Names | A Journal of Onomastics

Mapping Digital Discourses of the Capital Region of Finland: Combining Onomastics, CADS, and GIS

Jarmo Harri Jantunen
University of Jyväskylä, FINLAND

Terhi Ainiala
University of Helsinki, FINLAND

Salla Jokela
Tampere University, FINLAND

Jenny Tarvainen
University of Jyväskylä, FINLAND

Abstract

This article discusses the three Finnish city names *Helsinki*, *Espoo*, and *Vantaa,* and the urban discourses that surround them. The study reveals patterns of socio-spatial differentiation by examining what meanings people attach to these capital region cities and investigating how these meanings are expressed in online discourses. Using the methodological approach of corpus-assisted onomastics (CAO), this study incorporates onomastics, geographical information systems (GIS), and corpus linguistics. This interdisciplinary research also examines how corpus-assisted discourse studies (CADS) and GIS can be combined to reveal and visualize the contextual information and discursive patterns of toponyms. For this investigation, a data set of 2.7 billion words was collected from the Suomi24 Corpus, one of the biggest discussion fora in Finland. The use of social media corpus as a data source increases the authenticity of this research, as the data was not collected specifically for this study. The analysis reveals that the most frequent digital discourses about the cities refer to places and directions, housing, and mobility. The occurrences differ quantitatively and qualitatively between each city. This paper paves the way for future onomastic studies to research actual name usage using this new methodology. Knowledge gained from such research may not only enrich the field of onomastics, but also facilitate more socially sustainable urban planning.

Keywords: corpus-assisted onomastics, corpus-assisted discourse studies, CADS, geographic information systems, GIS, discourse prosody, toponyms, corpus linguistics, Finland

Introduction

This article brings together onomastics, geographical information systems (GIS), and corpus linguistics. In their research on discourses related to the toponyms *Hesa* and *Stadi* (slang terms for the city of Helsinki), Ainiala and Jantunen (2019, 2018) introduce the concept and method of corpus-assisted onomastics (CAO). They define CAO as

> a field in onomastic research, where electronic databases, i.e., corpora, are employed as data, where analysis is based on corpus research methods (such as wordlist analysis, concordancing and keyword analysis) and where the subject of the research is the prevalence of names, their usage in textual contexts (e.g., in collocations and other phraseological relations, and genres), as well as regional and local variation. (2019, 58; Ainiala & Jantunen's translation [2022]).

Corpus linguistics and onomastics have also been used together by other researchers, most recently by Motschenbacher (2020a, 2020b). In 2020a, Motschenbacher presents a theoretical discussion of the subject and introduces corpus methods that can be used in onomastic studies, such as the collocation analysis used by Ainiala and Jantunen (2019), and keyword analysis used in the present study. Ainiala and Jantunen (2019) point out that big data opens pathways to research that cannot be as easily explored through qualitative methods. Together with Motschenbacher (2020a), they also claim that corpus linguistic methods have great potential to enrich name studies research. Corpora compiled from authentic discourse can, for example, provide important insights into real language use, while statistical analyses can be effectively applied to large amounts of quantitative data to make the findings more reliable.

The method used in the present study is corpus-assisted discourse studies (CADS). This approach combines qualitative and corpus methodologies to examine language patterns such as collocations and keyword frequencies to reveal discourse types (Partington et al. 2013). Despite its potential utility, with the exception of Ainiala and Jantunen (2019) and Poole (2018), CADS has rarely been used in onomastics, however. Another relatively recent approach that combines onomastics and electronic data is "Visual GISting" (Gregory & Hardie 2011) or "Geoparsing" (Gregory et al. 2015). This method exploits corpora and spatial data provided by geographical information systems (GIS) to provide information about language use. In this study, we combine CADS and Visual GISting by first identifying discourses in Suomi24, a social media corpus of authentic Finnish discourse. These stretches of discourses are then analyzed and visualized on the basis of the associated spatial information.

Our study focuses on discourses that relate to the "Finnish Capital Region", a functional urban area that consists of four municipalities in Southeastern Finland: Helsinki, Espoo, Vantaa, and Kauniainen. Helsinki is the capital of Finland and is the biggest city in the nation-state with a population of 0.6 million. Espoo and

Vantaa are the nearest cities to Helsinki, and are the country's second and fourth most populous municipalities, respectively (Statistics Finland 2021). Altogether, the Capital Region currently has a population of 1.2 million. The total area of the region is approximately 1,500 km², half of which is made up of water. Although the Capital Region makes up only 0.2% of Finland's total area, it is home to approximately one fifth of Finland's inhabitants (Statistics Finland 2021). The population of the Region is also growing rapidly due to the expanding labor markets and good traffic connections to other urban centers around the Baltic Sea.

Using online discussion forum data, our study investigates the ways people attach meanings to the toponyms used for the neighboring cities of Helsinki, Espoo, and Vantaa. Our analysis focused on discursive associations made with the official names of these three cities (*Helsinki*, *Espoo*, and *Vantaa*) as well as two commonly used slang names for Helsinki, *Hesa* and *Stadi* (see Ainiala & Lappalainen 2017). Information about how these official and unofficial city names are used will contribute to a more nuanced understanding of the discursive processes shaping this region. With this aim in mind, this study posed four research questions:

1. What are the online digital discourses that relate to the names of these three capital region cities?
2. How do the discourses differ from one another?
3. What do these discourses reveal about the patterns of socio-spatial differentiation in Finland's capital region?
4. How does combining CADS and GIS work to reveal and visualize the contextual information and discoursal patterns of toponyms?

As the above research questions allude, it is hoped that the multidisciplinary approach used in this onomastic study can help shed new light on not only the meanings attached to the toponyms but also the various stances taken towards them (see, e.g., Ainiala & Östman 2017). This methodological approach is also relevant to geography and other social sciences, as it may allow us to identify and analyze the discursive processes of place-framing (Martin 2003). In doing so, it may be possible to gain new knowledge about "urban imaginaries" or "interpretive grids through which we think about, experience, evaluate, and decide to act in the places, spaces, and communities in which we live" (Soja 2000, 324). Knowledge about these processes is important, because it can help us understand how the construct of place forms the basis of both collective identities and political action. At the same time, it may help reveal how the discursive framing of such places connects with the material reality of those localities.

Background: Onomastics, Corpora, GIS, Discourses, and Geography

Corpus Data in Onomastics

Comprehensive text corpora have seldom been used as data source in onomastics. Instead, the digital data most often used in this field has taken the form of name lists varying in size and quality (see Motschenbacher 2020b). The subjects that have been addressed in these studies include the morphophonological and lexical structure of personal names (e.g., MacAulay et al. 2019; Lappe 2002; Pagan 1998); combinations of descriptive appositives and proper names (Bjorge 2003); the topographical words included in place names (Nurminen 2012); surnaming patterns (Laversuch 2011); and most recently, US toponyms that mark the history of Native American peoples (Nick 2017).

Increasingly, the internet has also been used as a source of digital data. The increasing popularity of this choice for data collection is understandable, since the internet may provide a huge amount of onomastic material. Ohlander and Bergh (2004), for example, examined the use of the name *Taliban* in US and British newspapers; Tanaka (2016) studied Japanese surnames and their varying written forms in social media; and Hämäläinen (2019) looked at internet usernames. Meanwhile, in literary onomastics, researchers have increasingly compiled their material from digitalized versions of literary works. These studies have considered, for instance, the uses and functions of proper names in translated novels (see, e.g., Dalen-Oskam 2013; Tuñón 2013).

There are also examples of text corpora being used for onomastic research. Sinclair (2004), for example, used the Bank of English to analyze the choice of preposition in the titles of organizations (e.g., *the European Society of Phraseology*). Likewise, Pierini (2008) used the British National Corpus (BNC) to examine the occurrence of personal and place names in various phraseological units (e.g., idioms, binomials, formulae) across different linguistic registers, while Tse (2004) relied on the BNC to explore the use of definite articles in personal name classes. Most recently, the Corpus of Contemporary American English was used by Motschenbacher (2020b) to analyze the use of definite articles in the names of countries.

Text corpora featuring languages other than English have also been analyzed for onomastic research. A large Norwegian newspaper corpus of 150 million words, for instance, was used by Halverson and Engene (2010) to study the metonymic use of *Maastricht* and *Schengen*, while collocations featuring *America* in a Lithuanian newspaper corpus was examined by Vaičenonienė (2001); and in a similar study, *Arab* and *Saudi*

Arabia was investigated by Kamandulytė (2006) with regard to semantic preferences and collocations. Finally, a large Finnish corpus of online texts, Suomi24, was used by Ainiala and Jantunen (2019) to explore the discourse prosodies of the slang toponyms of *Hesa* and *Stadi* for the Finnish capital of Helsinki.

