8-16-2024 9:00 AM

Extracting Prototypes From Lexical Feature Norms for Settlement Concepts

Chelsea-Leigh M. McKenzie, *The University of Western Ontario*

Supervisor: Dr. John Paul Minda, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Psychology

Abstract

The present study explored whether people share a common understanding of different settlement concepts despite individual variation. Participants completed a property listing task where they were asked to generate features for 57 settlement concepts. Hierarchical cluster analysis identified distinct clusters based on shared features. Central tendencies extracted from clusters at different levels of abstraction revealed featural prototypes and an overall family resemblance structure. To probe the effects of regional context on conceptual structure, subsequent cluster analyses used a subset of participants who were long-term residents of Canada or the United States. Prototypical features varied regionally, suggesting an effect of geographical region on conceptual structure. However, the results should be interpreted cautiously, as more data are needed to understand such differences in representation. Findings centralize the utility of semantic feature norms in understanding how people collectively think about where they live, and the importance of context effects on representations of settlements.

Keywords: Concepts, settlements, semantic features, prototypes, family resemblance

Summary for Lay Audience

Most people know what a city is, but when asked to describe a city, there are likely differences in the features that people name. For instance, one individual who primarily takes the subway might say a city has public transportation, but an individual who commutes by car might focus on traffic and parking. Despite these differences, these two individuals are not confused when they talk together about cities. Are these two individuals really talking about the same thing; is there a universal city concept that is shared across North American English speakers? To better understand overall similarities in how people think about settlements, participants (*N* = 122) were asked to generate features for 57 types of settlement concepts. A clustering method that groups similar items was applied to the features participants generated, revealing several distinct groupings of concepts. The clustering approach helped show the similarities among the groupings and the typical features of settlement concepts that grouped together. As an extension, the data were split by region to compare the features generated by long-term residents of Canada with features generated by long-term residents of the United States. Canadians and Americans tend to speak the same language, and because the two countries are geographically close, they generally have similar linguistic labels for settlement concepts. However, the Canadian and American environments are different, and as such, the features generated by Canadians and Americans for these settlement concepts should differ, too. My results showed distinct clusters for Canadians and Americans, supporting my prediction that people from different regions would generate different features for the same concepts. However, more information is needed to fully grasp these regional differences. This empirical

description of human settlement concepts helps clarify how people collectively think about and understand the built environment and how such representations can change depending on where they live.

Acknowledgements

First and foremost, I would like to thank my supervisor and mentor, Dr. John Paul Minda. This project would not have been possible without Paul's guidance, expertise, and willingness to assist along the way. Thank you for your trust and for encouraging me to pursue my interests in language and concepts, and for cultivating a lab environment where curiosity, learning, and growth are nurtured and valued. I would also like to thank my committee members, Dr. Ken McRae and Dr. Stefan Köhler for their advice and feedback at various stages of this project.

Thank you to all members (both past and present) of the Minda Lab who offered their feedback, support, and encouragement. A special thanks go out to Dr. Ana Ruiz Pardo, Tim Qiu, and Julia Ignaszewski for their general support while we were in the lab together, and for their friendship. I would also like to extend a big thank you to Liza Kholwadwala, Urooj Anees, Lydia Tzianas, Nicole Carrier, and Alexia Romita who helped me sort through and code thousands of features during the preprocessing phase. In addition, the supportive and empowering dynamic of the Brain and Mind at WIRB (special shout out to the 5th floor) and within the CDBS program would not be what it is without the colleagues and friends I have met here. I feel so incredibly fortunate to work alongside such a wonderful group of people.

Lastly, my most sincere gratitude is owed to my family, especially my partner Camilo and our dogs. My Jules, who had no idea what any of this meant but intuitively knew when I needed a break; who regularly checked in by resting her face on my lap during many phases of this work. And Lucie, whose bids to play provided much-needed respite as I wrapped up. You have made all the difference.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

1 Introduction

Imagine you are planning a trip to visit a friend in a new place, and they describe their new locale as a *Tourist Town*. What features would you expect to find there? You might anticipate locally-themed souvenir shops, seasonal restaurants, and bicycle or scooter rentals. Conversely, if your friend refers to their new environment as an *Industrial City*, you might expect pollution, factories, and noise. This is an example of an inductive inference, which depends on concepts. Communication of concepts and inductions depends on some degree of shared common knowledge. But how do you know you and your friend are both thinking of the same things? Whether it be a tourist town or an industrial city, the features called to mind are part of a conceptual representation that shapes understanding, expectations, and behaviours in relation to the place.

