Article

Urban Analytics and City Science

EPB: Urban Analytics and City Science
2023, Vol. 50(7) 1916–1931

DOI: 10.1177/23998083221150240
journals.sagepub.com/home/epb
Sage

Ethnic segregation on linguistic landscapes

Seong-Yun Hong and Yeorim Kim
Department of Geography, Kyung Hee University, Seoul, Republic of Korea

Yongchae Lee
Department of Housing and Interior Design, Kyung Hee University, Seoul, Republic of Korea

Abstract

This work presents a novel approach to studying ethnic segregation from the perspective of linguistic landscapes. Numerous street-level images accumulated over the last two decades have enabled the exploration of linguistic landscapes at a larger scale than ever before. Since the prevalence of a specific language in a public space implies the linguistic group inhabiting the area, its careful evaluation can reveal the degree of segregation between linguistically different ethnic groups. To demonstrate the effectiveness of the proposed approach, we applied it to the linguistic landscape of Seoul, South Korea. Using a large set of street-level images collected from an online map platform, we measured the levels of segregation between Korean and Chinese signs from 2010 to 2018. The levels of segregation on the street-level images were different to a certain extent from those of residential segregation. While residential segregation gradually increased between 2010 and 2018, except for 2011, more fluctuations were observed in linguistic segregation. This finding is likely because a linguistic landscape is shaped mainly by advertising signs, banners, and billboards in commercial areas, and such commodified urban spaces change more dynamically to attract inhabitants and visitors. These results suggest that the proposed approach can offer an alternative way of understanding the complex socio-demographic phenomenon from a new perspective, as with other emerging data sources in the era of big data.

Keywords

Ethnic segregation, linguistic landscape, street-level images, urban analytics, Chinese community, activity spaces

Introduction

Immigrants and ethnic minorities are often concentrated in certain parts of cities and form residential and business clusters. Until the 2000s, such ethnic segregation was perceived as a negative social

Corresponding author:
Seong-Yun Hong, Department of Geography, Kyung Hee University, 26 Kyungheedae-ro, Dongdaemun-gu, Seoul 02447, Republic of Korea.
Email: syhong@khu.ac.kr

process that hinders the integration of immigrants into mainstream society and causes various forms of inequalities (Alba and Logan, 1993; Massey, 1985). There are, however, some recent empirical cases that demonstrate positive aspects of ethnic settlements (Li, 2006). As the causes and consequences of ethnic segregation have diversified, a careful examination of this multidimensional phenomenon has become crucial for understanding the integration and inclusion of ethnic minorities in our society.

Census data have arguably been the most popular source of information for studying ethnic segregation. A census is conducted at a regular interval in many countries (decennially in the U.S., but at shorter intervals in most other countries), and it contains a comprehensive enumeration of populations and dwellings. The availability of and easy access to such complete demographic data has enabled researchers to measure the level of segregation at a reasonable geographic scale and investigate its relationship to social, cultural, and economic variables.

However, a critical argument has been raised recently regarding the use of census data, namely, that the reliance on such data only uncovers ethnic segregation in residential neighbourhoods (Östh et al., 2018; Wong and Shaw, 2011). Residential segregation is an important indicator of social inequalities because where one lives often determines one's access to quality education, job opportunities, and safe housing. Nevertheless, individuals' life experiences are not limited to their place of residence but are instead influenced by all aspects of their daily routines (Silm and Ahas, 2014; Van Ham and Tammaru, 2016). As the social exclusion and isolation of minority population groups may occur in other social contexts, such as workplaces, schools, or leisure activities, census data alone cannot convey the complete picture of all intergroup interactions in society.

To gain a richer understanding of this complex socio-spatial process, considerable efforts have been made in recent years to capture segregation outside of the residential context. Much research has been conducted on the so-called activity space segregation using newly emerged datasets, such as mobile phone trajectories (Östh et al., 2018; Silm et al., 2018), smart card data (Zhang et al., 2021), and web-based travel diaries (Li and Wang, 2017). The results obtained from these works have demonstrated that the actual level of segregation people experience in their daily life can substantially differ from the level depicted by the census; this discrepancy calls for further research on the causes, patterns, and consequences of segregation from multiple angles (Müürisepp et al., 2022).

However, although technological advances have paved the way for a more urban analytic approach to analysing segregation, detailed socio-demographic information from big data sources is still not easily accessible to individual researchers (Hong, 2020b). Unlike census data, data held by private companies are seldom open to the public or are available only to a limited spatial extent. Acquiring customised records that fit a specific research purpose involves significant costs, thereby posing another layer of challenges for studying ethnic segregation.

Therefore, in this work, we propose an alternative way of exploring ethnic segregation in cities using more accessible data, namely, street-level images on the Internet. The proliferation of online map services and social media platforms has led to the accumulation of enormous street-level images, and these images can be readily viewed from everyday devices, such as smartphones, tablets, and laptops. Since this type of data has a rich amount of information on the physical environment of cities, various applications have been proposed in urban studies (Hong, 2020a; Seiferling et al., 2017; Yin et al., 2015), and it also has the potential to unveil ethnic segregation from a unique perspective.

The primary purpose of this work is to demonstrate how a large volume of street-level images can be utilised to enhance our understanding of ethnic segregation. The proposed approach is based on the concept of the linguistic landscape. The linguistic landscape refers to the diversity of languages in public spaces, and it 'serves as a distinctive marker of the geographical territory inhabited by a given language community' (Landry and Bourhis, 1997: p. 25). The overrepresentation of a particular language in an urban scene is likely an indication of the potential clustering or segregation of the people using that language.

Of course, the linguistic landscape is not entirely determined by the demographic composition in a place. It is also affected by local regulations, economic motives, cultural trends, and political contexts (Ben-Rafael et al., 2006; Cenoz and Gorter, 2006; Landry and Bourhis, 1997). Therefore, it would be critical to assess the implication of segregation in linguistic landscapes and understand its link to conventional residential segregation. Although there were attempts to use the linguistic landscape and its numerical summary as an indicator of the ethnic population distribution (Hong, 2020a; Jeong et al., 2021), its validity has not yet been examined. To fill this gap in the literature, we compare the geographic distributions of Chinese signs and their level of segregation to various statistics for the Chinese population in Seoul. The findings from this empirical analysis demonstrate the advantages and limitations of the proposed approach.

