avoidance constraint
$C_2$, any substitution form is possible, even when it does not satisfy lexical
pressure constraint ($C_3$), size constraint ($C_4$) or (exceptionally) neither. In
Table 12, all adjective-based EPNs in column 1 fail to satisfy $C_2^-$. Some of
the substitution forms displayed in column 2 violate $C_3$, so that they do not
match lexical pressure (e.g. GABONITÉ); among them, the insertion of the
epenthetic consonant /l/ in CONGOLITÉ (based on the proper noun CONGO)
leads to two remarks: (1) this insertion allows the resulting EPN both to meet
the prosodic constraint (/kõ.go.li.te/ is quadrisyllabic) and to avoid otherwise
vowel hiatus (*/kõ.go.i.te/); (2) the chosen epenthetic consonant is the same
as that what is inserted during the ethnic adjective formation: CONGOLAIS.
The rest of the substituted forms in Table 12 contradict size constraint $C_4$, e.g.
THAÏTÉ. In a few cases, the preferred form violates the dissimilation principle
(CHARENTITÉ: /ʃa.kœ.ti.te/, RÉUNIONITÉ: /ʁe.y.njo.ni.te/). Finally, the fact that
THAÏTÉ be preferred to THAÏLANDITÉ shows that, when there are two
candidates that satisfy $C_2$, preference is given to the -ité rule application.

@@ Insert Table 12 here
Table 12: Collected EPNs and avoidance constraint C₂

Conversely, when dissimilation and avoidance constraints (C₁ and C₂) do not apply (and when the adjectival base can be chosen), the co-existence of two constructions can be explained by the activation of preference constraints. This is what examples in Table 13 show.

@@ Insert Table 13 here

<table>
<thead>
<tr>
<th>Avoided form: #occ</th>
<th>Collected forms: # [violated constraint]</th>
</tr>
</thead>
<tbody>
<tr>
<td>BURUND AISITÉ: 0</td>
<td>BURUNDITÉ: 20 [^C₃]</td>
</tr>
<tr>
<td>CONGOLAISITÉ: 0</td>
<td>CONGOLITÉ: 16700[^C₃]</td>
</tr>
<tr>
<td>GABONAITÉ: 0</td>
<td>GABONITÉ: 82[^C₁]</td>
</tr>
<tr>
<td>JAPONAITÉ: 3</td>
<td>JAPONITÉ: 216[^C₃]</td>
</tr>
<tr>
<td>FINLANDAISITÉ: 0</td>
<td>FINLANDITÉ: 12[^C₃]; FINNITÉ: 7[^C₄]</td>
</tr>
<tr>
<td>IRLANDAISITÉ: 0</td>
<td>IRLANDITÉ: 97[^C₃]; IRLANDEITÉ: 1[^C₄]</td>
</tr>
<tr>
<td>CHARENTAISITÉ: 0</td>
<td>CHARENTITÉ: 14[^C₁;^C₃]</td>
</tr>
<tr>
<td>CAMARGUAISITÉ: 0</td>
<td>CAMARGUITÉ: 7[^C₃]</td>
</tr>
<tr>
<td>RÉUNIONAISITÉ: 0</td>
<td>RÉUNIONITÉ: 91[^C₄]</td>
</tr>
<tr>
<td>THAÏLANDAISITÉ: 0</td>
<td>THÄITÉ: 2[^C₃]; (THAÏLANDITÉ: 0)</td>
</tr>
<tr>
<td>CAMEROUNDAISITÉ: 2</td>
<td>CAMEROUNITÉ: 403[^C₃]</td>
</tr>
<tr>
<td>RWANDAISITÉ: 0</td>
<td>RWANDITÉ: 46[^C₃;^C₄]; RWANDÉITÉ: 3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>-itè rule application: # occ</th>
<th>Denominal EPN: # occ [satisfied constraint]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALGÉRIANITÉ: 2360</td>
<td>ALGÉRITÉ: 300 [C₅; C₄]</td>
</tr>
<tr>
<td>ALSACIANITÉ: 134</td>
<td>ALSACITÉ: 176 [C₅; C₄]</td>
</tr>
<tr>
<td>AUSTRALIANITÉ: 51</td>
<td>AUSTRALITÉ: 10 [C₅; C₄]</td>
</tr>
<tr>
<td>BELGITE: 33</td>
<td>BELGICITÉ: 112 [C₅; C₄]</td>
</tr>
<tr>
<td>Language</td>
<td>Score</td>
</tr>
<tr>
<td>------------</td>
<td>-------</td>
</tr>
<tr>
<td>BRESILIANITÉ: 106</td>
<td>BRESILITÉ: 13 [C₃; C₄]</td>
</tr>
<tr>
<td>COMORRIANITÉ: 50</td>
<td>COMORRITÉ: 2 [C₄]</td>
</tr>
<tr>
<td>COREANITÉ: 67</td>
<td>CORÉITÉ: 4 [C₃; C₄]</td>
</tr>
<tr>
<td>ÉTHIOPIANITÉ: 44</td>
<td>ÉTHIOPITÉ: 1 [C₄]</td>
</tr>
<tr>
<td>ÉTHIOPICITÉ: 32</td>
<td></td>
</tr>
<tr>
<td>GUINÉANITÉ: 3</td>
<td>GUINÉITÉ: 2 [C₃; C₄]</td>
</tr>
<tr>
<td>ISRAÉLIANITÉ: 1</td>
<td>ISRAÉLITÉ: 5 [C₃]</td>
</tr>
<tr>
<td>NORVÉGIANITÉ: 20</td>
<td>NORVÉGITÉ: 10 [C₄]</td>
</tr>
<tr>
<td>PROVENÇALITÉ: 315</td>
<td>PROVENCITÉ: 4 [C₄]</td>
</tr>
</tbody>
</table>

