

# Exploring dynamic on-line gazetteers to map variation in the syntax of Italian urbanonyms

*Giuseppe Samo*

(Beijing Language and Culture University)

*Francesco-Alessio Ursini*

(Central China Normal University)

**Abstract:** The interaction between linguistics and geographical information systems becomes crucial when empirical evidence extracted from big data enriches the theoretical landscape of microvariation. In this paper, we explore dynamic on-line gazetteers (e.g., OpenStreetMap) to quantify the distribution of partitive/relational structures (e.g., the presence of *della* ‘of-the’ *Viale della Libertà* ‘Freedom avenue’) in Italian urbanonyms (e.g., Venetian *calle*, Genoan *crêuza*, Florentine *viuzzo*). We collect and discuss some relevant examples, and show that partitive structures often correlate with urbanonyms and generic terms of local (i.e., dialectal) origins. In passing, we also discuss other examples of complex urbanonym structures (e.g., coordinated generic terms). We conclude by suggesting that our results could be harvested by theoretical syntacticians mapping the syntactic cartography of the DP.

## 1. Introduction

The emergence of “big data” has become a cornerstone of research in geographic information systems in the last two decades or so (Hamzei et al. 2020). Geolinguistics, in particular, represents a sub-branch of linguistics in which big data research has become an invaluable source of evidence (Rabanus 2017). The interaction between linguistics, big data, and geographical information systems becomes crucial, for example, when researchers attempt to study two given linguistic phenomena attested in geographically disjoint areas (Di Nunzio & Poletto 2016). The intertwining of national and local varieties is also a well-documented and amply studied research path in Italian toponomastics and Geolinguistics (Granucci 2004). It is generally known that thousands of Italian *toponyms* (i.e., place names, Cresswell 2014: Ch. 1) have origins in local dialects or substrates (Granucci *ibid.*). Yet, there are aspects of this interaction between varieties that still seem mostly underexplored: the majority of works in

toponomastics tend to address morpho-syntactic matters only in passing (e.g., Marcato 2010–2011).

In this paper, we investigate the distribution and the syntax of generic terms in *urbanonyms*, toponyms that name places belonging to the urban landscape (Vannieuwenhuze 2007; David 2011; Köhnlein 2015; Stolz et al. 2017). We focus on Italian urbanonyms, on which only a few works have shed light so far, and only on their lexical/semantic properties (Ursini & Samo 2022a,b). Italian urbanonyms, like most toponyms, are built on a *generic term* (see section 2, e.g., *via* ‘street’ in *Via Garibaldi* ‘Garibaldi street’), being the head of the compound, followed by a specific term (*Garibaldi* in *via Garibaldi*). The same syntactic configuration is found in urbanonyms emerging from dialectal substrata such as Venetian *calle* ‘narrow street, alley’ in *Calle San Biasio* ‘Saint Basil’s alley’ (Calafiore 1975; Doria 1982; Rizzo 1983; Granucci 1988; Gasca Queirazza et al. 1990; Marcato 2009). At a preliminary glance, this sub-type of toponyms thus seems to follow the typical structure of phrasal compounds (e.g., Bisetto & Scalise 2009; Schlucker & Ackermann 2017).

More complex configurations are however represented by the so-called partitive/relational structures<sup>1</sup>, in which a preposition “mediates between generic and specific terms” (Ursini & Samo 2022a: 270), such as *della* ‘of-the’ in *Viale della Libertà* ‘avenue of-the Freedom’. The presence of the preposition is tendentially triggered by the second nominal entity (i.e., the specific term), but there are cases in which the same lexical material triggers two different structures. Compare for example, two street names having opposite dimensions with respect to the presence of the *di* ‘of’, such as Rome’s *Via di San Gregorio* ‘street of Saint Gregory’ and Milan’s *Via San Gregorio* ‘Saint Gregory street’. No relevant empirical studies of these nuanced patterns exist: therefore, a theoretical gap of what are the fine-grained morpho-syntactic properties of urbanonyms seems to arise.

Our goal is to map this kind of micro-variation in Italian urbanonyms and quantify the role of local generic terms (e.g., Venetian *calle*, Genoan *crêuza*, Florentine *viuzzo*) in triggering the presence of prepositions. We thus wish to fill the empirical and theoretical gaps involving urbanonyms in their role as possibly “local terms” for urban places. We achieve this goal by first offering a compact overview of this category (Section 2). We then discuss the results of an

---

<sup>1</sup> Within the literature on urbanonyms it is possible to find the label “partitive” to indicate the presence of a *d*-element preposition. Urbanonyms share only some properties of partitives (Stickey et al. 2013; Pinzin & Poletto 2022 and the recent special issue of *Studia Linguistica - volume 76* edited by Francesco Pinzin and Cecilia Poletto for an overview), pseudo-partitives (Alexiadou, Haegeman & Stavrou 2007) and possessives (Szabolcszi 1983, 1994; Cardinaletti 1998; Longobardi 2000; Si 2017). Nevertheless, these compound-like structures should be treated as complex nominals (Acquaviva 2019).

observational study on data retrieved from the large-scale online repository *OpenStreetMap* (henceforth, OSM: <https://www.openstreetmap.org/>; Section 3). In section 4, we summarize our results and discuss some relevant patterns. Section 5 concludes.

