



ARTICLE



<https://doi.org/10.1057/s41599-026-07178-z>

OPEN

The semantics of romance geonyms: cross-linguistic variation meets context sensitivity

Giuseppe Samo^{1,2}✉ & Francesco-Alessio Ursini³✉

The goal of this paper is to offer a model of the semantics of geonyms occurring in toponyms from Romance languages (e.g. Italian geonym *piazza* 'square' in toponym *Piazza Grande* 'Main Square'). The paper defines this model along the two dimensions of cross-linguistic and context-sensitive variation. Cross-linguistically, the paper focuses on five Romance languages: French, Italian, Portuguese, Romanian, and Spanish. For each language, data are collected involving toponyms and geonyms expressing the place concepts 'square', 'street' and 'alley' across different contexts of use. Building on these data, three key results are presented. First, it is shown that individuals use these geonyms to express wide arrays of contingent features that they ascribe to these concepts in general contexts (e.g. in conversations). However, uses become restricted to culturally, temporally (more) stable features forming place concepts in technical contexts (e.g. Wikipedia articles, Wikidata dictionary definitions). Second, it is shown that even if uses vary considerably, shared semantic features define language-specific prototypical senses that these place concepts express. Third, it is shown that cross-linguistic semantic similarity becomes quantifiable. For instance, French *place* has similar context-free uses to Italian *piazza* and Spanish *plaza*, since language-specific prototypical senses for these place concepts share most of their constituting semantic features. The paper concludes by discussing consequences for theories of toponyms' semantic properties across different disciplines.

¹Idiap Research Institute, Martigny, Switzerland. ²Department of Linguistics & Institute for Chomskyan studies, Beijing Language and Culture University, Beijing, China. ³School of Chinese Language and Literature, Central China Normal University, Wuhan, China. ✉email: giuseppe.samo@idiap.ch; samo@blcu.edu.cn; randorama@outlook.com

Introduction

Linguistic studies have displayed renewed interest in the grammatical (i.e., morphosyntactic) structure and semantic properties of toponyms (respectively Schlücker and Ackermann 2017, Stolz and Levkovich 2020, Romanova and Piridonov 2018; and Rezsegi 2012, Kijania-Placek 2021, Ursini and Long 2020). In a parallel line of research, psychologically informed *Geographic Information Science* (hereafter: GIS) studies assume that individuals can develop forms of emotional, social, cognitive attachment to geographical objects, i.e., mind-external *places*, based on their understanding of the properties of places (Cresswell 2014, Malpas 2018). Hence, individuals develop *place concepts*, i.e. mind-internal concepts for places.

A wealth of GIS works has analysed how individuals can develop place concepts that mentally represent places (Hu 2018, Purves et al. 2019, Hamzei et al. 2020), and how toponyms *qua* names for places can convey these place concepts. Within GIS and psychology, most studies consider toponyms as richly descriptive labels for language-general place concepts (Tenbrink 2019). However, within linguistics, this interest focuses on language-specific patterns. Therefore, a disconnect exists between the linguistic analysis of toponyms and the GIS/psychology analysis of the place concepts that toponyms express.

This disconnect emerges from the still ample empirical gaps across languages, cultures, and contexts. There exist taxonomic analyses on the semantic features¹ assigned to toponyms in Australian English (e.g., 'descriptive' or 'eponymous' features: Tent and Blair 2011). However, they gloss over the possible status of this analysis as capturing speakers' intuitions on toponyms' senses. GIS works have extracted data from social media and web-based corpora to explore the domains of semantic features (Zhang and Gelernter 2014, Zhou et al. 2014). These studies suggest that different communicative registers can affect the use of toponyms in conversations. For example, individual tourists can post reviews of the places they visit on FourSquare or Instagram. Crucially, these studies focus mainly on English-based toponyms and thus on place concepts typical of Anglophonic culture and communication (e.g., 'roadway'). Research efforts covering toponyms in other languages seem still scarce (cf. Li 2024, Choo 2015 for some exceptions).

Research on toponyms perhaps presents a linguistically and culturally Anglo-centric situation that seems to motivate several research questions (cf. Ursini and Samo 2022 for discussion). In this paper, however, we focus on three questions about cross-linguistic variation, due to the dearth of relevant data of this type. The first question is *what type* of semantic variation can be found when analysing the senses associated with toponyms and the place concepts they express. The second question is *how* different contexts and communicative purposes (e.g. daily conversations vs. official dictionaries) can modulate toponyms' place concepts, their use, and their constituting semantic features. The third question is *to what extent* toponyms across different contexts and languages may express similar place concepts. These questions have apparently not received substantial answers in previous literature from GIS, linguistics and toponomastics. Therefore, their respective answers could shed novel light on the semantic properties of toponyms.

The goal of this paper is to answer these questions by analysing the cross-linguistic expression of three place concepts that we approximate via the glosses 'square', 'street', and 'alley'. In English, the toponyms referring to these place concepts form the most frequent sub-set in texts (e.g. *Leicester Square*, *Alphabet Street*, *King's Alley*: Ursini and Samo 2025). We answer these questions by analysing these concepts in the five principal Romance languages: French, Italian, Portuguese, Romanian and Spanish. We choose these languages from this linguistic family for four theoretical and methodological reasons.

First, cross-linguistic semantic/lexical variation can sometimes involve sharp differences in how genetically unrelated and culturally distant languages express concepts (Koptjevskaja-Tamm et al. 2024). By choosing related languages from a single linguistic group, Romance, we delimit our research domain to commensurable forms of semantic similarity/dissimilarity. Second, these Romance languages may find their diachronic origins in the common Latin of the late Roman empire (Clackson 2016), but have followed distinct evolutionary paths over the centuries (Bossong 2016). We can thus study how these temporal and cultural factors might have converged to synchronically shape language-specific forms of the 'square', 'street', 'alley' place concepts. Third, these five languages act as official languages in their respective countries, although they co-exist with other official, regional Romance languages (e.g. Catalan in Spain, Sardinian in Italy).² Fourth, the sources we can find for these five languages are relatively comparable in size. Therefore, we can base our analysis on well-documented languages that also offer a wealth of toponomastic and linguistic resources for our inquiry.

We pursue this goal by reviewing previous work and motivating our study (Section "Literature review"). We then present our methodology and results (Sections "Methodology and Materials-Results"). Our methodology involves the extraction of data from three textual sources representing contextual uses of increasing precision. These are heterogeneous, extremely large-scale corpora of web data; Wikipedia articles; Wikidata items and lemmas offering official definitions for the target geonyms. We first provide the geographic distribution of these data from *OpenStreetMap* (OSM) (Ursini and Samo 2025). We then show that geonyms (and toponyms) in these languages express language-specific instances of the 'square', 'street' and 'alley' concepts. However, we individuate prototypical senses (i.e. combinations of features) instantiating these concepts across these languages and their contexts of use. We show that such prototypical senses display high degrees of semantic similarity, contextual and linguistic variation notwithstanding. We conclude by discussing the model's theoretical implications (Sections "Discussion-Conclusions").

Literature review

Our goal in this section is to provide an overview of the well-established grammatical properties of toponyms. We then outline the evidence supporting the content-rich, mind-internal analysis of place concepts and the senses individuating these contexts (Section "Grammatical and semantic properties of toponyms"). We establish how different contexts of use shape these senses, while also underlying that studies combining these insights focus mostly on English-based data (Section "The role of context in the use of toponyms"). We thus conclude showing that this dearth of data motivates our research questions, and by offering a set of predictions as answers to these questions **RQ1-RQ3** (Section "Target research questions and predicted answers").

Grammatical and semantic properties of toponyms. Research on the grammatical properties of toponyms has focused on widely spoken Asian and European languages. For instance, Li (2024) studies the grammatical and lexical properties of Chinese toponyms, and suggests that they are noun phrases of variable internal complexity (e.g. *Bei-Jing* lit. 'North-Capital', *Xiong-chu Gang Luo* lit. 'Bear-place Main street'). Similarly, Choo (2015) shows that Korean toponyms can include mono-morphemic (e.g. Seoul) and multi-morphemic toponyms (e.g. *Gang-nam Dong* lit. 'Southern-River District'). Similar patterns can be found among several European languages (e.g. German: *Berlin* and *Alexander-platz*,

Schlücker 2024; Dutch: *Ghent* and *Doleen-strasse*: Köhnlein 2015). The Romance languages in our sample confirm these observations (e.g. Italian, Samo and Ursini 2023; French, Morabito 2020; Portuguese, Carvalhinhos 2009; Romanian, Boamfă 2018; Spanish, Gordón Peral 2010). Examples include French *Paris*, Portuguese *Rua Central*, Italian *Via Delle Streghe*, among others.

Two grammatical generalisations emerge from this small sample of languages. First, toponyms may correspond to mono- or multi-morphemic proper names, thus distributing as “bare” N(oun)P(hrase)s. Second, when toponyms correspond to proper names also distributing as NPs, they usually include *geonyms*, classifying nouns that refer to place types (e.g. *rue* ‘street’ in French). Syntax-wise, they may precede or follow a specific name for a place (e.g. Köhnlein 2015, Samo and Ursini 2022). Liaison morphemes (e.g. -s ‘of’ in German *Donau-es-schagen*) and relational adpositions (e.g. *de* ‘of’ in Spanish *Calle de las Américas*) may mediate among these constituents. The Anglophonic literature employs the label “generic term” for this part of speech (cf. Tent and Blair 2011). However, “geonym” and translational equivalents occur in Romance research traditions (cf. Samo and Ursini 2024; Morabito 2020, Carvalhinhos 2009, Boamfă 2018, Gordón Peral 2010). We therefore use the Romance-oriented technical term, given its cross-linguistic appeal.

Semantic and conceptual generalisations, however, still seem a matter of controversy. Within toponomastics, philosophically oriented analyses have argued that toponyms lack any mind-internal content: they are “rigid designators” for mind-external places (Coates 2013, 2017). These analyses suggest that toponyms in discourse (e.g. the French *Place de la Concorde*) directly refer to mind-external places (the famous square in Paris). However, etymological and historical research has offered evidence that toponyms carry content that reflects how individuals may categorise places, even if this content may be synchronically opaque (Nyström 2016). Recent works have proposed semantic type taxonomies for toponyms in Australian English based on binary semantic features (e.g. ‘±descriptive’, ‘±eponymous’: Tent and Blair 2011, Blair and Tent 2015, 2021). Aside an extension to Japanese Sign Language toponyms (George 2023), this analysis remains restricted to Anglophonic data, and therefore neutral on whether these features define mind-internal concepts.