1.1 Concepts: Development, Organization, and Structure

Concepts are central to human cognition and a fundamental aspect of complex thought. People rely on concepts to organize and make sense of a vast amount of information by categorizing, drawing inferences, remembering, learning, and making decisions. The structure underlying various kinds of concepts can differ depending on context (e.g., culture, language, location), expertise, and purpose. There is a need to empirically examine and validate the characteristics of basic-level categories across different domains to understand how the world is organized and perceived. Cognitive scientists employ a range of behavioural paradigms and methods to probe the nature of these representations. Although people can come to understand concepts in several ways, the words people use to describe concepts seem uniquely suited to reveal how people perceive the content of the environment.

Over a lifetime, our spatial, social, cultural, and environmental context expands. Unless our caregivers constantly travel and expose us from infancy, humans tend to experience the environment gradually. Our first experiences of the world then usually take place in our homes where people are exposed to caregivers, food, and household objects. This context slowly expands to the surrounding neighbourhood and, eventually, our hometown. During this time, people become familiar with the smells, sounds, objects, events, people and language they are regularly exposed to. Our perceptual and linguistic components together then become part of our overall knowledge representation. These representations dynamically reflect the evolution and expansion of experiences acquired over space and time and are unique from one individual to the next.

There are regularities in the way humans acquire and organize conceptual information. When imposing a structure on a set of concepts, research has shown that concepts tend to be treated hierarchically with each level representing varying degrees of abstraction or inclusiveness (Collins & Quillian, 1969; Grill-Spector & Weiner, 2014; Mahon & Caramazza, 2009; Murphy & Lassaline, 1997; Rosch et al., 1976). The hierarchical account of classification described by Collins and Quillian accounts for people’s behaviour and does not literally imply that concepts are represented this way in the brain. Nonetheless, the hierarchical structure of conceptual organization aids classification, allowing us to employ induction to form generalizations from specific observations, and to use deduction to derive specific conclusions from general principles, facilitating the formation of behavioural equivalence classes and schemas for entities belonging to a class. The hierarchical representation is comprised of three levels of classification: the subordinate, basic, and superordinate levels.

The subordinate level of the hierarchy is the most specific of the three levels. Exemplars at the subordinate level are specific instantiations of concepts at the basic level (e.g., [Sup] plant > [B] tree > [Sub] maple, [Sup] animal > [B] dog > [Sub] golden retriever, [Sup] settlement > [B] city > [Sub] Toronto, etc.). At the subordinate level, there is a high degree of within-category and between-category similarity (e.g., city > Toronto shares features with city > Los Angeles and with other kinds of basic-level settlements, i.e., town > Milton). However, due to the high degree of feature overlap at the subordinate level, similarity between exemplars is not always a reliable indicator of category membership. A more robust, typical representation of each category at the basic level, i.e., a prototype, can help classify new instantiations of subordinate concepts.

The basic level of categorization possesses properties that distinguish it from both the superordinate and subordinate levels. Concepts at the basic level allow for reliable inductive inferences due to having high within and low between-category similarity (Gelman & Davidson, 2013; Gelman & Markman, 1986; Sloman, 1993). People also learn concepts at the basic level first (Brown, 1958; Markman & Hutchinson, 1984; Mervis & Crisafi, 1982). Consider, for example, how a young child might initially generalize all animals with fur, paws, ears, a nose, and a tail, to be members of the *cat* category. Perhaps this is due to limited exposure to animals, with the only exposure being restricted to their pet cat. Imagine the first time that child sees a dog and calls that dog a cat. The parent would likely correct them by stating the correct category, *dog*. However, the distinction between what a dog is and what a cat is might not fully crystallize until the child observes another dog and has the opportunity to observe the correct classification or to apply it themselves. When learning to classify dogs and cats, consider how specific instantiations or exemplars – at the subordinate level – share some similar features (Medin & Schaffer, 1978). However, people begin to form distinctions at the basic level of that hierarchy (Rosch et al., 1976) as it is more challenging to distinguish at the subordinate level (Jolicoeur et al., 1984; Zhuang & Lingnau, 2022). It is this exposure to other basic-level concepts that eventually allows us to make correct classifications via both distinction and similarity (Goldstone, 1998). In this case, with experience, that child will eventually come to classify dogs correctly because they look like other dogs but also because they share a qualitatively different set of features that distinguishes them from cats. In essence, people initially classify using information based on what they know and what they do not know, but with experience, people refine this knowledge utilizing feedback, updating their model until classifications approximate being correct.