## 2. Urbanonyms and their syntax

Urbanonyms, such as *Piazza Italia* ‘Italy square’, are often the result of a merge of a generic term (e.g., *piazza*) and a specific term (*Italia*), at least when no prepositions mediate between the two terms. Specific terms act as referring labels for these places (Blair & Tent 2015, 2021), frequently having a commemorative function related to historical, social, and geographical factors (e.g., David 2011). They often have a historically dynamic status, since different contexts may lead to a new specific term. An Italian example, among many, could be the “renaming” of *Via Toledo* in *Via Roma* in the urban context of Naples (Doria 1967). On the other hand, generic terms classify the place types they name. In many toponyms (cf. Gasca Queirazza et al. 1990), generic terms may have century-old origins, possibly rooted in local varieties evolving autonomously (e.g., via phonological changes, Stolz, Levkovitch & Urdze 2017) from previous stages of a language. A quintessential example can be the generic term *rua* in Ascoli Piceno, which can be considered a virtually unchanged form of the Latin term (De Stefani 2004).

The Italian landscape is worth exploring because of historical and dialectal variability with respect to the generic terms attested in urbanonyms (Rizzo 1983; Pellegrini 1990, 1992; Granucci 2004). This variability has been explored in recent works, but often with a focus on topics other than grammar. For instance, Ursini & Samo (2022a) offered a classification and a lexicographic analysis of Italian urbanonyms based on data extraction from the *Pagine Gialle* directory. In this study, dedicated urbanonyms are isolated and discussed as encoding cultural and linguistic facets, especially in local varieties (e.g., *calle* in Venice, *crosta* in Genoa, *viuzzo* in Florence). They thus showed that generic terms represent linguistic data that can encode specific features of the spatial elements they denote.

Urbanonyms thus represent complex nominals: their syntax is similar to the one discussed for other proper names (Ursini 2017; Acquaviva 2019: 71). However, experimental studies (Ursini & Samo 2022b) found asymmetries with the latter types with respect to binding (anaphoric) properties. Furthermore, forms of ellipsis of the generic term are also allowed, although extremely rare in Romance. For instance, Bochnakowa (2006: 12) reports the French example of *Avenue des Champs-Élysées* ‘Elysian fields avenue’ becoming *les/aux* ‘the/at-the’

*Champs-Élysées* in the opportune contexts.<sup>2</sup> Not only is the omission related when the specific name involves another place name (e.g., French *champs* ‘fields’, but also *parco* ‘park’ or *giardini* ‘gardens’), but also in cases when the specific name denotes a commemorative proper name, especially in the usage adopted by locals. Some cases include the naturally occurring examples exacted from web-searches related to the Neapolitan place names of *Toledo* (referring *via Toledo/via Roma*), *Caracciolo* (denoting *via Caracciolo*) and *Garibaldi* (Piazza Garibaldi): *Spesso e volentieri dopo una passeggiata a Toledo mi fermavo con la mia ragazza a sorseggiare qualcosa di fresco* ‘After a walk at Toledo, I often stopped [there] with my girlfriend to drink something fresh’; *Mister allo prima cos a sfugliatell'e'mery sotto a galleria unberto. [sic] per smaltire una bella passeggiata a caracciolo* ‘Coach, so, first of all [you should try] Mary’s sfugliatella [a Neapolitan pastry] inside Umberto’s Gallery and then make a walk at Caracciolo’; *Nella prenotazione indicate se ci raggiungete a Garibaldi o a Villa dei Misteri* ‘Please, in the booking system please indicate whether you reach us at Garibaldi or at Villa dei Misteri’.<sup>3</sup> Thus, urbanonyms seem to display grammatical (i.e., morpho-syntactic) properties specific to this category; however, these properties appear still understudied.

An exception to this dearth of data is Ursini & Long (2020). This work suggested that in partitive/relational configurations, different morpho-syntactic elements can “intervene” between the generic and the specific term. This is usually a *d*-element, such as *della* ‘of-the’ in *Viale della Libertà*. Tendentially, *d*-elements emerge when the specific term is a family name denoting a group (e.g., *Via de’ Tornabuoni* ‘street of Tornabuonis’), a set of social activities (*Via della Fornace* ‘street of-the furnace’, *Piazza del Duomo* ‘square of-the cathedral’), or a common noun (*Via delle Rose* ‘roses’ street’). On the other hand, if the specific name is commemorative, there is a general tendency in Italian to avoid constructions with prepositions, such as in *Via Garibaldi* ‘Garibaldi’s street’ or *Piazza XX Settembre* ‘September the 20<sup>th</sup> square’.

However, it is not rare to find *d*-element constructions with “commemorative” specific names beyond the format in which they are reported in gazetteers or atlases (Cantile 2004,

<sup>2</sup> Naturally occurring examples with the two syntactic configurations can be found in the same text (an on-line sport article on professional cycling on the French newspaper *Ouest-France*, <https://www.ouest-france.fr/>), here reported: *Wout van Aert s'impose sur les Champs-Élysées* ‘Wout van Aert wins on the Champs-Élysées’ and *Wout Van Aert (Jumbo-Visma) s'est imposé en costaud ce dimanche sur l'avenue des Champs-Élysées* ‘Wout Van Aert (Jumbo-Visma) won, without rivals, this Sunday on the Champs-Élysées’ Avenue’ (short link, [https://t.ly/\\_Zmw](https://t.ly/_Zmw), last accessed 13.06.2022).