Recent empirical evidence on toponyms’ semantic content has however become increasingly relevant to the debate. Works in psycho- and neuro-linguistic research (e.g. Peng et al. 2020, Desai et al. 2023) and in cognitive linguistics (e.g. Rezsegi 2020a, b) suggest that individuals associate toponyms to places’ descriptions (e.g. cities like Rome) and place types (e.g. cities in general). These descriptions may consist of specific aspects or features that can describe a place’s physical properties (e.g. a city’s location in a country) but also an individual’s subjective conceptualisation of this place (e.g. a city being an individual’s hometown). Individuals can thus use toponyms to exchange information about their personal place concepts in discourse (De Stefani 2016). If individuals talk about *Place de la Concorde*, then they compare their personal place concepts or “views” about this famous Parisian place. Such conceptual comparisons, these linguistic works suggest, pass through the semantic features constituting place concepts.

Cognitively oriented GIS literature also plays a key role in this debate via two strands of research. In the first strand, studies propose models of place concepts by extracting information on places that individuals offer across online textual sources (e.g. corpora, geographic documents, gazetteers; Purves et al. 2019, Hamzei et al. 2020, Davies 2013, Ballatore 2016, Hamzei et al. 2021). Individuals describe places by using physical and/or

geographical descriptions of varying accuracy, e.g. their spatial coordinates, geometrical extension, and so on. Individuals can also offer inherently subjective descriptions, e.g. whether these individuals may find places easy to reach, or relaxing at night. These studies thus suggest that place concepts involve *facets*: mind-internal representations only indirectly representing the mind-external properties of places (Canter 2012). Individuals can thus entertain place concepts that correspond to complex and personal descriptions of places, and that combine multiple facets: semantic features, in linguistics parlance.

In the second strand, works analyse individuals’ use of toponyms on social media. They show that descriptions connect to geo-locational information via acts of “digital reference” (Zhang and Gelernter 2014, Zhou et al. 2014, Hollenstein and Purves 2010, Vasardani et al. 2013, Chen et al. 2018a). For instance, individuals can review *Federation Square* in Melbourne on TripAdvisor, and link their multi-modal reviews to this place’s GPS coordinates. These coordinates refer to the actual place’s location via reference mediated via a place concept. “Collective”, shared definitions are also possible: some individuals may love Federation Square; others may instead hate it. Individuals acquainted with this place may therefore have a collective/shared, though not necessarily coherent concept of this place. Overall, cognitively oriented GIS works offer evidence that toponyms denote potentially unique mind-internal place concepts referring to mind-external places via semantic features. These principles seem to apply across different text types and media.

The role of context in the use of toponyms. The many GIS works supporting semantically rich approaches to place concepts also indirectly support the observation that different modalities and communicative contexts affect how individuals describe places via their personal perspectives. For instance, while X/Twitter may force individuals to offer brief place descriptions limited to 140 characters, FourSquare allows individuals to write up richly detailed place reviews. Hence, several studies offer indirect evidence that individuals may share personal place concepts according to the rules of the genres and contexts in which they communicate these concepts (Davies 2013, Hollenstein and Purves 2010, Vasardani et al. 2013, Chen et al. 2018a). Such correlation between semantic richness of place concepts and context of use, only sketched in these GIS works, finds an analysis in Terminology and Lexicography. The core assumptions to explain this correlation and its relevance to our research questions are as follows.

Terminology researchers distinguish *general*, *specialised* and (*technical*) *term* uses of a word (ten Hacken 2015, 2018). General uses mostly occur in informal contexts, e.g. daily conversations. For instance, Roman citizens may use *Piazza Navona* in conversation to simply refer to this square as a meeting point. Specialised uses occur in contexts presupposing specific goals and knowledge. Tourists may offer content-rich descriptions of Piazza Navona when writing TripAdvisor reviews as a specialised genre, in their language(s) of choice. Technical term uses involve definitions based on necessary and sufficient conditions that classify places, in turn based on official reference texts. For the “City of Rome” administration, *Piazza Navona* is the official name of a given square. Squares, in turn, are legally defined as a gathering open place surrounded by buildings and street. The toponym *Piazza Navona* in its technical term use thus establishes that this administration classifies this place as a square.

Terminology and Lexicography studies conceive general, special and term uses as a cline of increasingly refined but potentially flexible uses for words. Recent proposals model this context-based flexibility via the notion of “prototype” (San

Table 1 Research questions defining our paper.	
Current problem	Research question
Cross-linguistic evidence on the semantic properties of place concepts and toponyms expressing them is scarce.	What type of semantic features can associate with toponyms and the place concepts they express, cross-linguistically?
Most studies do not analyse contextual constraints on place concepts' uses.	How can context modulate the use of place concepts, within and across languages?
Dearth of cross-linguistic data prevents a comparison of place concepts and the geonyms/toponyms expressing these concepts.	To what extent can geonyms/toponyms and the place concepts they denote be similar, across contexts and languages?

Table 2 Geonyms in the different languages under investigation.					
Concept/ Language	French	Italian	Portuguese	Romanian	Spanish
'Street'	rue	via	rua	stradă	calle
'Square'	place	piazza	praça	piață	plaza
'Alley'	ruelle	vicolo	beco	alee	callejón

Martin 2022a, b). In this approach, (technical) term uses correspond to uses in which the prototypical features of a concept form a term's definition. For geonyms, these prototypical features can represent how individuals collectively conceive a place concept (San Martín et al. 2020, San Martín et al., 2017, L'Homme et al. 2022). For instance, a term definition for *piazza* can include the features 'gathering place', 'open space' and other features simultaneously defining the 'square' concept. Specialised uses (e.g. touristic platforms' reviews) can include most but not all these features (e.g. that *Piazza Navona* refers to a 'gathering place'). General uses may be the least prototypical: tourists visiting Rome may use this toponym without even knowing that they are referring to a square.

Terminology and Lexicography studies thus suggest that prototypical place concepts may be semantically related or even similar to differing degrees. GIS works focusing on place (formal) ontologies study how the notions of semantic relatedness and semantic similarity apply to place concepts (Ballatore, Wilson and Bertolotto 2013, Zhu et al. 2017, Ballatore, Bertolotto and Wilson, 2014, 2015, Chen et al. 2018b). According to these assumptions, *river* and *water* refer to 'water'-based concepts and potentially to places; they are thus semantically related geographic nouns. These two nouns also have similar semantic content, since *river* refers to 'flowing' bodies of water, while *lake* refers to 'static' bodies (Ballatore, Bertolotto and Wilson 2015, Chen et al. 2018b). Hence, these works show that geographic nouns can involve similar and related senses, thus forming delimited semantic domains. However, these works concentrate on single languages and their geonyms (e.g. English for Ballatore, Bertolotto and Wilson 2015, Chinese for Chen et al. 2018b), and on specialised contexts of use.

Studies that attempt context-sensitive analyses are Ursini and Samo (2023a, 2025), Samo and Ursini (2024). The first two studies investigate the context-sensitive place concepts associated to English toponyms including geonyms *square*, *street* and *alley*. The third study investigates Italian toponyms including geonym *piazza* 'square'. These studies show that these place concepts in general uses include the most extensive lists of semantic features; in specialised uses, they involve smaller though still extensive lists. Place concepts in term uses include compact sets of physical-geographical, measurable features that identify places and place types. These works thus suggest that context/use-sensitive features represent more "subjective" views about places (e.g. a place's 'beauty'); they express (more) contingent features defining

place concepts. Features representing inherent (e.g. physical) properties of places act as culturally/temporally (more) stable features (e.g. a place's 'shape' or 'functions'), across different contexts of use. Crucially, these works do not address cross-linguistic variation, and the possible existence of prototypical place concepts across different contexts. Hence, our research questions remain outstanding.

Target research questions and predicted answers. The picture emerging from previous research across different disciplines suggests that toponyms have typologically flexible yet homogeneous grammatical structures, qua proper names for places. Furthermore, they have richly descriptive senses denoting complex place concepts. Individuals can use these place concepts according to the communicative needs of a given context. Crucially, this picture also confirms that we must fill three empirical gaps pertaining to our research questions, summarised in Table 1: By answering these questions, we aim to reach our research goal. We aim to propose a semantic model that captures cross-linguistic variation and similarity of the 'square', 'street' and 'alley' concepts across the five target Romance languages. The remainder of this paper provides the answers to these questions.

Methodology and materials

Our goal was to extend the methodology and materials from Ursini and Samo (2023a, 2025), Samo and Ursini (2024) to the present study and target languages; we operated as follows.

First, we individuated which geographic nouns in each target language denote the three concepts 'square', 'street' and 'alley'. We thus first consulted Wikidata as a multilingual dictionary for each language in the sample (Vrandecic 2013, Turki et al. 2017) and verified that each noun would also regularly occur as a geonym in toponyms.³ If a definition in a given language is absent from Wikidata, reference is instead made to a monolingual dictionary. We used O(pen)S(treet)M(ap) as a gazetteer for this geographical verification task. For instance, we individuated *place* as a French geographic noun expressing the concept 'square'. We confirmed that it regularly occurs in French toponyms (e.g. *Place de la Concorde*) by verifying in OSM that we mostly extracted toponyms including this geonym originating in standard French, instead of e.g. Belgian French toponyms. Table 2 presents the list of each attested geonym in each language, while Fig. 1 (next page) visualizes their geographic distribution in terms of latitude and longitude. Certain forms appear in other geographical areas due to bilingual situations (e.g., Aosta Valley) or, as in the case of *calle* in specific urban contexts like Venice, reflect well-studied local variants in geo-linguistics (see Samo and Ursini 2023 and references therein).

