

Ora Matushansky, SFL (CNRS/Université Paris-8)/UiL OTS/Utrecht University  
 email: [O.M.Matushansky@uu.nl](mailto:O.M.Matushansky@uu.nl)  
 homepage: <http://www.let.uu.nl/~Ora.Matushansky/personal/>

**THE DEFINITE ARTICLE IN PROPER PLACES**  
**Workshop on the semantic contribution of Det and Num. (In)definiteness, genericity and**  
**referentiality. UAB, May 27-28, 2016**

## 1. INTRODUCTION

Received wisdom: proper names differ syntactically from common nouns in that they lack the definite article, which would have been expected otherwise.

Taken as evidence for not treating proper names as definite descriptions, whereas non-bare proper names in argument positions in languages such as Catalan, Modern Greek, etc., are perceived as evidence for this view (see Matushansky 2008 for an overview)

Non-bare proper names have also given rise to the N-to-D approach for deriving anarthrous proper names (Longobardi 1994, Borer 2005)

Actual facts are more complicated:

- |     |    |  |  |
|-----|----|--|--|
| (1) | a. | the Campbells, the Yorks                           | familial or political clan               |
|     | b. | the Beatles, the Rolling Stones, the Monty Pythons | band, company                            |
|     | c. | the Mets, the Mikes, the Tigers                    | baseball or football team, Army regiment |
| (2) | a. | the Alps, the Rockies                              | mountain chains                          |
|     | b. | the Hebrides, the Orkneys                          | archipelagoes                            |
|     | c. | the Netherlands                                    | conglomerate countries                   |
|     | d. | the Pleiades, the Hyades                           | constellations                           |

All syntactically plural proper names require the definite article (*Athens* is not plural)

Certain lexical-semantic classes of proper names are obligatorily definite:

- |     |    |                                   |                              |
|-----|----|-----------------------------------|------------------------------|
| (3) | a. | the Seine, the Erie, the Atlantic | water bodies                 |
|     | b. | the Titanic, the Lusitania        | vehicle names                |
| (4) | a. | the Milky Way, the Broadway       | former definite descriptions |
|     | b. | the Bronx, the Ukraine            | idiosyncratic cases          |

No languages with definite articles available where proper names are only anarthrous

Why non-bare proper names? A diacritic feature **or a phi-feature?**

Matushansky 2015: in **German**, proper names are bare iff they are not **specified for number or gender features**

Near-ideal **double dissociation: all non-neuter toponyms require an overt definite article and there is no toponym without the definite article that is not neuter:**

- |     |    |                             |           |
|-----|----|-----------------------------|-----------|
| (5) | a. | der Irak, der Jemen...      | masculine |
|     | b. | die Schweiz, die Türkei...  | feminine  |
|     | c. | die USA, die Niederlande... | plural    |

---

**Acknowledgments:** Many thanks to the audiences at the *TIN-dag 2015* (February 7, 2015), Frankfurt University (April 30, 2015), *Proper Names Workshop* (CEU, Budapest, May 18-19, 2015), *Séminaire de LaGraM* (Paris 8, June 8, 2015), *Syntax-interface meetings* (Utrecht, September 28, 2015), and Workshop “Namengrammatik”, (Delmenhorst, March 17-18, 2016), where parts of this research were presented.

(6) München, Berlin, Frankfurt am Main... no gender (neuter)

(7) (der) Hans, (die) Maria... semantic gender always available

One may hope that there is always a formal system behind language-internal variability.

## 2. PUZZLE: FRENCH PROPER NAMES

Proper names of humans are unexceptional, but toponyms vary.

**Cities vs. countries:**

- (8) a. Paris, Nice, Londres  
b. la France, le Canada, l’Egypte...

Core distinction: locative contexts vs. all others

**Table 1: The locative-argument divide**

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

Puzzles to be addressed:

- city names with an apparent definite article: its synchronic status
- the city/country divide: the presence of the definite article in argument positions
- locative (and especially **ablative**) **proper names**: what happens to the article with feminine countries (and countries beginning with a vowel)?
- the gender factor: why does it happen where it happens?

Further issues:

- exceptions to the main rule: bare countries and definite cities
- 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-features as the main factor determining overttness of the definite article**

Conclusion: definite articles do not realize just definiteness

## 3. CITY NAMES, BARE AND DEFINITE

Received wisdom: three determination options for city names:

- (9) a. (à/de) Paris, Londres, New York  
b. (à/de) Le Havre, Le Mans (variation: *au Havre, au Mans*)  
c. (à/de) La Rochelle, (à/de) La Nouvelle-Orléans

What is the difference between bare and definite city names?