<sup>3</sup> The relevant examples are extracted from the following webpages: *Toledo* ([https://it.tripadvisor.ch/Restaurant\\_Review-g187785-d8665101-Reviews-Bistrot\\_Toledo-Naples\\_Province\\_of\\_Naples\\_Campania.html](https://it.tripadvisor.ch/Restaurant_Review-g187785-d8665101-Reviews-Bistrot_Toledo-Naples_Province_of_Naples_Campania.html), last access 23.09.2022); *Caracciolo* (<http://www.rafabenitez.com/web/in/forum/what-are-the-ten-essential-places/112/pag/56/>, last access 24.09.2022); *Garibaldi* (<https://iltaccodibacco.it/campania/eventi/166742.html>, last access 23.09.2022). We opted to leave the current graphemes adopted (e.g., *unberto* in Galleria Umberto).

2013). For example, although gazetteers and maps report *Via Garibaldi* for the city of Siena, naturally occurring examples in web-searches show that users can adopt partitive/relational structures such as *Nella Farmacia Fiore, in Via di Garibaldi, la passione per le spezie divenne anche passione per la ricerca e per la storia dei dolci senesi* ‘In Fiore pharmacy, in Street of Garibaldi, the passion for spices also became a passion for research and for the history of Senese pastry’.<sup>4</sup> Similarly we can find examples, as mentioned in section 1, in which the same urbanonym can appear via two slightly different realisations, such as Rome’s *Via di San Gregorio* ‘Street of Saint Gregory’ and Milan’s *Via San Gregorio* ‘Saint Gregory Street’. The presence or absence of *d*-elements seems thus subject to possibly several micro-factors of variation, which we aim to explore with a quantitative study in the remainder of the paper.

### 3. Materials & Methods

Our central conjecture in carrying out our study can be defined as follows. Local (i.e., dialectal) traces in generic terms of urbanonyms should be found within the administrative regions in which a dialect can act as a substrate of Italian (Berruto 2012). These local traces may represent the connection between generic term types and partitive/relational structures: *di* and possibly other elements tend to be minimized in “Italian” urbanonyms. Crucially, Italian geo-linguistic regions representing distributions of a dialect and administrative regions only loosely coincide (Lameli et al. 2010: Ch. 3; Rabanus 2017). For example, Neapolitan varieties may have Naples as a centre of propagation (Sornicola 1997), although they extend beyond the Campania region in which Naples is located (Ledgeway 2009: ch. 1–3). We first discuss which generic terms have been taken into consideration and then the methodology we adopted for reliably retrieving local forms.

Ursini & Samo (2022a) isolated Italian generic terms that are distributed in specific regions. The supplementary data of their paper represent the starting point of our data analysis.<sup>5</sup> The authors list the distribution of 218 generic terms out of 107,387 tokens across 91 gazetteers (each one encoding a specific Italian province) extracted from the Pagine Gialle website (<https://www.paginegialle.it>). In the supplementary files they split the data in five class intervals (A, B, C, D, E) in terms of percentage of the total of gazetteers in which generic terms were found. An item marked has been marked as [A] when a term is found in exactly all gazetteers,

---

<sup>4</sup> <https://www.fioresiena.it/dal-1827/>, last access 30.06.2022.

<sup>5</sup> <https://www.tandfonline.com/doi/suppl/10.1080/13875868.2021.1954186?scroll=top> (last access, 29.06.2022)

[B] when a term occurs in at least the 80% of the gazetteers, [C] when a term is found within from 40% to 80% of the gazetteers, [D] when a term occurs from 10% to 40% of the gazetteers.

In this paper, we only analyze those generic terms which are given in less than <10% [E] of the gazetteers, since they can be automatically localized. If a generic term was only found in a province or in a set of neighboring areas, they would represent cases of a local generic term. For example, a target type would be the generic term *fondaco* ‘merchants’ storage room’ since is found in Naples, Salerno, and Campobasso’ gazetteers, and not, for example, *banchina* which is found in Ancona, Bari, and Cagliari. We explored the dynamic online gazetteer OSM.<sup>6</sup> OSM operates following the philosophy labelled as “Volunteered Geographic Information” (e.g., Keßler et al. 2009, Sui & Goodchild 2011). Users of the platform can spontaneously add and edit place names for places on maps. Eventual disagreements can be resolved via the platform’s discussion forums. When a critical number of users converge upon a specific input name, the place name is officially recorded. OSM thus represents grassroots, citizens’ bottom-up knowledge of citizens’ concerning places and their names. The dynamism is represented by that the fact that the dataset can be potentially growing or being edited/updated on a daily base. In a nutshell, such type of gazetteers offers us linguistic data which are “closer” to local knowledge.

Our dataset also serves as a comparison with previous studies. These open-source repository, differently from the *Pagine Gialle* gazetteers discussed in Ursini & Samo (2022a), also contains smaller municipalities, for a total of 340,381 tokens (Ursini & Samo 2022a: 276 operated on 107,387 tokens extracted from *Pagine Gialle* gazetteers).<sup>7</sup> We operate our counts on the all the available generic terms labelled as [E] and regional in Ursini & Samo (2022a), which are also present in our dataset extracted from OSM (38 types). Table 1 summarizes the lexical entries we quantitatively and qualitatively investigated.