Corpus Data and Geographic Information Systems (GIS)

Paterson and Gregory (2018, 37) state that "traditionally, corpus linguists have not explored the geographies within texts that they have analysed". However, visual GISting, which involves identifying place names within textual data and linking them to spatial coordinates, has been used in studies dating back to as early as 2010.[1] In the spatial humanities, Murrieta-Flores et al. (2014), for example, automatically extracted place names in texts, assigned them geographic coordinates, and produced visualizations of the toponymic data. Although this approach can help to determine which places are mentioned in the data, it disregards what is said about them (Gregory & Hardie 2011, 305). In other words, this strategy fails to take into account the discourses linked to the places identified.

Discourse-oriented studies seeking to answer the "what" have frequently taken advantage of CADS research methods, though this practice is not always explicitly highlighted. For example, Gregory and Hardie (2011) relied on the automatic extraction of semantic tags to identify the discursive association between the construct of MONEY and the place name *Dunkirk*, while Donaldson et al. (2017) used a similar method for their geographical collocation analysis of the association between Lake District names and language users' associations with the constructs of BEAUTY and PICTURESQUE. Combining corpus linguistics and GIS techniques in the spatial humanities is also addressed by Poole (2018) in the investigation of environmental discourses related to the "Rosemont Copper Mine Debate" in Arizona. In this work, Poole concluded that "social media provides a vast geo-referenced data source and could be used to further display the importance of place and its discursive production in environmental communication" (538). Although Poole is specifically referring to environmental communication, the assertion is generalizable to other discussions as well.

Discourses and Geography

Human geographers have theorized about the connection between discourse and space since the "textual" or "discursive" turn of the 1980s, when their attention turned towards the importance of language beyond everyday meaning-making. At the time, scientific attention was also directed toward the ways in which power manifests itself in texts, images, and other cultural products (e.g., Barnes & Duncan 1992).

It is now widely acknowledged that people's ideas about spaces and places are both discursively produced and connected to material reality. Shared ideas "become the frame of mind for social agents as well as being the outcome of the historical and contextual conditions under which they are articulated" (Richardson & Jensen 2003, 15). One reason why discourses matter is that they are inextricably bound to the social processes through which spatial entities such as cities and neighborhoods are (re)produced. Discourses also play an important role in the everyday (re)production of urban space, so that some conceptions of the city become hegemonic by virtue of their already being mainstream, while others are marginalized. As several studies have shown, the naming of places, together with different place-making perspectives and representations, can create socio-spatial distinctions, borders, and place identities based on a collective understanding of the urban space (e.g., Scott & Sohn 2019; de Koning 2015; Martin 2003). Scott and Sohn (2019), for instance, demonstrated the ways in which the sharing and naming of geographical representations (through, for example, maps, blogs, visitor information) have contributed to the emergence of new neighborhoods such as *Kreuzkölln* in Berlin and *Nyócker* in Budapest.

The use of big data from digital sources has facilitated the convergence of geography and humanities (Gregory & Hardie 2011). This convergence has manifested itself as the "spatial turn" in digital humanities (Presner & Shepard 2015, 247) or the emergence of the "digital geohumanities" (Crang 2015, 351). We follow the tradition of these studies, but also acknowledge the criticism that empiricist approaches may mistakenly presume that data can speak for itself, and is somehow "free of human bias or framing" (Kitchin 2013, 265). Therefore, to help to draw more reliable conclusions from the data, we utilize our "situated knowledge" (Haraway 1996), as Finns who are familiar with the sociocultural context of place-based meaning-making related to Finland's metropolitan region.

Data and Methodology

Data

Our data was extracted from messages posted on Suomi24 discussion forum. One of the most popular Finnish websites, as of 2019, this platform has approximately 2.2 million visits per month (FIAM). Suomi24 has 24 sub-fora and covers thematic areas such as traveling, health, and family. The Suomi24 Corpus was made available through the Language Bank of Finland (META-SHARE). The material collated for this investigation represents postings made between 2001 and 2016, and consists of 2.7 billion words.

For our study, we extracted data for the Vantaa and Espoo sub-corpora that included the names (or lemmas) *Vantaa* and *Espoo*. For the Helsinki sub-corpus, we compiled data from postings with the lemmas *Helsinki*, *Hesa,* and *Stadi*. The sizes of the three sub-corpora were 1.9 million tokens for *Vantaa*, 2.8 for *Espoo*, and 5.1 for *Helsinki*. The reference corpus for the keyword analysis was compiled using systematic sampling in which messages posted in the discussion forum were extracted at one minute past each hour around the clock. In this way, temporal thematic bias in the reference data were largely avoided.

Methodology: from Keywords to Discourses

For this methodology, keywords first needed to be determined. According to Scott and Tribble (2006), "keywords" are those words that occur in the data significantly more often than one would expect them to otherwise occur by chance. In the present study, the keywords were used to determine the semantic context forum users employed to talk about the three capital region cities. The keyword analysis was done using the Keyword List program within the AntConc corpus toolkit (Anthony 2017). The statistical measure chosen for the analysis was the Log-Likelihood test which is suitable and valid for frequency profiling (Rayson & Garside 2000). The first 300 statistically significant keywords for each of the three cities were then analyzed. A frequency of occurrence threshold was set at 50 to eliminate the noise which inevitably results from the repetition of identical postings. This precaution also helped us focus on established usage and meanings.

To uncover the subject of the discussions that made mention of the cities, we analyzed the "discourse prosodies": the repeatedly occurring associations made between the city names and their sets of semantically-related items (see, e.g., Jantunen 2018; Baker et al. 2008; Stubbs 2001). This was done by grouping the keywords into data-driven semantic categories which can then be grouped into either patterns of discourse (Baker 2006) or a coherent discourse function (Hunston 2007). The multiple meanings and referents of the keywords were determined by analyzing concordance lines in the data. Two of this article's authors undertook the analysis. In cases where the meaning was not immediately clear, the entire research team discussed the possible options and came to a consensus. On the rare occasion when no consensus could be reached, the keyword was classified as "other". The classification process resulted in 17 discourse prosody groups (see Figure 1).

After this classification was completed, following the example of previous GIS studies (e.g., Donaldson et al. 2017), we mapped and visualized the discourse prosodies based on geographical information contained in the social media. Finally, after a close reading, we examined the three most wide-ranging discourse prosodies. In this discourse analysis, we thereby complemented the statistical analysis with an interpretation of the possible deeper cultural meanings of the utterances derived from the data.

Results

An Overview of the Discourse Prosodies Associated with the Cities in the Capital Region

The categories and distributions, which emerged from the data, are illustrated in Figure 1. The largest discourse prosody is, by far, the category "places and directions" with 288 keywords (32.00%), followed by "mobility" (107, 11.89%) as the second largest. The difference is statistically significant (p < 0.001^2, z = 10.31) and confirms Ainiala and Jantunen's (2019) finding that place names are typically associated with not only toponyms of cities, towns, neighborhoods and villages, but also citizens, locations (specific and non-specific), spatial adverbs, and points on the compass (see also Table 1). The category "places and directions" was found to be similarly large for all three cities: 32.00% for *Helsinki* and *Espoo*, and 31.00% for *Vantaa*. After "mobility", the third largest category was "housing", and the fourth, "politics and organizations". The percentage shares were less similar in these categories: 14.33% "mobility" for *Helsinki*, 12.00% for *Vantaa,* and 9.33% for *Espoo*; 7.00% "housing" for *Helsinki,* 10.00% for both *Vantaa* and *Espoo*; and 9.00% "politics and organizations" for Helsinki, 6.67% for *Vantaa*, and 10.33% for *Espoo* (see Figure 1). However, the differences were not statistically significant.

In contrast, "crime and justice" (e.g., "poliisi" 'the police') was found to have major discourse prosody for *Vantaa* and *Espoo* with a statistically significant differences compared to *Helsinki* (*Vantaa-Helsinki*: p < 0.001, z = 4.31; and *Espoo-Helsinki*: p < 0.01, z = 2.54). In fact, "crime and justice" was an almost non-existent discourse prosody in the Helsinki data examined (Figure 1). Almost as low in the Helsinki sub-corpus was "commerce", which was also significantly different to both *Vantaa* (p < 0.01, z = 2.80) and *Espoo* (p < 0.01, z = 2.60). Conversely, "people and individuals" was a discourse category significantly associated with *Helsinki*, when compared to *Vantaa* (p < 0.001, z = 5.31) or *Espoo* (p < 0.001, z = 5.05). The other categories proved less common overall. Consequently, there was less difference between cities. The following section will present the findings obtained for the three largest categories of keywords: "places and directions", "mobility", and "housing". Coates, Richard. 2006b. "Properhood."