Second, we extracted toponyms including these geonyms to develop the materials for our analysis. As discussed in the Literature Review Section, Romance toponyms tend to display similar morphological structures involving compound NPs with geonyms preceding specific names (e.g. *Rua Central* in Portuguese), possibly via the mediation of prepositions (e.g. *de* in

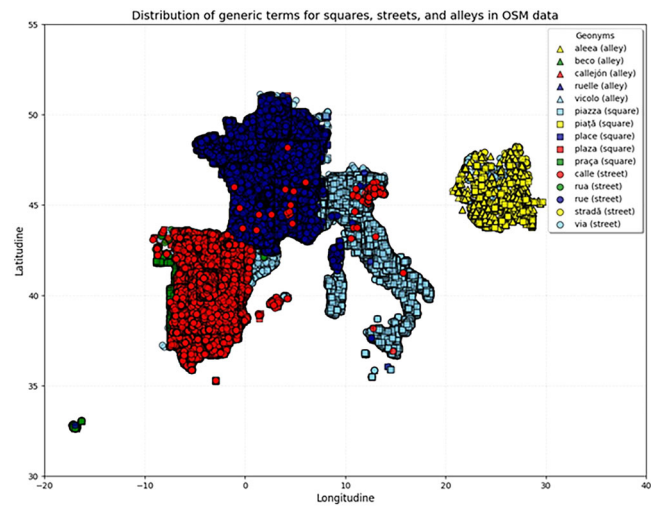


Fig. 1 Geographical distribution of the investigated geonyms in terms of latitude and longitude.

Table 3 Size, in token words, of the sketch engine corpora.		
Language	Corpus	Size (tokens)
French	Fr-tenten23	23 billions
Italian	Web2020	12 billions
Romanian	ELEXIS Web2021	1 billion
Spanish	Es-tenten18	16 billions
Portuguese	Web2020	13 billions

French *Place de la Concorde*). Thus, we extracted these two toponym structure types across our language sample. We used large-scale corpora data for general use types; Wikipedia entries for specialised use types; Wikidata dictionary definitions focusing on geographic, architectural and urbanistic senses for technical term types. We motivate these choices in the remainder of this section.

Corpora (general uses). We collected data from corpora for two reasons. First, corpora offer naturally occurring examples of linguistic data that cover a variety of genres and registers Gries (2009).⁴ Second, corpora represent the language output of individuals who may participate in conversations that usually involve minimal common ground assumptions, and few communicative goals shared among participants. For instance, corpora can offer examples of residents giving directions to lost tourists who wish to reach their hotels, but may only have the hotel’s name as a reference for this place. Thus, corpora examples can represent language use in general contexts and, for our study, provide evidence for geonyms and toponyms’ general uses.

We collected general uses of the triplets of geonyms via SketchEngine, a tool that allows users to access language-specific corpora and resources via a language-universal interface (Kilgariff et al. 2014, Wearne and Riedy 2024). The key properties of these corpora are in Table 3:

We removed three types of false positive hits. The first type involves uses of geonyms as common nouns (e.g. Spanish *calle* in *la primera calle a la derecha* ‘the first street on the right’). The second type involves false positive tokens, e.g. other names often occurring in toponyms but occurring in isolation in our queries. One example was *De Gaulle* as a hit for French *place*; *Place de Gaulle* is a common toponym and the surname occurred in queries involving French toponyms. The third type involves

unrelated uses due to polysemy, e.g., Italian *via libera* literally “free street,” an expression that describes a path devoid of any obstacles, or the Romanian *piată*, which has a particular meaning in the stock market. We favoured searching for discriminative casing patterns to minimize ambiguities and make the automatization process easier to replicate.

For each language and each target geonym, we used the advanced search configuration “concordances” (examples of the use context) on the Sketch Engine interface, querying only the lemma and part of speech (POS) restricted to nouns, and selecting a context size of 30 words. We downloaded the first 10,000 results in a.txt format where the target words are annotated by enclosure in <coll> tags. With respect to the sub-corpus, when possible, we isolated selected domains (e.g., fr for French, pt for Portuguese) in order to reduce the influence of non-standard spelling forms of the target word and to avoid sources such as Wikipedia (.org domains), which would represent a different population under investigation. These files are included in the supplementary materials.

In the next step, we performed a strict version of the *Pointwise Mutual Information* (PMI, Church and Hanks 1990) analysis on our corpora using a custom Python script (adapted from Ursini and Samo 2023a, b, 2025). Our goal in computing a strict PMI’s for each geonym was to obtain a list of semantic features for each geonym, i.e. the do- main of semantic relations between each target word and the words collocating with them (Recchia and Jones 2009, Bullinaria and Levy 2007).

For each text file in the input folder, the procedure consisted of several steps. First, all expressions enclosed in <coll> tags were analysed for the PMI. Only collocations occurring more than a minimum frequency threshold ($min_coll_freq = 10$) were retained for further analysis. Next, for each retained collocation, co-occurring words within a symmetric context window of size 5 were identified. Co-occurrences with a frequency lower than $min_cooccurrence = 10$ were discarded. Our model then counted the occurrences of all words in the text and of each collocation, as well as the co-occurrence frequencies between collocations and context words. For each collocation– context word pair, our PMI was computed as:

$$PMI = \log_2 \frac{p(\text{collocation}, \text{context})}{p(\text{collocation}) \cdot p(\text{context})}$$

Pairs with PMI values below a minimum threshold ($min_pmi = 2.0$) were discarded. For each collocation, context words were ranked by descending PMI and co-occurrence frequency. This procedure ensures that only statistically significant and sufficiently frequent co-occurrences are considered, providing a robust measure of association between collocations and their contexts. Finally, we implemented a language-specific text processing pipeline using SpaCy (version 3.8.7) with pre-trained models for Italian, Spanish, French, Portuguese, and Romanian. PMI results were processed to extract only nouns and adjectives.

We selected nouns and adjectives collocating with geonyms for further analysis for a semantic reason. When items belonging to these two categories collocate with toponyms, they describe properties of the places to which toponyms refer. For instance, in the Italian Noun Phrase *la famosa Piazza Navona* ‘the famous Piazza Navona’, the attributive adjective *famosa* ‘famous’ collocates with the toponym *Piazza Navona*, and thus with the geonym *piazza*. The adjective *famosa* introduces a property that can be ascribed to this and to any other square: it thus describes a feature of ‘square’ as a place concept. Symmetrically, the sentence *Piazza Navona è un famoso luogo turistico* ‘Piazza Navona is a famous touristic place’ includes the noun *luogo* ‘place’ and the adjectives *famoso* ‘famous’ and *turistico* ‘touristic’. Nouns thus can introduce features associated with place concepts when occurring in predicative structures.

Table 4 List of research answers/predictions for our paper.	
Research question	Answer/Prediction (and supporting literature)
RQ1 <i>What type of semantic features can associate with toponyms and the place concepts they express, cross-linguistically?</i>	Language speakers/users can associate both contingent and temporally/culturally stable semantic features with place concepts.
RQ2 <i>How can context modulate the use of place concepts, within and across languages?</i>	Contingent features should decrease from general to specialised and term uses; stable features should increase. However, features that occur across uses define proto- typical senses associated to concepts.
RQ3 <i>To what extent can geonyms/toponyms and the place concepts they denote be similar, across contexts and languages?</i>	Geonyms should provide high degrees of similarity the place concepts they express, though identity may be rare.

On the other hand, we did not include verbs and prepositions in our analysis because these categories introduce relational features. For instance, the verb *parked* in *I parked in Trafalgar Square* indirectly gives us information that the ‘square’ concept can also include the feature ‘parking place for vehicles’. Thus, to fully define this feature we would need to define the ‘vehicles’ concept. The preposition *in*, instead, would require the introduction of an ‘inclusion’ relational concept connecting objects and their locations. Overall, relational features connect concepts to other concepts in the lexicon. The computation of sense similarity relations, and thus the answering of our third question, would become remarkably challenging if we would include the relational feature type (cf. Ballatore, Bertolotto and Wilson 2015, Chen et al. 2018b). For this reason, we left these data types aside.

Wikipedia articles (specialised uses). Terminology studies suggest that encyclopaedic entries and other topic-oriented texts (e.g. reviews, forum posts) can provide instances of specialised uses for two reasons (Syed et al. 2021, Giacomini 2023, 2024). First, authors writing these texts usually target audiences pursuing detailed information about some topic: they have specific communicative goals. Furthermore, these texts usually provide definitions of key words, and introduce thorough descriptions of a given topic. Second, authors can also offer bibliographic references in support of the information they present. Authors can thus share contextual, co-textual information and potentially create a common ground with readers. Crucially, Wikipedia articles lack an official status, as in the case of cadastral documents, gazetteers, dictionaries and encyclopaedic entries. Thus, they cannot provide us with technical term uses.

To extract the linguistic data, we developed a Python-based scraper leveraging the official Wikipedia API (please see supplementary file, the “Python scripts” file). For each language–geonym pair, the script queried Wikipedia through its search endpoint, retrieving up to 500 articles containing the target lexical item. When possible, we aligned texts cross-linguistically by taking articles about famous streets, squares, alleys. The plain-text extracts of the corresponding pages were then downloaded via the API’s extracts property. To obtain a more standardized dataset, the retrieved texts were cleaned by removing non-informative markup such as reference tags, brackets, and redundant newlines.

We pursued this approach to analyse whether genre (here, Wikipedia articles) and topic (articles about particular places) could be factors influencing the content of place concepts in this type of specialized use. The same evaluation procedure, using strict PMI-based feature isolation (see section “Corpora (general uses)”) – with the difference that the target words are not annotated under the label <coll> – was carried out. To identify common character-based patterns across the other textual registers (sub-section “Corpora (general uses)” and sub-section “Wikidata and dictionary entries (technical term uses)”), we implemented a Python script that systematically computes

character n-grams. For this dataset as well, we isolated nouns and adjectives.

Wikidata and dictionary entries (technical term uses). For technical term uses, we consulted Wikidata entries for each target geonym (see footnote 3). We then selected definitions related to geographical, architectural, legal and urban planning domains for manual analysis. We operated in this manner for two reasons. First, we could have used official dictionaries of geographic terms for each language used within scientific communities (e.g. Klaus-Jürgen et al. 2010). However, they usually lack legal status, e.g. they may not be used for the coining of new toponyms. Wikidata entries act as encyclopaedic entries and as such, they report which senses/entries within definitions may have legal and technical (term) status. Second, Wikidata entries also report potential etymologies and naming functions for geonyms, and by extension for toponyms including these geonyms. Geographic dictionaries usually overlook such linguistic aspects, as they focus on place concepts under description. Thus, Wikidata entries permitted us an accurate analysis of term uses.

We extracted the definitions for target terms and operated a manual analysis of these definitions. Each author has a degree of fluency in each target language. We could thus map the components of these definitions into semantic features in English as the meta-language of this paper. For instance, the Wikidata’s definition of French *rue* includes the phrase *voie de circulation bordée de bâtiments et située dans une agglomération*, ‘a traffic way bordered by buildings and located in a built-up area’. Via this method of semantic analysis, we assigned this part of a definition to the semantic features ‘circulation’, ‘urban’, ‘bordered-by-buildings’. We thus operated this analysis in an exhaustive manner, i.e. by mapping all parts of definitions to semantic features.