GENERIC TERM	ENGLISH GLOSS	GAZETTEER
<i>alzaia</i>	‘towpath’	MI
<i>androne</i>	‘entrance’	GO, TN, TS
<i>archivolto</i>	‘aurch-vault’	GE, SV, SS
<i>atrio</i>	‘lobby’	BA, CZ, SR

<sup>6</sup> We retrieved our data querying the dedicated platform <https://overpass-turbo.eu/> (last access, 05.05.2022) by implementing the following code: `[out:csv ("name")][timeout:2500]; {{geocodeArea:Italy}}->.searchArea; (way["highway"]["name"])(area.searchArea); ); for (t["name"]) { make street name=_val; out;}. The output is then transformed in a .csv format, available as supplementary file (see Ursini & Samo 2023).`

<sup>7</sup> We manually tested the reliability of the data. Quantitatively, we compared the frequencies of local generic term of one gazetteer with the OSM data. For example, the Venice generic term *calle* occurs 1211 times in *Pagine Gialle* and 1200 times in OSM. The divergence is statistically non-significant (margin of error <1%).

<i>baglio</i>	‘court’	PA
<i>baluardo</i>	‘rampart’	NO
<i>calle (var.)</i>	‘alley’	VE (also MO, TN, TV)
<i>campiello (var.)</i>	‘little field’	VE (also TN)
<i>canto</i>	‘side, corner’	FI, PT
<i>chiasso (var.)</i>	‘alleys’ nexus’	AQ, FI, GR, LU, PI, RO, TE
<i>contra'</i>	‘quarter’	VI
<i>costa (var)</i>	‘cote’ (i.e. street coasting the walls of a city)	AQ, GR, FI, SI
<i>crosa (var)</i>	‘‘crêuza’ (alley going to the centre of a quarter, in Genoa)	GE (also AL)
<i>cupa</i>	‘country road’	NA, AV
<i>fondaco</i>	‘merchants's storage room’	NA, SA, CB
<i>fondamento/a</i>	‘foundations’	VE
<i>fondo</i>	‘merchants's storage room’	PA, ME
<i>piaggia</i>	‘slopped terrain’	AR, MC, PE
<i>piscina</i>	‘pool’	VE
<i>ramo</i>	‘branch’	VE
<i>rampari</i>	‘steps’ (Old Italian)	FE, IS
<i>ratto</i>	‘steep street’	TS
<i>rio</i>	‘canal’	VE
<i>ronco</i>	‘blind alley’	SR
<i>ruga</i>	‘shop street’	VE
<i>salizada</i>	‘cleaning place’	VE
<i>scesa</i>	‘descent’	TA
<i>sortita</i>	‘exit’	LU
<i>sottoportego</i>	‘under-porch’	VE
<i>spianata</i>	‘esplanade’	GE, IM
<i>supportico</i>	‘under-porch’	AV, BA, CZ, NA
<i>tresanda</i>	‘alley’	BS
<i>vallone</i>	‘small wall’	NA, ME
<i>viella</i>	‘alley’	CE
<i>viuzzo</i>	‘alley’	FI

<i>volta</i>	‘vault’	FE, FI, PR, PT
<i>volto</i>	‘vault’	FE, PD, VR
<i>voltone</i>	‘big vault’	BO

Table 1: Local Generic term, English gloss and gazetteer indicated with the abbreviation of the cities.

We have decided to assimilate different elements in terms of their spelling, signaled with (var.) in Table 1. For example, ‘*crova* (var.)’ contains all the occurrences of *creuza*, *crêuza*, *crova* and *crovino*. Similarly, *costa* and *costarella* and the occurrences of *calle* and *callesela/e* have been investigated as a unique generic term, marked with ‘(var.)’ in Table 1. Dimensions of geo-linguistic variation can be found. In Table 1, indeed, we have two realizations of lexical entries such as *sottoportego* ‘underporch’ found in Venice and *supportico* ‘under-porch’ found in Southern regions. However, we only discuss these geo-linguistic forms of variation when they also play a role in syntactic matters, in Section 4.

We then compared our results on the distribution of partitive/relational of the generic term *via* ‘street’. *Via* is the most frequent generic term in Ursini & Samo (2022a), representing the 72% of the total of tokens. We found 194,327 occurrences of *via* in OSM, and only 10,877 (5.6%) were in a partitive/relational construction (e.g., *Via dei Vespucci* ‘Street of-the Vespuccis’). This datum serves as a baseline and a form of “control” group concerning the probability that the specific name would trigger the partitive/relational construction. We also manually investigated the data. We present the raw data in which the generic terms occur in a partitive/relational construction (henceforth P/R), followed by a specific term, or followed by another generic element with or without conjunctions. We also collect the nature of the functional element and the type of generic. Results and discussions are given in section 4.

#### 4. Results and Discussion

We operate our analysis on 4,201 tokens. Table 2 summarizes the results. The table reads as follows: the “Generic Term” column indicates the term, and the “Total” column the number of tokens. The “Generic + Specific Term” column indicates the lone presence of a specific term (e.g. *Vicolo Rossi* ‘Rossi alley’); the “P/R” column, of a partitive element (e.g., *Vicolo dei/di Rossi* ‘alley of Rossi’); the “Ø/e + Generic” (e.g., *Via Ø/e arco Rossi* ‘Rossi street and arch’, *Vicolo a/su via Rossi* ‘Street at/on Rossi street’) column refers to occurrences in which the generic term is followed by another generic term, in isolation (Ø) or preceded by a conjunction

(conj, e.g., *e* ‘and’). The rightmost columns report the type of the retrieved functional element (e.g., *d*-elements) and the type of generic term in a coordinated construction or introduced by a functional element.