Figure 1: Distribution of Discourse Prosodies Between the Cities of *Helsinki*, *Espoo*, and *Vantaa*
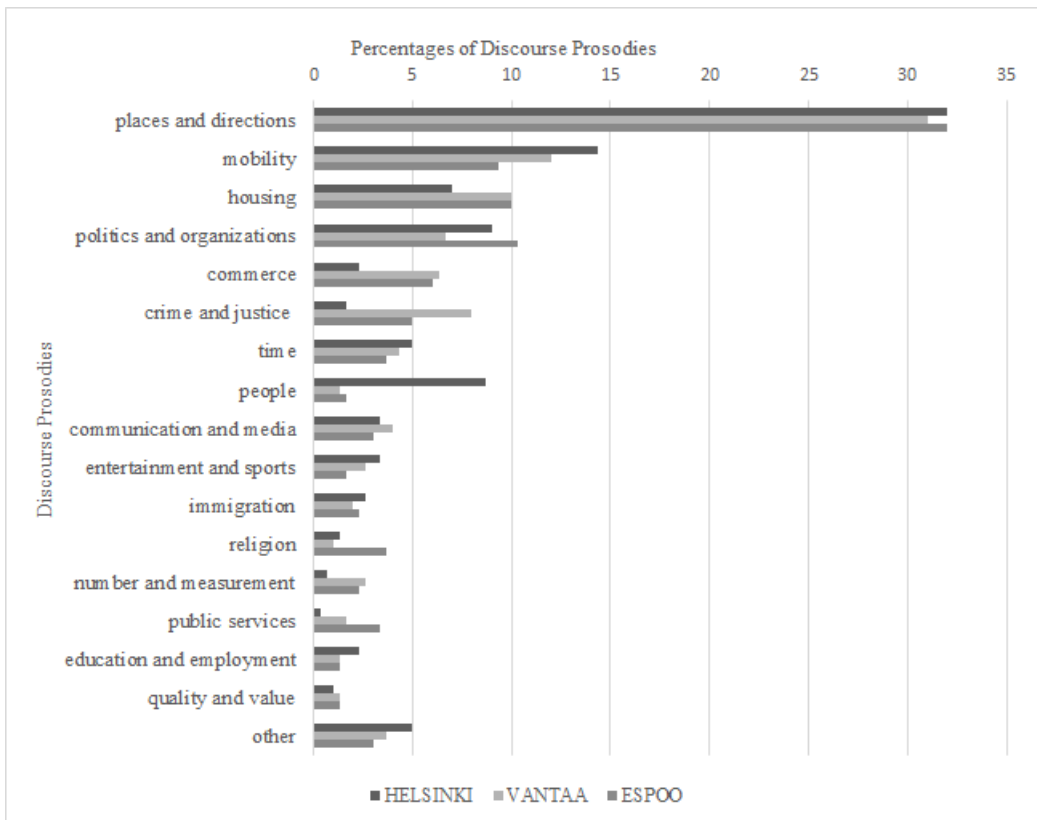
Mapping the Three Most Significant Discourse Prosodies

Places and Directions

In the discourse prosody "places and directions", there were several differences found between some of the sub-categories and keywords they shared. Examples of these sub-categories and keywords are represented below in Table 1. The number of keywords in each subcategory are provided in Table 2. For each sub-category in Table 1 (as in Tables 3 and 5), illustrative examples of keywords are provided. The last column contains keywords that were shared by all three cities, while the other columns show keyword examples that were typical for only one or two of the cities.

Table 1. Sub-Categories of "Places and Directions" and Examples of the Keywords

Sub-Categories	Helsinki	Vantaa	Espoo	Shared Keywords

| Neighborhoods	*Töölö, Hakaniemi, Eira* [neighborhoods of Helsinki], ydinkeskusta 'nuclear center', kantakaupunki 'central urban area'	*Tammisto, Korso* [neighborhoods of Vantaa]; *Malmi, Kontula, Vuosaari* [neighborhoods of Helsinki]	*Tikkurila* [neighborhood of Vantaa]; *Jorvi, Olari, Otaniemi* [neighborhoods of Espoo]; *Kallio, Kamppi, Lauttasaari* [neighborhoods of Helsinki]	*Kontula, Vuosaari* [neighborhoods of Helsinki], keskus 'center', keskusta 'downtown', kaupunginosa 'neighborhood'
Cities, Towns, and Municipalities	*Hanko, Riihimäki, Kotka, Lontoo* 'London', *Tallinna* 'Tallinn', *Tukholma* 'Stockholm', *metropoli* 'metropolis'	*Mäntsälä, Seinäjoki, Loviisa*, maalaiskunta 'rural municipality'	*Siilinjärvi, Forssa, Siuntio*, pääkaupunkiseutu 'capital district'	*Tampere, Turku, Oulu, Lahti*, kaupunki 'town', pääkaupunki 'capital', kunta 'municipality'
Provinces	*Lappi* 'Lapland', *Savo* 'Savonia', *Pohjanmaa* 'Ostrobothnia'	-	-	*Uusimaa* [the southernmost region in Finland]
Countries	*Suomi* 'Finland', *Ruotsi* 'Sweden', maa 'country'	-	*Suomi* 'Finland'	-
Directions	edustalla 'in front of', keskellä 'in the middle of', ulkopuolella 'outside'	vieressä 'next to', sijainti 'location'	vieressä 'next to', sijainti 'location', suunta 'direction'	lähellä 'near', päin 'towards' siellä 'there', täällä 'here'
Unspecified Areas	maaseutu 'countryside', lähistö 'surroundings', lande 'countryside' [slang]	perukka 'far end', lähialue 'neighborhood'	maaseutu 'countryside', lähialue 'surroundings', lähistö 'surroundings', lande 'countryside' [slang]	alue, 'area', paikka 'place', paikkakunta 'locality'
People from a Specific Area	*suomalainen* 'Finn', *helsinkiläinen* 'Helsinkian', *savolainen* 'Savonian'	-	*stadilainen* 'Stadi dweller'	-
Nature Areas	puisto 'park', meri 'sea'	*Vantaanjoki* [river name]	*Bodom* [lake name]	-
Compass Points	pohjoinen 'north'	länsi 'west'	eteläinen 'southern', länsi 'west'	etelä 'south', itä 'east'
Other	*Kauppatori* 'Market Square'	*Myyrmanni* [shopping center name], *Peijas* [hospital name]	*Areena* [Espoo Metro Areena, stadion name], *Jorvi* [hospital name]	*Itäkeskus* [shopping center name]

Table 2. Frequency of Keywords for Each Sub-Category of "Places and Directions"

Sub-Categories	Helsinki	Vantaa	Espoo	Total

Neighborhoods	14	23	27	64
Cities, Towns, and Municipalities	39	43	41	123
Provinces	5	2	2	9
Countries	3	0	1	4
Directions	15	8	9	32
Unspecified Areas	7	6	8	21
People from a Specific Area	7	0	1	8
Nature Areas	2	3	2	7
Compass Points	3	3	4	10
Other	2	5	3	10
Total	**97**	**93**	**98**	**288**

As shown in Table 2 (examples in Table 1), in all three cities the most keywords fall into the sub-category of Cities, Towns, and Municipalities, and there are no statistically significant differences. The differences in the second largest sub-category, Neighborhoods, on the other hand are statistically significant, *Espoo* having more mentions of neighborhoods than *Helsinki* ($p < 0.05$, z = 2,25).

Espoo (n = 23) and *Vantaa* (n = 20) both appeared more often with local neighborhood names than *Helsinki* (n = 9) (see Figure 2). While the difference between *Vantaa* and *Helsinki* was almost statistically significant ($p < 0.05$, z = 2.31), the difference between *Espoo* and *Helsinki* did reach statistical significance ($p < 0.01$, z = 2.71). The neighborhoods mentioned were clearly within the boundaries of each respective city, as illustrated below in Figure 2. This finding suggests that, in comparison to *Helsinki*, *Espoo* and *Vantaa* were more closely linked in our investigation to toponyms that are meaningful to local communities, but less well known outside them. This may be due to the strong neighborhood identities that are reputedly connected to the polycentric urban structure of both Espoo and Vantaa.