Concept of the linguistic landscape

The most quoted definition of the term ‘linguistic landscape’ probably comes from Landry and Bourhis (1997): ‘Linguistic landscape refers to the visibility and salience of languages on public and commercial signs in a given territory or region’ (p. 23). Although it is a relatively recent area of research, the use of languages in public spaces and their implications have been of consistent interest over the decades (Backhaus, 2007). In particular, many surveys and empirical studies have been conducted around bilingual or multilingual environments (Bourhis, 1994; Spolsky and Cooper, 1991; Tulp, 1978).

In essence, the linguistic landscape reflects the composition of language groups inhabiting a region (Landry and Bourhis, 1997). The language of the majority population in a specific area often appears most dominantly in the locality, and thus, it can delineate the territorial limits of linguistically distinctive ethnic people to a certain extent. Retail shops, cafés, and restaurants in commercial areas tailor their signs and banners to attract the target customers, sometimes regardless of the language spoken by the business owner (Leeman and Modan, 2009). Therefore, a careful examination of the varying densities of signs in a particular language can reveal how urban spaces are occupied and consumed by the linguistic group in everyday life.

It would be important to emphasise here, however, that this approach requires a careful examination based on the socio-spatial context. The linguistic landscape is shaped by various other factors, such as the economic and political status of language groups and the social tolerance towards the public display of foreign languages (Landry and Bourhis, 1997; Spolsky and Cooper, 1991). Ben-Rafael and colleagues (2006), for example, argued that the linguistic landscape of Jerusalem reflects the relational power differences between Israeli Jews and Palestinian Israelis. They found that the Arabic language is nearly invisible in the areas mainly inhabited by the Jewish population, whereas Hebrew signs are frequently observed in Israeli-Palestinian communities.

The linguistic landscape is also regulated by legislation to preserve cultural and linguistic identities or revitalise endangered languages in many countries and cities. Quebec is one of the cases in point (Bourhis, 1994). As a response to the decline in the francophone population in Quebec from 1921 to 1941, the provincial government has made efforts to redress the status of French by implementing laws and regulations on the use of the language in the public and private sectors. The language policy of Quebec requires the predominant display of French on signs, and thus it practically determines the linguistic landscape in the region.

It may be worth noting that the visibility of English in the linguistic landscape should be interpreted with care. English signs are often prevalent in international travel destinations, such as Bangkok (Huebner, 2006), Athens (Nikolaou, 2017), and Lithuanian and Polish resort towns (Ruzaitė, 2017). However, this is usually a strategic choice that is motivated by economic purposes and to attract international tourists; it is not an indicator of native English-speaking inhabitants. Furthermore, it is not unusual that English characters written outside retail shops merely reflect 'a desire to project a cosmopolitan, sophisticated, and trendy outlook' (Nikolaou, 2017: p. 160). If a specific language of interest is chosen for a symbolic value instead of for its informative function in society, then the relationship between the visibility of the language and ethnic segregation would be weak.

Nevertheless, many empirical observations across cities support the link between the public display of minority languages and the presence of the ethnic population in urban spaces. While commercial establishments whose signs are written in, say, Chinese are not necessarily owned by co-ethnic Chinese people, such establishments are still likely to be places where Chinese residents and visitors can work and consume using their language. In this regard, the segregation in the linguistic landscape has a similar characteristic to that in activity spaces, except that its focus lies more on the physical environments than on individuals' movements.

The literature on the linguistic landscape has demonstrated that the longitudinal analysis of linguistic landscapes can reveal how our cities are changing through various social processes, such as globalisation (Vandenbroucke, 2015) and urban development (Papen, 2012). This dynamic visual component of urban spaces reflects the real-time changes in society (Gorter and Cenoz, 2015); thus, it has the potential to be a valuable data source for segregation research. A systematic evaluation of the clustering and isolation in linguistic landscapes and its careful interpretation can enhance our understanding of ethnic segregation, as we will demonstrate through an example of Chinese segregation in Seoul.

Ethnic segregation in Seoul

The Chinese population in South Korea

In this work, we will investigate the degree of ethnic segregation in Seoul, South Korea, using linguistic landscapes on street-level images. Ethnic segregation and linguistic landscapes have received not much attention in South Korea until recently, as the proportion of the foreign population is relatively marginal in comparison to that in Western countries. However, the increased globalisation and transnational mobilities have led to the inflow of immigrants and international students, mainly from other Asian countries. The capital city, Seoul, and its vicinities are the most popular destinations for these individuals; as a result, the number of foreign residents in the city has risen from 61,920 in 2000 to 242,623 in 2020[1] (Seoul Institute, 2022).

The Chinese population is the largest foreign group in the country, but it comprised less than 1% of the total population as of 2020 (Statistics Korea, 2022). More than half of the members of this group are *Joseonjok* (or *Chaoxianzu*), that is, Chinese people of Korean origins, but Han Chinese migrants and international students also make up a considerable proportion. Although short-term visitors from China, including tourists, are usually not perceived as a part of the ethnic community, their number is not negligible as they affect co-ethnic businesses to a great extent. Over 5.5 million Chinese people visited the country for various purposes each year between 2016 and 2019 on average (Korea Tourism Organization, 2022), and they played an important role in maintaining the economic vitality of the Chinese community.

In Seoul, the largest and most visible Chinese community can be found in the southwest areas. Of the 25 districts in the city, the Chinese population is clustered throughout two districts, namely, Guro and Yeongdeungpo. As of 2020, there were 26,136 Chinese nationals living in Guro, followed by 25,579 in Yeongdeungpo (Statistics Korea, 2022). Considering the total population of Chinese in Seoul (i.e. 149,620), these figures indicate that over a third of the long-term residents are concentrated in these ethnic neighbourhoods.

The residential concentration of the Chinese population in these two districts has been mainly driven by economic reasons. Until the 1980s, the Guro district was occupied by many labour-intensive textile manufacturers and low-cost flats for factory workers. The access to job opportunities and affordable housing brought the early Joseonjok migrants to this area, and the ethnic settlement has expanded to the adjacent district Yeongdeungpo over the last two decades (Chung and Lee, 2014). Daerim-dong, a local town in Yeongdeungpo, is now the centre for ethnic organisations and businesses. A full range of shops, from butcheries, cafes, and restaurants to medical and legal services, are located along the so-called Yanbian Street, which has stimulated the growth of the community (Chung and Lee, 2014).

Residential segregation of Chinese

Despite its growing visibility and increasing attention from the media, relatively little scholarly attention has been given to the Chinese community in Seoul. That said, several studies, mostly in local academic outlets, have investigated the causes and implications of residential segregation in a descriptive manner (Jun et al., 2013; Kim, 2021; Yun et al., 2014). Although these studies employed different methodologies ranging from in-depth interviews and field surveys to more quantitative analysis, the findings essentially appear parallel. The Chinese population in Seoul has been concentrated in the southwestern districts since the 1990s, but the inflow of Han Chinese migrants and international students has led to the emergence of smaller Chinese communities to accommodate different needs.