GENERIC TERM	TOTAL	GENERIC + SPECIFIC TERM	GENERIC + P/R + SPECIFIC TERM	Ø/CONJ + GENERIC	TYPE OF FUNCTIONAL ELEMENT	TYPE OF GENERIC (Ø/CONJ + GENERIC)
<i>alzaia</i>	11	10	1	0	d-	-
<i>androne</i>	3	3	0	0	//	-
<i>archivolto</i>	26	13	13	0	d-	-
<i>atrio</i>	8	4	4	0	d-	-
<i>baglio</i>	4	4	0	0	//	-
<i>baluardo</i>	2	0	2	0	d-	-
<i>calle (var.)</i>	1203	611	587	5	a-, d-, va, adj	<i>campiello, quartier, (+e) ramo, salita</i>
<i>campiello (var.)</i>	158	75	82	1	d-	-
<i>canto</i>	12	4	8	0	d-, a-	-
<i>chiasso (var.)</i>	104	66	37	1	d-	<i>porta</i>
<i>contra'</i>	114	83	20	11	d-	<i>piazza, ponte, porta</i>
<i>costa (var.)</i>	105	57	47	1	d-	(+e) <i>arco</i>
<i>crosa (var.)</i>	17	4	13	0	d-	-
<i>cupa</i>	50	39	11	0	d-, adj	-
<i>fondaco</i>	9	0	9	0	d-	-
<i>fondamenta</i>	271	135	131	5	d-, su	<i>rio, rielo</i>
<i>fondo</i>	34	33	1	0	d-	-
<i>piaggia</i>	23	11	12	0	d-	-
<i>piscina</i>	9	7	2	0	d-	-

<i>ramo</i>	300	164	129	7	a, -d, va	<i>calle, campiello, ponte, salizada</i>
<i>rampari</i>	3	0	3	0	d-	-
<i>ratto</i>	3	0	2	1	d-	<i>canale</i>
<i>rio</i>	83	75	7	1	d-	<i>frazione</i>
<i>ronco</i>	988	942	46	0	a-	<i>corso, costa, via.</i> *30/46 <i>d</i> -element between two generic terms
<i>ruga</i>	20	8	12	0	d-, adj	-
<i>salizada</i>	38	26	12	0	d-	-
<i>scesa</i>	27	20	6	1	d-	<i>porta</i>
<i>sortita</i>	4	0	4	0	d-	-
<i>sotoportego</i>	228	85	141	2	d-	(+e) <i>calle</i> , (+e) <i>corte</i>
<i>spianata</i>	4	1	2	1	d-	<i>borgo</i>
<i>supportico</i>	16	7	8	1	d-, num	<i>fondo</i>
<i>tresanda</i>	6	2	4	0	d-	-
<i>vallone</i>	41	27	14	0	a-, d-	-
<i>viella</i>	15	12	3	0	d-, dietro	-
<i>viuzzo</i>	52	6	46	0	d-	-
<i>volta</i>	12	4	8	0	d-	-
<i>volto</i>	15	8	7	0	d-	-
<i>voltone</i>	3	1	2	0	d-	-

Table 2: Raw frequencies of local generic terms in the relevant syntactic configurations.

Focusing on tokens with more than 30 occurrences, we observe that only two generic terms found in Sicily, *fondo* (Palermo/Messina in Table 1; total tokens 34, P/R tokens 2.9%) and *ronco* (Syracuse in Table 1; total tokens 988, P/R tokens 46, 4.7%) show a distribution of P/R lower than the one the control group of *via* (5.6%). The group of Venice generic terms display forms of asymmetries, instead. The generic term *rio* ‘side canal’ (total tokens 83, P/R tokens 7, 8.4%) represents the lowest distribution of P/R tokens. On the other hand, the other generic terms related to Venice display similar distributions (~30 - ~60%: *calle* (var.) 1203 tokens,

P/R 48.8%; *campiello* (var.) 158 tokens, P/R 51.9%; *fondamento/a* 271 tokens, P/R 48.3%; *ramo* 300 tokens, 43%; *salizada* 38 tokens, 0, 31%, *sottoportego* 228 tokens, 61.8%; *ruga* 20 tokens; P/R 12 tokens; *piscina* 9 tokens, P/R 2 tokens). Thus, terms likely originating from the local dialects (or older, local varieties of Italian) seem to attract the presence of *d*-elements.

We also observed with local generic terms that similar lexical entries (e.g., in hagionyms, i.e., Saints' names) trigger two different syntactic configurations, such as the case of *Via di San Gregorio* in Rome and *Via San Gregorio* in Milan. Relevant examples are *Chiassetto di San Marco* 'Small alleys' nexus of Saint Mark' in Pisa and *Chiassetto San Sabino* 'Saint Sabino' small alleys' nexus' in Barete (L'Aquila); *Piaggia di San Francesco* 'slopped terrain of Saint Francis' in Assisi (Perugia) and *Piaggia Sant'Agostino* 'Saint Augustine slopped terrain' in Pollenza (Macerata). The cross-dialectal variation between *sottoportego* and *supportico* 'under-porch' cannot be compared because of the different sizes of tokens. However, out of the 8 occurrences (out of 16) for *supportico*, (i.e., the "Southern" variant in Table 1), five are instances or numerals (e.g., *supportico I di Via Indipendenza* 'first under-porch of Independence street'). Beyond numerals, other modifiers such as adjectives can appear (e.g., *larga* in *Calle Larga Giacinto Gallina* 'Wide alley Giacinto Gallina', *vecchia* 'old' in *Ruga vecchia San Giovanni* 'Saint John's old shop-street', both in Venice), also co-occurring with the *d*-element (e.g., *Calle Larga del Clero* 'wide alley of-the clergy' in Venice). Thus, the presence or absence of *d*-elements can also be linked to regional varieties, i.e., varieties between the "local" and the "national" level.