Figure 2: Place Names in the Capital Region and Surrounding Municipalities
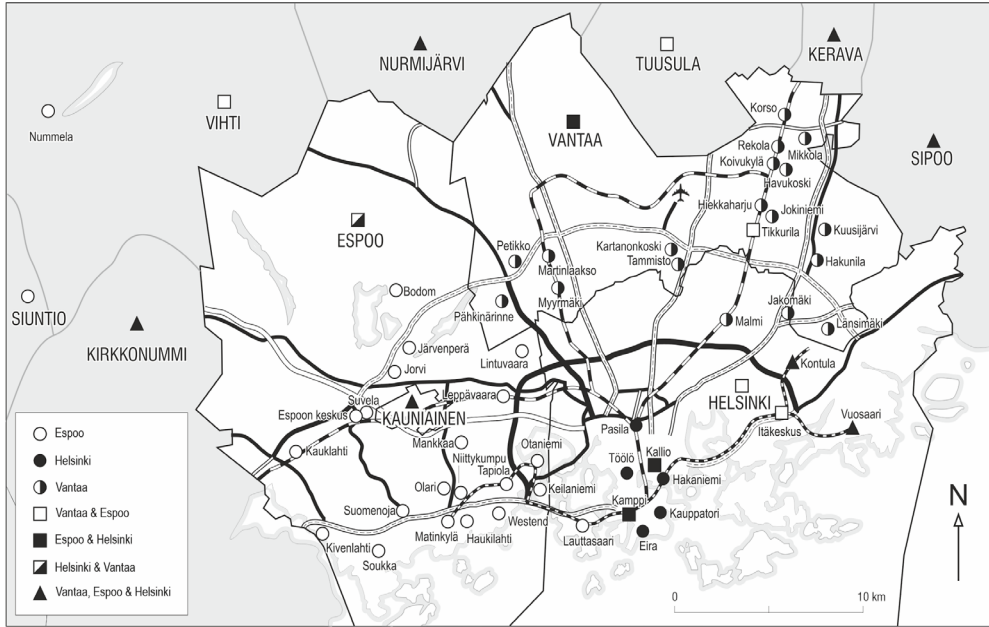
In excerpt 1 displayed below, the Suomi24 user describes the Espoo neighborhood of Leppävaara as the only real urban area in a city that they clearly see as their own ("meidän kaupunki' 'our city'). This excerpt illustrates how some neighborhoods were only discussed in relation to the city where they were located. Other neighborhoods, however, were discussed in connection to all three cities. In excerpt 2 below, the neighborhood of *Vuosaari*, located in the east of Helsinki, is compared to areas in *Espoo*. This passage demonstrates how place-framing can draw on and perpetuate socio-spatial differences by drawing parallels and distinctions between places. Vuosaari and the residential areas adjacent to the Länsiväylä Road are similar to each other in that they are both suburban with good traffic connections to the center of Helsinki. However, they differ demographically. For example, in Vuosaari, the proportion of residents whose mother tongue is not one of Finland's official languages is substantially higher (26%) than in Lauttasaari (7%) on the Länsiväylä Road (Ulkomaalaistaustaiset Helsingissä 2021).

> **Excerpt 1.**
> Leppävaara on Espoossa ainoa kunnollisen kokoinen taajama, meidän kaupungin ainoa todellinen City-alue.
> [Leppävaara is the only decent-sized urban area in Espoo, the only part of our city that's a real urban area.][3]
>
> **Excerpt 2.**
> [...] lisaksi espoosta ei loydy yhtaan niin epamaaraisat asuinaluetta kuin vuosaari, lisaksi liikenne pelaa lansivaylan varressa paremmin.
> [... and there are no areas that are as dodgy as Vuosaari in Espoo, and the traffic's better on the Länsiväylä Road.]

The toponyms were also prevalent when broader geographical units were discussed. City, town, and municipality toponyms occurred in the data 28 times for *Helsinki*, 37 for *Vantaa*, and 36 for *Espoo*—including those mentioned in connection with at least two of the three cities. The Finnish provinces of *Savo* 'Savonia', *Pohjanmaa* 'Ostrobothnia', and *Lappi* 'Lapland' were also mentioned, but they were only discussed in connection with *Helsinki*. Such discursive pairings may point to a deep-rooted geographical dichotomy between the capital city and the 'rest of the country'. The toponyms mentioned at the municipal and regional scale are shown in the maps in Figure 3.

Figure 3: Place Names in Finland and Europe
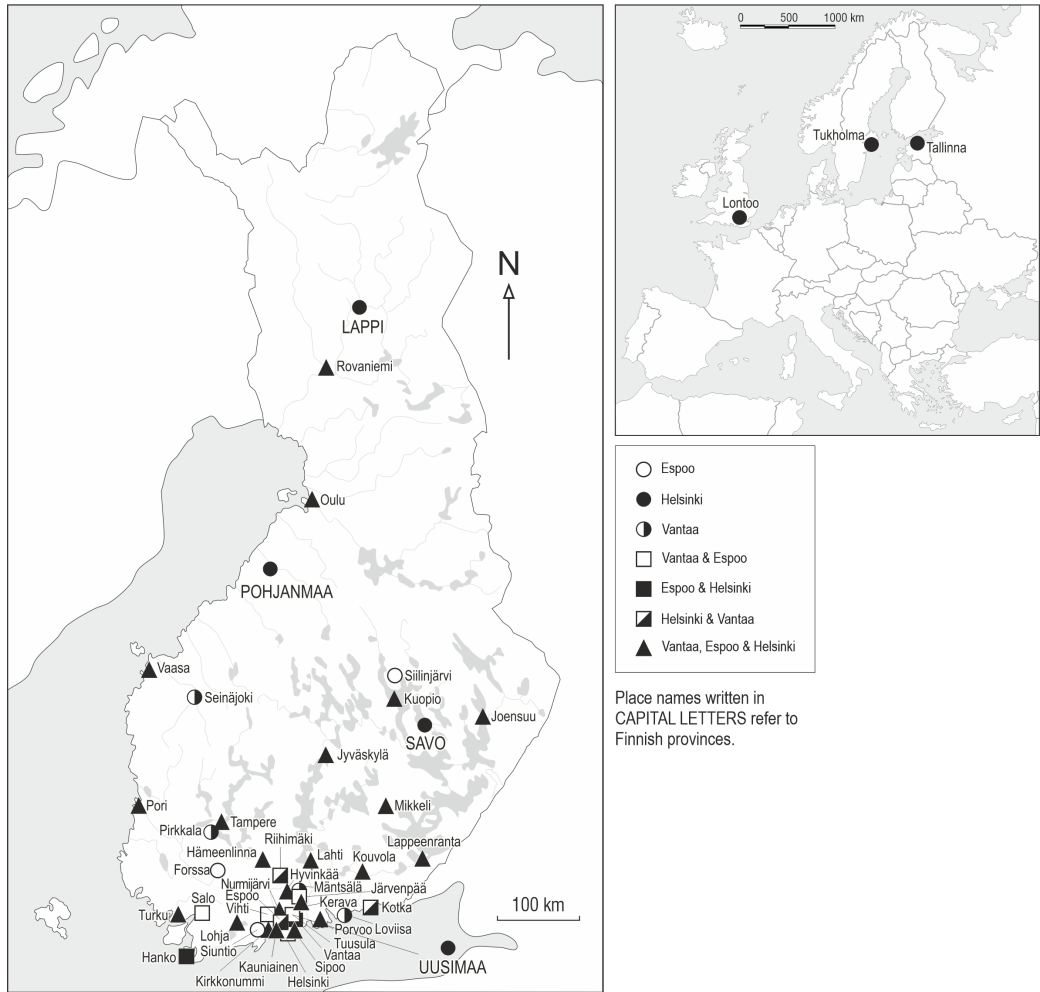
The maps show the hierarchical structure of place names in the online discourses. As a capital city, Helsinki is connected to other capitals in Europe, as well as Finnish provinces, whereas Espoo and Vantaa are only connected to other Finnish cities. In excerpt 3, the northernmost province of *Lappi* is compared to *Helsinki*. This comparison demonstrates how the spatial order of Finland is based on real and imagined socio-spatial differences between the nation's 'core' (*Helsinki*) and 'periphery' (*Lappi*). This excerpt also highlights the ways in which meanings attached to places may evoke mental images that reinforce a collective identity and encourage political actions, which seek to, for example, address the real and/or perceived spatial inequalities at the national level.