**Table 13: Preferences in EPN formations**

Two more phenomena are worth noticing, as far as constraint competition is concerned.

First, in some cases, the choice of a suppletive base (HELVÊTE ‘Helvetian’, IBÈRE ‘Iberian’, MAGYAR, NIPPON, HELÈNE ‘Hellene’, …) would allow to meet avoidance constraints C₁ and C₂, as well as preference constraints C₃ and/or C₄ and at the same time to instanciate -ité suffixation rule. However, this solution requires for the speaker to know this base. This explains why optimal forms such as HELLÉNITÉ or HELVÉ(T/C)ITÉ, which satisfy all constraints, are less frequent that their not-learnt counterparts GRÉCITÉ (‘Greek-ness’) and SUISSITÉ, which violate at least one constraint:

\[(12)\]

HELLÉNITÉ (< HELÈNE): 303 \hspace{1cm} GRÇITÉ (< GRÈCE): 842
Second, when several forms compete, a correlation is observed between the number of constraints fulfilled, the choice of the base category, and the number of EPN occurrences. Thus, in Table 14, EPNs are concerned neither by constraint $C_1$ nor by $C_2$. The differences in occurrence are related to the score -ité nouns obtain according to lexical pressure and size constraints, i.e. $C_3$ and $C_4$. For each noun, appropriatedness of the -ité general rule is also verified. The undisputable preference for HISPANITÉ ‘Hispany-ness’ (more than 10,000 occurrences) over IBÉRITÉ ‘iberianness’ (50 occurrences) requires further explanation, since both nouns have the same score with respect to $C_3 / C_4$. It certainly has to do with speakers common knowledge. In other words, the formal proximity between /ispani/ and /espaŋ/ (ESPAGNEPrN ‘Spain’), and, on the other hand, the formal distance between /espaŋ/ and /iber/, certainly are in favour of HISPANITÉ, and work against IBÉRITÉ. Another EPN pair raises an issue: that of ESPAGNOLITÉ ‘spanishness’ and ESPAGNITÉ ‘Spain-ness’. ESPAGNOLITÉ occurs 40 times – despite its violating both constraints $C_3$ and $C_4$ – whereas the unproduced ESPAGNITÉ only fails to meet constraint $C_3$. Here the explanation is related to categorial preference: ESPAGNOLITÉ is preferred to ESPAGNITÉ because the former is deadjectival (ESPAGNOLADJ), whereas the latter is detoponymic (ESPAGNEPrN).

@@ Insert Table 14 here
<table>
<thead>
<tr>
<th>Suppletive Base 1</th>
<th>Suppletive Base 2</th>
<th>Ethnic adjective</th>
<th>Toponym</th>
</tr>
</thead>
<tbody>
<tr>
<td>[*-ité rule; C₃; C₄]</td>
<td>[-ité rule; C₃; C₄]</td>
<td>[*-ité rule; *C₃; *C₄]</td>
<td>[*-ité rule; *C₃; C₄]</td>
</tr>
<tr>
<td>HISPANITÉ: 10,400</td>
<td>IBERITÉ: 50</td>
<td>ESPAGNOLITÉ: 40</td>
<td>ESPAGNITÉ: 0</td>
</tr>
</tbody>
</table>