What is the source of the variation in (9b)?

### 3.1. The gender of city names

Hasselrot 1943/1944, Edwardsson 1968: complex gender agreement patterns, high degree of confusion and optionality:

- (10) **Fondée** en 1869, Kemi est **devenu** un centre industriel important. E271  
founded.FSG in 1869, Kemi is become.MSG a center industrial important  
*Founded in 1869, Kemi became an important industrial center.*

Masculine agreement only (Hasselrot 1943/1944:215, Edwardsson 1968) for: *tout* 'all', *vieux* 'old' (in the sense of the older part of the city), *petit* 'small' and *grand* 'big'.

Feminine agreement only (Edwardsson 1968) for: definite postnominal epithet:

- (11) Tsaritsine **la blanche** est devenue "rouge". E285  
Tsaritsine the.FSG white.FSG is become.FSG red  
*Tsaritsyn the White has become red.*

Otherwise a lot of variation

Hasselrot 1943/1944: city names have no inherent gender, and have had no gender for quite a while (earliest stages of French examined, but no proper corpus analysis)

Edwardsson 1968: feminine agreement generally preferred (no syntactic analysis):

- (12) a. Predicate past participle (including depictives and appositives):  
Grand Larousse Encyclopédique (661): masculine 130 (20%), feminine 531 (80%)  
Grand Larousse (2158): masculine 885 (41%), feminine 1273 (59%)  
b. Predicate adjectives (including those in apposition):  
No corpus indicated (90): masculine 18 (20%), feminine 72 (80%).  
c. Nonrestrictive adjectives (*épithètes de nature*):  
No corpus indicated (44): masculine 13 (30%), feminine 31 (70%)  
d. Indefinite article (Hasselrot 1943/1944: masculine preferred)  
No corpus indicated (90): masculine 60 (66%), feminine 30 (33%)

Hasselrot 1943/1944 indicates that for villages (masculine in French), masculine is preferred

Even proper names with an apparent internal definite article may show gender disagreement:

- (13) a. **le** Caire, nostalgique et **orgueilleuse** E279  
the.MSG Cairo, nostalgic and proud.FSG  
*Cairo, nostalgic and proud*  
b. **Le** La Rochelle de mon enfance a disparu. Internet  
the.MSG the.FSG Rochelle of my childhood has disappeared  
*The La Rochelle of my childhood has disappeared.*

If *Le Caire* and *La Rochelle* have the gender indicated by their apparent definite articles, how can they trigger "wrong agreement"?

### 3.2. Is there a definite article in city names?

It is far from clear that what looks like the definite article in (13) is an article (cf. also (9b)):

- (14) [X] a bon espoir de faire de sa ville **la** La Rochelle de la droite.  
has good hope of do.INF of his city the.FSG the.FSG Rochelle of the right  
[X] *has a good hope of making out of his city the La Rochelle of the right.*

In the gender agreeing Internet example (14) the definite article is doubled, showing that it is a morphological part of the proper name

NB: For the native speaker allowing the pattern (14) that I have consulted postnominal modification of *Le Havre* is ineffable, while the native speaker with the pattern in (13b) removes the inner article in *le vieux* (\**Le*) *Havre* yet totally blocks \**la belle Rochelle*.

Hypothesis: in the current stage of French the apparent definite article inside a proper name is a morphological part of that proper name. Expecting correlation between doubling (13b) and the lack of liaison in the masculine (9b).

Remaining issue: inhabitants (e.g., *rochellois*)

Lomholt 1983:287-288: *la/le* in city names is not doubled with modifiers, contracts with the preposition for masculine and should be removed with other determiners (yet cites *du Vieux La Rochelle* and has mostly irrelevant examples)

### If city names have no inherent gender, agreement with them is purely semantic

The fluctuating agreement pattern results from the tension between two influencing factors:

- masculine is the default gender in French
- the corresponding sortal (*ville* 'city') is feminine

Support for the **influence of the sortal**: villages and planets (*Jupiter* is feminine (albeit with some difficulty, the gender of planets is also underdetermined) because the word *planète* is)

Edwardsson 1968: the final segment (-*e* or -*es*) facilitates the feminine

The situation is very different for rivers (which require the definite article in French like they do in English):

- (15) a. Combien d'espèces de poissonstrouve-t-on dans la/\*le Seine de nos jours?  
how.many of.kinds of fish finds.one in DEF.F/M Seine of our days  
*How many species of fish can be found in the Seine of today?*
- b. Ensemble, nous parlerons du/ \*de la Rhône d'hier.  
together we speak.FUT.1PL of+DEF.M of DEF.F Rhone of+yesterday  
*Together we will speak of the Rhone of yesterday.*

The gender of the article determines agreement of the proper name, the article contracts with the preposition and does not double.

