Ora Matushansky, SFL (CNRS/Université Paris-8/UPL)/UiL OTS/Utrecht University
email: Ora.Matushansky@cnrs.fr
homepage: http://www.trees-and-lambdas.info/matushansky/

EN FRANCE, À PARIS, LÀ-DEDANS - OÙ?
Frankfurt, May 18, 2021

1. THE OLD PUZZLE: COUNTRY NAMES, BARE AND DEFINITE

French toponyms are complicated in two ways:

Cities vs. countries:

(1) a. Paris, Nice, Londres
    b. la France, le Canada, l'Egypte...

Articles with locative prepositions:

(2) a. Huawei croit   en la   France.
       Huawei believes in DEF   France
       Huawei believes in France.

    b. Tu   te   crois   en France?
       you you.ACC   believe in France
       Do you believe yourself in France?

Table 1: The locative-argument divide

	city	feminine country	vowel country	masculine country
argument	Paris	la France	l'Egypte	le Canada
locative/directional	à Paris	en France	en Egypte	au Canada
ablative	de Paris	de France	d'Egypte	du Canada

Puzzles to be addressed:
- ➤ **locative and allative toponyms**: the *en/au* allomorphy
- ➤ **ablative proper names**: what happens to the article with feminine countries (and countries beginning with a vowel)?
- ➤ the phi-factor: why does it happen where it happens?

Further issues:
- ➤ other lexical-semantic classes and bare locations
- ➤ the locative preposition *dans*

Core proposals:
- ➤ **denotation of toponyms: entities vs. loci**
- ➤ **locative case in French for locus-denoting toponyms**
- ➤ **phi-dependent realization** *with declension classes*

Left for future research: what do definite articles realize?

2. THE LOCATIVE PORTMANTEAU

French country names are mostly definite. The article is marked for gender and number:

(3) a. la France, la Mauritanie…   feminine
    b. le Canada, le Pérou…   masculine
    c. l'Irlande, l'Egypte...   feminine/vowel
    d. l'Afghanistan, l'Angola, l'Iran...   masculine/vowel
    e. les Philippines, les Indes   plural

In locative/allative PPs involving country names the combination of *à*+DEF alternates with the portmanteau *en* in the function of phi-specification and phonology of the proper name

(4)   Je suis/vais…
      I   am/go

	à	la	maison	b.	au	bureau	c.	à	l'	aeroport
a.	at/to	DEF.F	home		at/to+DEF.M	office		at/to	DEF.M	airport

Cornulier 1972, Zwicky 1987, Miller, Pullum and Zwicky 1997: *en* is used if the proper name is **not a plural** and
➢   the proper name is **feminine** or
➢   the proper name **begins with a vowel**

Cornulier 1972: the same pattern with certain time expressions: *en hiver/au printemps*. Molinier 1990: there is a semantic difference (durative vs. punctual) partially correlated with the preposition. Lomholt 1983:22 provides some more pairs of the same kind: *en son nom et au mien*; *il croit en Dieu, il croit au diable*.

(5)   a.   en France, en Mauritanie…   feminine
      b.   au Canada, au Pérou…   masculine
      c.   en Irlande, en Egypte...   feminine/vowel
      d.   en Afghanistan, en Angola, en Iran...   masculine/vowel
      e.   aux/*en Philippines/Indes   plural

The combination of *à* with a country name does not give rise to *en* if *à* is not locative:

(6)   lié à la France/*en France

In combination with *en* the article is maintained if *en* is not locative (Lomholt 1983:37):

(7)   J'aurais   perdu ma   foi   en l'Amérique.
      I+would.have   lost   my   faith in   DEF+America
      I would have lost my faith in America.

Miller, Pullum and Zwicky 1997: The same is true for the ablative *de*:

(8)   a.   de France, de Mauritanie…   feminine
      b.   du Canada, du Pérou…   masculine
      c.   d'Irlande, d'Egypte...   feminine/vowel
      d.   d'Afghanistan, d'Angola, d'Iran...   masculine/vowel
      e.   des/*de Philippines/Indes   plural

The article remains if the preposition is not locative:

(9)   discuter de *(la) France

In all these cases **the masculine beginning in a vowel behaves like the feminine**

2.1.   Extending the picture: *dans* 'inside'

The lexical preposition *dans* 'inside' can be used with any toponym

With **restrictively modified country names** *dans* is used instead of *en/au* with the meaning 'in' (Lomholt 1983:126-135;145, see also Homma 2010):

(10) a.   dans l'Algérie/   la   France contemporaine/ d'aujourd'hui
           in   DEF+Algeria DEF France contemporary   of+today
           in contemporary/today's Algeria/France

      b.   dans l'Iran/   le   Canada contemporain/ d'aujourd'hui
           in   DEF+Iran DEF Canada contemporary of+today
           in contemporary/today's Iran/Canada

(11) a.   *en/✓dans l'Alsace libérée
      b.   *en/✓dans la Bretagne de mon enfance

In some situations *dans* and *en/au* seem to be in free variation (more on this below)

So the full picture includes **three possibilities, not two**, and all three can be attested within the same lexical-semantic class of **US states and Canadian provinces** (as well as with other compositional administrative units of federal states, see section 8 for the full picture):

(12) a.   en Californie, Caroline du Nord, Caroline du Sud…   *feminine states*
      b.   {en/[%]dans l'} Alabama, Alaska, Arizona, Arkansas…   *vowel-initial masculine states*
      c.   {dans le/au} Colorado, Connecticut, Delaware…   *masculine states*
      d.   à Hawaï   *island*

Unlike the functional prepositions *à* and *de*, **dans is a contentful lexical preposition, which never merges with the article**

2.2. The role of the lexical-semantic class

The distribution of *au/en* changes per lexical-semantic class (also noted in Miller, Pullum and Zwicky 1997; see section 8 for the full picture)

Table 2: Directional/locative summary

	feminine	**masculine V**	**masculine C**
country, continent	*en*	*en*	*au*
US state, Canadian province	*en/[%]dans la*	*en/[%]dans l'*	*au/dans le*
French province	*en/[%]dans la*	*en/[%]dans l'*	*[%]en/dans le*
département	*en/dans la*	*en/dans l'*	*dans le*
cities	*à*	*à*	*à*
oceans, seas, lakes, etc.	*dans la*	*dans l'*	*dans le*

The **lexical-semantic class** seems to be decisive: department names, names of US states and Canadian provinces, country names and city names all pattern differently

Issue 1: what is **the syntactic (and ontological) status of the lexical-semantic class**?
Issue 2: **variability** within lexical-semantic classes (systematic: islands and archipelagoes; unsystematic: the anarthrous country of *Israël* (taking *en/de*) or *Haïti* (taking *en/de*; taking *à* in some dialects), the definite archipelago of *le Dodécanèse* (taking *dans le/du*)...)