Predictions. The predictions for our study stem from the findings in previous literature reported in Section “Literature review”, and can be formulated as potential answers to our three research questions. We report these answers in Table 4:

Overall, these three predictions/answers should offer novel evidence that cross-linguistic and context-sensitive analyses should offer language-general evidence about toponyms, geonyms and the place concepts they express. Therefore, the findings in previous GIS (e.g. Hamzei et al. 2020, Davies 2013), toponomastic (e.g. Nyström 2016, Blair and Tent (2015)), and cognitive/psycholinguistic studies (George 2023, Peng et al. 2020) should be confirmed and extended via our cross-linguistic Romance sample. Terminology, Lexicology (e.g. San Martín 2022b), linguistics (cf. Ursini and Samo 2025), and GIS works (Ballatore, Bertolotto and Wilson 2014) findings should also be extended to context-sensitive uses. Consequently, these three predictions/answers should also connect the hitherto weakly connected strands of research on toponyms into an integrated, albeit preliminary, cross-linguistic model.

Table 5 Temporally/culturally stable semantic features for the 'square' concept.

Language	English gloss	French	Italian	Portuguese	Romanian	Spanish
General use	'street(s)',	<i>rue,</i>	<i>via,</i>	<i>rua,</i>	<i>stradă,</i>	<i>calle,</i>
	'beautiful',	<i>belle,</i>	<i>bella,</i>	<i>bonita,</i>	<i>frumoasă,</i>	<i>bonita,</i>
	'big',	<i>grand,</i>	<i>grande</i>	<i>grande,</i>	<i>mare,</i>	<i>grande,</i>
	'church',	<i>église,</i>	<i>chiesa,</i>	<i>igreja,</i>	<i>biserica,</i>	<i>iglesia,</i>
	'central',	<i>central,</i>	<i>centrale,</i>	<i>central,</i>	<i>central,</i>	<i>central,</i>
	'concert',	<i>concert,</i>	<i>concerto,</i>	<i>concerto,</i>	<i>concert,</i>	<i>concierto,</i>
	'hotel',	<i>hôtel,</i>	<i>hotel,</i>	<i>hotel,</i>	<i>hotel,</i>	<i>hotel,</i>
	'market (stores)',	<i>marché,</i>	<i>mercato,</i>	<i>mercado,</i>	<i>piată,</i>	<i>mercado,</i>
	'museum',	<i>musée,</i>	<i>museo,</i>	<i>museu,</i>	<i>muzeu,</i>	<i>museo,</i>
	'cathedral'	<i>cathédral,</i>	<i>cattedrale</i>	<i>catedral,</i>	<i>catedrală,</i>	<i>catedral,</i>
Specialised use	'public'	<i>public</i>	<i>pubblico</i>	<i>público</i>	<i>public</i>	<i>público</i>
	'street',	<i>rue,</i>	<i>Via,</i>	<i>rua,</i>	<i>stradă</i>	<i>calle,</i>
	'central',	<i>centrale,</i>	<i>centrale,</i>	<i>central,</i>	<i>centrală</i>	<i>central,</i>
	'church'	<i>église</i>	<i>chiesa,</i>	<i>igreja,</i>	<i>biserica,</i>	<i>iglesia,</i>
	'monument',	<i>monument</i>	<i>monumento,</i>	<i>monumento,</i>	<i>monumentul,</i>	<i>monumento,</i>
	'building',	<i>immeuble,</i>	<i>edificio,</i>	<i>edificio,</i>	<i>clădirea,</i>	<i>edificio,</i>
Term use	'public'	<i>public</i>	<i>pubblico</i>	<i>público</i>	<i>public</i>	<i>público</i>
	'urban'	<i>espace public</i>	<i>centro abitato</i>	<i>urbano</i>	<i>dintr-o localitate,</i>	<i>urban</i>
		<i>desservi</i>	<i>urbano</i>			
		<i>par</i>				
		<i>des</i>				
		<i>voies</i>				

Results

First prediction: semantic feature types. Our findings confirm the first prediction as follows. Most features occurring in general and specialised uses tend to occur in some but not all languages. Most adjectives and nouns co-occurring with toponyms including our target geonyms may thus express (more) contingent features, i.e. features less frequently associated with these place concepts. For instance⁵, the Italian adjective *pulita* 'clean' occurs in predicative structures including toponyms for streets and square as being (un)clean (also historically). Example sentences include *una curiosa iscrizione ottocentesca invita a tenere pulita la piazza* 'A curious 19th-century inscription invites people to keep the square clean'.⁶ We did not find adjectives expressing this concept (i.e. a 'square' place being 'clean') associated to toponyms naming squares, in the other four languages. Thus, the feature 'clean' only associates to the Italian instance of the 'place' concept, i.e. *piazza*, whence its contingent status. The other language-specific instances do not have share this association.

However, small sets of features occur in each language and express aspects of these concepts that seem (more) temporally and culturally stable. Few adjectives and nouns co-occur as language-specific forms expressing these features across the five languages of our sample. For instance, 'important' is a feature associated to the 'street' concept in each of the five target languages, as the first line of Table 5 shows. For instance, *La rue de la Loi (en néerlandais : Wetstraat) est une rue importante de Bruxelles* 'Rue de la Loi (in Dutch: Wetstraat) is an important street in Brussels'⁷ or *Se trata [Calle del Císter] de una importante calle del centro histórico* 'It is an important street in the historic center'⁸. In Spanish are sentences including adjectives *importante* 'important' in predicative position. As each of the five languages includes hundreds of such examples associating this feature to the 'street' place concepts via adjectives, we consider this conceptual association (more) temporally and culturally stable. We introduce the full sets of temporally/culturally stable features per each concept and context of use via Tables 5, 6 and 7:

Thus, (more) temporally and culturally stable semantic features find a specific realization in each language, usually as adjectives (e.g. *publique*, French; *pubblica*, Italian; *publica*, Spanish and Portuguese; *public*, Romanian). This result should not appear

surprising if we consider the fact that certain properties of places are relatively temporally and culturally stable because agents create places with certain social functions and geo-physical properties. For instance, squares in European societies have traditionally existed as places for public functions and activities. The concept 'square', as expressed via geonyms and toponyms across these five Romance languages, thus includes the semantic feature 'public' expressing in language this social and administrative (language-external) property of squares. Contingent features instead form extensive sets with hundreds of items. We invite readers to consult the "ReadMe" supplementary files for an overview and instructions, as inserting these lists in the main text would be impractical. One further example, however, is as follows.

Italian adjectives *bella*, *stretto* and French adjectives *mignonne*, *etroit*, respectively express the features 'beautiful', associated to 'square' and 'alley' in these languages (i.e. Italian *piazza*, *vicolo*; French *place*, *ruelle*) in their general uses. We did not find adjectives expressing these features as co-occurring with toponyms/geonyms expressing the 'street' concept. We also did not find adjectives expressing the 'beautiful' feature in Romanian, Portuguese and Spanish data involving the 'square' concept. Speakers of these languages may actually conceptualise squares as beautiful, thus sharing this more subjective or at least cognitively-oriented of squares with speakers of French and Italian. Our textual data do not however offer us evidence of this view. The contingent status of this and other features thus seem to reflect the fact that place concepts can also express individually-oriented conceptions of places.

Overall, Tables 5, 6 and 7 thus invite the conclusions that temporally and culturally stable features have also relatively stable cross-linguistic distribution. They appear as features of each place concept in each language in our sample. Conversely, contingent features, as reported in the supplementary files, tend to represent less socially and culturally relevant properties of the places they can describe, through the lenses of a specific language/culture, but also modes of (subjective) attachment to places. Furthermore, the tables invite the conclusion that one can find a wealth of semantic types and features expressing the place concepts 'square', 'street' and 'alley' in these Romance languages. This result thus confirms

Table 6 Temporally/culturally stable semantic features for the 'street' concept.

Language	English Gloss	French	Italian	Portuguese	Romanian	Spanish
General use	'center', 'house', 'city', 'school', 'museum', 'square', 'park', 'bridge', 'theatre', 'hospital', 'public', 'social'	<i>centre, maisons,</i> <i>ville,</i> <i>école,</i> <i>musée,</i> <i>place,</i> <i>parc,</i> <i>pont,</i> <i>théâtre,</i> <i>hôpital,</i> <i>public,</i> <i>social</i>	<i>centrale, casa,</i> <i>citta',</i> <i>scuola,</i> <i>museo,</i> <i>piazza,</i> <i>parco,</i> <i>ponte,</i> <i>teatro,</i> <i>ospedale,</i> <i>pubblico,</i> <i>sociale</i>	<i>centro, casa,</i> <i>cidade,</i> <i>escola,</i> <i>museu,</i> <i>prac, a,</i> <i>parque,</i> <i>teatro,</i> <i>hospital,</i> <i>público,</i> <i>social</i>	<i>centru, casă,</i> <i>oras,</i> <i>scoală,</i> <i>muzeu,</i> <i>piată,</i> <i>parc,</i> <i>pod,</i> <i>teatru,</i> <i>spital,</i> <i>public,</i> <i>social</i>	<i>centro, casa,</i> <i>ciudad,</i> <i>escuela,</i> <i>museo,</i> <i>plaza,</i> <i>parque,</i> <i>puente,</i> <i>teatro,</i> <i>hospital,</i> <i>public,</i> <i>social</i>
Specialised use	'central', 'commercial' city', 'house', 'construction' 'important', 'park', 'principal', 'transport'	<i>central,</i> <i>commercial</i> <i>cial,</i> <i>ville,</i> <i>maison,</i> <i>construction</i> <i>importan</i> <i>parc,</i> <i>principal,</i> <i>transport</i>	<i>centrale,</i> <i>commerciale</i> <i>cial,</i> <i>città,</i> <i>costruzione</i> <i>costruzione,</i> <i>importante,</i> <i>parco,</i> <i>principale,</i> <i>trasporto</i>	<i>central, comercial,</i> <i>cidade,</i> <i>casa,</i> <i>construção,</i> <i>importante,</i> <i>parque,</i> <i>principal,</i> <i>transporte</i>	<i>central, comercial,</i> <i>oras,</i> <i>casă,</i> <i>constructie,</i> <i>important,</i> <i>parc,</i> <i>principal,</i> <i>Transport</i>	<i>central, comercial,</i> <i>ciudad,</i> <i>casa</i> <i>construcción,</i> <i>importante</i> <i>parque,</i> <i>principal,</i> <i>transporte</i>
Term use	'circulation', 'urban'	<i>circulation,</i> <i>agglomération</i>	<i>centro abitato,</i> <i>segnaletica</i>	<i>pedestrese veículos,</i> <i>ambiente construído</i> <i>con-</i> <i>struído</i>	<i>interiorul unei-</i> <i>calităţi, trotuarele</i>	<i>circulación urbana</i>

Table 7 Temporally/culturally stable semantic features for the 'alley' concept.