While these previous studies offer valuable insights into the settlement patterns of the ethnic population, they seldom address various aspects of the intricate socio-spatial process. One of the important limitations is that most of the literature on Chinese segregation in Seoul only examines such segregation in the residential context. There have been a few recent attempts to quantify the levels of Chinese segregation more comprehensively (Choi et al., 2021; Jeong et al., 2021; Shin, 2022). Nevertheless, these studies were conducted at limited spatial and temporal extents or only for a specific subgroup due to data constraints.

It is also important to note that, to the best of our knowledge, there is a lack of systematic longitudinal research on the degree of ethnic segregation. Each empirical analysis has reported a significantly different result for the same spatial configuration of people, depending on the data and method adopted in the research. A comparison across studies is almost impossible, and it impedes our understanding of how social integration across ethnic minorities changes over time. This is in part due to the presence of numerous indices addressing various dimensions of segregation. At the same time, however, the difficulty in obtaining consistent long-term data has played a critical role in the discrepancies.

The proposed approach can utilise street-level images from all sources, including those on online map platforms, if they convey linguistic landscapes. Since the launch of Google Street View in 2007, over 220 billion panoramic scenes have been published (Russell, 2022), and the volume of geotagged images on social media is increasing rapidly. In the subsequent sections, we will demonstrate how we can use street-level images to explore ethnic segregation from an angle that was not previously reached by conventional data sources.

Data and methods

Data

In this work, therefore, we used street-level images on one of the local online map platforms, Kakao Road View (KRV), to explore the linguistic landscapes of Seoul, South Korea. KRV was the first to start a street-level image service in the country, and it arguably has the most extensive coverage both in spatial and temporal terms. To collect an unbiased sample of linguistic landscapes, we randomly selected 2,647 points from the road network of Seoul and captured snapshots of urban streetscapes at these locations. Figure 1 shows the geographic distribution of the sampling points; since the road network is dense in commercial areas, sparse in residential areas, and empty in mountainous outskirts, the points are not evenly spread across the city. This is, however, not considered a problem for the present study because linguistic landscapes are usually found in shopping and business districts.

From the 2,647 random sampling points, 12,381 street-level images were collected in total. Over 90% of the locations (or 2,398 out of 2,647) had at least two panoramic photographs of the street taken at different timeframes between 2008 and 2020, and more than a third of the locations (or 921 out of 2,647) offered six or more snapshots for the same period. The average number of images per point was approximately 4.67, with a maximum of 13 at one point around the central city.

It may be worth noting that these images were manually captured because the application programming interface (API) for KRV did not allow us to access historical versions at the time of data collection. To maintain a consistent quality of data, all the images were captured at a fixed resolution of 1280-by-760 pixels, and the viewing angle was carefully chosen to display as many signs and billboards as possible in a scene. Nonetheless, 13.5% of those images collected (or 1,668 out of 12,381) did not contain any text characters, so we removed them from the final dataset.

In the remaining 10,713 street-level images, we drew a bounding polygon for each object with text written on it, except for those on the sides of buses, pedestrians' t-shirts, or other moving objects. Then, we assigned an attribute value to each polygon according to the type of language, that is, 1 for Korean, 2 for English, 3 for Chinese, 4 for numbers, 5 for all other languages, and 6 for all unrecognisable characters, mainly due to the small text size. When a sign was written in several different languages, we broke it down into pieces so that each polygon contained only one language. Such multilingual signs, especially those with a combination of Korean and English, frequently appeared in commercial sectors, parallel to previous observations (Lawrence, 2012; Lee, 2019). It is assumed in this work that the relative areas of the polygons representing each language indicate a quantitative summary of the linguistic landscape in that place.

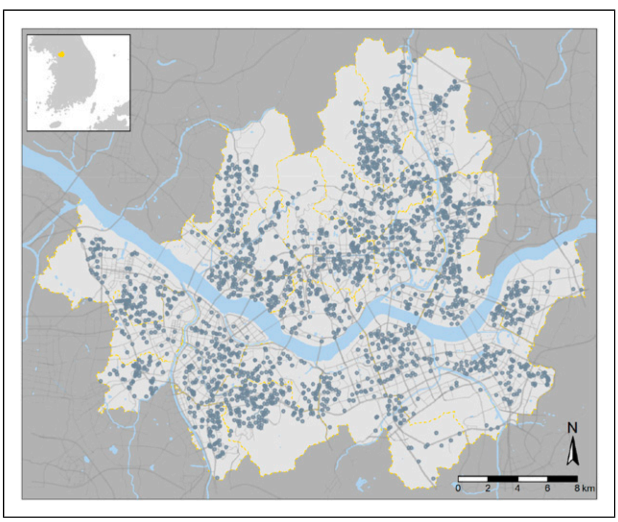

Figure 1. Randomly selected sampling points in Seoul.

This annotation process was carried out by hand to ensure the most accuracy. Indeed, the recent development in scene text detection and recognition techniques could help reduce the time for data preparation to a great extent at a reasonable level of errors. However, we believe that human annotation was necessary for this work to obtain a more confident dataset. This dataset may be utilised as a training dataset for future research that will focus on the efficient process of massive linguistic landscape photographs.

Table 1 shows some fluctuations in the number of linguistic landscapes obtained across years. While we could gather as many as 1,898 and 1,480 samples containing signs from 2018 and 2017, respectively, there were only 9 images available in 2019. To supplement insufficient samples for specific years, we calculated a 3-year running average of the linguistic diversity figures at each point, with missing values replaced by the average of other values in the 3-year time span. In this way, the linguistic landscape could be estimated if there were at least one image over three consecutive years, which resulted in enough samples to explore the temporal changes in the level of linguistic segregation.

Measure of segregation

Once the relative proportions of signs in each language were calculated for all images, we analysed the spatial distributions of those in Korean and Chinese using a measure of segregation. The results could indicate the degree of segregation in places between the two linguistic groups. In this work, we used one of the surface-based measures called the spatial information theory index (Reardon and O'Sullivan, 2004), as the data contain the exact location of each landscape.