Conclusion: **item-specific encoding**. How?

Factors of variation:
  ➢   denotation (locus rather than entity)
  ➢   topological properties (container vs. point)
  ➢   gender and number

Simplification: **productive spatial and temporal localization only**. For the broader picture of *en* see Waugh 1976, Guimier 1978, Katz 2002, Amiot and De Mulder 2011, among others

3.   ANALYSIS: THE PORTMANTEAU VARIANT AS CASE

Proposal: Unlike *dans*, which is a true preposition, the portmanteau morphemes involve **case**: locative/directional (*au/en*) and ablative (*du/de*)

Country names and their ilk denote loci (regions, contiguous sets of points in space, etc.)
City names and their ilk are object-denoting
No commitment as to the precise formalism (see, e.g., Creary, Gawron and Nerbonne 1989, Wunderlich 1991, Zwarts and Winter 2000, Kracht 2002, Bateman et al. 2010, etc.), but the dichotomy is essential

> Locus-denoting nouns do not need a preposition to denote a location

Matushansky 2019: locative case in a language can be limited to locus-denoting nouns

3.1. Restricted locatives

It turns out that locative cases frequently have restricted distribution:

Locative cases **restricted to toponyms and certain common nouns** (Latin; Biblical Hebrew locative *he*: Hoftijzer 1981, Waltke and O'Connor 1990, Arnold and Choi 2003, Medill 2013, etc., some remnants in Modern Hebrew; Itzaj Maya: Hofling 2000:219)
> ➢ only these denote loci

Locative case-marking optional or absent for toponyms and some common nouns (Biblical Hebrew: Waltke and O'Connor 1990; Tswana: Creissels 2009; Western Armenian: Guekguezian 2011; Yimas: Foley 1991:165, 170-171; Gurr-goni: Green 1995:35)
> ➢ only these denote loci

Special locative case forms for toponyms and some common nouns (Hungarian (a handful of toponyms and a few common nouns): Rounds 2001:118; Agul, Archi, Avar, Lezgian, etc.: Daniel and Ganenkov 2009; Basque)
> ➢ only these denote loci

The case paradigm for toponyms and certain common nouns restricted to locative cases and genitive (Bagvalal: Daniel and Ganenkov 2009, Diyari: Austin 2013:52)
> ➢ these denote only loci

Limiting cases: locative forms only available for demonstratives, simplex wh-words and their derivatives (e.g., the English *here*, *where*, *there*; also *home*)

> In a lot of languages there are locus-denoting nouns and toponyms that can be identified by their syntax (see also Haspelmath 2018 for an alternative view)

Proposal to be explored here: **French area names denote loci** and when they don't, they are coerced into object-denotation and take the definite article and combine with *dans*
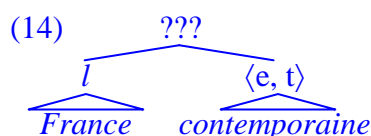
3.2. Modification

Reminder: modified country names require the article and the lexical preposition *dans*
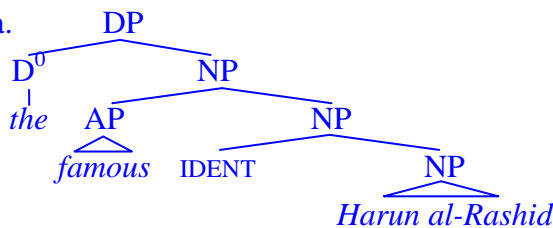
Straightforward explanation: type clash:

(13) a. $[\![\text{France}]\!] = \iota x \in D_l$. x is France   locus denotation
      b. $[\![\text{contemporain}]\!] = \lambda x \in D_e$. x is contemporary

Two issues: *France* is an entity and it is **of the wrong sort** (*l* instead of *e*):

(14)   ???
         _____
        l   ⟨e, t⟩
       ___   _____
   France   *contemporaine*

Modification of proper names, restrictive and non-restrictive, is known to be possible. This is how non-restrictive modification can be done:

(15) a.   DP   after Matushansky 2015

```
        DP
      /    \
    D⁰      NP
    |      /    \
   the   AP      NP
        /\      /    \
    famous  IDENT     NP
                     /\
              Harun al-Rashid
```
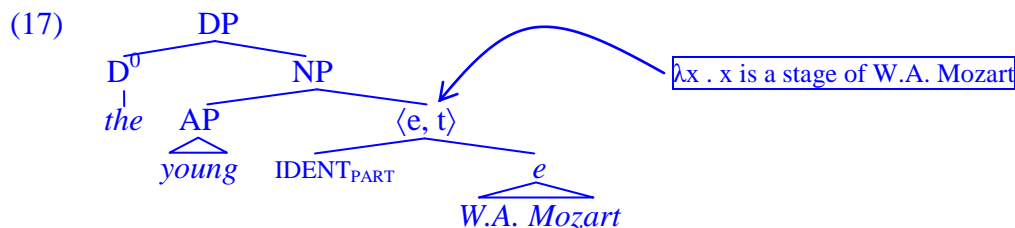
b. ιx . x is Harun al-Rashid and x is famous

In our case modification is restrictive and of a very specific kind (Paul 1994, Gärtner 2004, Jonasson 2005), coercing the proper name into a set of its **aspects** (cf. Landman 1989), a.k.a. **facets** (Kleiber 1981, 2005), temporal **stages** or **spatial parts** (cf. Carlson 1977):

(16) a.   **The upper Rhine** is polluted.   **material part**
b.   **The young W.A. Mozart** visited Paris.   temporal stage
c.   I will show you **the secret Paris**.   aspect/guise/facet
d.   **The Somerset Maugham that his nephew describes** is a lot more   proxy?
      disagreeable than the Somerset Maugham described by Somerset Maugham.