### 3.3. Summary

French city names are, contrary to appearances, **systematically anarthrous** (unless modified) and have no inherent gender

Prediction: contraction of the masculine definite article/lack of doubling should correlate with the gender for agreement (unverifiable for city names, competing grammars)

City names do not constitute an argument against the hypothesis that the presence of **an overt definite article on proper names is determined by inherent phi-feature specification!**

Topic for future research: the mechanism of agreement in the absence of inherent gender

## 4. COUNTRY NAMES, BARE AND DEFINITE

French country names require the definite article, irrespective of gender or number:

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

There are anarthrous country names, but they can be argued to be derived ones:

- (17) a. Bahreïn, Chypre, Cuba, Guam, Haïti... islands (not all)  
 b. Djibouti, Gibraltar, Hong Kong... other landmarks (including cities)  
 c. Israël individual exceptions

The gender of non-bare country names is stable  $\Rightarrow$  **inherent gender or number requires the presence of the definite article**

Two environments where the article vanishes (though **not with masculine consonant-initial or plural country names**):

- relational *de* (e.g., *les vins de (\*la) France*, but *du Pérou*)
- locative *de* (e.g., *venir de (\*la) France*, but *du Pérou*)

What is the formal difference between definite country names in (16) and anarthrous country names in (17)?

## 5. THE SYNTAX AND SEMANTICS OF LOCATIVE CASES

It turns out that locative case (and its derivatives) frequently have restricted distribution:

- locative case 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)
- 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)
- 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)
- case paradigm for toponyms and certain common nouns restricted to locative case (Bagvalal: Daniel and Ganenkov 2009, Diyari: Austin 2013:52)

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

Locative case is generally not compatible with human or animate nouns (which sort of makes sense)

### 5.1. Latin locative cases: overview

Latin: names of towns, cities, small islands and a few common nouns including *domus/domi* 'home', *rus/ruri* 'countryside' and *humus/humi* 'ground' have locative case.

Non-restrictive modification does not remove the ability to function as a locative (*meae domi* 'at my home' (Plautus, *Aulularia* 432 via Calabrese 2008); *proximae viciniae habitat* 's/he lives nearby' (Plautus, *Bacchidae* 2, 2, 27)), yet restrictive modification blocks the bare locative (Donaldson 1860:314).

- (18) a. *iacēre humi* Gildersleeve and Lodge 1876  
 lie.INF ground.LOC  
*to lie on the ground*
- b. *Militēs Albae cōstitērunt in urbe opportūnā.*  
 soldiers Alba.LOC halted in city.ABL convenient.ABL  
*The soldiers halted at Alba, a conveniently situated town.*

Exactly the same set of lexical items uses **accusative** case-marking for **allative** and **ablative** case-marking without a preposition **for the source**.

Directional accusative also surfaces with prepositions:

- (19) a. Multos annos Gallia **sub imperio Romano** fuit. locative  
many years Gaul under rule.LOC Roman.LOC be.PRET  
*For many years Gaul was under Roman rule.*
- b. **Sub imperium Romanum** Gallia cecidit. directional  
under rule.ACC Roman.ACC Gaul fall.PRET  
*Gaul fell under the Roman rule.*

Reasonable assumption: the accusative of direction results from the presence of an additional functional head for both bare and prepositional locatives

Bierwisch 1988, Zwarts 2005, 2006, den Dikken 2006, Lestrade 2006, 2010, etc.: same case superposition in other languages

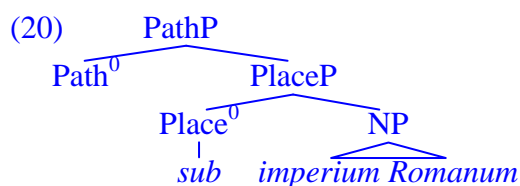
## 5.2. Paths and places (loci)

Bierwisch 1988, Koopman 2000, Tungseth 2003, Zwarts 2005, among others: directional PPs are more complex (semantically and syntactically)

Bierwisch 1988: directional prepositions are specified [+ dir]

Koopman 2000: for directional interpretation, a locative PP must be contained in the functional projection PathP

Zwarts 2005: directional PPs contain a Path function, in addition to the location



The Path component is the same in PPs and locative NPs

Core intuition: **prepositions are functions from entities to loci** (however defined). But if an NP already denotes a locus, the preposition is not necessary for either the locative function or any of its derivatives

It is also possible to hypothesize that locative is no different from the directional and ablative cases in that all three assigned by the corresponding null prepositions, but this approach has no explanatory power: the puzzle remains of why these specific lexical items can while others cannot combine with the relevant null preposition and get the appropriate case-marking. To say nothing of inelegance.