For a finite set of points representing linguistic landscapes, $X = [x_1, x_2 \ldots x_n]$, the spatial information theory index can be calculated as follows:

$$\tilde{H} = 1 - \frac{1}{TE} \sum_{i=1}^{n} \tau_i \tilde{E}_i$$

where τ_i indicates the sum of the sign areas in either Korean or Chinese at point x_i and T is the total sum of the sign areas in all street-level images collected (i.e. $T = \sum_{i=1}^{n} \tau_i$). Both E and $\tilde{E}_i$ refer to the uneven representation of the two languages; the former indicates it for the entire study area, and the latter for point x_i and its vicinity, as follows:

$$E = -(\alpha \log_2 \alpha + \beta \log_2 \beta),$$

$$\tilde{E}_i = -(a_i \log_2 a_i + b_i \log_2 b_i).$$

Table 1. Data collected from 2,647 sampling points in Seoul.

Year	2008	2009	2010	2011	2012	2013	2014
Images	143	1117	1199	485	1174	1245	506
With signs	116	969	1035	406	1008	1077	436
Year	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**Total**
Images	1266	1224	1480	1898	7	637	12,381
With signs	1092	1062	1264	1653	7	588	10,713

In the equations, α and β indicate the total areas occupied by Korean and Chinese signs, respectively, whereas a_i and b_i represent the weighted sum of Korean and Chinese signs in the vicinity of point x_i.

The vicinity, sometimes referred to as the local neighbourhood, of a particular point x_i can be defined in various ways. There is already a massive literature on the definitions of 'local', so we do not discuss further on this topic here. In this work, we adopt a negative exponential function with $p = 2$ as a distance decay function:

$$w(d) = e^{(-d_{ij} \times p)}$$

where d_{ij} denotes the Euclidean distance between two points i and j.[2] As a result, the vicinity was implicitly defined as a circle of 2 km radius: the linguistic landscapes at points further than this threshold were practically ignored during the estimation of $\tilde{E}_i$.

The index value can be interpreted in a similarly intuitive manner to the classical index of dissimilarity. It has a theoretical range from 0 to 1. If the value is closer to 1, it implies that the signs in different languages are spatially separated from each other. On the contrary, if the value is closer to 0, the two languages may be visible in similar relative proportions across the study area.

Results

Descriptive analysis

In Seoul, the proportion of Chinese signs increased rapidly between 2010 and 2013 and remained constant until 2018 (Figure 2). In 2010, the signs written in Chinese characters comprised approximately 0.8% of the linguistic areas on average, but the same figure grew to 1.3% in 2016, thereby representing an increase of approximately 58% compared to 2010.

This increase might be mainly attributable to the growth in the Chinese population during the same period, as indicated by the bar chart in Figure 2. In particular, the number of Chinese students

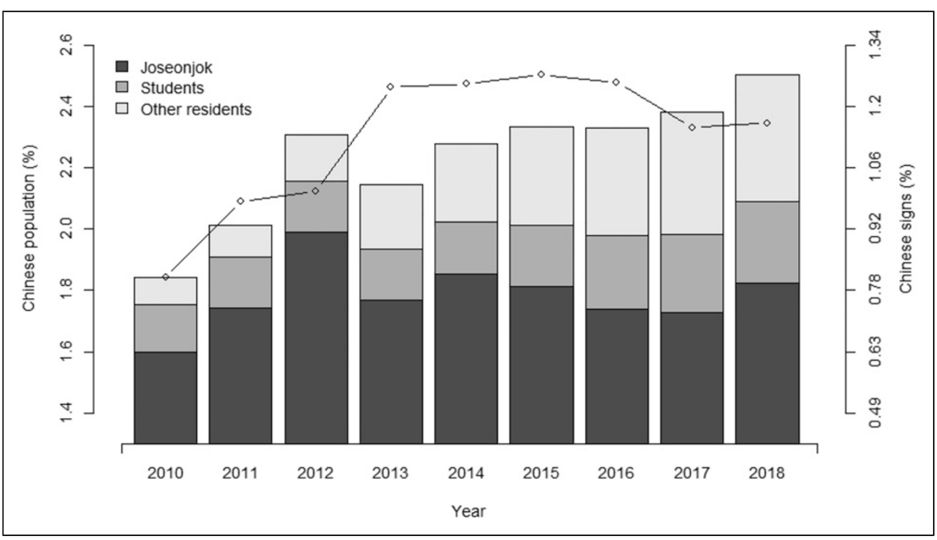

Figure 2. Proportions of Chinese population groups in Seoul and signs written in the Chinese language, 2010–2018.

has increased notably in recent years. Although the Joseonjok population has declined since its peak in 2012, the ethnic population could have expanded its size as the other subgroups have increased.

Figure 3 compares the distribution of the Chinese signs in Seoul to that of English signs. Clearly, the signs displaying the English language are scattered over the city rather than agglomerated in specific areas. This finding is consistent with previous observations (Lawrence, 2012; Lee, 2019) and is likely because the English language is widely used not only for its informative function but also as a symbolic feature. The pattern in Figure 3(a) can be another example of the symbolic use of the foreign language to convey a modern and trendy appeal; thus, the presence of English signs in Seoul should not be interpreted as an indicator of the English-speaking population.

In contrast, the signs in Chinese are concentrated in certain parts of the city. The Chinese signs are distributed similarly to the Chinese population; that is, there is one large cluster located in the southwestern corner of Seoul, while several smaller agglomerations are also found around the districts of Seongbuk, Jung, and Jongno. Since several private universities with well-established language programs are located in Seongbuk, many international students, including Chinese nationals, are known to reside in the district. On the other hand, the Jung and Jongno districts are popular destinations for Chinese visitors. According to a survey of Chinese tourists in South Korea, such tourists spend 77% of their travel expenditures on shopping, and more than two-thirds of them visit the Jung district during their stay.

It may be worth noting that Chinese signs are almost invisible in the Gangnam district. As with Jung and Jongno, Gangnam is also a popular shopping destination for international travellers. The number of Chinese people visiting this area is not negligible, as they allegedly account for over 20% of all foreign tourists. However, the Gangnam district contains more department stores and brand shops, so there might be less space for which the minority language is reflected in the linguistic landscape.

This interpretation can be further supported by Figure 4, which exhibits the distributions of the Chinese population and signs from 2010 to 2016. In general, the maps indicate that more Chinese signs are visible in the areas inhabited by the Chinese population. However, there is also apparent clustering of Chinese signs in the Jongno and Jung districts, where few Chinese people reside, and such clustering has intensified since 2014. The overrepresentation of the Chinese language in these areas is likely driven by short-term visitors. Therefore, the linguistic landscape in Seoul reflects not only the residential distribution of the ethnic minority but also the geographic pattern of their main economic activities for the co-ethnics.