Is this coercion of an entity into a set of its material, temporal or functional parts represented in syntax?

I think not, but for the sake of clarity let us represent it:

(17)

```
            DP
          /    \
        D⁰      NP
        |      /    \           λx . x is a stage of W.A. Mozart
       the   AP    ⟨e, t⟩
            /\     /    \
        young  IDENT_PART   e
                         /\
                    W.A. Mozart
```
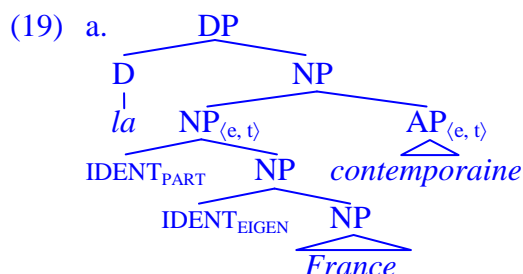
If *France* denoted an object, this would be the end of the story

We need to pass from a locus to the entity corresponding to this locus, and for this we need an appropriate function, which we will dub IDENT$_{EIGEN}$:

(18) IDENT$_{EIGEN}$: maps a locus to the unique entity located at this locus
      λl . ιx . EIGEN (x) = l,
      where EIGEN is a primitive function returning its location for a given object

The rest is straightforward, we once again represent the type-shift as part of the structure:

(19) a.

```
            DP
          /    \
        D       NP
        |      /        \
       la   NP_⟨e, t⟩      AP_⟨e, t⟩
            /    \        /\
      IDENT_PART  NP    contemporaine
                 /    \
          IDENT_EIGEN   NP
                       /\
                    France
```

b. ιy $\in$ D$_e$. y is a stage of France and y is contemporary

The IDENT$_{EIGEN}$ function is also used to pass from the locus denotation to the object denotation for argument uses of toponyms:

(20) a.   Huawei croit   en   la   France.
            Huawei believes in   DEF   France
            Huawei believes in France.

     b.   J'aurais   perdu ma   foi   en   l'Amérique.
            I+would.have   lost   my   faith in   DEF+America
            I would have lost my faith in America.

The presence of an article can be accounted for by assuming that $IDENT_{EIGEN}$ is projected or by hypothesizing that IDENT shifts return sets consisting of a single member and D^0 reflects the presupposition of uniqueness:

(21)
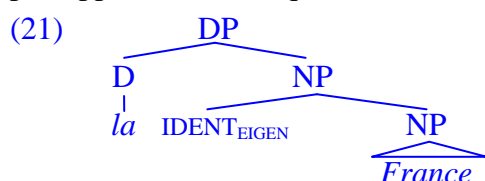
The use of the corresponding entity-correlate makes the portmanteau variant impossible

The resulting entity-correlates are **conceptualized as three-dimensional entities** (rather than points), thus the corresponding locative preposition is *dans*

Hypothesis: some (very few) toponyms have **both denotations at once**: locus and entity. As a result, they appear bare in argument positions and combine with the portmanteau morpheme (core instance *Israël*, also *Haïti*)

Now we can explain why certain modifiers allow both variants:

Among such modifiers are cardinal directions (*du Nord*, *Orientale*), colonial allegiances (*Britannique*, *Mineure*, *Grande/Petite*, *Nouvelle/Vieille*), administrative sub-divisions (*en Bourgogne nivernaise*, *dans la Campine anversoise*), etc.

(22) a.   Que le   probleme se   produise   **en Europe de l'Est.** Lomholt 1983:132
            that the problem REFL produce.SBJV in Europe of   the+East
            That the problem be produced in Eastern Europe

     b.   une allegresse inconnue **dans 1'Europe   de l'Est**
            a   joy   unknown   inside the+Europe of   the+East
            a joy unknown in Eastern Europe

Hypothesis: the variation is determined by whether the proper name is locus-denoting when combining with the modifier: $IDENT_{PART}$ can combine with a locus-denoting noun without the need for $IDENT_{EIGEN}$

> The appearance of the definite article indicates object-denotation

4.   INTERMEDIATE CONCLUSION

Hypothesizing that proper names can be locus-denoting as well as object-denoting yields:

➢   an account of **the city/country divide**: cities are entities, while countries are loci. Possible alternative: both are loci, but cities are points and countries are places (*à* is essive/allative, *en/au* is inessive/illative)

➢   a theory of **locative portmanteaus**: they realize locative cases

➢   an explanation for the emergence of **the lexical preposition *dans* with modified country names**: in order to be modified, country names must type-shift to denote objects, which makes locative cases impossible

What is the portmanteau?

Option 1: the portmanteau is the case-marked form of the proprial definite article
- ➢ pro: lots of languages mark case on the article only
- ➢ contra: what about vowel-initial names?

Option 2: the portmanteau is a case prefix of the proper name itself, i.e., such proper names are anarthrous
- ➢ pro: declension classes account for vowel-initial names
- ➢ contra: why is the case prefix in complementary distribution with the definite article? (Answer: because articles are there only for the object-denotation)
- ➢ pro: department names are anarthrous in argument positions

Missing: an approach to different lexical-semantic classes, explaining variation for masculine consonant-initial toponyms (old provinces, US states, etc.)

5. PHI-FEATURE SPECIFICATION AND THE MORPHOLOGY OF LOCATIVE CASES

A global issue: how can the lexical-semantic class influence the syntax of a proper name?