## 5.3. Explanatory power

Assuming that some NPs can denote loci explains the cross-linguistic restrictions on locative cases on the assumption that **locative cases can be interpretable or uninterpretable**:

- i. locative case restricted to toponyms and some common nouns: only these denote loci (see above)
- ii. locative case-marking optional or absent for toponyms and some common nouns: only these denote loci; for all others locative case-marking is either interpretable or indicates the presence of the corresponding null locative preposition
- iii. special locative case forms for toponyms and some common nouns: these denote loci and can combine with uninterpretable locative case as in (i). For all others the chosen locative case form is interpretable as in (ii).
- iv. case paradigm for toponyms and some common nouns restricted to locative case: these denote loci only and do not have entity denotations (the morpheme for the relevant type shift is not available)

Bonus: the appositive genitive (*the city of New York*) explained, diachronically at least

Extra bonus: bare NP-adverbials (e.g., *Monday*) may be temporal loci

## 5.4. Summary

As our ontology at any rate requires loci, it is unsurprising that there should be terminals that denote loci

Two points of variation:

- whether a language has locus-denoting nouns at all
- whether a given locative case (form) is interpretable

Now back to French

## 6. FRENCH IS THE NEW LATIN

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

Stolz, Lestrade and Stolz 2014:ch.4.1: bare street names used as locations:

- (21) 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.*

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

- (22) ... a loué une chambre à l' Hôtel des Cinq Continents avenue de la République  
has rented 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:

- (23) Quand le taxi s'arrêta 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

Novel proposal: **French does have overt locative case-marking**

### 6.1. The locative portmanteau

In locative/directional 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

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 some time expressions: *en hiver/au printemps*

- (24) 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*+DEF cannot be replaced by *en* if *à* is not locative:

(25) *lié à la France/\*en France*

Miller et al. 1997: The same is true for the locative and relational *de*:

- |         |                                    |                 |
|---------|------------------------------------|-----------------|
| (26) 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 cannot disappear if the preposition is not locative:

(27) *discuter de \*(la) France*

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

## 6.2. 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*, a portmanteau morpheme realizing two syntactic positions, P<sub>LOC</sub>+FSG; a special rule of referral, replacing the masculine form with the feminine one, is activated for proper names beginning with a vowel. However:

- elsewhere, elision has priority over contraction (*à l'* > *au*); with possessives and definite articles feminine is replaced with masculine (*mon amie*, *l'amie*)
- lexical exceptions (Grevisse: *Danemark/Portugal/Luxembourg*; *en Limousin*)

Miller 1992, Miller et al. 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?

## 6.3. Extending the picture

Two major factors:

- The distribution of *au/en* **changes in a different lexical-semantic class**
- The alternation between *au/en* and the **prepositional variant *dans*+DEF**

Miller et al. 1997: "Grevisse (1980:627ff.) notes that *Danemark*, *Portugal* and *Luxembourg* (consonant-initial masculines) take either *en* or *au*, and *de* or *du* (*Il est alle en/au Danemark* 'He went to Denmark'; *Il revient de/du Danemark* 'He came back from Denmark'). [...] Similarly, **names of old provinces can be preceded by *en*, even when they are consonant-initial masculines: *en Limousin*, *en Berry*, etc.**" (cf. Grevisse 2006:1506-1507)

With **restrictive modification** *dans* is used (cf. Lomholt 1983:126-135;145):

- |         |   |     |        |                |               |
|---------|---|-----|--------|----------------|---------------|
| (28) a. | dans l'Algérie/                               | la  | France | contemporaine/ | d'aujourd'hui |
|         | in DEF+Algeria                                | DEF | France | contemporary   | of+today      |
|         | <i>in contemporary/today's Algeria/France</i> |     |        |                |               |
| b.      | dans l'Iran/                                  | le  | Canada | contemporain/  | d'aujourd'hui |
|         | in DEF+Iran                                   | DEF | Canada | contemporary   | of+today      |
|         | <i>in contemporary/today's Iran/Canada</i>    |     |        |                |               |