Language	English Gloss	French	Italian	Portuguese	Romanian	Spanish
General use	'house' 'family' 'famous' 'hotel' 'light' 'wall' 'historical' 'typical' 'old' 'narrow'	<i>maison, famille, fameux,</i> <i>hôtel, lumière,</i> <i>mur,</i> <i>historique, typique, vieux,</i> <i>étroit</i>	<i>casa, famiglia, famoso,</i> <i>hotel,</i> <i>luce,</i> <i>muro, storico, tipico,</i> <i>vecchio, stretto</i>	<i>casa, família,</i> <i>famoso, hotel,</i> <i>luz,</i> <i>muro, histórico,</i> <i>típico,</i> <i>velho, estreito</i>	<i>casă, familie, faimos,</i> <i>hotel, lumină,</i> <i>mur,</i> <i>istoric,</i> <i>tipic,</i> <i>vechi,</i> <i>îngust</i>	<i>casa, familia,</i> <i>famoso, hotel,</i> <i>luz,</i> <i>muro, histórico,</i> <i>típico,</i> <i>viejo, estrecho</i>
Specialised use	'house' '(main) street' 'narrow'	<i>rue,</i> <i>étroit</i>	<i>strada, stretto</i>	<i>rua,</i> <i>estreito</i>	<i>principala, strict</i>	<i>calle,</i> <i>estrecho</i>
Term use	'small/tight'	<i>petite</i>	<i>stretta</i>	<i>estreita e curta</i>	<i>îngustă, i scurtă</i>	<i>estrecha</i>

our first prediction, regarding the existence of these two feature types as cross-linguistically connecting geonyms/toponyms and the place concepts they express. It also leads to the confirmation of the second prediction, as we discuss next.

Second prediction: influence of context on uses. Our findings confirm the second prediction as follows. Quantitatively, semantic feature lists for general uses include the highest number of contingent features for Italian, Portuguese and Spanish. The number decreases in specialised uses, for these three languages, and reaches a minimum in term uses. Specialised uses for French *place* 'square' and *rue* 'street', Romanian *strada* 'street', however, falsify this prediction. For both languages, nevertheless, term uses include the least number of features. Conversely, stable features reach the highest numbers in term uses, and (minimally) decrease in specialised and general uses. We take this as evidence that our second prediction is overall accurate: 42 out of 45 (three contexts, three concepts and five languages) possible semantic features lists

follow our quantitative prediction. Figure 2 provides an overview confirming this first dimension of the second prediction:

Our second prediction is cross-linguistically confirmed at a qualitative level in a more exhaustive manner, as Tables 5–7 have partially foreshadowed. This is the case because certain features occur in each context of use, even if this “context-free” distribution applies to language-specific geonyms and place concepts. Therefore, these features define the prototypical senses associated to each geonym, as a language-specific expression of each place concept. Again, prototypical senses (and uses) correspond to uses that individuals can collectively use to define a concept (e.g. San Martín et al. 2020). Consider Tables 8, 9 and 10:

The tables suggest that some features forming prototypical senses may be cross-linguistically stable, and form prototypical senses in a single language. For instance, the feature *pubblica* 'publica' occurs in the general, specialised and term uses of Italian *piazza* 'square'. This feature appears in each target language, and forms the prototypical sense of the geonyms and related toponyms expressing this concept: Italian *piazza*, French *place*,

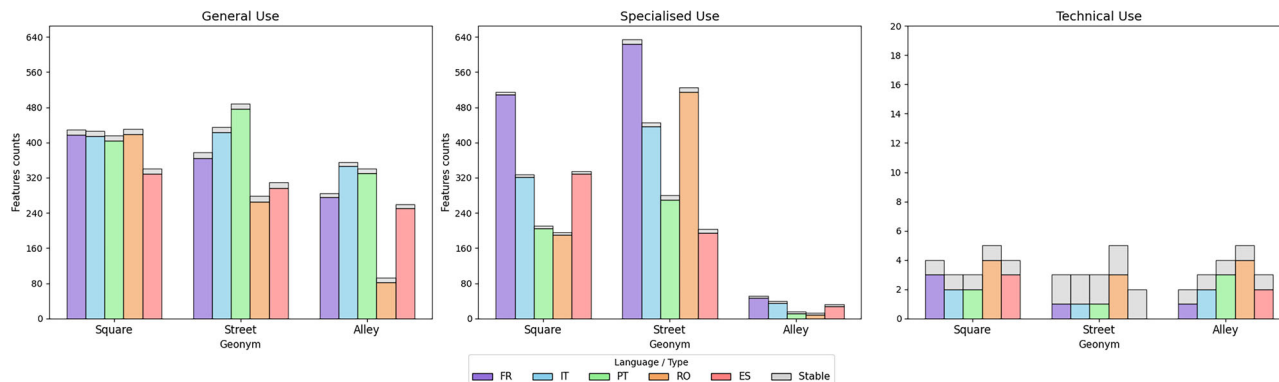


Fig. 2 Distribution across datasets and languages of contingent and stable features. The y-axis for the Technical Use panel appears in reduced scale to improve readability.

Table 8 Prototypical senses for the geonyms expressing the 'square' concept.

Language	French: <i>place</i>	Italian: <i>piazza</i>	Portuguese: <i>praça</i>	Romanian: <i>piață</i>	Spanish: <i>plaza</i>
English glosses	public square, central location, often with monuments, churches, markets, civic buildings, pedestrian space	public square, central location, surrounded by churches, monuments, civic buildings, marketplace	public square, central gathering place, with monuments, churches, parks, civic buildings	public square, central location, monuments, parks, surrounding streets, civic functions	public square, central square, civic, religious buildings, monuments, markets, events
WiC	26	200	110	78	124

WiC words in common in lists.

Table 9 Prototypical senses for the geonyms expressing the 'street' concept.

Language	French: <i>rue</i>	Italian: <i>via</i>	Portuguese: <i>rua</i>	Romanian: <i>stradă</i>	Spanish: <i>calle</i>
English gloss	Urban street, main or secondary, buildings, shops, pedestrian and vehicle access, central or peripheral location	Urban street, main or secondary, buildings, shops, traffic, location, sometimes historical	Urban street, main or secondary, buildings, shops, residential or commercial, central or peripheral	Urban street, main or secondary, buildings, shops, pedestrian and vehicle access, neighborhood	Urban street, main or secondary, buildings, shops, pedestrian and vehicle access, city center or neighborhoods
Wic	178	190	118	230	84

Table 10 Prototypical senses for the geonyms expressing the 'alley' concept.

Language	French: <i>ruelle</i>	Italian: <i>vicolo</i>	Portuguese: <i>beco</i>	Romanian: <i>alee</i>	Spanish: <i>callejón</i>
English gloss	Narrow alley, small passage, residential buildings, sidestreet, sometimes historic, community-focused	Narrow alley, small street, buildings, corners, pedestrian access, often historic	Narrow alley, small passage, houses, side access, local community	Narrow path, small alley, limited access, pedestrian, part of residential area	Narrow alley, small passage, side street, buildings, vehicle access, historic or residential
WiC	26	200	9	7	20

Portuguese *praça*, Romanian *piată* and Spanish *plaza*. It is thus a stable feature that highlights how “Romance” speakers may conceive squares as places of public relevance and use, and thus how Italian speakers use *piazza* and related toponyms as carrying the matching feature. Overall, prototypical senses act as language- and culture-specific senses, with a degree of semantic variation stemming from which features the speakers of a language associate with a concept. Nevertheless, prototypical senses also provide evidence that context modulates part of a geonym’s sense, though a semantic “core” connects all contexts of use and, for features such as ‘public’, also languages cross-linguistically.

Interestingly, these results also suggest that stable features tend to form the majority of proto- typical senses, and that term uses mostly reduce to these features. Indeed, term uses *qua* normative definitions aim to list necessary and sufficient conditions individuating the referents they describe (here, places and place concepts). However, the presence of contingent features in prototypical senses suggests that cultural/temporal aspects also play a key role in shaping place concepts. Quantitative and qualitative results thus combine to define a nuanced confirmation of our second prediction. Contexts of use can modulate place concepts’ uses by letting speakers use contingent and thus culture-specific, context-sensitive expressions of place concepts. Nevertheless, variation seems to involve language-specific prototypical (i.e. “context-free”) senses possibly including contingent features. We discuss the third prediction to fully address this aspect.

Semantic similarity and prototypical senses. Before we discuss the results confirming the third prediction, we need to offer a short preamble regarding how we measured semantic similarity. As mentioned in the Literature Review, there is a wealth of research on semantic similarity across disciplines (e.g. computational linguistics, formal ontologies) and applications (e.g. information retrieval, text analysis, ontology development: Rada et al. (1989), Li et al. (2003), Harispe et al. (2015)). Thus, there are dozens of mathematical methods to measure the semantic similarity of words, sentences or even texts (Navigli and Martelli 2019, Reshmi and Shreelekshmi 2023, Vulić et al. 2020).