Desideratum: there is no such thing in syntax as a lexical-semantic class. There are some connections between the lexical-semantic class of a particular lexical item and its phi-feature specification. It is the phi-feature specification that is syntactically active

Core intuition: belonging to a given lexical-semantic class may translate into a particular phi-feature specification (e.g., in Latin names of trees are feminine), which in turn gives rise to a particular morphosyntactic pattern

Core assumption: masculine agreement in French, being the Elsewhere case, can correspond to more than one phi-feature specification

Core empirical generalization: only three options, with the apparent optionality in, e.g., US state names or French province names coming from individual phi-feature specifications (or different encoding of the entire lexical-semantic class)

In all classes feminine is phi-specified (as is plural, where available)

Evidence for the three categories: lexical-semantic classes that fit into only one of them (i.e., consonant-initial department names are only compatible with *dans*; no optionality for months or seasons):
Google searches reveal an abundance of *en* with such department names as *Deux-Sèvres* or *Pas-de-Calais*

(23) a. {en/dans la}   Savoie
     b. {en/dans l'}   Aveyron/Isère
     c. dans les   Deux Sevres

There are two lexical-semantic classes that follow two patterns:
- French provinces: *en* throughout or *en* vs. *dans le*
- US states: *en/au* or *en* vs. *dans le*

The patterns within the category have to be distinct by one (value of a) feature

The rest of the observed picture:
- ➢ cities and city-like islands denote objects and combine with *à*
- ➢ French departments may also start out as object-denoting and combine with *dans* throughout (because perceived as containers, unlike cities which are perceived as points)
- ➢ all plural toponyms combine with *aux*

Rivers, oceans, seas, lakes, mountains, etc., generally denote objects

5.1. Vocabulary Insertion rules and phi-feature specifications

Handling vowel-initial toponyms: declension classes

Core intuition: vowel-initial = feminine

(24) declension class I: [+ feminine] or vowel-initial
      declension class II: [– plural]

The realization of locative with **masculine consonant-initial loci** depends on how they are lexically specified for number, gender and animacy:

Table 3: Locative/directional featural specifications

	feminine	**masculine V**	**masculine C**	**masculine ϕ**
country **US state (1)** *season*	[+feminine, –animate]	[–feminine, –animate]	*au*	[–feminine, –animate]
French province (old pattern) *month*	[+feminine, –plural]	[–feminine, –plural]	*en*	[–feminine, –plural]
US state (2) *department* **French province (new pattern)**	[+feminine, –animate, –plural]	[–feminine, –animate, –plural]	*dans le*	[–feminine, –animate, –plural]

The distribution of *en* suggests that it is the elsewhere case, though feminine isn't default

(25) [–feminine, –animate, –plural] → Ø / __[locative][declension II]

This means that after impoverishment masculine consonant-initial proper names in the purple cell lack ϕ-feature specification:

	feminine	**masculine V**	**masculine C**	**masculine ϕ**
country **US state (1)** *season*	[+feminine, –animate]	[–feminine, –animate]	*au*	[–feminine, –animate]
French province (old pattern) *month*	[+feminine, –plural]	[–feminine, –plural]	*en*	[–feminine, –plural]
US state (2) *department* **French province (new pattern)**	[+feminine, –animate, –plural]	[–feminine, –animate, –plural]	*dans le*	Ø

Now for the Vocabulary Insertion rules:

(26) in the context of the features [+proper]

a. [locative][+plural] → aux
    [locative][–F][–animate] → au
    [locative]ϕ → en

b. [ablative][+plural] → des
    [ablative][–F][–animate] → du
    [ablative]ϕ → de

Ineffability results for proper names that lack ϕ-feature specification

The prepositional variant must then be used with concomitant switch to the entity-correlate

Several issues here:

➢   is the homophony with [common] locatives for plural and masculine accidental?
➢   can the lack of declensional specification yield syntactic effects? (yes)
➢   how artificial is this solution? The ineffable cell is the one with the default values (but most richly specified)

Alternative intuition: feminine is the default (semantic?) gender for toponyms. To pursue?

5.2.   Variant specification for some lexical-semantic classes

Variant behavior of **French provinces** (taking *en* in the old pattern or *en/dans le* in the new one): surface-masculine can correspond to the presence or absence of the animacy feature

Variant behavior of **US states**, Canadian provinces, etc. (taking *au* or *dans le*): surface-masculine can correspond to the presence or absence of the number feature

Variant behavior of **French departments** (taking *en/au* or *dans la/le*): this lexical-semantic class can be conceptualized as object-denoting (*dans* throughout) or locus-denoting

Depending on the speaker an individual lexical item might be specified differently from the lexical-semantic class it belongs to

Islands are either locus-denoting and specified for number, like countries, or object-denoting and unspecified for phi-features, like cities

6.   CONCLUSION

The syntax of locative portmanteaus in French is accounted for by the assumption that they realize locative case

A theory of locative case is provided linking constraints on its distribution to semantic type: **uninterpretable locative case is only available for locus-denoting NPs**

The variant realization of the locative cases in different lexical-semantic classes is linked to their phi-feature specification

Could all of this have been done with articles?
No: declension classes cannot be specified on articles

Is there independent evidence that different lexical-semantic classes of toponyms can vary in their phi-feature specification?

7.   INDEPENDENT EVIDENCE FOR VARIANT PHI-FEATURE SPECIFICATION

In Russian close apposition the toponym can have the **case assigned to the entire xNP** (case-agreement) or appear in the **default** (nominative) case

The distribution of case-agreement depends on the lexical-semantic class:

(27) a.   na   ulice   Jakimanka/Jakimanke   streets
            in   street.FSG.LOC   Yakimanka.FSG.NOM/LOC
            on the Yakimanka street

      b.   ot   stancii   Moskva/*Moskvy   stations
            from   station.FSG.GEN   Moscow.FSG.NOM/GEN
            from the station Moscow

Sometimes the non-agreeing nominative is the only option, sometimes case-agreement is:

➢   Animates require case-agreement
➢   Man-made objects (books, ships, hotels...) disallow case-agreement

Case agreement can be conditioned by phi-feature congruence and/or by the lexical category of the proper name
When not preceded by a common noun, proper names and kind names in argument positions are assigned case by regular mechanisms

7.1. Number congruence and optional case-agreement

In close apposition involving **toponyms** both the sortal and proper name are inanimate

Graudina, Ickovič and Katlinskaja 1976:141 (GIK): plural toponyms are incompatible with singular sortals:

(28) a.   v   gorode   Gagry/*Gagrax
      in   city.MSG.LOC   Gagry.PL.NOM/LOC
      in the city of Gagry

    b.   v   gorode   Velikie Luki/*Velikix Lukax
      in   city.MSG.LOC   Velikie Luki.PL.NOM/LOC
      in the city of Velikie Luki

For other lexical-semantic categories, the lexical category and formal features (φ-features) of the proper name can affect case-agreement:

> ➢ gender congruence not required (city, country, river names)

> ➢ gender congruence required (street names, syntactically complex city names with internal agreement)

> ➢ only with phi-congruent adjectival proper names (railway station, cape, peninsula, etc., names)

Note: there is notable cross-speaker variation in assigning different lexical-semantic categories of toponyms to these classes. There is also variation for individual proper names

Lack of familiarity makes case-agreement less likely.