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

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 Lomholt 1983):

- (30) 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**

#### 6.4. 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 entity-denoting

##### 6.4.1. Argument positions

Locus-denoting toponyms cannot appear in regular argument positions with locus-denotation, hence a type-shifting rule is required transforming a locus into the entity that comprises this locus

The resulting entity-correlates have a number of properties:

- they are **conceptualized as three-dimensional entities** (rather than points; this is probably the default for inanimates), thus the corresponding locative preposition is *dans*
- they are **fully specified for phi-features** (hypothetically, as the result of the type-shifting operation), hence the obligatory definite article

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 *Haiti*)

##### 6.4.2. Modification

Straightforward explanation: modifiers (whether type  $\langle e, t \rangle$  or  $\langle \langle e, t \rangle, \langle e, t \rangle \rangle$ ) are incompatible with locus-denotation

The corresponding entity-correlate is used, which makes the portmanteau variant impossible, and the lexical preposition *dans* must be used

##### 6.4.3. Morphosyntax of the portmanteau variant

What is the portmanteau? Options:

- the portmanteau is the case-marked form of the proprial definite article (lots of languages mark case on the article only)
- the portmanteau is a case prefix of the proper name itself, i.e., such proper names are bare

In both cases, no need for "phrasal inflection"

Assuming that the presence of inherent phi-features on a proper name triggers the realization of the definite article, it must be concluded that **the portmanteau is the definite article with the appropriate case-marking**.

**Table 2: Prenominal allomorphy**

	possessive	article	demonstrative	adjective
feminine C	<i>ma, ta, sa</i>	<i>la</i>	<i>cette</i>	<i>belle, nouvelle, vieille</i> , etc.
V	<i>mon, ton, son</i>	<i>l'</i>	<i>cet</i>	<i>bel, nouvel, vieil</i> , etc.
masculine C	<i>mon, ton, son</i>	<i>le</i>	<i>ce</i>	<i>beau, nouveau, vieux</i> , etc.

Assuming that the *en/au* alternation is a **liaison phenomenon**, the consonant-final variant is chosen for vowel-initial contexts

The locative article patterns with adjectives and demonstratives (the liaison form is identical to the feminine)

NB: Caveat: liaison does not always give rise to a form identical to the feminine (Tranel 1990 et seq., etc.)

## 6.5. Summary

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

- an account of **the city/country divide**: cities are entities, while countries are loci
- a theory of **locative portmanteaus**: they are definite articles marked for locative or ablative 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 entities, which makes locative cases impossible

The unified theory attributing the emergence of the definite article with proper names to their inherent phi-feature specification can be maintained

Still in the need of explanation:

- bare country names (most likely, not locus-denoting or not denoting loci only)
- bare street names (one possibility is, indeclinable locus-denoting noun phrases)

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

## 7. LEXICAL-SEMANTIC CLASSES OF LOCUS-DENOTING TOPONYMS

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

### 7.1. 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

Core proposal: 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

### 7.1.1. 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)

#### (31) **French provinces and regions**

- |    |  |                           |
|----|--|---------------------------|
| a. | <i>en</i> Picardie, <i>en</i> Normandie...                                 | feminine provinces        |
| b. | % <i>en/dans</i> le/*au Béarn, Poitou...                                   | masculine provinces       |
| c. | <i>en</i> /* <i>dans</i> l’Aquitaine, <i>en</i> /* <i>dans</i> l’Alsace... | feminine/vowel provinces  |
| d. | <i>en/dans</i> l’Artois, <i>en/dans</i> 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

### 7.1.2. French department names

**French department** names, irrespective of gender or initial segment, combine with *dans*, but *en* is possible for (Grevisse 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 is a standard description, but all of these cases are **feminine or begin with a vowel**

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)

### 7.1.3. 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 (no match found yet):

- |         |  |                |
|---------|--|----------------|
| (32) a. | <i>la</i> Corse, <i>la</i> Sicile, <i>la</i> Tasmanie... | feminine       |
| b.      | <i>le</i> Groenland, <i>le</i> Spitzberg                 | masculine      |
| c.      | <i>l’</i> Irlande, <i>l’</i> Islande...                  | feminine/vowel |
| (33) a. | <i>en</i> Corse, <i>en</i> Sicile, <i>en</i> Tasmanie    | feminine       |
| b.      | <i>au</i> Groenland, <i>au</i> Spitzberg                 | masculine      |
| c.      | <i>en</i> Irlande, <i>en</i> 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:

- (34) 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*):

- (35) a. *le Svalbard (au/du), le Vanuatu (au/du)*  
 b. *Madère (à/de), Zanzibar (à/de)*  
 c. *le Dodécanèse (dans le/du, just like the island of l'Etac de Sercq 'Little Sark')*  
 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 empirically 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*

#### 7.1.4. Summary and 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:

- (36) 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 repêché 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*.