The superordinate level represents the highest and most general tier of classification, where similarity between and within categories is minimal. This broad level of categorization facilitates the grouping of distinct types (e.g., tree, dog, city) into general domains (e.g., plant, animal, settlement) by allowing activation to spread bidirectionally from specific instances to more general concepts (Collins & Loftus, 1975). The experiments by Rosch et al. (1976) on categorization at different taxonomic levels revealed differences in how participants identified and listed features at different levels of abstraction. In their study, they found that objects categorized at the superordinate level had the fewest common features identified by participants, indicating a lower cue validity (i.e., whether a given feature is a reliable cue for determining category membership).

The hierarchical system discussed above makes assumptions about the degree of specificity at each level. A related question is, how are the concepts represented at each level? The basic level tends to be represented by *prototypes*, defined as central tendencies of typical or ideal features (Rosch et al., 1976). An example of a prototype for the dog category could be a golden retriever because it possesses typical features but also a set of ideal features possessed by other members of the dog category. Because of these unique properties as a prototype, golden retrievers are recognized faster and with greater ease than less prototypical members of the dog category (McCloskey & Glucksberg, 1979).

The clusters of features that represent a particular concept are not strictly necessary and sufficient to justify placing them in a particular category. Concepts tend to possess a more flexible quality with fuzzy boundaries, known as a family resemblance. When a concept possesses a *family resemblance* structure, it has a cluster of features that identifies it as being a part of a category, but there's no single feature that perfectly predicts category membership (Rosch & Mervis, 1975; Wittgenstein, 2009). Continuing with our dog example, there is not one particular feature that distinguishes kinds of dogs or distinguishes dogs from other animals. It is generally a cluster of features that come together and settle on the classification of being members of the dog category due to some degree of structural alignment between the member to be classified and the prototype for the overall category (Petkov & Petrova, 2019).

Both prototype and exemplar models predict category typicality based on family resemblance, but the models differ in the approach used to achieve this prediction (Voorspoels et al., 2011). In the exemplar model, every encountered instance of a category is retained in memory with each exemplar contributing individually to the overall family resemblance (Nosofsky, 1986; Stanton et al., 2002). Therefore, typicality is assessed by comparing new items to many stored exemplars. In contrast, the prototype model maintains a single abstracted representation that consists of the most common features of the category members (Minda & Smith, 2011; Posner & Keele, 1968; Rosch & Mervis, 1975). Despite these operational differences, exemplar and prototype models converge in their prediction that greater family resemblance leads to higher typicality (Dieciuc & Folstein, 2019). Moreover, both models rely on features (Medin & Schaffer, 1978; Minda & Smith, 2002; Smith & Medin, 1981). Overall, people tend to utilize common features to group entities, though theoretical differences lie in whether such features are abstracted into a single prototype or remain distributed among many remembered exemplars.

The literature discussed so far adds strength to the assertion that prototypes are well-suited to studying global representations of concepts. Instead of focusing on the fine-grained detail of every stored exemplar feature, prototypes represent the average featural tendency, providing an overall representation of a concept. Given that prototypes are characteristic of the basic level (Rosch et al., 1976), and basic-level categories are the first to be formed when perceiving the environment (Brown, 1958), the most commonly used in language, and used to draw inductive inferences (Barsalou, 1985), prototypes serve as effective and efficient means of understanding conceptual structure.