7.2. Case-agreement on the condition of number congruence

For syntactically simplex city/town names, as well as for the names of countries and rivers, **gender congruence is not required for case agreement**, which is **optional**:

(29) a.   v   gorode   Moskva/Moskve   masculine sortal, feminine PN
      in   city.MSG.LOC   Moscow.FSG.NOM/LOC
      in the city of Moscow

    b.   v   gorode   Tallinn/Tallinne   masculine sortal, masculine PN
      in   city.MSG.LOC   Tallinn.MSG.NOM/LOC
      in the city of Tallinn

Gender-congruent toponyms are more likely to agree
and for some native speakers only gender-congruent city, country and river names can

7.3. Case-agreement on the condition of gender congruence

Street names and syntactically complex toponyms do not agree in case unless congruent in gender (GIK:142):

(30) a.   na ulice   Jakimanka/Jakimanke   ✓phi-congruent
      in   street.FSG.LOC   Yakimanka.FSG.NOM/LOC
      on the Yakimanka street

  b. na ulice   Balčug/*Balčuge   ✗phi-congruent
      in   street.FSG.LOC Balčug.MSG.NOM/LOC
      on the Balčug street

7.4. Case agreement with phi-congruent adjectival proper names only

For some categories of proper names case agreement is possible only with **morphologically adjectival** toponyms on the condition of both gender and number congruence:

(31) a. do   stancii   Bologoe/*Bologogo   ✗phi-congruent, ✓adjective
      until station.FSG.GEN Bologoe.NSG.NOM/GEN
      until the station Bologoe

   b. na   stancii   Moskva/*Moskvy   ✓phi-congruent, ✗adjective
      on station.FSG.GEN Moscow.FSG.NOM/GEN
      on the station Moscow

   c. na   stancii   Tixoreckaja/Tixoreckoj ✓phi-congruent, ✓adjective
      on station.FSG.GEN Tixoreckaja.FSG.NOM/GEN
      on the station Tixoreckaja

An incomplete list of such proper names includes boroughs (*mestečko*), villages (*selo*), ports, lakes, bays, volcanoes, hills (*sopka*), mountains, planets and railway stations. Prescriptive grammars may insist that case-agreement is impossible with such proper names or include in this list islands, republics, etc. Thus toponyms preceded by the common nouns *aúl* 'a village in the Caucasus and Central Asia' and *kišlák* 'a village in Central Asia' are claimed to never agree for case, but this most likely is due to the fact that the names of such villages are extremely unlikely to be adjectival: when they are, case-agreement becomes possible on the condition of phi-congruence

7.5. Summary: lexical semantics and phi-congruence

For lexical-semantic categories of proper names allowing case-agreement in close apposition there is a hierarchy of phi-features:

Table 4: Case-agreement with toponyms

	number congruence	**gender congruence**	**adjectival name**
cities, countries, rivers	required	optional	optional
streets, complex names	required	required	optional
residue	required	required	required

Animate proper names cannot not agree in case

7.6. Proposal: case-agreement as agreement

Matushansky 2021: the following features are underlyingly set for different lexical-semantic categories of proper names

Table 5: Case-agreement with toponyms
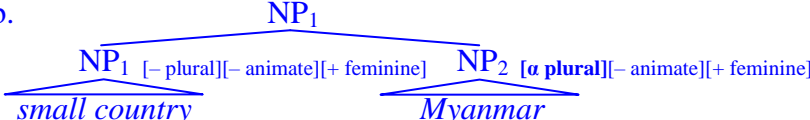
	number	**gender**	**animacy**	**congruence**
cities, countries, rivers	unvalued	valued	[– animate]	number
streets, complex names	valued	unvalued	[– animate]	number & gender
residue	unvalued			adjectival

Features not valued in this table can always be set according to nominal declension classes (formal features, no case-agreement) or by the corresponding sortal ((semantic) agreement of sorts, allows case-agreement)

Inanimates can be specified for all phi-features. When they are, agreement fails:

(32) a.   NP$_1$ no agreement

NP$_1$ [– plural][– animate][+ feminine]   NP$_2$ **[– plural]**[– animate][+ feminine]

small country   *Myanmar*

When one feature is unvalued, agreement becomes possible:

b.   NP$_1$ case agreement

NP$_1$ [– plural][– animate][+ feminine]   NP$_2$ **[α plural]**[– animate][+ feminine]

small country   *Myanmar*

When the number feature is valued as [+plural], case-agreement is impossible (toponyms are all inanimate and there is no gender in the plural)

The same phi-features are at play

8.   LEXICAL-SEMANTIC CLASSES OF LOCUS-DENOTING TOPONYMS

8.1.   Lexical-semantic class patterns

First impression: in function of the lexical-semantic class, the distribution of portmanteau variants and the availability of other options change

8.1.1.   French provinces and regions

Two issues at once: the portmanteau pattern does not distribute as it does with countries (no *au* altogether) and for masculine provinces the contentful lexical preposition *dans* is also possible:

The variation between *en* and *dans* for masculine province names seems partly historical (*en* is the more archaic variant, dealing with the older feudal province rather than a modern region), partly pragmatic (*dans* also has the "somewhere in" interpretation)

(33) **French provinces and regions**
    a.   en Picardie, en Normandie…   feminine provinces
    b.   [%]en/dans le/*au Béarn, Poitou…   masculine provinces
    c.   en/*dans l'Aquitaine, en/*dans l'Alsace...   feminine/vowel provinces
    d.   en/dans l'Artois, en/dans l'Angoumois...   masculine/vowel provinces

The older obligatory *en* pattern with masculine proper names resurfaces with month names (*en février*) and means of transportation (*en bus*, *en ski*)