Excerpt 3.
Edellisessä ohjelmassa oli juttua Helsingin kalliista vuokratasosta, pitäisi tulla julki ettei asuta ilmaiseksi täälläkään pitkien välimatkojen lapissa.
[In the previous program, they talked about the high rents in Helsinki, but it is important to mention that we don't live for free here in Lapland either, what with all the long distances.]

Perhaps for similar capital city reasons, the foreign capitals of London (*Lontoo*), Stockholm (*Tukholma*), and

Tallinn (*Tallinna*) are only mentioned in conjunction with *Helsinki*. In excerpt 4, for instance, London and Helsinki are both described as being good for shopping.

> **Excerpt 4.**
> Käyn vaateostoksilla Lontoon Harrodsilla ja harvoin Helsingin Stockmannilla.
> [I buy my clothes at Harrod's in London and occasionally from Stockmann's in Helsinki.]

As well as referring to geographical units in terms of their relative size, the category of "places and directions" also includes utterances that depict the people living in particular areas as being different or special. For instance, one of the most striking differences found between Helsinki and its neighboring cities was that the Helsinki inhabitants were often juxtaposed with the rest of the country, whereas no such contrast was drawn with the residents of Espoo or Vantaa. This finding might mirror not only the views of many people living in Helsinki, but also the wider national geographical imagination which is continuously reinforced by Finland's mainstream media. Excerpt 5 also highlights this perceived contrast between Helsinki and elsewhere in Finland. This passage was taken from a forum discussion about the difficulties people face when they move to the capital.

> **Excerpt 5.**
> En tiedä miten ongelma tulisi ratkaista, ehkä parasta olisi jos maalaiset tullessaan Helsinkiin edes yrittäisivät opetella huomioimaan muut ihmiset.
> [I don't know how this problem can be solved, but maybe the best thing would be that people from the country start by learning to take other people into consideration when they come to Helsinki.]

The above suggestion may point to norms of formal politeness commonly associated with dense urban environments—the impersonal, transitory, superficial social interactions thought typical of big cities (Wirth 1938). Taken together, the findings of this study revealed that Helsinki was discussed in relation to the rest of the nation and other world capitals, whereas discussions about Espoo and Vantaa focused more on their inner provincial characteristics. This difference may highlight the special role of the capital city in the spatial order of the nation-state (Jokela & Linkola 2013). Helsinki is a node that links different geographical spaces to each other and, in some cases, acts as a metonym for the entire capital region, including Espoo and Vantaa. With its prominent architecture, state institutions, and urban culture, Helsinki is also, however, connected to some negative western myths of the 'big city'. In stark contrast with the seemingly more inferior rural settings of the provinces, Helsinki appears as superior, with political power and aloof urban sophistication (Jokela & Linkola 2013).

Housing

Within the discourse prosody "housing", the most important differences between the three cities relate to housing type. This difference was mainly evident in the qualitative analysis (Table 3), but was also to some extent present in the quantitative analysis as well (Table 4).

Table 3. Sub-Categories of "Housing" and Examples of Keywords

Sub-Categories	Helsinki	Vantaa	Espoo	Shared Keywords
Residential Area	-	-	-	lähiö 'suburb', slummi 'slum', asuinalue 'residential area', asuinpaikka 'residence'
Housing in General	hima 'home' [slang], asunnoton 'homeless'	neliöhinta 'price per square meter', osoite 'address'	neliöhinta 'price per square meter', osoite 'address'	asua 'to live', asunto 'residence', asukas 'resident', neliö 'square meter', talo 'house', kotoisin 'to come from'
Housing Type	-	kolmio 'two-bedroom apartment', kerrostalo 'apartment building', omakotitalo / okt 'detached house' / [abbreviation for a detached house], *VVO* [Finnish rental housing group]	rivitalo 'row house', kivitalo 'stone house', kolmio 'two-bedroom apartment', kerrostalo 'apartment building', omakotitalo / okt 'detached house '/ [abbreviation for a detached house], *VVO* [Finnish rental housing group]	yksiö 'studio', kaksio 'one-bedroom apartment', kämppä 'place, flat' [slang], vuokra 'rent', vuokrata 'to rent'
Building	-	-	kaavoittaa 'to zone'	rakentaa 'to build', tontti 'lot'
Moving In/Out	-	muutto 'removal'	muutto 'removal'	muuttaa 'to move'

Table 4. Frequency of Keywords for Each Sub-Category of "Housing"

Sub-Categories	Helsinki	Vantaa	Espoo	Total
Residential Area	4	4	4	12
Housing in General	9	9	9	27
Housing Type	5	12	12	29
Building	2	2	3	7
Moving In/Out	1	2	2	5
Total	**21**	**29**	**30**	**80**

Studios, called "yksiö" in Finnish, and two-room apartments or "kaksio" appeared in discourses compiled for all three cities, whereas three-room apartments, "kolmio", and certain types of buildings (e.g., apartment block "kerrostalo"; detached house "omakotitalo" and row house "rivitalo") were only found in the sub-corpora for *Espoo* and *Vantaa* but not in the *Helsinki* sub-corpus. These differences might reflect generally held ideas about housing patterns (Clapham 2005) in which young people leave home, start their adult life in the city, and then raise a family in bigger homes further away from the city centre. In Finland's capital region, Kallio and Punavuori, two traditional working-class districts located next to Helsinki's historical center, are popular among young adults and students who both favor small flats. This trend might also partly explain the prevalence

of the word "asunnoton" 'homeless' (see excerpt 6), as some of the discussions in the forum concerned young people in search of accommodation.

> **Excerpt 6.**
> Kaikki tietävät, että nuorisotyöttömyys on Helsingissä huipussaan ja etsi siinä sitten työtä kun olet samaan aikaan vielä asunnoton.
> [Everybody knows that youth unemployment is at its highest in Helsinki, so how can you look for a job if you're homeless at the same time.]

Although Helsinki's inner-city neighborhoods also attract middle-class families (Lilius 2018), in our study, discourses surrounding housing patterns frequently constructed *Espoo* as being the ultimate place for pursuing the "middle-class dream" (see Ikonen 2010). This idea may stem from the relatively high socioeconomic status of Espoo as compared to Helsinki and Vantaa (Vilkama et al. 2014). Espoo was also commonly culturally framed as home to both the national elite and the most important information and communication technology companies in Finland. In the following passage, a user describes a typical resident of stone house "kivitalo" in Espoo:

> **Excerpt 7.**
> Toisella miehistä on siisti ulkonäkö, iso kivitalo Espoossa meren rannalla, uusi Audi, purjevene ja hyvä toimeentulo. Kaikin puolin ongelmaton kaveri.
> [The other guy is decent-looking, has a big stone house in Espoo by the sea, a new Audi, a sailing boat, and a proper salary. In every regard, an unproblematic guy.]

As shown above in excerpt 7, a person is mockingly described as exemplifying the stereotypes commonly associated with people who live in Espoo.

Mobility

Although there were no significant differences between the cities in terms of the quantity of "mobility", the

qualitative properties for each city seemed to differ. In Table 5, the qualitative differences between the cities are highlighted with the help of examples of the keywords under various sub-categories.

Table 5. Sub-Categories of "Mobility" and Examples of the Keywords

Sub-Categories	Helsinki	Vantaa	Espoo	Shared Keywords
Public Transport	*VR* [railway company name], kuski 'driver', joukkoliikenne 'public transport'	seutulippu 'regional ticket', kuljettaja 'driver', taksiautoilija 'taxi driver', parkkipaikka 'parking place'	*HSL* [Helsinki Region Transport], joukkoliikenne 'public transport'	liikenne 'traffic', ruuhka 'rush hour', ajaa 'to drive', ajella 'to drive around', työmatka 'way to work'
Traveling	lento 'flight', lentää 'to fly', matkustaa 'to travel'	lento 'flight'	-	matka 'trip, journey'
Vehicles	ratikka 'tram', dösä 'bus [slang]', spora 'tram [slang]', laiva 'ship´		länsimetro 'the Western subway', *Audi*	juna 'train', bussi 'bus', taksi 'taxi', auto 'car', metro 'subway'
Movement in General	lähteä 'to leave', käydä 'to visit', kulkea 'to go, walk', saapua 'to arrive'	virrata 'to flow', käydä 'to visit', saapua 'to arrive'	-	
Roads	katu 'street', kaista 'lane', moottoritie 'highway', linja 'line', reitti 'route'	*3* [Ring road three], rata 'railroad', kehärata 'ring rail line'	*Länsiväylä* 'the Western highway', kaista 'lane', *3* [Ring road three], moottoritie 'highway', linja 'line', rata 'railroad'	kehä 'ring road'
Traffic Station	rautatieasema 'railway station', kenttä '[air]port', lentokenttä 'airport', satama 'harbor'	lentoasema 'airport', lentokenttä 'airport', kenttä '[air]port'		asema 'station'
Distance	kilsa 'klick' [slang for km]	-	kilometri 'kilometer'	km 'km'

As shown in the Table 5, sub-categories of Public transport and Roads are common to all three cities, whereas the other categories are existent only for two of the cities.