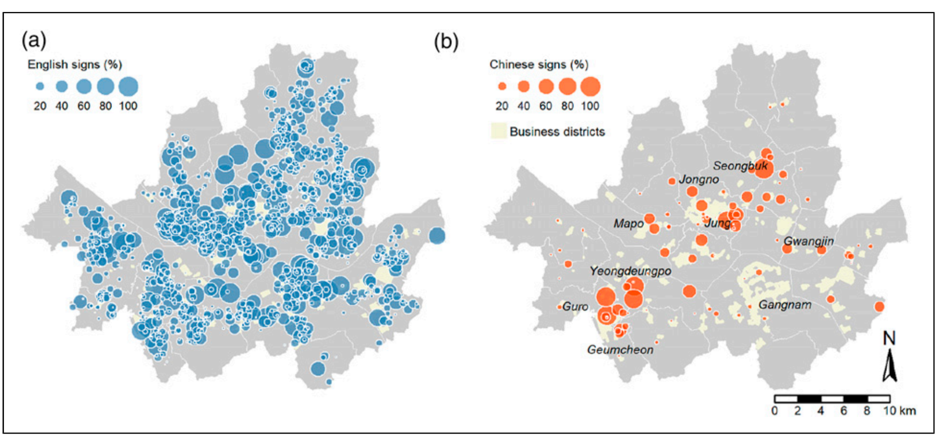

Figure 3. Geographic distribution of signs written in English and Chinese, 2018.

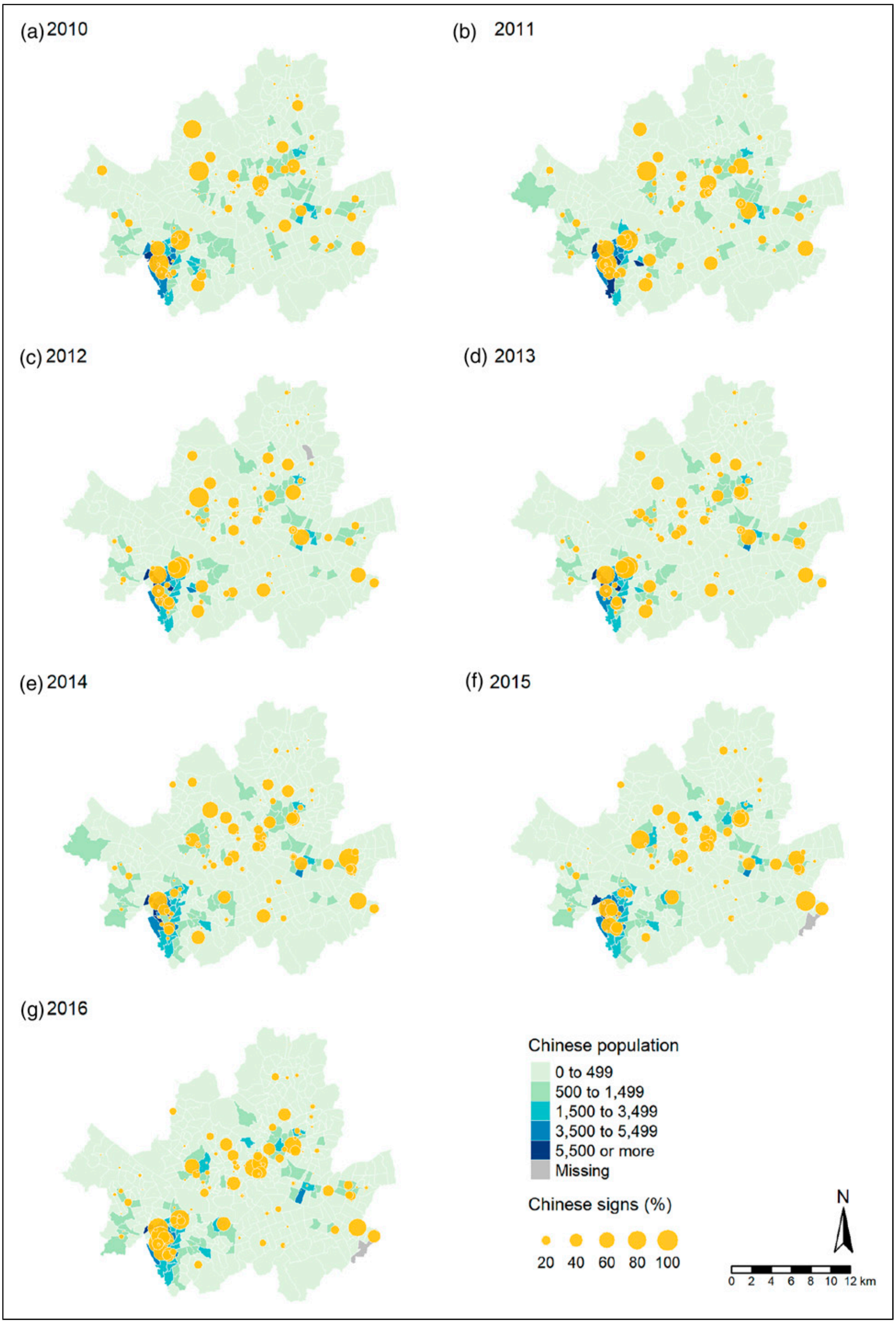

Figure 4. Geographic distribution of Chinese residents and signs written in the Chinese language, 2010–2016.

Levels of segregation

Figure 5 presents how the Chinese population changed over the period 2010–2018 by subgroup, along with the levels of their segregation measured by the spatial information theory index. For all subgroups, except the Chinese students, there are positive correlations between the population changes and the levels of segregation. In particular, the relationship is strong for Joseonjok and Chinese visitors, with correlation coefficients of 0.324 and 0.580, respectively.

The indices suggest that Chinese signs remained highly segregated in Seoul for the entire study period. This result is plausible considering their geographic distribution shown in Figure 4. However, a rapid increase and decrease in the segregation level have been observed, which may be worth mentioning. The spatial information theory index increased rapidly from 0.606 in 2010 to 0.657 in 2011. Over this period, the numbers of all Chinse subgroups in Seoul, especially Chinese visitors, have increased, which may have resulted in the emergence of more Chinese signs in the commercial sectors, such as Jongno and Jung.

On the other hand, the level of segregation declined considerably from 0.649 in 2013 to 0.601 in 2014. In a similar vein to the abovementioned increase, the decline also seems to be attributable to the Chinese population. In 2014, the number of Chinese visitors dropped by over 20,000. While it is difficult to conclude in this work that there is a causal relationship between the ethnic population and the linguistic landscape, the lower inflow of Chinese visitors may have resulted in the lower display of the Chinese language at retail shops, cafés, and restaurants to attract consumers.