**Table 3: Directional/locative summary**

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

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

## 7.2. Puzzle

First impression: locative encoding in function of the lexical-semantic class

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 even when denoting a country), the definite archipelago of *le Dodécanèse* (taking *dans le/du*)...)

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

Factors of variation:

- denotation (location 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

## 7.3. Phi-feature specification and the realization of locative cases

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

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 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

**Table 4: Locative/directional variation restatement**

	feminine	masculine V	masculine C
<b>French province (old)</b> <i>month</i>	<i>en</i>	<i>en</i>	<i>en</i>
<i>country</i> <b>US state (1)</b> <i>season</i>	<i>en</i>	<i>en</i>	<i>au</i>
<b>US state (2)</b> <i>département</i> <b>French province (new)</b>	<i>en</i>	<i>en</i>	<i>dans le</i>

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)

The rest of the observed picture:

- cities and city-like islands denote entities and combine with *à*
- French departments may also denote entities and combine with *dans* (because perceived as containers, unlike cities which are perceived as points)
- all plural toponyms combine with *aux*

### 7.3.1. Vocabulary Insertion rules and phi-feature specifications

Assuming that the realization of locative cases for masculine vowel-initial toponyms can be taken care of by whichever mechanism is normally responsible for liaison:

(37) in the context of the features [definite] and [locative]:

[pl] → aux  
[gender] → en  
[number] → au

**Plural cannot be unspecified**

**Feminine gender cannot be unspecified**, which means that all feminine loci yield *en*

The variation in the realization of locative cases with masculine consonant-initial loci is due to whether they are specified for number and gender (as loci; type-shifting to entity yields full phi-specification):

- (38) a. months: gender feature present as unvalued, [masculine] or [-feminine] → *en*  
b. countries, seasons: gender feature absent, number feature unvalued, [singular] or [-plural] → *au*  
c. French departments: no gender or number features

Crucially, **there is no Elsewhere condition**.

In the absence of number and gender features there is **no default allomorph** that could be used. **Ineffability results**, except for vowel-initial proper names (the third row in the table above)

### 7.3.2. Variant specification for some lexical-semantic classes

Variant behavior of **French provinces**: surface-masculine can correspond to the total lack of phi-features or to a present gender feature

Variant behavior of **US states** (Canadian provinces, etc.): surface-masculine can correspond to the total lack of phi-features or to a present number feature

Variant behavior of **French departments**: this lexical-semantic class can be conceptualized as entity-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

There is no reason to postulate country-like and city-like islands: either they are locus-denoting (and specified for number) or they are entity-denoting (and unspecified for phi-features)

Necessary stipulation: **type-shifting to entity yields full phi-specification**

## 7.4. **Bare street names**

It seems inconceivable that proper names containing a sortal should lack phi-features (and the article is in fact present in argument positions)

Zero realization is usually the Elsewhere clause -- but we don't want an Elsewhere clause for the locative case

Two syntactic options:

- m-merger (Matushansky 2006) of the sortal with the locative article
- N-to-D movement (Longobardi 1994)

Problem: *Grande rue (Saint-Jacques)*, unless perceived as unanalyzable (cf. *Grand'rue*)

Alternative: such proper names have no gender despite the presence of a sortal. Their gender is computed with the shift to the entity-denotation

The status of a sortal forming part of a proper name remains unclear even for semantic purposes (and incidentally, **titles also lack the definite article when appended to proper names**)

## 8. PROPER NAMES OF HUMANS

No phi-specification: the gender of [animate] proper names is purely semantic (determined by the gender of the referent); the attribution is determined by convention

## 9. CONCLUSION

The correlation between inherent gender and the presence of the definite article on a proper name seems to be valid for French as well

- country names and their ilk are systematically non-bare and have gender
- city names are bare and lack gender
- the article-like segment in apparently non-bare city names does not behave as a definite article should with respect to modification and contraction

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**

French provides further evidence for treating **locus denotation as a basic type** rather than a result of type-shifting: locus-denoting toponyms are less specified formally than their entity-denoting counterparts