1.2 Approaches to Studying Representation

Several methods exist to determine the contents of conceptual representation. For example, categorization tasks, rating-based methods (e.g., semantic differential via bipolar feature ratings, similarity ratings, typicality ratings, etc.), priming, eye tracking, and lexical decision. Another method used to study concepts' internal structure is to obtain feature production norms. Two feature norming methods that have been used are behavioural-based and corpus-based. However, behavioural feature norms are considered the gold standard approach (Buchanan et al., 2019; Devereux et al., 2014; Garrard et al., 2001; Kivisaari et al., 2023; McRae et al., 2005; Vinson & Vigliocco, 2008).

Behavioural-based methods have used free association norms or semantic feature production norms. Free association norms ask participants to list the first word that comes to mind for a given concept (Nelson et al., 2004), whereas semantic feature production norms ask participants to list its features (McRae et al., 2005). Further, semantic feature production norming paradigms prompt participants in the instructions with kinds of attributes (e.g., how it smells, what it looks like, what it sounds like, etc.) having the effect of eliciting a broad cross-modal range of features. In both cases, the words or features generated for concepts are quantified in terms of frequency of mention, perceived importance, distinctiveness to the concept, and relational or associative strength. Behavioural feature norms and their resulting statistics can then be used to create experimental stimuli that test theories about the structure of semantic memory, behaviour, and how these relate to various contextual factors (McRae et al., 2005).

1.3 The Current Study

Most of the research uncovering the internal structure of categories and concepts has included entities such as objects, events and activities, colours, directions, odours, and both natural and artificial stimuli. But what about the different categories of settlement; the background context in which all exposure to these entities is framed? Recent research in cognition and geocomputation has approached this idea through the study of landscape concepts (Purves et al., 2023; Striedl et al., 2024; van Putten et al., 2020) and cognitive mapping of urban space (Hou et al., 2021; Jang & Kim, 2019). Though previous research has systematically studied representations of objects (Buchanan et al., 2019; Kivisaari et al., 2023; McRae et al., 2005; Valério et al., 2023), colours (Berlin & Kay, 1969), and events (Vinson & Vigliocco, 2008), including the collection concreteness (Brysbaert et al., 2014; Muraki et al., 2023), and socialness (Diveica et al., 2023) ratings for these data, and despite the closely related work on landscape concepts (Purves et al., 2023; Striedl et al., 2024; van Putten et al., 2020), no norms have been collected for settlement concepts in this way.

There are several reasons why the underlying semantic content of settlement concepts is important. For one, every person, object, colour, event, and experience takes place within the backdrop of some form of settlement context; they are a central, enduring, and formative part of everyday human experience. And yet, an empirical set of semantic content has not been derived for this category, hindering a full understanding of how settlements are conceptualized across individuals and regions. A technical report released in 2018 by the United Nations Human Settlements Programme details the need for extracting a consensus on what it means to be a *city*, and how to distinguish cities from other kinds of settlement concepts (UN-Habitat, 2018). Further, the lack of agreement on what *city* means poses a challenge to achieving Sustainable Development Goal 11 of making cities and human settlements inclusive, safe, resilient, and sustainable (UN-Habitat, 2018). The methodological approach within UN-Habitat's 2018 technical report suggests spatial analysis using satellite imagery to assess settlement morphology and density. However, there is no mention of inquiring how citizens conceptualize settlements. How can defining a city based on measuring the degree of urbanization by looking at population size and density alone, without asking citizens across regions what these concepts mean to them, make cities more inclusive, safe, resilient, and sustainable for the people who live there?

A prototype extracted for different kinds of settlement concepts could complement the UN-Habitat proposed approach by standardizing meaning across different regions, allowing cities to be compared while affording enough flexibility to adapt to local contexts. Prototype models appear to be particularly relevant for large and complex categories (Minda & Smith, 2001; Murphy & Wisniewski, 1989), such as settlements. According to Minda and Smith (2001), as the size and complexity of the category increases, the effectiveness of prototype-based learning is enhanced, simplifying learning and categorization processes by emphasizing common features and minimizing cognitive overload. Incorporating cognitive science into the UN-Habitat's proposed approach to achieving Goal 11 could be used to enhance urban planning and policy development by ensuring that meaning and strategies align with the perceptions and experiences of citizens.