To examine the relationship between ethnic segregation in the linguistic landscape and traditional residential segregation, Figure 6 compares the results from this work to the levels of Chinese residential segregation over the period 2010–2016. The census data pertaining to registered foreigners were used for the calculation of residential segregation, but the comparison is limited to 2016, as the data are available only until that year. At first glance, the most notable difference between the two approaches is that the levels of residential segregation are much lower than those of linguistic segregation, ranging from 0.18 to 0.22. This is, however, presumably because the census data are aggregated into administrative units, which means that small-scale clustering was not properly reflected during the estimation of residential segregation.

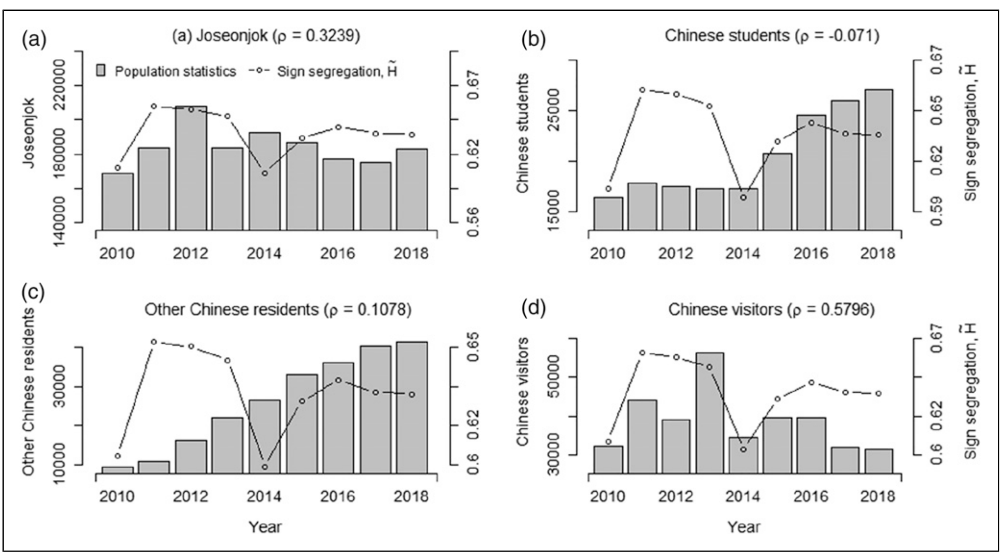

Figure 5. Correlations between Chinese population groups and the level of sign segregation.

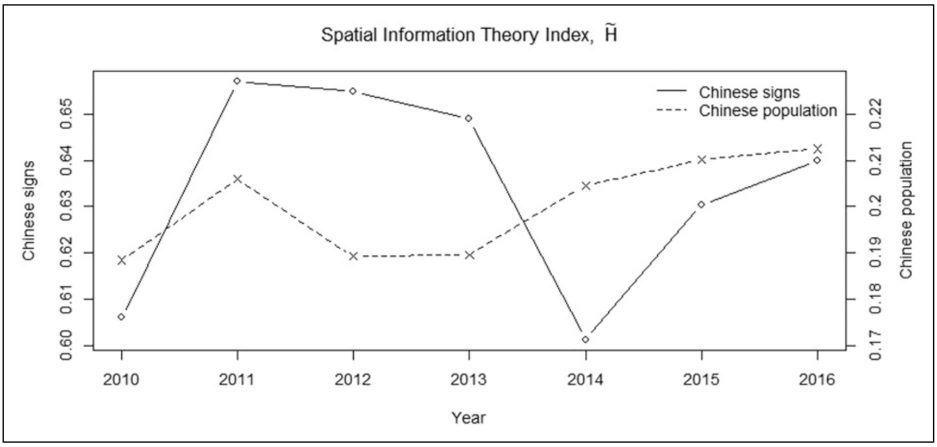

Figure 6. Levels of segregation for the Chinese population and signs, 2010–2016.

Looking at the temporal trends in the two results, the segregation in the linguistic landscape shows more fluctuation than the residential segregation. One possible explanation is that the linguistic landscape is mainly shaped by economic activities and changes more dynamically, which means that the level of segregation can also vary year by year to reflect the social reality. While census-based residential segregation provides a fundamental indicator regarding the social integration of long-term migrants, linguistic segregation on street-level images can shed light on the more dynamically changing economic aspect of the phenomenon, which other data sources have not addressed.

Conclusions

In this work, we proposed an alternative approach to studying ethnic segregation from the perspective of linguistic landscapes. A large volume of street-level images accumulated over the last two decades has enabled the exploration of linguistic landscapes at a larger scale than ever before. Since the prevalence of a specific language in public space implies the linguistic group inhabiting the area, its careful evaluation can reveal the degree of segregation between linguistically distinguishable ethnic groups at varying scales and resolutions.

To demonstrate the effectiveness of the proposed approach and examine its association with traditional residential segregation, we analysed the languages visible in the streetscape of Seoul. We collected over 10,000 images from an online map platform and calculated the relative proportions of Korean and Chinese signs in each urban scene. The degrees of segregation between the two linguistic groups were estimated using the spatial information theory index, and the results were compared to those of ethnic segregation based on the census.

Our empirical analysis showed that the signs written in the Chinese language were not randomly spread over the city; rather, they were concentrated in the southwest part of Seoul. Overall, this spatial pattern aligns with the distribution of Chinese residents reported in previous studies. However, linguistic clustering was also observed around the central business districts and universities, where many businesses and language schools for Chinese visitors and international students are located. While this clustering would attract more Chinese people to spend their daytime hours in these areas, it could have been overlooked in residential segregation, thereby consequently distorting their exposure to and isolation from other population groups in Seoul.

The levels of Chinese linguistic segregation on the street-level images were distinctive from those of residential segregation in two aspects. First, while residential segregation gradually increased between 2010 and 2018, except for 2011, there were more fluctuations in linguistic segregation. A linguistic landscape in a city is shaped mainly by advertising signs, banners, and billboards in commercial areas; therefore, it is market-driven to a certain extent. Unlike the residential neighbourhoods formed by long-term residents, commodified urban spaces change more dynamically to attract potential customers; this could be a reason for the varying levels of segregation in Chinese signs across the years.

Second, the values from the spatial information theory index were much higher for linguistic segregation than residential segregation. The former ranged between 0.601 and 0.657, whereas the latter only reached approximately one-third of those values, that is, from 0.189 in 2010 to 0.212 in 2016. However, this outcome is likely due to the data resolution rather than the actual distributions of the residential population and signs. Since the census data in this paper are aggregated into administrative units, any clustering that is smaller than this scale might not have been adequately addressed.