Reviewing all these methods here would lead us too far afield, so we leave this aspect aside. Here we describe how we adapted a method that features prominently in GIS works that perform sense similarity analyses via the use of semantic features (e.g. Samo and Ursini 2024). To quantify the similarity between groups of words based on their initial morphemes (roughly roots), we implemented the following computational procedure in Python. First, we defined an *initial letter profile* function, which computes the normalized distribution of the initial morphemes across all words in each set. Formally, for a word set W , the function extracts the relevant characters of each instance and returns a probability distribution $P(I)$ over the alphabet, with values normalized to the $[0, 1]$ range. Second, we measured similarity between two such profiles using *cosine similarity*. For two distributions P_1 and P_2 , represented as vectors over the union of their keys, the cosine similarity is computed as:

$$\text{sim}(P_1, P_2) = P_1 \cdot P_2$$

$$\|P_1\| \|P_2\|$$

With a similarity of 0 assigned if either vector is the zero vector. Third, we constructed a *similarity matrix* for all groups of words under comparison. Each entry (i, j) of the matrix corresponds to the cosine similarity between the profiles of groups i and j . Scores thus correspond to values between “0” (no similarity: no features match between lists) to “1” (identity: lists

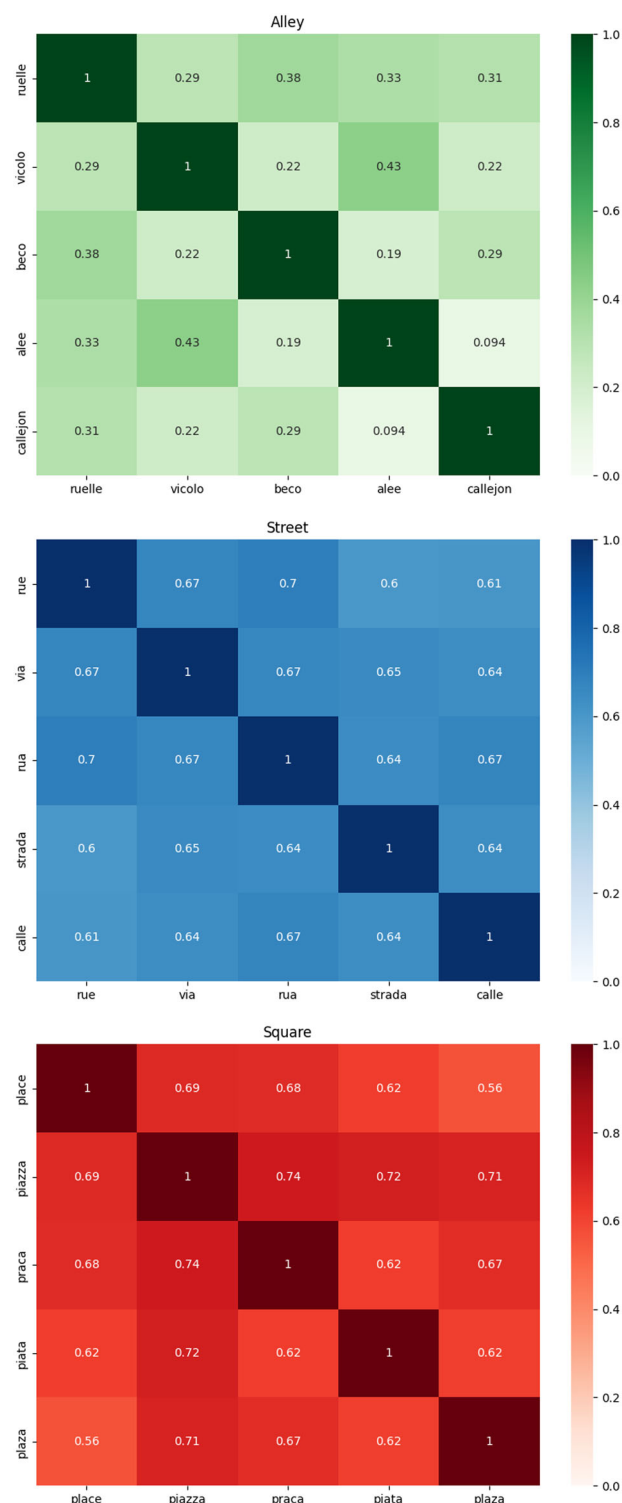


Fig. 3 Similarity coefficients for the geonyms/toponyms expressing the prototypical senses.

match on a 1:1 ratio). We performed this analysis for each concept, as we show in Fig. 3:

The results can be summarised as follows. For the geonyms expressing the ‘square’ prototypical concepts, their prototypical senses tend to be similar but to heterogeneous values, with values oscillating between 0.56 (Spanish *plaza* and French *place*) and 0.74 (Italian *piazza* and Portuguese *praça*). For geonyms expressing the ‘street’ prototypical concepts, values are more

homogenous, as they are enclosed in the interval between 0.6 (e.g. Italian *strada* and French *rue*) and 0.7 (e.g. French *rue* and Portuguese *rua*). Geonyms expressing the 'alley' prototypical concepts include the lowest, more heterogeneous values: between 0.094 (French *alee* and Spanish *callejón*) and 0.43 (Italian *vicolo* and French *alee*).

More generally, these results confirm that language-specific prototypical senses are similar but not identical, given the language-specific feature lists forming these senses. The fact that degrees of similarity tend to be relatively high for the concepts 'square' and 'street' suggests that speakers across languages and contexts of use tend to conceive these concepts in converging manners. For the concept 'alley', speakers have language-specific prototypical senses with low degrees of similarity. What counts as an 'alley' and can be described and named as such thus seems a matter of cultural variation and personal perspective that reflects how speakers associate different language-specific senses with the corresponding geonyms. Overall, speakers of these Romance languages seem to converge on how they conceive squares and streets, but seem to have more subjective views on alleys. The uses of geonyms and toponyms expressing these concepts across our sample reflect these complex patterns of cultural, linguistic, and individual variation.

We also observe that the fact that degrees of similarity do not approach semantic identity suggests that speakers associate contingent, culture- and language-specific features to these concepts. In other words, speakers use geonyms (e.g. French *rue* for 'street', Italian *piazza* for 'square', Portuguese *beco* for 'alley') and toponyms including these geonyms to express nuanced views of language-general concepts. While these views tend to converge for some concepts (here, 'square' and 'street'), they apparently are dissimilar for other concepts (here, 'alley'). Nevertheless, once we quantify and measure the feature lists forming these senses, we can also quantify their degree(s) of (dis)similarity. Our third prediction thus confirms the possibility of identifying the semantic properties connecting these geonyms and place concepts across languages and contexts.

Discussion

We believe that we can summarise the theoretical and empirical import of our study in four points.

First, our predictions suggest that our target geonyms, and thus the toponyms including these geonyms, express richly descriptive place concepts as combinations of possibly high numbers of semantic features. One can then distinguish between contingent and culturally/temporally stable features as two minimal types forming these concepts. One possible interpretation of these findings is as follows. Speakers can associate a wide domain of properties to places and place types such as squares, alleys, and streets. Cultural, social, psychological, and possibly individual factors influencing place concept definition can thus find an expression in language as those features that we have labelled as contingent. However, certain features repeatedly occur across languages and across contexts of use, defining (more) stable or "core" features of these concepts. They thus determine "collective", prototypical, language-specific place concepts and the geonyms/toponyms expressing them, i.e. place concepts shared among speakers of a language.

Our study thus confirms and extends the findings in cross-disciplinary cognitivist proposals (e.g. Rezsegi 2012, 2020a, b, Peng et al. 2020, Desai et al. 2023, Purves et al. 2019, Hamzei et al. 2020), because they show that individuals use place concepts as nuanced semantic representations of places. These results also confirm and refine the classical model of "place" as a complex concept (e.g. Cresswell 2014, Malpas 2018), and suggest that this

model cross-linguistically applies to toponyms from a broader set of Romance languages. Place concepts can vary considerably across languages from one linguistic family, for they can express the highly contingent viewpoints of speakers of a language about places. At the same time, however, they share some central culturally and socially stable semantic features constituting these concepts. Geonyms and toponyms expressing these concepts thus reflect this similarity in their uses. Our study confirms this model of place concepts, but also extends it to cross-linguistic and cross-contextual dimensions of analysis, which were hitherto mostly unexplored aspects of the semantics of toponyms.

Second, our predictions suggest that the rich descriptive content of the target geonyms correlates with contexts of use in a strong but nuanced manner. General uses tend to correlate with lists in which contingent feature types are numerous and prevalent; specialised uses include lower numbers of contingent features, and term features include few such numbers. Conversely, stable features represent the most numerous types in term uses, but progressively decrease in relevance and numbers in specialised and general uses. Nevertheless, for each language, one can individuate prototypical senses denoting place concepts as combinations of context-general features. The place concepts 'square', 'street' and 'alley' thus have culture-, language- and context-specific declinations that reflect how speakers of these Romance languages flexibly interpret these concepts. Speakers can name and describe squares, streets, and alleys in subtly or even considerably different manners across different modes of communication.

These data thus confirm the cross-disciplinary claims that (textual) context of use can affect place concepts' use and prototypical aspects (e.g. Terminology, Lexicology, ten Hacken 2015, 2018, San Martín 2022b, San Martín et al. 2020; GIS and ontology, Ballatore, Bertolotto and Wilson 2014, 2015). However, our view of prototypical uses partially diverges from recent proposals in Lexicology studies (e.g. San Martín 2022b, San Martín et al. 2020). These and other proposals suggest that prototypical uses consist of stable semantic features, with term uses (e.g. encyclopaedic definitions) potentially providing ideal examples. We instead propose that prototypical uses can emerge when one looks at what language-specific semantic features crop up across contexts, thus potentially including contingent features. Our reasoning is that, if speakers of a language invariably associate a contingent feature to a concept (e.g. 'beautiful' for 'square' in Italian and French), then this feature develops a dual nature. It is prototypical (i.e. context-free) in a language, though culture-specific and thus contingent cross-linguistically.

Third, our predictions suggest that the complexity of place concepts also modulates their degree(s) of semantic similarity. We acknowledge that we have focused only on prototypical senses to analyse similarity and variation in a clear-cut manner, leaving more fine-grained comparisons aside (e.g. similarity among general uses of 'square' concepts). However, we observe that this more restricted view provides us with one important result. We can show that language-specific prototypical senses for geonyms tend to be highly similar, since they share cross-linguistic, (more) stable semantic features. Variation seems to stem from (more) contingent features providing culture- and language-specific contributions to the senses that these geonyms and toponyms can express. Thus, variation and similarity act as two sides of the same phenomenon: cross-linguistically motivated semantic relatedness.

These data thus confirm claims in GIS works focusing on place ontologies that geographic nouns, geonyms and toponyms can display forms of sense similarity (Ballatore, Bertolotto and Wilson 2014, 2015, Zhu et al. 2017, Ballatore, Wilson and Bertolotto 2013, Chen et al. 2018b, Ursini and Samo 2022). They however

extend these claims along the two novel dimensions of cross-linguistic and cross-contextual (semantic) analysis. The results also provide evidence that one can investigate the semantic properties of toponyms/geonyms via data extraction from textual sources, with cross-linguistic (empirical) coverage: in our case, SketchEngine, Wikipedia and Wikidata. They also provide evidence that textual sources can operate as different types of contexts of use. Thus, semantic analysis can find important allies in multi-lingual, multi-contextual source types.