Helsinki is often being discussed in terms of air travel ("lento" 'flight', "lentää" 'to fly'). Because Helsinki's airport is geographically located in Vantaa, the keyword "airport" also features heavily in the context of Vantaa. Furthermore, in our investigation, *Vantaa* often appeared together with *Helsinki* in toponym *Helsinki-Vantaan lentokenttä* or *Helsinki-Vantaa lentosasema* ('Helsinki-Vantaa airport'), the last being the official Finnish name for the Helsinki airport. In the discussions, the role of Helsinki as the place for departing and arriving flights was often emphasized (see excerpt 8), even though the airport itself is not located within municipal boundaries of Helsinki.

> **Excerpt 8.**
> Itse jokin aika sitten lensin Havanaan Helsingistä Moskovan kautta.
> [Some time ago I flew to Havana from Helsinki via Moscow.]

In addition to air travel, car travel was a common theme in the discussions. 'Roads' was a prominent subcategory for all three cities (Figure 4), which may point to the important role of private motoring and the polycentric urban structure of the cities we examined. Interestingly, the only make of car mentioned here in the corpus was Audi, a brand that is a high-status symbol in Finland (as was also illustrated in excerpt 7). The mention of this brand name probably corresponds to the relatively high socio-economic status of Espoo residents as compared to those in Vantaa and Helsinki. The semantic classes within the discourse prosody "Mobility", and the total number of keywords referring to each per city are presented in Figure 4. The size of each city's pie chart on the map is proportional to the total number of "Mobility" keywords for each city. In this way, the pie charts also show the relative proportions of semantic class based on the number of keywords in each. The figures around the charts correspond to the total number of keywords per color-coded semantic class.

Figure 4: Proportional Symbol Map Showing Relative Number of "Mobility" Keywords and Relative Proportion of Semantic Classes in this Discourse Prosody for Helsinki, Vantaa, and Espoo.

Discussions on different modes of transport were accompanied with references to movement in general. Verbs expressing visiting "käydä", leaving "lähteä", and arriving "saapua" appeared in both the *Helsinki* (4 verbs) and *Vantaa* (3 verbs) sub-corpora, but there were no such instances whatsoever detected in the *Espoo* sub-corpus.

One explanation might be that *Espoo* was less seen as a point of departure or arrival than either *Helsinki* or *Vantaa*.

One detail here, as in many of the other discourse prosodies, related to the occurrence of slang words associated with *Helsinki* (e.g., "dösä" for 'bus', "spora" for 'tram', "kilsa" for 'kilometer'). Helsinki slang is a unique variety that developed at the turn of the 20th century among the working classes. From the 1950s onwards, this urban speech became a common street language of young people throughout the Helsinki region (Ainiala & Lappalainen 2017). Even though slang words and names are used in Espoo and Vantaa as well, in our study, slang was most prominently used to convey users' 'Helsinkiness'. The connection between Helsinki and slang was further evident by the fact that Helsinki was the only one of our three cities to have slang names— *Hesa* and *Stadi* (see also Ainiala & Jantunen 2019; Ainiala & Lappalainen 2017).

Conclusions

This article investigates the digital urban discourses surrounding the names used for three cities in Finland's Capital Region: *Helsinki*, *Espoo*, and *Vantaa*. We used onomastics, geographical information systems (GIS), and corpus-assisted discourse analysis (CADS) to detect the potential differences in online discourses that made reference to these cities. Our analysis revealed stark patterns of socio-spatial differentiation and thus confirmed the perceived special role of capitals in the spatial order of nation-states. The knowledge gained from research such as ours may serve urban planners and branding experts who strive to understand the ways people attach meanings to particular cities and neighborhoods; to plan meaningful and efficient enquiries or participatory processes; and to anticipate potential conflicts in socially sustainable ways.

However, by concentrating only on the three most frequent discourse prosodies, we have barely scratched the surface. Moreover, because a frequency-based method for data analysis was used, less frequent phenomena that may have been more discernible with traditional methods of discourse analysis were probably overlooked. Nevertheless, analyzing keywords by grouping them into discourse prosodies helped to reveal what people in the discussion forum were saying about Helsinki, Espoo, and Vantaa. One of the benefits of this methodological approach is that data was not collected expressly for this study, but as part of a larger research objective. As a result, potential biases that may be introduced when researchers collect and observe data with a particular study in mind may have been reduced. But since the demographic groups using social media platforms vary greatly (Sadah et al. 2015), and because there are many Finns who do not use Suomi24, we must bear in mind that this forum is only partially representative. Furthermore, the corpus did not provide information about the relation of users to the area (e.g., whether they live in the area or not). For a more complete picture, other data would need to be analyzed. Another limitation of this study involves the ethical concerns of using social media to gather data without obtaining user consent. It is fair to say that many users in this study were not aware that their personal posts and information were being used for research purposes (Williams et al. 2017).

Despite these limitations, as Motschenbacher (2020b) notes, corpus data and corpus linguistic methods are a powerful way to study names in actual use. By using this approach, the study answers the call for onomastics to reveal more about actual name usage in authentic spoken and textual contexts (e.g., Ainiala & Östman 2017; De Stefani 2016). In this way, our research provides a stronger foundation for other new and innovative onomastic investigations. By using corpus-assisted mixed method approaches such as presented here, future big data studies can contribute to increasingly multidisciplinary study of names.

Endnotes

[1] For more on the spatial humanities and Visual GISting, see, for example, Gregory et al. (2015) and Murrieta-Flores et al. (2011).

[2] Statistical significances: $p < 0.001$ = extremely significant, $p < 0.01$ very significant and $p < 0.05$ significant.

[3] All excerpts are translated by the authors.

Acknowledgements

Bibliography

Ainiala, Terhi and Jarmo Harri Jantunen. 2018. "Hur slangnamnen *Hesa* och *Stadi* används i en diskussionsgrupp: en öppning för korpusassisterad onomastik." [How the Slang Names *Hesa* and *Stadi* are Used in a Discussion Group: An Opening for Corpus-assisted Onomastics.] Presentation in 48th NORNA symposium *Names in Writing* in Göteborg November 29–30 2018. Accessed June 16, 2020. https://docplayer.se/124285330-Norna-s-48-e-symposium-i-goteborg-november-namn-i-skrift-names-in-writing.html

Ainiala, Terhi and Jarmo Harri Jantunen. 2019. "Korpusonomastinen tutkimus slanginimistä *Hesa* ja *Stadi* digitaalisissa diskursseissa." [Corpus Onomastic Research on the Slang Names *Hesa* and *Stadi* in Digital Discourses.] *Sananjalka* 61, no 61: 57–79. Accessed June 16, 2020. https://journal.fi/sananjalka/article/view/80312

Ainiala, Terhi and Hanna Lappalainen. 2017. "Orienting to Norms: Variability in the Use of Names for Helsinki." In *Socio-onomastics: The Pragmatics of Names*, edited by Terhi Ainiala and Jan-Ola Östman, 130–153. Amsterdam: John Benjamin Publishing Company. Accessed June 16, 2020. https://benjamins.com/catalog/pbns.275.06ain

Ainiala, Terhi and Jan-Ola Östman. 2017. "Introduction: Socio-onomastics and Pragmatics." In *Socio-onomastics: The Pragmatics of Names*, edited by Terhi Ainiala and Jan-Ola Östman, 1–18. Amsterdam: John Benjamin Publishing Company. Accessed June 16, 2020. https://benjamins.com/catalog/pbns.275.06ain

Anthony, Laurence. 2017. *AntConc 3.5.0*. Tokyo: Waseda University. Accessed June 16, 2020. https://www.laurenceanthony.net/software/antconc/

Baker, Paul. 2006. *Using Corpora in Discourse Analysis*. London: Continuum.