Hypothesis: two competing grammars: the older one with *en* throughout the paradigm and the newer one with *dans* for masculine province names (I know at least one native speaker with this latter pattern)

This latter pattern is obvious with French department names

8.1.2.   French department names

French department names, irrespective of gender or initial segment, combine with *dans*, but *en* is possible for (Grevisse and Goosse 2006:1507 and various sources):

➢   departments that have the same name as provinces (*Dordogne*, *Gironde*, *Vendée*, *Vaucluse*, *Savoie*, *Aveyron*)

➢   composite singular toponyms (*Haute-Corse*, *Corse-du-Sud*, *Haute-Marne*, *Haute-Saône*, *Meurthe-et-Moselle*, *Indre-et-Loire*, *Seine-et-Marne*)

This list is a standard description, but all of these cases are **feminine or begin with a vowel**
Google searches reveal an abundance of *en* with such department names as *Deux-Sèvres* or *Pas-de-Calais*

Non-French European provinces and regions appear to vacillate between the French province pattern and the French department pattern (see Lomholt 1983:160-162 claiming that it makes a difference whether the region or province in question is Francophone)

8.1.3. Islands and archipelagoes

It turns out that islands do not behave uniformly with respect to the presence of the article or locative syntax

Vikner 1970, Lomholt 1983:235-245: country-like and city-like islands:

Country-like islands: overt definite article, *en* in the feminine and perhaps with vowel-initial names (I know of no islands that are masculine, definite and begin with a vowel):

(34)   a.   la Corse, la Sicile, la Tasmanie…   feminine
       b.   le Groenland, le Spitzberg   masculine
       c.   l'Irlande, l'Islande…   feminine/vowel

(35)   a.   en Corse, en Sicile, en Tasmanie   feminine
       b.   au Groenland, au Spitzberg   masculine
       c.   en Irlande, en Islande   feminine/vowel

When modified, they appear with the locative preposition *dans* (Vikner 1970:240)

City-like islands: no article, the locative preposition is *à*; the article is absent in the ablative; may contain a definite article as part of the proper name itself:

(36)   a.   à/de Terre-Neuve, à/de Belle-Ile, à/de Bornholm
       b.   à/de Madagascar, à/de Malte, à/de Bornéo
       c.   (à/de) la Réunion, (à/de) la Nouvelle-Amsterdam, (à/de) la Grenade

I am aware of no masculine city-like islands with the definite article

What about **the gender of city-like islands**?

Grammars assert that some city-like islands are masculine, but do not agree on which islands are (cf. Lomholt 1983:237-240). Neither do native speakers

The picture is reversed for archipelagoes (most of which are plural, anyway): they are, to the best of my knowledge, masculine when bare (as is the word for archipelago, *archipel*):

(37)   a.   le Svalbard (au/du), le Vanuatu (au/du)
       b.   Madère (à/de), Zanzibar (à/de)
       c.   le Dodécanèse (dans le/du)
       d.   la Côte-Froide (à/de)
       e.   la Nouvelle-Zélande (en/de)

To the best of my knowledge, no one treats archipelagoes as a separate lexical-semantic class, even though realistically they are very different from islands
Lexical exceptions: some islands allow both: *en/à la Martinique*, *en/à la Gouadeloupe*, but also *en/à Haïti*, sometimes for the same author within the same text (Vikner 1970:238); with the ablative *de* the article cannot be omitted (ibid.), but Lomholt 1983:244 is more cautious, claiming simply that the drop of the definite article is not as frequent as the use of *en*

8.1.4.   The residue

Continents behave like loci; as they are all feminine and begin with a vowel, *en*.

Rivers generally take the lexical prepositions *dans* and *sur*, as well as *à*; feminine ones may accept *en*. No information on vowel-initial river names in Lomholt:

(38) a.   Celle-ci   évita   d'être […] jetée   en Loire.   Lomholt 1983:285
           this.FSG-PROX avoided of+be.INF thrown in Loire
           This one avoided being thrown into the Loire.

     b.   le   corps d'un   inconnu repéché   dans la   Seine. Lomholt 1983:285
          the body of+INDEF unknown fished.out in   the.FSG Seine
          an unidentified body fished out in the Seine

Most other toponyms (e.g., **oceans, seas, lakes, bays**, **mountain chains**, etc.) take *dans*.

8.2.   Further support: bare urban landmarks

French has bare locatives, i.e., **French has locus-denoting NPs**

Stolz, Lestrade and Stolz 2014:ch.4.1: bare urban landmarks used as locations:
More research in Palm 1989, but I don't have it

(39) a.   Vous êtes allé   hier   rue Notre-Dame-de-Lorette.
           you.PL be.2PL go.PRTCPL yesterday street Notre-Dame-de-Lorette
           You went yesterday to the Notre-Dame-de-Lorette street.

     b.   Ils   arrivèrent   dans la   rue de la   Grande-Turanderie.
          they arrive.PAST.3PL in   the.FSG street of the.FSG Grande-Turanderie
          They arrived [somewhere] in the street of Grande-Turanderie.

The presence of internal *de* seems orthogonal (cf. Bosredon and Tamba 1999)

NP-internal and/or locational use is also possible:

(40) … une chambre à l' Hôtel des Cinq Continents **avenue de la République**   SLS2014
       a   room   to the Hôtel des Cinq Continents avenue de la République
       [Your Marcel has rented] a room at the Hôtel des CC on avenue de la R.

The default use of street names as locations or goals is **without a preposition**, and without an article, which resurfaces in argument positions and with lexical prepositions:

(41) Quand le taxi s'arreta sur le boulevard Richard-Lenoir…
     when the taxi stopped on the boulevard Richard-Lenoir
     When the taxi stopped on the boulevard Richard-Lenoir…

This is how we expect locus-denoting NPs to behave in a language that has no morphological case on nouns
Remember for the future: the behavior of the article!