Not only *d*-elements, but also *a*-elements can be present if other generic terms such as *campo* 'field' and *ponte* 'bridge', appear in complex nominals, such as the Venice cases of *Calle al campo de l'arsenal* 'Alley to the Arsenal's field', *Calle al Ponte de l'Anzolo* 'Alley to Anzolo's bridge', or locative elements, such as *fianco* 'side' in *Calle a fianco la Chiesa* 'Alley to the side of the Church'. An interesting case is represented *Calle agli incurabili* 'Calle at-the incurables' in Venice, in which the generic term *ospitale* 'hospital' (*Ospitale degli incurabili*, Tassini 1863: 332) has been omitted and the determiner of the second nominal entity merges with the higher preposition. *A*-elements can be possibly preceded by *va*, originated from the verb 'to go' (*Calle va al forno* 'Calle goes to furnace' in Venice). *Va*-elements can precede spatial prepositions, such as *in* 'in' as in *Ramo va in campo* 'Branch goes into field'. Other spatial prepositions that can occur are *su* 'on' as given in *Fondamenta su la laguna* 'Foundations on the lagoon' in Venice or *dietro* 'behind' preceding a generic term for a human artifact as *croce* 'cross' in *Viella dietro croce* 'little street behind cross' in Gricignano d'Aversa (Caserta). Once more, Venetian but also Campanian data offer us important insights and

evidence on partitive/relational structures, and thus on *d-/a*-elements and other spatial prepositions.

The syntax of *d*-elements and *a*-elements can be explored in further detail, thanks to our wealth of data. For example, Sicilian *ronco* ‘blind alley’ displays an interesting asymmetry with the Venetian data in OSM database. Indeed, 29 out of 31 of occurrences of *ronco di* are followed by a generic term, such as in *ronco di via Damone* (*Damone’s Street blind alley*) in Syracuse. The numeral can precede the *d*-elements following this term, such as the case of *Ronco decimo di Via Roma* ‘Rome’s street tenth blind alley’ in Carlentini (Syracuse). In this case, we possibly have either a case of a possessive relation, since the blind alley represents branches “belonging” to the main road; or of a partitive relation, since every *ronco* is part (in terms of extension) of the main road.

Another interesting case is the one of *Ronco di via Maria Montessori* (alley of Maria Montessori Street) in Belvedere (Syracuse). OSM data reports it as *di*, while other on-line repositories, such as *inter alia*, Google maps (Google n.d.) reported it as *Ronco a via Maria Montessori* (alley at Via Maria Montessori). Finally, conjunctions connecting two generic terms can be found (8 occurrences out of 38, in Table 2), such as the case of *costa ed arco San Flaviano* ‘Saint Flaviano’s cote and arc’ in L’Aquila and *Calle e Corte de la Raffineria* ‘alley and court of the refinery’ in Venice. Though rare, these data suggest that generic terms display some properties of lexical items, one being the possibility of occurring in coordinated phrases. Along with their ability of combining with *d*- and other functional elements, this fact suggests that urbanonyms may form complex nominal structures with fine-grained array of constituting elements.

Before we conclude, we can observe that our dataset has some limitations: for instance, the diachronic dimension appears hard to explore. It is impossible to automatically detect when a specific syntactic construction was used and if there have been changes. In online gazetteers, this could be automatically partially achieved by investigating the specific term. An interesting datum is given by the urbanonym *Crosa Fabrizio de Andrè* ‘Fabrizio de Andrè alley’ [a 20<sup>th</sup> century Italian artist] in Savignone (Genoa), which represents one of the four occurrences of *crosta* (and its variants) followed by a specific term without a *d*-element. Future studies should improve the methodology adding, when possible, a diachronic dimension beyond the geographical one.

## 4. Conclusions

In this paper we quantitatively and qualitatively explored the linguistic data of local Italian urbanonyms extracted from an online dynamic gazetteer. We introduced a methodology that may permit researcher to understand the geographical distribution and quantitative properties. The empirical evidence from complex nominals such as urbanonyms can enrich the landscape on the cartography of the DP layer (Poletto 2015; Cinque 2022 and references therein), from a synchronic point of view in mapping microvariation, and, possibly, from a diachronic point of view. Specifically, how generic terms, specific terms, and other functional elements intertwine, in the spirit of Ursini (2017) and Acquaviva (2019). The use of *a*-elements and *d*-elements in Standard Italian, regional and local varieties, as well as a generalized use of *di* for other locatives could be further explored (see, e.g., Franco & Manzini 2017 and Manzini et al. 2017). However, we leave to future studies a detailed fine-grained analysis of the phenomena discussed in this contribution, which would involve native speakers of Italian and Italian varieties to confirm or enrich the results of our observational study.