The present study sought to bridge this ontological gap by empirically deriving the conceptual structure of settlement concepts using the property listing task, originally described by Rosch et al. (1976) though using the methodology employed by McRae et al. (2005). The aggregate features obtained from the property listing task were used to derive overall feature production frequencies for each of the concepts, indicating which features are central to a concept and identifying co-occurrences of attributes in clusters of related concepts. The latter characterizes the overall family resemblance for each cluster via a prototypical set of features. By characterizing the conceptual structure of settlements using behavioural feature norms, prototypical features can be identified and, thereby, aspects of similarity and differences in how people think about the general categories used to describe where they live. Since the dataset was obtained from North American participants and there was a roughly even split between Canadians and Americans, I was able to test the hypothesis that there would be regional differences in representations of settlement concepts depending on where participants currently lived. If differences are found in feature type and production frequency as a function of region, then effects of context on feature saliency and typicality are implied.

Chapter 2

2 Constructing Representations of Settlements

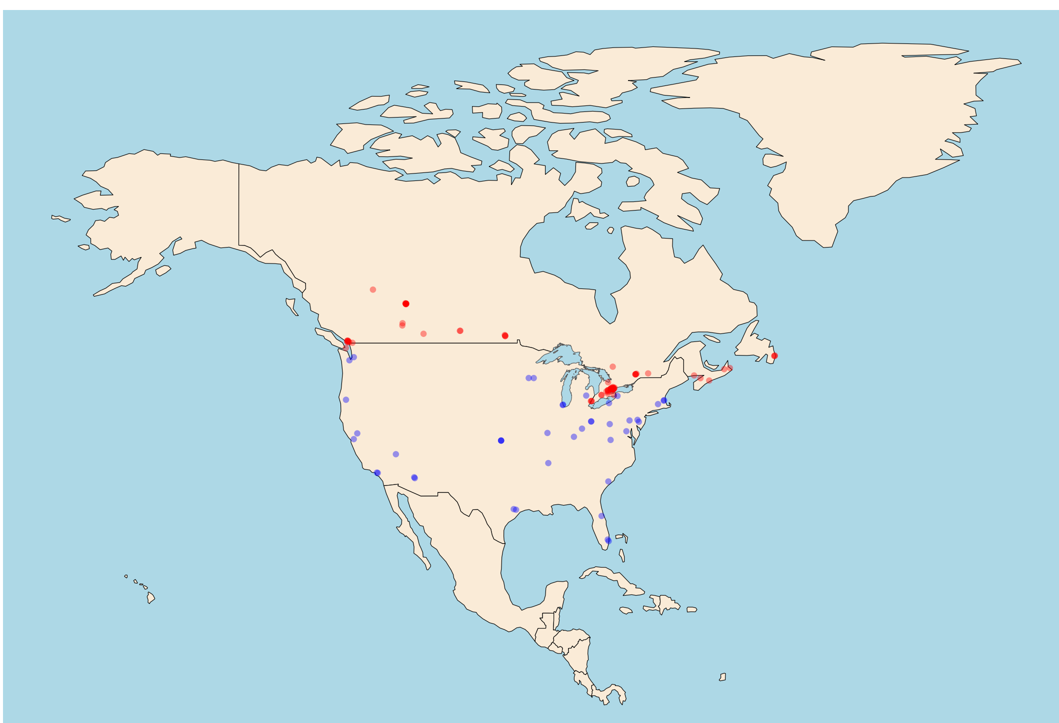
2.1 Methods

2.1.1 Participants

One hundred and twenty-two participants were recruited for an online study through Prolific (Prolific, 2023, Retrieved July 7, 2023, from www.prolific.com) in exchange for GBP £6.00/hr. Participants' ages ranged from 18 to 72 years, with a median age of 33 years (IQR: [27.25, 40]). All participants were residents of Canada or the United States and were fluent in English (see Appendix A for a summary of participant demographic characteristics). Participants were eligible to participate if they could read and understand the Letter of Information, including the purpose and methods of research, and choose to participate.

Figure 1

Geographic Distribution of Participants



Note: Each point represents a participant's location. Participants in Canada are represented by red points, and participants in the United States of America in blue.

2.1.2 Materials and Design

The familiarity rating task (Appendix B), property listing task (Appendix C), and demographics questionnaire (Appendix D) were created using Qualtrics software (Qualtrics, Provo, UT), Version July 2023, Copyright © 2023 Qualtrics.