Hypothesis: they denote loci

These toponyms are not specified for declension class and thus cannot be case-marked

The corresponding entity-correlates must have the article

Possibility: the definite article only occurs with the entity-denotation (and then *en/au* is a case marker on anarthrous proper name)

Or: both anthroponyms and toponyms can be anarthrous, so loci-denoting toponyms also can be

8.3. Prior treatments of these facts

Cornulier 1972: **the definite article remains iff it is phonologically incorporated into the preposition** (with *au* and *aux*). This is counter-cyclic and non-explanatory

Zwicky 1987: *en*, like *au* and *aux*, is a portmanteau morpheme realizing two syntactic positions, P$_{LOC}$ + FSG; there is a special rule of referral, replacing the masculine form with the feminine one, that is activated for proper names beginning with a vowel. However:
> ➤ elsewhere, elision (*à l'*) has priority over contraction (*au*); with possessives and definite articles feminine is replaced with masculine (*mon amie*, *l'amie*)
> ➤ there are lexical exceptions (*Danemark*, *Portugal* & *Luxembourg* used to take *en*; and old provinces still do so sometimes, as in *en Limousin*)

Fahlin 1942 via Molinier 1990: while there is historical development (en le → el → eu → ou → au), its timing (XIII c.) does not support the hypothesis that the underlying representation is *en* + DEF rather than *à* + DEF
Grevisse and Goosse 2008:1351: *à la* was used with feminine country names up to the XIX century

Miller 1992, Miller, Pullum and Zwicky 1997: French determiners and the prepositions *à*, *de* and *en* must be analyzed not as syntactic words but as **phrasal inflections** which are lexically realized on the first word of the NP

Theoretical issue: what is phrasal inflection?

Homma 2010: punctual objects are masculine, extended ones are feminine (explicitly ignores phonology)

Major problem: modularity

And none of them has looked at the full empirical picture

9. LATIN LOCATIVE CASES: THE FACTS

The **locative case in Latin** only appears with names of towns, cities, small islands and a few common nouns (42b) including *domus/domi* 'home', *rus/ruri* 'countryside' and *humus/humi* 'ground' (henceforth, *L-nouns*). All other toponyms and common nouns require a preposition

(42) a. iacēre humi   Gildersleeve and Lodge 1876:266
         lie.INF ground.LOC
         to lie on the ground

    b. Mīlitēs Albae   cōnstitērunt in urbe   opportūnā.
        soldiers Alba.LOC halted   in city.ABL convenient.ABL
        The soldiers halted at Alba, a conveniently situated town.

(43) a. Pompeius in Thessaliam   pervenit.   Woodcock 1959:4
         Pompey   in Thessaly.ACC arrived
         Pompey arrived in Thessaly.

    b. Me   potius in Hispania fuisse   tum quam Formiis! Woodcock 1959:36
        I.ACC   able   in Spain.ABL be.PERF.INF then than   Formiae.LOC
        To think of my having been in Spain at that time rather than at Formiae!

Cannot be a morphological restriction on the distribution of the locative case suffix (which is syncretic with other cells in the paradigm anyway): exactly the same set of lexical items uses bare **accusative** case-marking **for allative** and bare **ablative** case-marking **for the source**:
NB: The directional accusative can appear also with some country names (Woodcock 1959:4-6). No explanation yet

(44) a.　Missī　lēgātī　Athēnās　sunt.　　　　　Gildersleeve and Lodge 1876:214
　　　　sent.PL envoys Athens.ACC are
　　　　Envoys were sent to Athens.

　　b.　Innumerābilēs (philosophī) numquam domum revertērunt.
　　　　innumerable philosophers never home.ACC returned
　　　　Innumerable philosophers never returned home

(45) a.　(Verrēs) omnia　domō　ēius　abstulit.　　Gildersleeve and Lodge 1876:249
　　　　Verres everything house.ABL his took.away
　　　　Verres took everything away from his house.

　　b.　Dolābella Dēlō　proficīscitur.　　　　Gildersleeve and Lodge 1876:251
　　　　Dolabella Delos.ABL depart
　　　　Dolabella sets out from Delos.

And it is not the case that the use of a preposition is excluded:

(46) a.　ut　　a　Mutina　　discederet　　　　Latin sources, via Allen et al.
　　　　so.that from Modena.ABL retire.SBJ
　　　　that he should retire from Modena (which he was besieging)

　　b.　ad Alesiam　　proficiscuntur
　　　　to Alesia.ACC advance.3PL
　　　　they set out for Alesia

Since only L-nouns can make use of accusative and ablative cases to function as sources or goals, only L-nouns denote loci

10.　BIBLIOGRAPHY

Amiot, Dany and Walter De Mulder. 2011. L'insoutenable légèreté de la préposition *en*. *Studii de lingvistică* 1, pp. 9-28.

Arnold, Bill T. and John H. Choi. 2003. *A Guide to Biblical Hebrew Syntax*: Cambridge University Press.

Austin, Peter K. 2013. *A Grammar on Diyari, South Australia*. London: SOAS, University of London.

Bateman, John A., Joana Hois, Robert Ross, and Thora Tenbrink. 2010. A linguistic ontology of space for natural language processing. *Artificial Intelligence* 174, pp. 1027-1071.

Bosredon, Bernard and Irène Tamba. 1999. Une ballade en toponymie : de la rue Descartes à la rue de Rennes. *Linx* 40 pp. 55-69.

Carlson, Greg N. 1977. Reference to Kinds in English, Doctoral dissertation, University of Massachusetts, Amherst.

Cornulier, Benoit de. 1972. A peeking rule in French. *Linguistic Inquiry* 3, pp. 226-227.

Creary, Lewis G., Jean Mark Gawron, and John Nerbonne. 1989. Towards a theory of locative reference. In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, pp. 42-50: Association for Computational Linguistics.

Creissels, Denis. 2009. Spatial cases. In *The Oxford Handbook of Case*, ed. by Andrej Malchukov and Andrew Spencer, pp. 609-625. Oxford: Oxford University Press.

Daniel, Michael and Dmitry Ganenkov. 2009. Case marking in Daghestanian: limits of elaboration. In *The Oxford Handbook of Case*, ed. by Andrej Malchukov and Andrew Spencer, pp. 668-685. Oxford: Oxford University Press.

Fahlin, Carin. 1942. *Étude sur l'emploi des prépositions en, à, dans au sens local*. Uppsala: Almqvist & Wiksell.