## References

- Acquaviva, Paolo (2019). 'Two studies on the internal syntax of complex names' *Italian Journal of Linguistics* 31(2): 3-36.
- Alexiadou, Artemis, Liliane Haegeman, Melita Stavrou (2007). *Noun Phrase in the Generative Perspective*. Berlin: Mouton de Gruyter.
- Berruto, Gaetano (2012). *Sociolinguistica dell'Italiano Contemporaneo* (2nd ed). Roma: Carocci Editore.
- Blair, David, Jan Tent (2015). 'Feature terms for Australian toponymy'. *ANPS technical paper*. no. 2., New South Wales: Placenames Australia.
- Blair, David, Jan Tent (2021). 'A revised typology of Place Names'. *Names* 69(4): 1-15.
- Bochnakowa, Anna (2006). 'Parcourir la ville, en français et en polonais' in: J. Brzozowski, *Traduire la ville*, Krakow: Wydawnictwo Uniwersytetu Jagiellońskiego, pp. 11-17.
- Calafiore, Giovanni (1975). *Termini geografici dialettali in Italia*, Roma: Pubblicazioni dell'Istituto di Geografia dell'Università di Roma - Serie C (Miscellanea).
- Cantile, Andrea (2004). *Linee Guida per la Normalizzazione dei Nomi Geografici ad Uso degli Editori di Cartografia ed Altri Editori, 3za edizione*. Florence: Istituto Geografico Militare.
- Cantile, Andrea (2013). *Lineamenti di Storia della Cartografia italiana*. Geoweb.

- Cardinaletti, Anna (1998). 'On the deficient/strong Opposition in Possessive Systems' in: A. Alexiadou, C. Wilder, *Possessors, Predicates and Movement in the Determiner Phrase*, Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 17-54.
- Cinque, Guglielmo (2022). 'On Two Sub-projections of the Nominal Extended Projection: Some Romance Evidence' *Probus* <https://doi.org/10.1515/probus-2022-0005>
- Cresswell, Tim (2014). *Place: An introduction, second edition*. London: Wiley Blackwell.
- David, Jaroslav (2011). 'Commemorative place names – Their specificity and problems' *Names* 59(4): 214-224.
- De Stefani, Elwys (2004). 'La terminologia nell'onomastica' *Rivista italiana di onomastica* 10(1): 53-71.
- Di Nunzio, G. M., Cecilia Poletto (2016). 'A Love-Hate Relationship for Big Data and Linguistics: Present Issues and Future Possibilities' in P.Clough, P. Goodale, M. Agosti, S. Lawless, *Proceedings of the First International Workshop on Accessing Cultural Heritage at Scale co-located with Joint Conference on Digital Libraries 2016 (JCDL 2016)*, Newark: CEUR-ws.org, pp. 1-4.
- Doria, Gino (1967). *Via Toledo*. Cava dei Tirreni: Di Mauro.
- Doria, Gino (1982). *Le Strade di Napoli: Saggio di Onomastica Storica*. Naples: Grimaldi & Co. Editori.
- Franco, Ludovico and Maria R. Manzini (2017). 'Genitive/'of' arguments in DOM contexts' *Revue Roumaine de Linguistique* 62(4): 427-444.
- Gasca Queirazza, Giuliano, Carla Marcato, Giovan Battista Pellegrini, Giulia Petracco Sicardi, Alda Rossebastiano (eds.) (1990). *Dizionario di Toponomastica Italiana*. Turin: Utet.
- Google. (n.d.). [Ronco a Via Maria Montessori 96100 Belvedere SR, Italia]. Retrieved 13.06.2022, from <https://goo.gl/maps/Z6NCrXoSrFw8YbfD9>
- Granucci, Fiorenza (2004). 'Categorie toponomastiche ed uguaglianze linguistico-morfologiche'. in V. V. Aa. (ed), *Atlante dei tipi geografici*, Florence: Istituto Geografico Militare, 715-719.
- Hamzei, Ehsan, Stephan Winter, Martin Tomko (2020). 'Place facets: A systematic literature review' *Spatial Cognition and Computation* 20(1): 33-81. doi:33–81. 1080/13875868.2019.1688332.
- Keßler, Carsten, Krzysztof Janowicz, Mohamed Bishr (2009). 'An agenda for the next generation gazetteer: Geographic information contribution and retrieval' *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in Geographic Information Systems*, pp. 91-100.