The syntax of locus-denoting proper names provides **evidence for syntactic realization of semantic type-shifts** (i.e., KP, etc.):

- full phi-specification after shifting from loci to entities
- languages without this type-shift (Bagvalal, Diyari – see above)

A proper formal account therefore requires **a real theory of phi-feature specification** (which is independently required for other reasons -- semantic agreement, among others)

The realization of the definite article with proper names seems to be governed by morpho-syntactic factors (phi-specification) rather than by semantic factors

Hypothesis for English: the presence of the definite article indicates a specified [number]

Another phi-feature that might be relevant for the definite article realization is animacy, yet to account for the languages we are dealing with the marked value would have to be [inanimate], which would be suspicious

The syntax of the definite articles with proper names clearly indicates no correlation between actual definiteness (all proper names are definite in argument positions) and the presence of an overt definite article

Can the definite article be treated as the realization of features present on an NP ("phrasal inflection") with no semantics whatsoever?

(39) *And then Elvis shows up. THE Elvis. Elvis Presley.*

The only justification for the presence of the definite article here is focus, but to the best of our knowledge focus operates on semantic content.

This consideration also excludes Longobardi's "expletive definite articles"



## 10. APPENDIX: ITALIAN PROPER NAMES

The realization of definite articles and spatial prepositions with Italian toponyms is not that different from French

The definite article is obligatorily absent for (bare) city names and obligatorily present for all other toponyms (Proudfoot and Cardo 2002:15-16) in argument positions:

- (40) a. Firenze 'Florence', Londra 'London'  
b. le Alpi 'the.PL Alps', il Tamigi 'the.MSG Thames', la Italia 'the.FSG Italy'

As in other languages, restrictive modification triggers the presence of the definite article:

- (41) la Firenze del Settecento  
the.F Florence of.the eighteenth.century

As in French, the definite article may fail to appear on the surface in locative uses, although in Italian the effect is limited to the locative/directional prepositions *a* and *in* 'in'

The difference between cities and regions translates into the choice of a preposition: *a* vs. *in* (both translating into the same prepositional variant with modification)

- (42) a. a Roma 'in/to Rome' vs. nella Roma imperiale 'in/to Imperial Rome'  
b. in Italia 'in/to Italy' vs. nell'Italia meridionale 'in/to southern Italy'

The locative + definite combination is realized as *in* with feminine region names, unless they are restrictively modified:

- (43) a. in/\*nella Italia 'in/to Italy'  
b. nell'Italia meridionale 'in/to southern Italy'

With bare masculine regions both variants are allowed. With plurals, only the composite one:

- (44) a. in/nel Veneto/Lazio masculine  
b. nei/\*in Paesi Bassi, nelle Marche plural

Again, restrictive modification makes *in* impossible

Syntactically complex country names pattern with restrictively modified country names in allowing the preposition-determiner combination *nel/nella*, but *in* is also sometimes possible:

- (45) a. in/nella Nuova Guinea, in/nella Unione Sovietica, in/nella Corea del sud feminine  
b. \*in/nella Guinea Equatoriale, \*in/nella Guyana Francese

- (46) nel/?in Timor Oriental, nel/\*in Regno Unito, nel/\*in Dakota del sud masculine

Islands may be feminine and bare (*Rodi*, *Miconos*, *Cipro*), feminine and definite (*la Corsica*, *la Sardegna*) or masculine and definite (one example: *il Madagascar*). There also seems to be a correlation with the realization of the locative preposition (*in Sardegna*, *a Rodi*)

So far, same picture as in French

For proper names of humans overt definite articles turn out to correlate with gender:

- in many (northern?) dialects of Italian feminine given names are preceded by the definite article; masculine given names appear with the definite article in a subset of these dialects
- in modern standard Italian women's surnames used in argument positions require the definite article (Longobardi 1994)

The presence of gender features on the proper names of humans is optional since always correlated with reference