In summary, the results suggest that the proposed approach can provide additional insight into ethnic segregation, which is not offered by other conventional data sources. While recent studies exploiting mobile phone trajectories or smart card data have focused on the separation and isolation of individuals, street-level images would be more suitable for evaluating the segregation of places. In this respect, this type of data shares a similarity with the census, which is another place-oriented dataset. However, the proposed approach places more attention on commercial areas, where various social and economic activities occur not only by immigrants but also by short-term visitors and international students. This feature could make segregation through the lens of the linguistic landscape distinctive and thereby enable the exploration of a new dimension of the complex socio-demographic phenomenon.

Notwithstanding, it might be worth noting that the empirical results presented in this paper could be limited to the specific settings of our study area. The relationship between linguistic and residential segregation could be different in other configurations, such as in cities where long-term residents comprise the majority in the foreign population composition or in areas where ethnic minorities are linguistically assimilated. Furthermore, the collection and annotation of the street-level images were primarily manual processes, and the languages were classified into only several categories. Cantonese and Mandarin Chinese characters, for example, were not distinguished in this work; therefore, spatial patterns related to subgroup populations could not be addressed.

Further research is indeed required to supplement these limitations. However, it is challenging even for state-of-the-art technologies to detect and recognise texts from natural scene images, such as street-level images. Nevertheless, scene text detection and recognition are rapidly developing fields of computer vision, and the applications of neural network-based models have resulted in significant progress over the last several years. The annotated images used in this work can be considered useful training data for the automatic classification process in the future, which will eventually increase the reproducibility and transparency of the proposed approach.

More empirical analyses in various settings should also be conducted to understand the implications of linguistic segregation. There are still many things unknown about what street-level images truly convey about the experiences of people in society. Since street-level images essentially provide information on places instead of people, the degree of segregation obtained from linguistic landscapes should be interpreted with care. Certainly, the use of street-level images for evaluating ethnic segregation cannot replace traditional data sources, such as censuses. However, this approach can offer an alternative way of understanding the complex socio-demographic phenomenon from a new perspective, as with other emerging data sources in the era of big data.

Acknowledgements

The authors would like to thank the undergraduate and graduate students in the Department of Geography at Kyung Hee University, especially Woosung Choi, for his help in preparing the figures.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT) (NRF-2021R1C1C1009849).

ORCID iD

Seong-Yun Hong https://orcid.org/0000-0001-5049-8810

Notes

1. It is worth noting that the number of foreign residents in Seoul did not constantly increase from 2000 to 2020. Rather, it increased rapidly until the peak of 279,095 in 2011. Since then, however, the number has remained stable or even decreased, especially during the global pandemic.
2. This is the default setting for the R package 'seg'. We utilised this package to compute the spatial information theory index for linguistic landscapes.

References

Alba RD and Logan JR (1993) Minority proximity to Whites in suburbs: an individual-level analysis of segregation. *The American Journal of Sociology* 98(6): 1388–1427. https://www.jstor.org/stable/2781825

Backhaus P (2007) *Linguistic Landscapes a Comparative Study of Urban Multilingualism in Tokyo*. Clevedon: Multilingual Matters http://site.ebrary.com/lib/auckland/Doc?id=10170609

Ben-Rafael E, Shohamy E, Hasan Amara M, et al. (2006) Linguistic landscape as symbolic construction of the public space: the case of Israel. *International Journal of Multilingualism* 3(1): 7–30. DOI: 10.1080/14790710608668383.

Bourhis RY (1994) Introduction and overview of language events in Canada. *International Journal of the Sociology of Language* 105-106(1): 5–36. DOI: 10.1515/ijsl.1994.105-106.5.

Cenoz J and Gorter D (2006) Linguistic landscape and minority languages. *International Journal of Multilingualism* 3(1): 67–80. DOI: 10.1080/14790710608668386.

Choi W, Ha J and Hong S-Y (2021) Changes in the level of ethnic segregation during the spread of COVID-19: a case study of Garibong-dong and Daerim 2-dong in Seoul. *Journal of the Korean Cartographic Association* 21(1): 53–64. DOI: 10.16879/jkca.2021.21.1.053.

Chung S-Y and Lee J-H (2014) The development of foreigners' residential concentrations by nationality through migration in Korea: a case study of the Chinese immigrants in Daelim-dong, Seoul (in Korean). *The Geographical Journal of Korea* 48(1): 93–107.

Gorter D and Cenoz J (2015) Translanguaging and linguistic landscapes. *Linguistic Landscape* 1(1–2): 54–74. DOI: 10.1075/ll.1.1-2.04gor.

Hong S-Y (2020a) Linguistic Landscapes on Street-Level Images. *ISPRS International Journal of Geo-Information* 9(1): 57. DOI: 10.3390/ijgi9010057.

Hong S-Y (2020b) Open-source tools for the measurement of socio-spatial segregation in activity spaces. *Transactions in GIS* 24(5): 1248–1263. DOI: 10.1111/tgis.12640.

Huebner T (2006) Bangkok's linguistic landscapes: environmental print, codemixing and language change. *International Journal of Multilingualism* 3(1): 31–51. DOI: 10.1080/14790710608668384.

Jeong Y, Ha J, Choi W, et al. (2021) Chinese segregation in linguistic landscapes: a case study of Seoul. *Journal of the Korean Geographical Society* 56(1): 113–121. DOI: 10.22776/kgs.2021.56.1.113.

Jun M-J, Ha S-K and Jeong J-E (2013) Spatial concentrations of Korean Chinese and determinants of their residential location choices in Seoul. *Habitat International* 40: 42–50. DOI: 10.1016/j.habitatint.2013.02.002.

Kim T-S (2021) Center and margin on the margin: a study of the multilayered (Korean) Chinese Migrant Neighborhood in Daerim-dong, South Korea. *Geoforum* 120: 165–175. DOI: 10.1016/j.geoforum.2021.01.027.

Korea Tourism Organization (2022) *Tourism Statistics*. Retrieved 2022-06-13 from https://datalab.visitkorea.or.kr/

Landry R and Bourhis RY (1997) Linguistic landscape and ethnolinguistic vitality: an empirical study. *Journal of Language and Social Psychology* 16(1): 23–49. DOI: 10.1177/0261927x970161002.

Lawrence CB (2012) The Korean English linguistic landscape. *World Englishes* 31(1): 70–92. DOI: 10.1111/j.1467-971X.2011.01741.x.