Fourth, our results can license further predictions regarding place concepts, because they define a linguistic model of place concepts and their semantic properties. We propose three such predictions. First, if other languages (e.g. English, Chinese) include these place concepts, then their language-specific features would belong to the model (e.g. English 'square' may include the 'beautiful' feature). Second, if other languages include these place concepts, then their language-specific features could expand the model (e.g. Chinese *guǎngchǎng* 'square' may express the feature 'immense'). Third, if place concepts can change over time, then they can involve the inclusion of missing features from the model (e.g. Portuguese *praça* associated with a 'private' feature in the past, or developing this feature in the future). We do not test these predictions here, however; we move to the conclusions.

Conclusions

The goal of this paper has been to offer a model of semantic similarity among toponyms denoting the same place concepts. We offered a cross-linguistic and context-sensitive analysis of three place concepts (i.e. 'square', 'street', 'alley') and their realisation in five Romance languages: French, Italian, Portuguese, Spanish, and Romanian. We thus have obtained a better understanding of these concepts by answering three research questions about feature types in place concepts, their modulation by the context, and their cross-linguistic similarity. We acknowledge that our proposal has a shortcoming in providing a narrow focus on three concepts, even if we have operated at a cross-linguistic, cross-contextual level. Note, however, that a broader empirical coverage (i.e. the analysis of more place concepts) would require more resources but also space for the presentation of such broader data sets. Length requirements have prevented us from reaching such coverage.

Nevertheless, we believe that by reaching this goal, we have proven the validity of expanding toponym-based research beyond its usual Anglo-centric focus. Future research building on this paper's cross-linguistic model will possibly offer further evidence in support of this claim. We believe that aside offering a broader coverage of empirical data, our study provides future avenues of research intersecting with current research lines in semantics, among various fields. For instance, the predictions outlined at the end of Section "Discussion" suggest that our model of place concepts shares certain properties with the "semantic map" models found in lexical and semantic typology (e.g. Georgakopoulos and Polis 2022, Ursini and Zhang 2025). This convergence is not surprising, since we ultimately build maps of these concepts that are structurally like those found in semantic maps models. Given the complexity of the topic, however, we defer explorations to future work.

Data availability

The datasets generated during and/or analysed during the current study are available in the OSF repository: <https://doi.org/10.17605/OSF.IO/DSVVKR>.

Received: 14 September 2025; Accepted: 24 March 2026;

Published online: 15 April 2026

Notes

- 1 We define a *semantic feature* as a basic unit of meaning that form the concepts that linguistic units can express. See Ursini and Zhang (2025) for discussion and references. Following standard linguistic practice, we use single quotes to mark semantic features and/or concepts that linguistic items (e.g. toponyms) may express.
- 2 Geo-dialectal dialects/standards (e.g. Belgian French, Brazilian Portuguese) often include subtly distinct uses of toponyms whose study would lead us beyond our goal. Hence, we leave these forms of variation aside.
- 3 We used the definitions for these concepts provided in Wikidata. These are: *Alley*: <https://www.wikidata.org/wiki/Q1251403>; *Street*: <https://www.wikidata.org/wiki/Q79007>; *Square*: <https://www.wikidata.org/wiki/Q174782>. Last accessed September 5, 2025.
- 4 Purpose-specific corpora certainly exist (e.g. the TRAINS corpus of guided dialogues: Allen et al. 2000). We can however safely leave them aside, as they do not play a role in our study.
- 5 In the following examples we only report data scraped automatically from Wikipedia pages since their source could be easily cited and retrieved.
- 6 [https://it.wikipedia.org/wiki/Piazza_dei_Signori_\(Padova\)](https://it.wikipedia.org/wiki/Piazza_dei_Signori_(Padova)) last access,
- 7 https://fr.wikipedia.org/wiki/Rue_de_la_Loi (last access, 31.12.2025).
- 8 https://es.wikipedia.org/wiki/Calle_C%C3%ADster (last access, 31.12.2025).

References

- Allen J, Ferguson G, Miller BW, Ringger EK, Teresa SZ (2000) Dialogue systems: From theory to practice in trains-96. In Dale R, Moisl H, Somers H (eds.), *Handbook of natural language processing*, 347–376. New York: Marcel Dekker
- Ballatore A, Bertolotto M, Wilson DC (2014) An evaluative baseline for geo-semantic relatedness and similarity. *Geoinformatica* 18:747–767
- Ballatore A, Bertolotto M, Wilson DC (2015) A structural-lexical measure of semantic similarity for geo-knowledge graphs. *ISPRS International Journal of Geo-Information* 4:471–492
- Ballatore A, Wilson DC, Bertolotto M (2013) Computing the semantic similarity of geographic terms using volunteered lexical definitions. *International Journal of Geographic Information Sciences* 27:2099–2118
- Ballatore A (2016) Prolegomena for an ontology of place. In Onsrud H, Kuhn W (eds.), *Advancing geographic information science*, 91–103. GSDI Association Press
- Blair D, Tent J (2015) Feature terms for Australian toponymy. ANPS Technical Proposal 2
- Blair D, Tent J (2021) Feature terms for Australian toponymy. *Names* 69. <https://doi.org/10.5195/names.2021.2260>
- Boamfă I (2018) Considerations related to the mapping of Romanian anthroponyms and toponyms. *Diacronia* 8:118–119. <https://doi.org/10.17684/i8A118en>
- Bosson G (2016) Classifications. In A Ledgeway (ed.), *The Oxford guide to the Romance languages*, Oxford. <https://doi.org/10.1093/acprof:oso/9780199677108.003.0006>
- Bullinaria JA, Levy JP (2007) Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods* 39:510–526
- Canter D (2012) *Facet theory: Approaches to social research*. Springer Science Business Media
- Carvalhinhos P, de J (2009) Variantes lexicais na toponímia portuguesa: os elementos genéricos (entidades geográficas) denominados. estudo de caso: Diferenças terminológicas entre português do Brasil e português europeu. [lexical variants in Portuguese toponymy: generic terms (geographical entities) denominated. Study of a case: terminological differences between Brazilian and European Portuguese]. *Cadernos do CNLF* 11:177–194
- Chen H, Vasardani M (2018a) Georeferencing places from collective human descriptions using place graphs. *Journal of Spatial Information Science* 17:31–62. <https://doi.org/10.5311/JOSIS.2018.17.417>
- Chen Z, Song J, Yang Y (2018b) An approach to measuring semantic relatedness of geographic terminologies using a thesaurus and lexical database sources. *ISPRS International Journal of Geo-Information* 7:98. <https://doi.org/10.3390/ijgi7030098>
- Choo S (2015) Elements of cultural heritage in Korean geographical names. In Sungjae Choo (ed.), *Geographical names as cultural heritage*, 139–154. Seoul
- Church K, Hanks P (1990) Word association norms, mutual information, and lexicography. *Computational Linguistics* 16:22–29
- Clackson J (2016) Latin as a source for the Romance languages. In A Ledgeway (ed.), *The Oxford guide to the Romance languages*, Oxford. <https://doi.org/10.1093/acprof:oso/9780199677108.003.0001>
- Coates R (2013) Place-names and linguistics. In Carroll J, Parsons D (eds.), *The Oxford guide to the Romance languages*, Nottingham: English Place-Name Society
- Coates R (2017) The meaning of names: A defence of the pragmatic theory of properhood (tptp) addressed to van Langendonck, Anderson, Colman and McClure. *Onoma* 52:7–26
- Cresswell T (2014) *Place: An introduction*, 2nd edition