Baker, Paul, Costas Gabrielatos, Majid Khosravinik, Michal Krzyżanowski, Tony McEnery, and Ruth Wodak. 2008. "A Useful Methodological Synergy? Combining Critical Discourse Analysis and Corpus Linguistics to Examine Discourses of Refugees and Asylum Seekers in the UK Press." *Discourse & Society* 19, no. 3: 273–306. Accessed June 16, 2020. https://doi.org/10.1177/0957926508088962

Barnes, Trevor J. and James S. Duncan, eds. 1992. *Writing Worlds: Discourse, Text and Metaphor in the Representation of Landscape*. London: Routledge.

Bjorge, Anne Kari. 2003. "The Explicitation of Proper Names in Expository News Texts." *International Journal of Applied Linguistics* 13, no. 1: 115–138. Accessed June 16, 2020. https://doi.org/10.1111/1473-4192.00040

Clapham, David. 2005. *The Meaning of Housing: A Pathways Approach*. Bristol: Policy Press.

Crang, Mike. 2015. "The Promises and Perils of Digital Geohumanities." *Cultural Geographies* 22, no. 2: 351–360. Accessed June 16, 2020. https://doi.org/10.1177/1474474015572303

Dalen-Oskam, Karina. 2013. "Names in Novels: An Experiment in Computational Stylistics." *Literary and Linguistic Computing* 28, no. 2: 359–370. Accessed June 16, 2020. 10.1093/llc/fqs007

de Koning, Anouk. 2015. "This Neighbourhood Deserves an Espresso Bar too': Neoliberalism, Racialization, and Urban Policy." *Antipode* 47, no. 5: 1203–1223. Accessed June 16, 2020. https://doi.org/10.1111/anti.12150

De Stefani, Elwys. 2016. "Names and Discourse". In *The Oxford Handbook of Names and Naming,* edited by Carole Hough, 52–66. Oxford: Oxford University Press.

Donaldson, Christopher, Ian N. Gregory, and Joanna E. Taylor. 2017. "Implementing Corpus Analysis and GIS to Examine Historical Accounts of the English Lake District". In *Historical Atlas: Its Concepts and Methodologies,* edited by Peter Bol. 152–172. Seoul: Northeast Asian History Foundation. Accessed June 16, 2020. https://www.semanticscholar.org/paper/Implementing-corpus-analysis-and-GIS-to-examine-of-Donaldson-Gregory/d1bf6298e62bc4c5be273ec22a58abcdc0e31eb1

FIAM (website). *Finnish Internet Audience Measurement: Tulokset* [Results]– *Comscore MMX MP*. Accessed June 16, 2020. https://fiam.fi/tulokset/.

Gregory, Ian N, and Andrew Hardie. 2011. "Visual GISting: Bringing Together Corpus Linguistics and Geographical Information Systems." *Literary and Linguistic Computing* 26, no. 3: 297–314. Accessed June 16, 2020. https://doi.org/10.1093/llc/fqr022

Gregory, Ian N., Christopher Donaldson, Patricia Murrieta-Flores, and Paul Rayson. 2015. "Geoparsing, GIS, and Textual Analysis: Current Developments in Spatial Humanities Research." *International Journal of Humanities and Arts Computing* 9, no. 1: 1–14. Accessed June 16, 2020. 10.3366/ijhac.2015.0135

Halverson, Sandra L. and Jan Oska Engene. 2010. "Domains and Dimensions in Metonymy: A Corpus-Based Study of Schengen and Maastricht." *Metaphor and Symbol* 25, no. 1: 1–18. Accessed June16, 2020. https://doi.org/10.1080/10926480903538456

Haraway, Donna. 1996. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Human Geography: An Essential Anthology*, edited by J. Agnew, D. Livingstone, and A. Rogers. Malden, 108–128. MA: Blackwell Publishing. Accessed June 16, 2020. https://doi.org/10.2307/3178066

Hunston, Susan. 2007. "Semantic Prosody Revisited." *International Journal of Corpus Linguistics* 12, no. 2: 249–268. Accessed June 16, 2020. 10.1075/ijcl.12.2.09hun

Hämäläinen, Lasse. 2019. "Nimet verkossa: Tutkimus verkkoyhteisöjen käyttäjänimistä ja virtuaalisen minigolfpelin radannimistä." [Names on the Internet: A Study of User Names in Online Communities and Level Names in a Virtual Minigolf Game.] Doctoral Thesis, The University of Helsinki. Accessed June 16, 2020. http://urn.fi/URN:ISBN:978-951-51-5259-6

Ikonen, Hanna-Mari. 2010. "Muutu ihmiseksi, joka olet: Maalla-lehti jälkimodernin työntekijän ja   paikan kuvaajana." [Become the Person You Are: A Postmodern Employee and Place as Presented in the Country Living Magazine.] *Alue ja ympäristö* [Region & Environment] 39, no. 2: 39–52. Accessed June 16, 2020. https://www.researchgate.net/publication/303685665_Muutu_ihmiseksi_joka_olet_-_Maalla-lehti _jalkimodernin_tyontekijan_ja_paikan_kuvaajana

Jantunen, Jarmo Harri. 2018. "Korpusavusteinen diskurssintutkimus (CADS): analyysiesimerkki homouden ja heterouden digitaalisista diskursseista." [Corpus-assisted Discourse Studies (CADS): Digital Discourses of Homosexuality and Heterosexuality as a Research Example]. *AFinLA-e* 11, 20–44. Accessed June 16, 2020. https://doi.org/10.30660/afinla.69259

Jokela, Salla, and Hannu Linkola. 2013. "'State Idea' in the Photographs of Geography and Tourism in Finland in the 1920s." *National Identities* 15, no. 3: 257–275. Accessed June 16, 2020. 10.1080/14608944.2013.779644

Kamandulytė, Laura. 2006. *Identity Analysis Based on the Methodology of Corpus Linguistics*. Kaunas: Vytautas Magnus University. Accessed June 16, 2020.

Kitchin, Rob. 2013. "Big Data and Human Geography: Opportunities, Challenges and Risks." *Dialogues in Human Geography* 3, no. 3: 262–267. Accessed June 16, 2020. https://doi.org/10.1177/2043820613513388

Lappe, Sabine. 2002. "Monosyllabicity in Prosodic Morphology: The Case of Truncated Personal Names in English." In *Yearbook of Morphology,* edited by G. E. Booij and Jaap van Marle, 135–186. Springer-Verlag: Netherlands. Accessed June 16, 2020. https://doi.org/10.1007/0-306-48223-1_6

Laversuch, I. M. 2011. "'May Change Name and Pretend to be Free': A Corpus Linguistic Investigation of Surnames Adopted by Fugitive Slaves as Advertised in Colonial American Newspapers Between 1729 and 1818." *Names* 59, no. 4: 191–203. Accessed June 16, 2020. https://doi.org/10.1179/002777311X13095112507404

Lilius, Johanna. 2018. *Reclaiming Cities as Spaces of Middle Class Parenthood*. Singapore: Springer Verlag.

MacAulay, Lauren, Daniel Siddiqi, and Ida Toivonen. 2019. "The Age of Aidans: Cognitive Underpinnings of a New Trend in English Boys' Names." *Names* 67, no. 3: 160–174. Accessed June 16, 2020. https://doi.org/10.1080/00277738.2018.1490523

Marchi, Anna, and Charlotte Taylor. 2018. "Introduction: Partiality & Reflexivity." In *Corpus Approaches to Discourse. A Critical Review*, edited by Charlotte Taylor and Anna March, 1–15. London: Routledge.