2.1.3 Stimuli

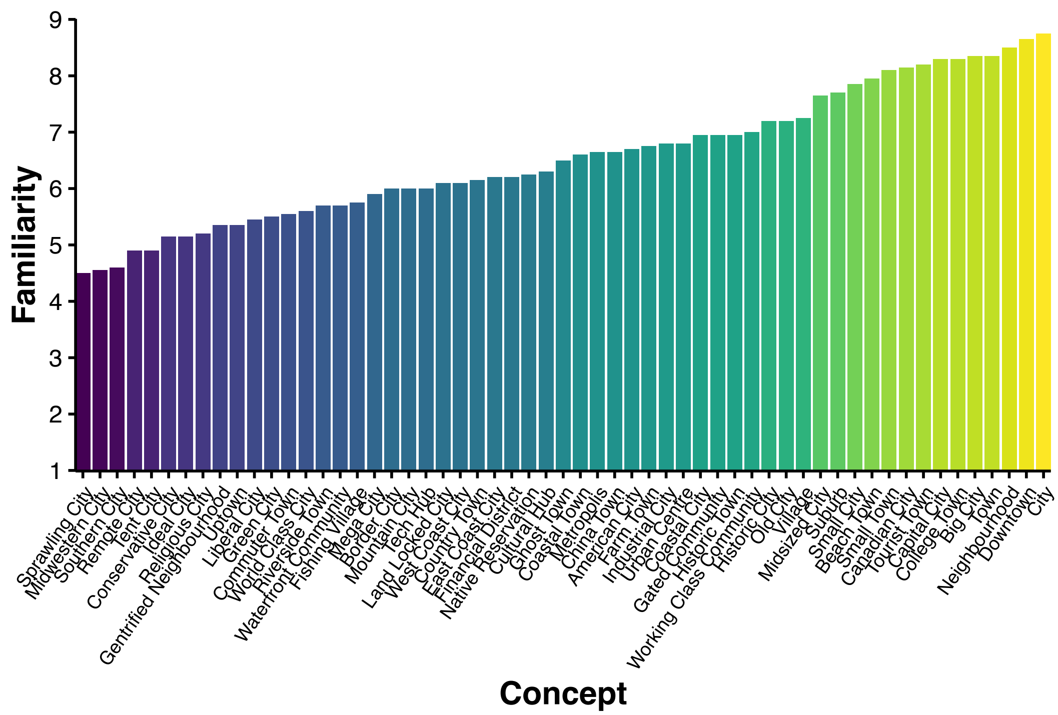
The concepts used for the property listing task were developed in two phases. The first phase involved brainstorming a list of different kinds of settlements, and the second phase in which I obtained familiarity ratings for the brainstormed list. The Corpus of Contemporary American English (COCA) word frequencies were obtained for each of the concepts from english-corpora.org. Since the set of concepts included collocations, where possible, I used existing word frequencies. Collocated concepts that were not present in COCA were averaged to obtain the overall frequency for the concept. Frequency per million (FPM) was used to normalize word counts to render them comparable across texts of different sizes in the corpus, which involved dividing the total frequency of the word in the corpus by the total number of words in the corpus multiplied by one million. One was then added to the FPM before taking the log to ensure there were no negative values. See Appendix F for all other statistics for the 57 concepts.

2.1.3.1 Familiarity Ratings.

Twenty colleagues in the Psychology and Neuroscience graduate programs at the University of Western Ontario were asked to indicate their familiarity with each of the 60 brainstormed concepts. The cut-off criterion for inclusion in the final set of concepts used in the property listing task was a mean familiarity rating of ≥ 4.50. The following three concepts received low average familiarity ratings, reducing the set from 60 to 57: *Military Town* (*M* = 4.35, *SD* = 2.28), *The Boonies* (*M* = 4.40, *SD* = 3.07), and *Rust Belt City* (*M* = 2.85, *SD* = 2.35). See Figure 2 for the average familiarity ratings of the final set of 57 concepts.