- Davies C (2013) Reading geography between the lines: Extracting local place knowledge from text. In Tenbrink T, Stell J, Galton A, Wood Z (eds.), *International conference on spatial information theory*, Berlin: Springer. <https://doi.org/10.1007/978-3319-01790-718>
- De Stefani E (2016) Names and discourse. In Carole Hough (ed.), *The oxford handbook of names and naming*, Oxford: Oxford Academic. <https://doi.org/10.1093/oxfordhb/9780199656431.013.48>
- Desai, Rutvik H, Tadimeti U, Riccardi N (2023) Proper and common names in the semantic system. *Brain Struct Funct* 228:239–254. <https://doi.org/10.1007/s00429-022-02593>
- Georgakopoulos, Thanasis, Polis S (2022) New avenues and challenges in semantic map research (with a case study in the semantic field of emotions). *Zeitschrift für Sprachwissenschaft* 41:1–30
- George J (2023) Using the anps typology to unearth the relationship between Japanese sign language (jsl) endonymic toponym distribution and regional identity. *Names* 71:1–19. <https://doi.org/10.5195/names.2023.2563>
- Giacomini L (2023) Specialised language and conceptual knowledge in lexicographic portals. *Lexicographica* 29:143–164
- Giacomini L (2024) Ontologies and knowledge representation in terminology: Present and future perspectives. *Applied Ontology* 19:1–15
- Gordón Peral M (2010) *Toponimia de España: Estado actual y perspectivas de la investigación* Berlin: De Gruyter. <https://doi.org/10.1515/9783110233490>
- Gries, Stephan TH (2009) What is corpus linguistics? *Language and Linguistics Compass* 3:1–17
- Hamzei E, Winter S, Tomko M (2020) Places facets: A system-atic literature review. *Spatial Cognition & Computation* 20:33–81. Doi:33–81. 1080/13875868.2019.1688332
- Hamzei E, Winter S, Tomko M (2021) Templates of generic geographic information for answering where-questions. *Int J Geograph Inform Sci* <https://doi.org/10.1080/13658816.2020.1869977>
- Harispe S, Ranwez S, Janaqi S, Montmain J (2015) *Semantic similarity from natural language and ontology analysis*. In ? 2–7. Nîmes, France: Morgan Claypool
- Hollenstein, Livia, Purves RS (2010) Exploring place through user-generated content: Using flickr tags to describe city cores. *Journal of Spatial Information science* 1:21–48
- Hu Y (2018) Geospatial semantics. In Bo Huang, Thomas J. Cova, Ming-Hsiang Tsou & et al. (eds.), *Comprehensive geographic information systems*, 80–94. Oxford: Elsevier. <https://doi.org/10.1016/B978-0-12-409548-9.09597-X>
- Kijania-Placek K (2021) Names of places. *Semiotica* 240:187–210. <https://doi.org/10.1515/sem-2021-0020>
- Kilgariff, Adam, Baisa V, Bušta J, Jakubíček M, Kovár V, Michelfeit J, Rychlý P, Vít S (2014) The sketch engine. *Lexicography* 1:7–36
- Klaus-Jürgen E, Ballard E, Elswö DE (eds.) (2010) *Encyclopedic dictionary of landscape and urban planning multilingual reference book in english, spanish, french, and german*. Berlin: Springer
- Koptjevskaja-Tamm M, Miestamo M, Börstell C (2024) A cross-linguistic study of lexical and derived antonymy. *Linguistics* <https://doi.org/10.1515/ling-2023-0140>
- Köhnlein B (2015) The morphological structure of complex place names: The case of dutch. *Journal of Comparative Germanic Linguistics* 1:183–212. <https://doi.org/10.1007/s10828-015-9075-0>
- L'Homme M-C, Marshman E, Martín AS (2022) Environment terms and translation students: A reading based on frame semantics. *Babel. Revue internationale de la traduction / International Journal of Translation* 68:1–NN. <https://doi.org/10.1075/babel.00254.lho>
- Li Y, Bandar ZA, McLean D (2003) An approach for measuring semantic similarity between words using multiple information sources. *IEEE Trans. Knowl. Data Eng.* 15:871–881
- Li Z (2024) Chinese place names. In L Jiao (ed.), *The routledge handbook of chinese language and culture*, 1–60. <https://doi.org/10.4324/9781315167800-15>
- Malpas J (2018) *Place and experience: A philosophical topography*. Sydney: Routledge
- Morabito JS (2020) *La géonymie originelle de la celtique primitive de l'atlantique à l'adriatique*. Paris: Edilivre
- Navigli R, Martelli F (2019) An overview of word and sense similarity. *Natural Language Engineering* 25:693–714
- Nyström S (2016) Names and meaning. In Carole Hough (ed.), *The oxford handbook of names and naming*, Oxford: Oxford Academic. <https://doi.org/10.1093/oxfordhb/9780199656431.013.26>
- Peng J, Strijker D, Wu Q (2020) Place identity: How far have we come in exploring its meanings?. *Frontiers in Psychology* 11:nn–nn. <https://doi.org/10.3389/fpsyg.2020.00294>
- Purves R, Winter S, Kuhn W (2019) Places in information science. *Journal of the Association for Information Science and Technology* 70:1173–1182
- Rada R, Mili H, Bicknell E, Blettner M (1989) Development and application of a metric on semantic nets. *IEEE Tran. Syst. Man Cybern.* 19:17–30
- Recchia G, Jones N (2009) More data trumps smarter algorithms: Comparing pointwise mutual information with latent semantic analysis. *Behavior research methods* 41:647–656
- Reshmi N, Shreelekshmi R (2023) Sentence similarity using modified latent semantic analysis and semantic relations. *XX* <https://doi.org/10.3233/FAIA230766>
- Rezsegi K (2012) Cognitive approaches to hungarian toponymy. *Onoma* 47:367–379
- Rezsegi K (2020a) Name — culture — identity. the semantic structure of names. In Luisa Gaizazzo, Richard Coates & Maria Azaryahu (eds.), *Naming, identity and tourism*, 154–167. Cambridge Scholars Publishing
- Rezsegi K (2020b) Toponyms and spatial representations. *Onomastica* 64:23–40. <https://doi.org/10.17651/ONOMAST.64.4>
- Eugenia R, Dmitry P (2018) Syntactic features of proper names: The generativist approach(es) to properhood. *[Onomastics Questions]* 15:7–35
- Samo G, Ursini F-A (2022) Exploring dynamic on-line gazetteers to map variation in the syntax of Italian urbanonyms. In Tommaso Balsemin, Irene Caloi, Jacopo Garzonio, Nicolas Lamoure, Francesco Pinzin, Emanuela Sanfelici (Eds.), *Quaderni di lavoro ASIT* 24. 407–423. (<http://asit.maldura.unipd.it/papers.html>)
- Samo G, Ursini F-A (2023) Geographical Maps meet Place Names where Languages meets Dialects: The Case of Italian. *Forum Italicum* 57:1019–1040
- Samo G, Ursini F-A (2024) The Semantics of Generic Terms in Toponyms: Formal Semantics meets multi-source Big Data. In Hu X, Hu Y, Jens K, Stock K (eds.), *Proceedings of GeoExt2024 (=lecture Notes in Computer Science XXXX)*, 79–91. Berlin: Springer
- San Martín A (2022a) Contextual constraints in terminological definitions. *Frontiers in Communication* 7:1–nn. <https://doi.org/10.3389/fcomm.2022.885283>
- San Martín A (2022b) A flexible approach to terminological definitions: Representing thematic variation. *International Journal of Lexicography* 35:53–74. <https://doi.org/10.1093/ijl/ecab013>
- San M (2017) Recent advances in ecollexicon. *Dictionaries: Journal of the Dictionary Society of North America* 38:96–115. <https://doi.org/10.1353/dic.2017.0004>
- San M (2020) Contextual constraints in terminological definitions. *Presente y futuro de la base de conocimiento terminológico* *EcoLexicon* 49:174–202. <https://doi.org/10.7764/onomazein.49.09>
- Schlucker B (2024) Von donaustrum zu donauwelle. die entwicklung der eigennamenkomposition von 1600–1900. *Zeitschrift für Germanistische Linguistik* 2:238–268. <https://doi.org/10.1515/zgl-2020-20020>
- Schlucker B, Ackermann T (2017) The morphosyntax of proper names: An overview. *Folia Linguistica* 51:309–339. <https://doi.org/10.1515/flin-2017-0011>
- Stolz T, Levkovych N (2020) Grammatical versus onymic classifiers: First thoughts about a potentially interesting topic. In Levkovych N, Nintermann J (eds.), *Aspects of the grammar of names. empirical case studies and theoretical topics*, 140–165. Munich: LINCOM Europa
- Syed Z, Finin T, Joshi A (2021) Wikipedia as an ontology for describing documents. *Proceedings of the International AAAI Conference on Web and Social Media* 2:136–144. <https://doi.org/10.1609/icwsm.v2i1.18627>
- ten Hacken P (2015) *Terms and specialized vocabulary: Taming the prototypes*. In ? Amster-dam Philadelphia: John Benjamins
- ten Hacken P (2018) Terms between standardization and the lexicon. *Roczniki Humanistyczne* 66:100–118. <https://doi.org/10.18290/rh.2018.66.11-4>
- Tenbrink T (2019) The language of place: towards an agenda for linguistic platial cognitive research. In Westerholt R, Mocnik F-B (eds.), *Proceedings of the 2nd international symposium on platial information science (platial'19)*, 5–12. <https://doi.org/10.5281/zenodo.3628849>
- Tent J, Blair D (2011) Motivations for naming. The development of a toponymic typology for australian placenames. *Names* 59:67–89
- Turki H, Denny Vrandečić HH, Adel I (2017) Using wiki- data as a multi-lingual multi-dialectal dictionary for arabic dialects. In 2017 IEEE/ACS 14th international conference on computer systems and applications (aiccasa), 437–442. <https://doi.org/10.1109/AICCSA.2017.115>
- Ursini F-A, Long H (2020) Urbanonyms in Italian: Morpho-Semantic Properties and Distribution. *Names: A journal of onomastics* 55:1–15. <https://doi.org/10.1080/00277738.2020.1811049>
- Ursini F-A, Samo G (2022) Names for urban places and conceptual taxonomies: The view from Italian. *Spatial Cognition & Computation* 22:264–292. <https://doi.org/10.1080/13875868.2021.1954186>
- Ursini F-A, Samo G (2023a) Extracting toponyms from OpenStreetMap: A cross-linguistic perspective. In Hu, Xuke, Yianji Hu, Jens Kerstens, Kirstin Stock (eds.), *Proceedings of GeoExt2023 (=lecture Notes in Computer Science 4321)*, 110–120. Berlin: Springer
- Ursini F-A, Samo G (2023b) A semantic model for generic terms and place nouns. In AA VV. (eds.), *Proceedings of SDSS 2023*, 1–10. <https://doi.org/10.25436/E2V30M>
- Ursini F-A, Samo G (2025) Extracting Toponyms from OpenStreetMap and Other Gazetteers: Comparing representational accuracy in multilingual contexts. *HSSCOMMS* 12:1–16. <https://doi.org/10.1057/s41599-025-05025-1>

- Ursini F-A, Zhang Y(S) (2025) Syntactic Cartography meets semantic maps: charting the polysemy of Italian simple prepositions. *The Linguistic Review* 42:1–49. <https://doi.org/10.1515/tlr-2025-0009>
- Vasardani M, Winter S (2013) Locating place names from place descriptions. *International Journal of Geographical Information Science* 27:2509–2532
- Vrandečić D (2013) The rise of wikidata. *IEEE Intelligent Systems* 28:90–95
- Vulić I, Ponti I, Leviant E, Majewska O, Malone M, Reichart R, Baker S, Petti U, Wing K, Bar E, Poibeau T, Korhonen A (2020) Multi-simlex: A large-scale evaluation of multilingual and cross-lingual lexical semantic similarity. *Computational Linguistics* 46:1–73
- Wearne S, Riedy C (2024) Whose “place” is it? using corpus-based techniques to sketch place-based sustainability discourses in public and academic forums. *Sustainability Science* 19:883–904
- Zhang W, Gelernter J (2014) Geocoding location expressions in twitter messages: A preference learning method. *J Spat Inform Sci* 47–90. <https://doi.org/10.5311/JOSIS.2014.9.170>
- Zhou S, Stephan Winter MV, Zhou S (2014) Place descriptions by landmarks. *Journal of Spatial Science* 62:47–67. <https://doi.org/10.1080/14498596.2016.1196623>
- Zhu Y, Zhu A-X, Song J, Zhao H (2017) Multidimensional and quantitative interlinking approach for linked geospatial data. *International Journal of Digital Earth* 10:1–21

Author contributions

G.S. wrote the manuscript text and collected the data. F.-A. U. designed, wrote the manuscript and the prepared the references. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1057/s41599-026-07178-z>.

Correspondence and requests for materials should be addressed to Giuseppe Samo or Francesco-Alessio Ursini.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026