Foley, William A. 1991. *The Yimas Language of Papua New Guinea*. Stanford, California: Stanford University Press.

Gärtner, Hans-Martin. 2004. Naming and economy. In *Empirical Issues in Formal Syntax and Semantics 5*. Available at http://www.cssp.cnrs.fr/eiss5.

Gildersleeve, Basil L. and Gonzalez B. Lodge. 1876. *Latin grammar*. London: Macmillan Education Ltd.

Graudina, Ljudmila Karlovna, Viktor Aleksandrovič Ickovič, and Lia Pavlovna Katlinskaja. 1976. *Grammatičeskaja pravil'nost' russkoj reči. Stilističeskij slovar' variantov*. Moscow: Nauka.

Green, Rebecca. 1995. A Grammar of Gurr-goni (North Central Arnhem Land), Doctoral dissertation, Australian National University.

Grevisse, Maurice and André Goosse. 2006. *Le bon usage. XIII édition*. Paris-Gembloux: Duculot.

Grevisse, Maurice and André Goosse. 2008. *Le bon usage. XIV édition*. Paris-Gembloux: Duculot.

Guekguezian, Peter Ara. 2011. Bare locatives in Western Armenian. Ms., USC.

Guimier, Claude. 1978. *En* et *dans* en francais moderne: Etude semantique et syntaxique. *Revue des Langues Romanes* 83, pp. 277-306.

Haspelmath, Martin. 2018. Differential place marking and differential object marking. Ms., Leipzig University.

Hofling, Charles Andrew. 2000. *Itzaj Maya Grammar*. Salt Lake City: University of Utah Press.

Hoftijzer, J. 1981. *A Search for Method: a Study in the Syntactic Use of the H-Locale in Classical Hebrew*. Studies in Semitic Languages and Linguistics 12. With the collaboration of H.R. van der Laan and N.P. de Koo. Leiden: Brill.

Homma, Yukiyo. 2010. Etude sur l'emploi de "en" devant les noms de territoire en français. In *Verbe, Préposition, Locution, Didactique*, ed. by Danielle Leeman, pp. 35-54. Available at https://halshs.archives-ouvertes.fr/halshs-00450823.

Jonasson, Kerstin. 2005. La modification du nom propre dans une perspective contrastive. *Langue française* 146, pp. 67-83.

Katz, Eva. 2002. Systématique de la triade spaciale *à*, *en*, *dans*. *Travaux de linguistique* 44, pp. 35-49.

Kleiber, Georges. 1981. *Problèmes de référence: descriptions définies et noms propres*. Paris: Klincksieck.

Kleiber, Georges. 2005. Les noms propres „modifiés" par *même*. *Langue Française* 146, pp. 114-126.

Kracht, Marcus. 2002. On the semantics of locatives. *Linguistics and Philosophy* 25, pp. 157-232.

Landman, Fred. 1989. Groups, II. *Linguistics and Philosophy* 12, pp. 723-744.

Lomholt, Jørgen. 1983. *Syntaxe des noms géographiques en français contemporain*. Copenhagen: Munksgaard.

Matushansky, Ora. 2015. The other Francis Bacon: on non-bare proper names. *Erkenntnis* 80, pp. 335-362.

Matushansky, Ora. 2019. The case of restricted locatives. In *Proceedings of Sinn und Bedeutung 23*, vol. 2, ed. by M. Teresa Espinal, Elena Castroviejo, Manuel Leonetti, Louise McNally, and Cristina Real-Puigdollers, pp. 161-178: Open Journal Systems.

Matushansky, Ora. 2021. *Die Stadt Leipzig* and other fun places in Russian. Paper presented at *FDSL 14*, Leipzig (Online), June 2-4, 2021.

Medill, Kathryn McConaughy. 2013. Directional strategies in Biblical Hebrew: influences on the use of locative *hey*. Ms., Indiana University.

Miller, Philip H. 1992. *Clitics and Constituents in a Phrase Structure Grammar*. New York: Garland.

Miller, Philip H., Geoffrey K. Pullum, and Arnold M. Zwicky. 1997. The principle of phonology-free syntax: four apparent counterexamples in French. *Journal of Linguistics* 33, pp. 67-90.

Molinier, Christian. 1990. Les quatre saisons : à propos d'une classe d'adverbes temporels. *Langue française* 86, pp. 46-50.

Palm, Lars. 1989. *"On va à la Mouff'?" : étude sur la syntaxe des noms de rues en français contemporain*. Uppsala; Stockholm, Suède: [Université d'Uppsala] ; Distribué par Almqvist & Wiksell International.

Paul, Matthias. 1994. Young Mozart and the joking Woody Allen. Proper names, individuals and parts. In *Proceedings of Semantics and Linguistic Theory (SALT) 4*, ed. by Mandy Harvey and Lynn Santelmann, pp. 268-281. Ithaca, New York: CLC Publications, Department of Linguistics, Cornell University.

Rounds, Carol. 2001. *Hungarian. An Essential Grammar*. London: Routledge.

Stolz, Thomas, Sander Lestrade, and Christel Stolz. 2014. *The Crosslinguistics of Zero-Marking of Spatial Relations*. Berlin: De Gruyter.

Vikner, Carl. 1970. La syntaxe des noms d'îles en français moderne. *Revue Romane* 5, pp. 231-249.

Waltke, Bruce K. and Michael Patrick O'Connor. 1990. *An Introduction to Biblical Hebrew Syntax*. Winona Lake, Indiana: Eisenbrauns.

Waugh, Linda R. 1976. Lexical meaning: The prepositions *en* and *dans* in French. *Lingua* 39, pp. 69-118.

Woodcock, Eric Charles. 1959. *A New Latin syntax* 267 p. Cambridge: Harvard University Press.

Wunderlich, Dieter. 1991. How do prepositional phrases fit into compositional syntax and semantics? *Linguistics* 29, pp. 591-622.

Zwarts, Joost and Yoad Winter. 2000. Vector space semantics: a model-theoretic analysis of locative prepositions. *Journal of Logic, Language and Information* 9, pp. 169-211.

Zwicky, Arnold M. 1987. French prepositions: no peeking. *Phonology Yearbook* 4, pp. 211-227.