Table 14: Frequences of EPN based on “Espagne” (Spain) toponym

To end with the examination of constraints (C₁-C₄), we have to invoke a further reason, which contributes to explain the realization of such or such form. This motivation appeals to proportional analogy. It is illustrated here with IRAKITÉ, in example (13a). This noun was found with 112 occurrences, whereas IRAKIANITÉ has only 3. Now, according to the above descriptions, IRAKIANITÉ does not violates fundamental constraints (namely, it does not contradict avoidance constraints C₁ and C₂), and IRAKITÉ fails to fulfill all requisites (it is certainly quadrisyllabic, but fails to meet C₃). Therefore, other reasons have to be given to justify the former’s relatively high frequency, and, comparatively, the almost non-existence of the latter.

Now, Irak’s geography and political news proximity with Iran’s are obvious, as are the prosodic similitudes in French between these countries names: /i.ʁɑ̃/ and /i.ʁak/. And figures show that IRAKITÉ’s predominance over IRAKIANITÉ mirrors what happens with IRANPN based EPN. As expected, there is no occurrence of IRANIANITÉ, that infringes constraint C₁ (see line b, Table 8). On the other hand, IRANITÉ, found with 448 occurrences, meets all
constraints. Therefore, the frequency of IRAKITÉ (compared with IRAKIANITÉ) has very likely to do with the wish of echoing IRANITÉ; this analogic construction can be modelled by means of equation (13b):

\[
\text{(13)}
\]

\[
\begin{array}{|c|c|c|}
\hline
\text{Toponym} & \text{Toponym-based} & \text{-ité rule application:} \\
\hline
\text{EPN: } \# \text{ occ} & \# \text{ occ} \\
\hline
\text{IRA} \text{K: } \text{IRA} \text{KITE: } 112 & \text{IRA} \text{KIANITE: } 3 \\
\text{IR} \text{A: } \text{IRANITE: } 448 & \text{IRANIANITE: } 0 \\
\hline
\end{array}
\]

\[
\text{b} \quad \text{Iran/iranité: Irak/ } X \Rightarrow X = \text{ irakité}
\]

4.3 Experimentation: student survey

We built an experiment aiming to assess the above-mentioned assumptions among French native speakers: (1) the choice for such or such a base is not a matter of meaning but is form-governed (in other words, base formal variations across EPNs are not correlated with differences in meaning); (2) by default, EPNs are instantiations of the general -ité suffixation rule; (3) this default case can be (in)validated by formal constraints. To achieve this, we conducted a survey with 38 third year linguistics students in the following way.
We provided them with a list of 143 nouns of both French and foreign towns, and their corresponding ethnic adjective. The instructions were: “for each toponym/adjective pair, give the corresponding ending -ité EPN(s), when possible”. Most of these nouns (and their corresponding adjective) were chosen on the basis of phonetic and/or prosodic criteria, to test our hypotheses. For example, Milan/milanais (‘Milanese’) allowed us to test C2 (do students produce milanaisité?), C3 and C4 (milanité is quadrisyllabic and contains a well-represented final sequence). For this pair, we expected milanité to be preferentially produced instead of milanaisité. Another example is Parme/parmesan. In this case, our expectation was in favor of the adjective-based EPN: parmité and parmesanité both violate C4, but the latter instanciates the -ité rule application, and it ends with a well-represented final sequence.

The results:

– confirm the preference for -ité rule application, when possible,
– show a strong avoidance for -ais and -ois ending bases,
– indicate a clear preference for quadrisyllabic outputs,
– confirm what we called lexical pressure (§ 4.2.2).

Table 15 proposes a sample of our results:
In this sample, $C_1$ is quite always satisfied ($lisbonnités$ is preferred to $lisbonninités$, $madrilé(a)nités$ to $madridités$), even when the produced form fails to meet $C_2$ (see line (c): $lilloisité$ vs $lillités$).

Students also tend to apply $C_2$ (line (b): $barcelonité$ vs $barcelonaisités$, line (k): $robertvalité$ vs $robertvalloisités$), except when the toponym-based EPN would
be trisyllabic (line (a): albigeoisité is more frequent than albité /al.bi.te/; line
(e): lyon(n)aisité is preferred to lyon(n)ité /ljo.ni.te/).