- Köhnlein, Björn (2015). 'The morphological structure of complex place names: The case of Dutch.' *Journal of Comparative Germanic Linguistics* 18(1), 183-212. doi: 10.1007/s10828-015-9075-0
- Lameli, Alfred, Roland Kehrein, Stefan Rabanus (Eds.) (2010). *Language and Space. Vol. 2: Language Mapping*. Berlin/New York: De Gruyter Mouton.
- Ledgeway, Adam. (2009). *Grammatica diacronica del napoletano* (Beihefte zur Zeitschrift für romanische Philologie Band 350). Tübingen: Max Niemeyer Verlag.
- Longobardi, Giuseppe (2000). 'Postverbal Subjects and the Mapping Hypothesis' *Linguistic Inquiry* 31 (4): 691-702.
- Manzini, Maria R., Paolo Lorusso, Leonardo M. Savoia, L.M. (2017). 'a/bare finite complements in Southern Italian varieties: mono-clausal or bi-clausal syntax?' *Quaderni di Linguistica e Studi Orientali*, 3, 11-59.
- Marcato, Carla (2009). *Nomi di persona, nomi di luogo. Introduzione all'onomastica italiana*. Il Mulino.
- Marcato, Carla. (2010–2011). 'Toponimi' in R. Simone, *Enciclopedia dell'Italiano*. Roma: Istituto dell'Enciclopedia Italiana.  
[http://www.treccani.it/enciclopedia/toponimi\\_%28Enciclopedia-dell%27Italiano%29/](http://www.treccani.it/enciclopedia/toponimi_%28Enciclopedia-dell%27Italiano%29/)
- Pellegrini, Giovan Battista (1990). *Manuale di Toponomastica Italiana*. Milan: Hoepli.
- Pellegrini, Giovan Battista (ed) (1992). 'Toponomastica', in VV.AA., *Enciclopedia Italiana*, Appendice V SO-Z, Rome: Treccani, p. 515.
- Pinzin, Francesco, Cecilia Poletto (2022). 'Partitive Articles and Indefinites, Micro and Macrovariation' *Studia Linguistica* 76: 1-12.
- Poletto, Cecilia (2015). 'Word orders in the Old Italian DP' In U. Shlonsky, *Beyond Functional Sequence [The Cartography of Syntactic Structures]*, Oxford & New York: Oxford University Press, pp. 109-127.
- Rabanus, Stefan (2017). 'Dialect Maps' in C. Boberg, J. Nerbonne, D. Watt, *Handbook of Dialectology*, London: Wiley Blackwell, pp. 348-368.
- Rizzo, Tiziano (1983). *I Ponti di Venezia*. Newton Compton Editore.
- Romanova, Eugenia, Dmitry V. Spiridonov (2018). 'Syntactic features of proper names: The generativist approach(es) to Properhood', *Вопросы ономастики [Onomastics Questions]* 15(3): 7-35.
- Scalise, Sergio, Antonietta Bisetto (2009). 'The Classification of Compounds' in R. Lieber, P. Štekauer, *The Oxford Handbook of Compounding*, Oxford, Oxford University Press.

- Schlücker, Barbara, Tanja Ackermann (2017). 'The morphosyntax of proper names: An overview' *Folia Linguistica* 51(2). 309-339.
- Si, Fuzhen (2017). 'Two Types of Chinese Possessive Structures' in F. Si, *Studies on Syntactic Cartography*, Beijing: China Social Sciences Press, 197-217.
- Sornicola, Rosanna (1997). 'Campania' in M. Maiden, M. Parry, *The Dialects of Italy*, London: Routledge, pp. 330-338.
- Stickney, Helen, Chelsea Mafrika, Jordan Lippman (2013). 'Variation in the Syntax of the Partitive Structure' *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, 37. Retrieved from <https://escholarship.org/uc/item/6z58p2s1>
- Stolz, Thomas, Nataliya Levkovich, Aina Urdze (2017). 'When zero is just enough . . . in support of a special toponymic grammar in Maltese' *Folia Linguistica* 51(2): 453-482.
- Sui, Daniel Z., Michael F. Goodchild (2011). 'The convergence of GIS and social media: Challenges for GIScience' *International Journal of Geographical Information Science* 25(11): 1737-1748. doi: 10.1080/13658816.2011.604636
- Szabolczi, Anna (1983). 'The Possessor that Ran Away from Home' *The Linguistic Review* 3: 89-102
- Szabolczi, Anna (1994). 'The Noun Phrase' in F. Kiefer, K. É. Kiss, *The Syntactic Structure of Hungarian Syntax and Semantics*, New York: Academic Press, 179-274.
- Tassini, Giuseppe (1863). *Curiosità Veneziane ovvero Origini delle denominazioni stradali di Venezia*, Venice: Premiata Tipografia di Gio. Cecchini.
- Ursini, Francesco-Alessio (2017). 'On the Structure of Toponyms' in: L. Kortellessy, P. Stekauer, S. Valera, *Word-formation across Languages*, Cambridge: Cambridge Scholars Publishing, pp. 375-393.
- Ursini, Francesco-Alessio, Haiping Long (2020). 'Urbanonyms in Italian: Morpho-Semantic Properties and Distribution' *Names: A journal of onomastics* 55(1): 1-15. (<https://doi.org/10.1080/00277738.2020.1811049>).
- Ursini, Francesco-Alessio, Giuseppe Samo (2022a). 'Names for urban places and conceptual taxonomies: the view from Italian' *Spatial Cognition & Computation* 22 (3-4): 264-292.
- Ursini, Francesco-Alessio, Giuseppe Samo (2022b). 'The interpretation of urbanonyms in discourse: Reconciling theoretical accounts with experimental results' *Poznan Studies in Contemporary Linguistics* 58(2): 263-288.
- Ursini, Francesco-Alessio, Giuseppe Samo (2023). 'Extracting Toponyms from OpenStreetMap: A Cross-Linguistic Perspective', in: X. Hu, Y. Hu, B. Resch, J. Kersten,

K. Stock, *Proceedings of GeoExT 2023: First International Workshop on Geographic Information Extraction from Texts at ECIR 2023*, April 2, 2023, Dublin, Ireland.

Vannieuwenhuze, Bram (2007). 'The study and classification of Medieval urban toponymy: The case of late medieval Brussels (13th-16th centuries)' *Onoma* 42(1): 189-211.