When specified, as in dialectal German

## 11. BIBLIOGRAPHY

- Amiot, Dany, and Walter De Mulder. 2011. L'insoutenable légèreté de la préposition *en*. *Studii de lingvistică* 1:9-28.
- 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.
- Bierwisch, Manfred. 1988. On the grammar of local prepositions. In *Syntax, Semantik und Lexikon*, ed. by Manfred Bierwisch, Wolfgang Motsch and Ilse Zimmermann. *Studia Grammatica XXIX*, 1-65. Berlin: Akademie.
- Borer, Hagit. 2005. *In Name Only*. Structuring Sense 1. Oxford: Oxford University Press.
- Calabrese, Andrea. 2008. On absolute and contextual syncretism. Remarks on the structure of paradigms and on how to derive it. In *The Bases of Inflectional Identity*, ed. by Andrew Nevins and Asaf Bachrach, 156-205. Oxford: Oxford University Press.
- Cornulier, Benoit de. 1972. A peeking rule in French. *Linguistic Inquiry* 3:226-227.
- Creissels, Denis. 2009. Spatial cases. In *The Oxford Handbook of Case*, ed. by Andrej Malchukov and Andrew Spencer, 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, 668-685. Oxford: Oxford University Press.
- den Dikken, Marcel. 2006. On the functional structure of locative and directional PPs. Ms., CUNY Graduate Center.
- Donaldson, John William. 1860. *A Complete Latin Grammar for the Use of Students*. Cambridge: Deighton, Bell, & Co.
- Edwardsson, Roland. 1968. Le genre des noms de villes en français. *Studia Neophilologica* 40:265-316.
- Foley, William A. 1991. *The Yimas Language of Papua New Guinea*. Stanford, California: Stanford University Press.
- Gildersleeve, Basil L., and Gonzalez B. Lodge. 1876. *Latin grammar*. London: Macmillan Education Ltd.
- Grevisse, Maurice. 2006. *Le bon usage. XIII édition*. Paris-Gembloux: Duculot.
- Guekguezian, Peter Ara. 2011. Bare locatives in Western Armenian. Ms., USC.
- Guimier, Claude. 1978. *En et dans en français moderne: Etude sémantique et syntaxique*. *Revue des Langues Romanes* 83:277-306.
- Hasselrot, Bengt. 1943/1944. Le genre des noms de villes en français. *Studia Neophilologica* 16:201-223.
- 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.
- Katz, Eva. 2002. Systématique de la triade spatiale à, en, dans. *Travaux de linguistique* 44:35-49.
- Koopman, Hilda. 2000. Prepositions, postpositions, circumpositions, and particles. In *The Syntax of Specifiers and Heads*, ed. by Hilda Koopman, 204-260. London: Routledge.
- Lestrade, Sander. 2006. Adpositional case, MA thesis: Radboud University Nijmegen.
- Lestrade, Sander. 2010. Finnish case alternating adpositions: a corpus study. *Linguistics* 48:603-628.
- Lomholt, Jørgen. 1983. *Syntaxe des noms géographiques en français contemporain*. Copenhagen: Munksgaard.
- Longobardi, Giuseppe. 1994. Reference and proper names. *Linguistic Inquiry* 25:609-665.
- Matushansky, Ora. 2006. Head-movement in linguistic theory. *Linguistic Inquiry* 37:69-109.
- Matushansky, Ora. 2008. On the linguistic complexity of proper names. *Linguistics and Philosophy* 31:573-627.

- Matushansky, Ora. 2015. On being [feminine] and [proper]. In *NELS 45: Proceedings of the 45th Meeting of the North East Linguistic Society: Volume 2*, ed. by Thuy Bui and Deniz Özyıldız, 165-178. Amherst, Massachusetts: GLSA.
- 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* 33:67-90.
- Proudfoot, Anna, and Francesco Cardo. 2002. *Modern Italian Grammar: A Practical Guide*. London: Routledge.
- 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.
- Tranel, Bernard. 1990. On suppletion and French liaison. *Probus* 2:169-208.
- Tungseth, Mai. 2003. Two structural positions for locative and directional PPs in Norwegian motion constructions. In *Proceedings of the 19th Scandinavian Conference of Linguistics*, ed. by Anne Dahl, Kristine Bentzen and Peter Svenonius. *Nordlyd* 31.2, 473-487. Tromsø: University of Tromsø.
- Vikner, Carl. 1970. La syntaxe des noms d'îles en français moderne. *Revue Romane* 5:231-249.
- 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* 39:69-118.
- Zwarts, Joost. 2005. The case of prepositions: Government and compositionality in German PPs. Paper presented at *Israel Association for Theoretical Linguistics 21*, The Technion, Haifa, June 22-23, 2005.
- Zwarts, Joost. 2006. Case marking direction: The accusative in German PPs. Paper presented at *CLS 42*, Chicago, April 6-8, 2006.
- Zwicky, Arnold M. 1987. French prepositions: no peeking. *Phonology Yearbook* 4:211-227.