C₃, which gives preference to well-represented final sequences, is also
illustrated in Table 15: for instance, EPNs in /anite/ are frequently produced,
without regard to base categories (parmesanité, pavesanité are adjective-
based, milanité, taurignanité are toponym-based). Nanterrianité (as well as
nanterrien(n)ité, line (h)) constitutes an exception, but the alternative form,
nanterrité, contains another well-represented sequence (/erit/: see Table 10).

5. Conclusion

In this paper, we tried to demonstrate that a French speaker has access to two
orthogonal, but not mutually exclusive, construction ways to form an -ité
suffixed EPN:
– instanciate the general -ité rule which applies to adjectives and produces
nouns,
– apply -ité directly to the toponym.

We have shown that the choice between these two competing ways is a
matter of form, rather than a matter of meaning, since toponyms and
corresponding ethnic adjectives are semantically equivalent from the point of
view of EPN construction.
In French as well as in other languages, this formal competition is not exceptional. For instance, it can be observed for -isation French nouns (Table 16, lines a-c) and -ización Spanish nouns (Table 16, lines d-f)\(^7\).

More generally, we can conclude that two dimensions have to be accounted for to coin new lexemes: the first dimension considers word formation rules

<table>
<thead>
<tr>
<th>Toponym</th>
<th>Ethnic Adjective</th>
<th>Toponym-based Xisation (Xización): # occ</th>
<th>Adjective-based Xisation (Xización): # occ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>AUSTRALIE</td>
<td>AUSTRALIEN AUSTRALISATION: 23</td>
<td>AUSTRALIANISATION: 445</td>
</tr>
<tr>
<td>b</td>
<td>CAMEROUN</td>
<td>CAMEROUNAIS CAMEROUNISATION: 91</td>
<td>CAMEROUNAISISATION: 0</td>
</tr>
<tr>
<td>c</td>
<td>INDONÉSIE</td>
<td>INDONÉSIEN INDONÉSISATION: 0</td>
<td>INDONÉSIANISATION: 383</td>
</tr>
<tr>
<td>d</td>
<td>FINLANDIA</td>
<td>FINLANDÉS FINLANDIZACIÓN: 503</td>
<td>FINLANDESIZACIÓN: 0</td>
</tr>
<tr>
<td>e</td>
<td>INGLATERRA</td>
<td>INGLÉS INGLATERRIZACIÓN: 0</td>
<td>INGLESIZACIÓN: 9</td>
</tr>
<tr>
<td>f</td>
<td>PORTUGAL</td>
<td>PORTUGUÉS PORTUGALIZACIÓN: 57</td>
<td>PORTUGUESIZACIÓN: 144</td>
</tr>
</tbody>
</table>

Table 16: Competition for -isation and -ización nouns in French and in Spanish
and relations between an input and an output (that is, syntagmatic relations).
The second one considers the form of the output and, in some cases, the pressure the existing lexicon exerts on the coinage process (that is, paradigmatic relations). This observation connects with Burzio’s Output-to-Output faithfulness principle (e.g. Burzio 2002), according to which morphology has to be seen as a set of surface relations, and not (only) as a one-to-one relation between inputs and outputs. Finally, it witnesses paradigmatic morphology reemergence which has been stated in several recent works (e.g. Booij 1997 & 2007; Dal 2008).

References


Lignon, Stéphanie & Marc Plénat. (forthcoming). “Echangisme suffixal et contraintes phonologiques (Cas des dérivés en -ien et en -icien)”. Aperçus


* We are grateful to Franz Rainer for his comments, and to Cyril Auran for his linguistic corrections.

1 PRN indices stand for Proper Nouns.

2 In BELGICITÉ, pronounced /belʒɪsite/, /s/ results from the assibilation of the final /k/ in /belʒɪk/ (BELGIQUE).
For this experimentation, only the most likely allomorphs were generated. For instance, we chose to systematically apply /jan/ (resp. /eann/) allomorphy with -ien (resp. -éen) ending adjectives (e.g. ITALIEN > ITALIANITÉ; EUROPÉEN > EUROPÉANITÉ). Yet, a survey conducted with our students (§ 4.3) shows that this choice is disputable. And indeed, on the Internet, we can find examples such as italienité (6 occ.) or italienneité (21 occ.), européenité (5 occ.) or européennité (10 occ.).

As a further oddity, none of these three nouns is stored as a main dictionary entry: each of them appears as a subentry, of, respectively, germain, français and italien.

All contexts in (3) to (8) come from the Internet (June 2008).

Most of them also fail to satisfy the prosodic constraint C₄.

Spanish nouns and their frequencies on the Internet come from Franz Rainer (p.c.).