Lee JS (2019) Multilingual advertising in the linguistic landscape of Seoul. *World Englishes* 38(3): 500–518. DOI: 10.1111/weng.12427.

Leeman J and Modan G (2009) Commodified language in Chinatown: a contextualized approach to linguistic landscape. *Journal of Sociolinguistics* 13(3): 332–362. DOI: 10.1111/j.1467-9841.2009.00409.x.

Li F and Wang D (2017) Measuring urban segregation based on individuals' daily activity patterns: a multidimensional approach. *Environment and Planning A: Economy and Space* 49(2): 467–486. DOI: 10.1177/0308518x16673213.

Li W (2006) *From Urban Enclave to Ethnic Suburb: New Asian Communities in Pacific Rim Countries*. Honolulu: University of Hawai'i Press. Table of contents only http://www.loc.gov/catdir/toc/ecip066/2005037660.html

Massey DS (1985) Ethnic residential segregation: a theoretical and empirical review. *Sociology and Social Research* 69: 315–350.

Müürisepp K, Järv O, Tammaru T, et al. (2022) Activity spaces and big data sources in segregation research: a methodological review. *Frontiers in Sustainable Cities* 4: 861640. DOI: 10.3389/frsc.2022.861640.

Nikolaou A (2017) Mapping the linguistic landscape of Athens: the case of shop signs. *International Journal of Multilingualism* 14(2): 160–182. DOI: 10.1080/14790718.2016.1159209.

Östh J, Shuttleworth I and Niedomysl T (2018) Spatial and temporal patterns of economic segregation in Sweden's metropolitan areas: a mobility approach. *Environment and Planning A: Economy and Space* 50(4): 809–825. DOI: 10.1177/0308518x18763167.

Papen U (2012) Commercial discourses, gentrification and citizens' protest: the linguistic landscape of Prenzlauer Berg, Berlin1. *Journal of Sociolinguistics* 16(1): 56–80. DOI: 10.1111/j.1467-9841.2011.00518.x.

Reardon SF and O'Sullivan D (2004) Measures of spatial segregation. *Sociological Methodology* 34(1): 121–162. DOI: 10.1111/j.0081-1750.2004.00150.x.

Russell E (2022, 2022-05-24) Street View turns 15 with a new camera and fresh features. *The Keyword* https://blog.google/products/maps/street-view-15-new-features/

Ruzaitė J (2017) The linguistic landscape of tourism: multilingual signs in Lithuanian and Polish resorts. *Journal of Estonian and Finno-Ugric Linguistics* 8(1): 197–220. DOI: 10.12697/jeful.2017.8.1.11.

Seiferling I, Naik N, Ratti C, et al. (2017) Green streets − Quantifying and mapping urban trees with street-level imagery and computer vision. *Landscape and Urban Planning* 165: 93–101. DOI: 10.1016/j.landurbplan.2017.05.010.

Seoul Institute (2022) *The Seoul Rearch Data Service*. Retrieved 2022-06-13 from https://data.si.re.kr/data/

Shin EJ (2022) Spatial segregation of Chinese immigrants in Seoul, South Korea, during the COVID-19 pandemic: evidence from population data derived from mobile phone signals. *The Social Science Journal*: 1–22. DOI: 10.1080/03623319.2021.2014207.

Silm S and Ahas R (2014) Ethnic differences in activity spaces: a study of out-of-home nonemployment activities with mobile phone data. *Annals of the Association of American Geographers* 104(3): 542–559. DOI: 10.1080/00045608.2014.892362.

Silm S, Ahas R and Mooses V (2018) Are younger age groups less segregated? Measuring ethnic segregation in activity spaces using mobile phone data. *Journal of Ethnic and Migration Studies* 44(11): 1797–1817. DOI: 10.1080/1369183X.2017.1400425.

Spolsky B and Cooper RL (1991) *The Languages of Jerusalem*. Oxford: Clarendon Press. Publisher description http://www.loc.gov/catdir/enhancements/fy0605/91023018-d.html

Statistics Korea (2022) *Korean Statistical Information Service (KOSIS)*. Retrieved 2022-06-13 from https://kosis.kr/statHtml/statHtml.do?orgId=111&tblId=DT_1B040A9C&conn_path=I2

Tulp S (1978) Reklame en tweetaligheid. Een onderzoek naar de geografische verspreiding van Franstalige en Nederlandstalige affiches in Brussel [Advertising and bilingualism. An investigation into the geographical distribution of French-speaking and Dutch-speaking posters in Brussels]. *Taal en Sociale Integratie* 1: 261–288.

van Ham M and Tammaru T (2016) New perspectives on ethnic segregation over time and space. A domains approach. *Urban Geography* 37(7): 953–962. DOI: 10.1080/02723638.2016.1142152.

Vandenbroucke M (2015) Language visibility, functionality and meaning across various TimeSpace scales in Brussels' multilingual landscapes. *Journal of Multilingual and Multicultural Development* 36(2): 163–181. DOI: 10.1080/01434632.2014.909442.

Wong DWS and Shaw S-L (2011) Measuring segregation: an activity space approach. *Journal of Geographical Systems* 13(2): 127–145. DOI: 10.1007/s10109-010-0112-x.

Yin L, Cheng Q, Wang Z, et al. (2015) Big data' for pedestrian volume: exploring the use of Google Street View images for pedestrian counts. *Applied Geography* 63: 337–345. DOI: 10.1016/j.apgeog.2015.07.010.

Yun J-H, Kwon O-K and Ma K-R (2014) Changing spatial distribution of Chinese population in Seoul: a comparison between Korean-Chinese and non-Korean Chinese. *Journal of The Korean Regional Development Association* 26(2): 39–54.

Zhang T, Duan X, Wong DWS, et al. (2021) Discovering income-economic segregation patterns: a residential-mobility embedding approach. *Computers, Environment and Urban Systems* 90: 101709. DOI: 10.1016/j.compenvurbsys.2021.101709.

Seong-Yun Hong is an associate professor of geography at Kyung Hee University, South Korea. He is the principal investigator for the research project "Developing and implementing a GeoAI model for urban analytics", of which this paper is a part. His interdisciplinary research interests lie in using computational methods to explore the population dynamics in urban areas through the combination of theories and methods from geography and sociology.

Yeorim Kim is a graduate student in the Department of Geography at Kyung Hee University. Yeorim joined the research group in 2021 and participated in this project as a research assistant from the first semester of 2022.

Yongchae Lee is an undergraduate student in the Department of Housing and Interior Design at Kyung Hee University and is a prospective graduate student in geography. She participated in this project as a research assistant for the first semester of 2022.