Martin, Deborah G. 2003. "Place-Framing' as Place-Making: Constituting a Neighborhood for Organizing and Activism." *Annals of the Association of American Geographers* 93, no. 3: 730–750. Accessed June 16, 2020. https://doi.org/10.1111/1467-8306.9303011

META-SHARE (website). *The Suomi24 Corpus*. Accessed June 16, 2020. http://urn.fi/urn:nbn:fi:lb-2015120401

Motschenbacher, Heiko. 2020a. "Corpus Linguistic Onomastics: A Plea for a Corpus-Based Investigation of Names." *Names* 68, no. 2: 88–103. Accessed June 16, 2020. https://doi.org/10.1080/00277738.2020.1731240

Motschenbacher, Heiko. 2020b. "Greece, the Netherlands and (the) Ukraine: A Corpus-Based Study of Definite Article Use with Country Names". *Names* 68, no. 1: 1–16. Accessed June 16, 2020. ttps://doi.org/10.1080/00277738.2020.1731241

Murrieta-Flores, Patricia, Alistar Baron, Ian N. Gregory, Andrew Hardie, and Paul Rayson. 2015. "Automatically Analyzing Large Texts in a GIS Environment: The Registrar General's Reports and Cholera in the 19th Century." *Transactions in GIS* 19, no. 2: 296–320. Accessed June 16, 2020. https://doi.org/10.1111/tgis.12106

Nick, I. M. 2017. "Squaw Teats*, Harney Peak, and Negrohead Creek*: A Corpus-Linguistic Investigation of Proposals to Change Official US Toponymy to (Dis)honor Indigenous US Americans." *Names* 65, no. 4: 223–234. Accessed June 16, 2020. https://doi.org/10.1080/00277738.2017.1370067

Nurminen, Terhi Johanna. 2012. *Hill-terms in the Place-names of Northumberland and County Durham*. Newcastle: University of Newcastle Upon Tyne. Accessed June 16, 2020. http://hdl.handle.net/10443/1602

Ohlander, Solve, and Gunnar Bergh. 2004. "*Taliban*—A Rogue Word in Present-Day English Grammar." *English Studies* 85, no. 3: 206–229. Accessed June 16, 2020. 10.1080/00138380412331339120

Pagan, Joseph Martin. 1998. *A Morphological and Lexical Study of Personal Names in the Ebla Texts*. Rome: Missione archeologica italiana in Siria.

Paterson, Laura L., and Ian N. Gregory. 2018. *Representations of Poverty and Place: Using Geographical Text Analysis to Understand Discourse*. Palgrave MacMillan: Cham, Switzerland.

Partington, Alan, Alison Duguid, and Charlotte Taylor. 2013. *Patterns and Meanings in Discourse: Theory and Practice in Corpus-assisted Discourse Studies (CADS)*. Amsterdam: John Benjamin Publishing Company.

Pierini, Patrizia. 2008. "Opening a Pandora's Box: Proper Names in English Phraseology." *Linguistik Online* 36, no. 4: 43–58. Accessed June 16, 2020. https://doi.org/10.13092/lo.36.518

Poole, Robert. 2018. "Ecolinguistics, GIS, and Corpus Linguistics for the Analysis of the Rosemont Copper Mine Debate." *Environmental Communication* 12, no. 4: 525–540. Accessed June 16, 2020. 10.1080/17524032.2016.1275735

Presner, Todd, and David Shepard. 2015. "Mapping the Geospatial Turn." In *A New Companion to Digital Humanities,* edited by Susan Schreibman, Ray Siemens, and John Unsworth, 199–212. New Jersey: John Wiley & Sons. Accessed June 16, 2020. https://doi.org/10.1002/9781118680605.ch14

Rayson, Paul, and Roger Garside. 2000. "Comparing Corpora Using Frequency Profiling." In *WCC '00 Proceedings of the workshop on Comparing corpora* 9, edited by Adam Kilgarrif and Tony Berber Sardinha, 1–6. New Brunswick: The Association for Computational Linguistics. Accessed June 16, 2020. https://doi.org/10.3115/1117729.1117730

Richardson, Tim and Ole B. Jensen. 2003. "Linking Discourse and Space: Towards a Cultural Sociology of Space in Analysing Spatial Policy Discourses." *Urban Studies* 40, no. 1: 7–22. Accessed June 16, 2020. 10.1080/00420980220080131

Sadah, S. A., M. Shahbazi, M. T. Wiley, and V. Hristidis. 2015. "A Study of the Demographics of Web-Based Health-Related Social Media Users." *Journal of Medical Internet Research* 17, no. 8: e194. Accessed June 16, 2020. 10.2196/jmir.4308

Scott, James W., and Christophe Sohn. 2019. "Place-making and the Bordering of Urban Space: Interpreting the Emergence of New Neighborhoods in Berlin and Budapest." *European Urban and Regional Studies* 26, no. 3: 297–313. Accessed June 16, 2020. 10.1177/0969776418764577

Scott, Mike, and Christopher Tribble. 2006. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: John Benjamin Publishing Company.

Sinclair, John. 2004. *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

Smail, Rob, Ian Gregory, and Joanna E. Taylor. 2019. "Qualitative Geographies in Digital Texts: Representing Historical Spatial Identities in the Lake District." *International Journal of Humanities and Arts Computing* 13, no. 1–2: 28–38. Accessed June 16, 2020. https://doi.org/10.3366/ijhac.2019.0229

Soja, Edward. 2000. *Postmetropolis: Critical Studies of Cities and Regions*. Oxford: Blackwell.

Statistics Finland. 2021. "Population." Accessed May 22, 2021. https://www.tilastokeskus.fi/tup/suoluk/suoluk_vaesto_en.html

Stubbs, Michael. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.

Tanaka, Yu. 2016. "Phonotactically-Driven Rendaku in Surnames: A Linguistic Study Using Social Media." In *Proceedings of the 34th West Coast Conference on Formal Linguistics,* edited by Aaron Kaplan, Abby Kaplan, Miranda K. McCarvel, and Edward J. Rubin, 519–528. Somerville, MA: Cascadilla Proceedings Project. Accessed June 16, 2020. https://www.semanticscholar.org/paper/Phonotactically-Driven-Rendaku-in-Surnames%3A-A-Study-Tanaka/542dc8774610c9f435533f1ca003eff02a109122

Tse, Grace Y. W. 2004. "A Grammatical Study of Personal Names in Present-Day English: With Special Reference to the Usage of the Definite Article." *English Studies* 85, no. 3: 241–259. Accessed June 16, 2020. https://doi.org/10.1080/0013838041233133914O

Tuñón, Laura Cantora. 2013. *Proper Names as Cultural Referents in British Chick Lit: A Corpus-based Analysis of their Translations into Spanish and Italian*. Leeds: The University of Leeds.

Ulkomaalaistaustaiset Helsingissä [Population with Foreign Background in Helsinki]. 2021. "Alueellinen sijoittuminen" [Spatial Distribution of Population with foreign background.]) Accessed May 28, 2021. https://ulkomaalaistaustaisethelsingissa.fi/fi/alueellinensijoittuminen.

Vaičenonienė, Jūratė. 2001. "Using Corpora to Obtain Social and Cultural Information: A Case Study of America." *Kalbu Studijos / Studies about Language* 1, no. 1: 6–9. Accessed June 16, 2020. https://www.ceeol.com/search/article-detail?id=805610

Vilkama, Katja., Henrik Lönnqvist, Jenni Väliniemi-Laurson, and Martti Tuominen. 2014. *Erilaistuva pääkaupunkiseutu: Sosioekonomiset erot alueittain 2002–2012.* [Differentiating Capital Area: Socio-economic Differences by Region 2002–2012.] Helsinki: Helsingin kaupungin tietokeskus. Accessed June 16, 2020. https://www.hel.fi/hel2/helsinginseutu/liitteet/PKS_erilaistuva_seutu_tieke.pdf

Williams, Matthew L., Peter Burnap, Luke Sloan, Curtis Jessop, and Hayley Lepps. 2017. "Users' Views of Ethics in Social Media Research: Informed Consent, Anonymity and Harm." In *The Ethics of Online Research*, edited by Kandy Woodfield, 27–51. Bingley: Emerald Publishing Limited. Accessed June 16, 2020. https://doi.org/10.1108/S2398-60182018000002002

Wirth, Louis. 1938. "Urbanism as a Way of Life." *American Journal of Sociology* 44, no. 1: 1–24. Accessed June 16, 2020. https://www.journals.uchicago.edu/doi/abs/10.1086/217913

Notes on Contributors

Jarmo Harri Jantunen is a professor of Finnish Language at the University of Jyväskylä, Finland. His research interests include corpus linguistics, corpus-assisted discourse analysis, second language acquisition, queer linguistics, and corpus onomastics.

Terhi Ainiala is an adjunct professor and university lecturer of Finnish Language at the University of Helsinki, Finland. Her research interests include socio-onomastics, folk onomastics, corpus onomastics, sociolinguistics, and urban studies.

Salla Jokela is a university lecturer in Sustainable Urban Development at Tampere University, Finland. She has a PhD in human geography from the University of Helsinki. Her research interests include urban studies, identity politics, tourism, and visual studies.

Jenny Tarvainen (MA) is a doctoral student in Finnish Language at the University of Jyväskylä, Finland. Her research interests include corpus linguistics, natural language processing, second language acquisition, and corpus onomastics.

Correspondence to: Professor Jarmo Harri Jantunen, University of Jyväskylä, Department of Language and Communication Studies, PO Box 35, FI-40014 University of Jyväskylä. Email: mailto:jarmo.jantunen@jyu.fij