Figure 2

Average Familiarity per Concept



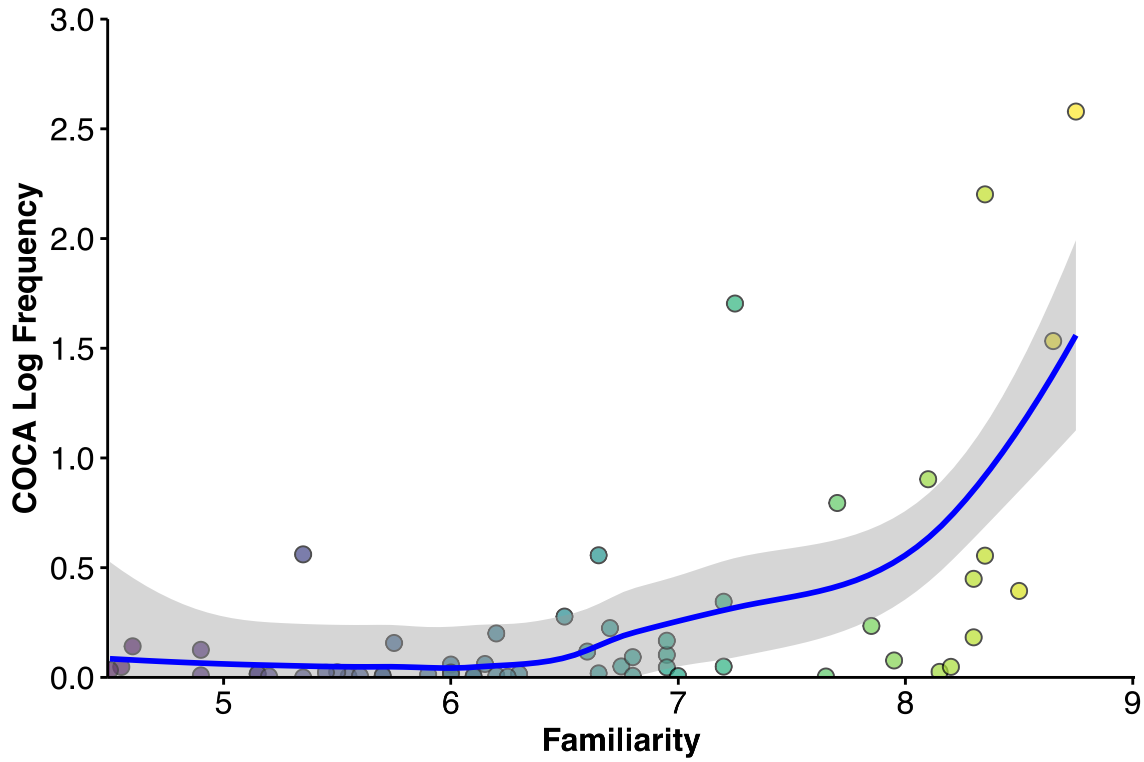
Note: Level of familiarity was assessed using a Likert scale with values ranging from 1 (not at all familiar) to 9 (extremely familiar). The average familiarity rating across the final set of 57 concepts indicated moderate to high familiarity overall; however, some concepts were more familiar to some participants than others (*M* = 6.56, *SD* = 1.94).

2.1.3.2 Familiarity and Corpus Word Frequency.

A Spearman's rank-order correlation was run to determine the relationship between rated familiarity and corpus word frequency. The Corpus of Contemporary American English (COCA; Davies, 2008) was selected as the measure of comparison against the rated familiarity of the concepts due to the COCA being a widely used corpus, and having vast and diverse contents of over one billion words. There was a moderate, statistically significant positive correlation between familiarity and COCA word frequency, $r_s(55) = .52$, $p < .001$ (see Figure 3). Words that appear more frequently in the COCA tend to be rated as more familiar. Therefore, the settlement concepts selected aligned with what is expected of the general population with a range of frequency and familiarity, and the words used for the settlement concepts were not unusual words.

Figure 3

Correlation Between Familiarity and COCA Word Frequency



Note: Each point represents a concept. COCA values were log-transformed + 1. Familiarity is the average rating per concept. The data were fit with a LOESS (Locally Estimated Scatterplot Smoothing) curve using locally weighted regression. The grey area represents the confidence interval.

2.1.3.3 Concreteness Ratings.

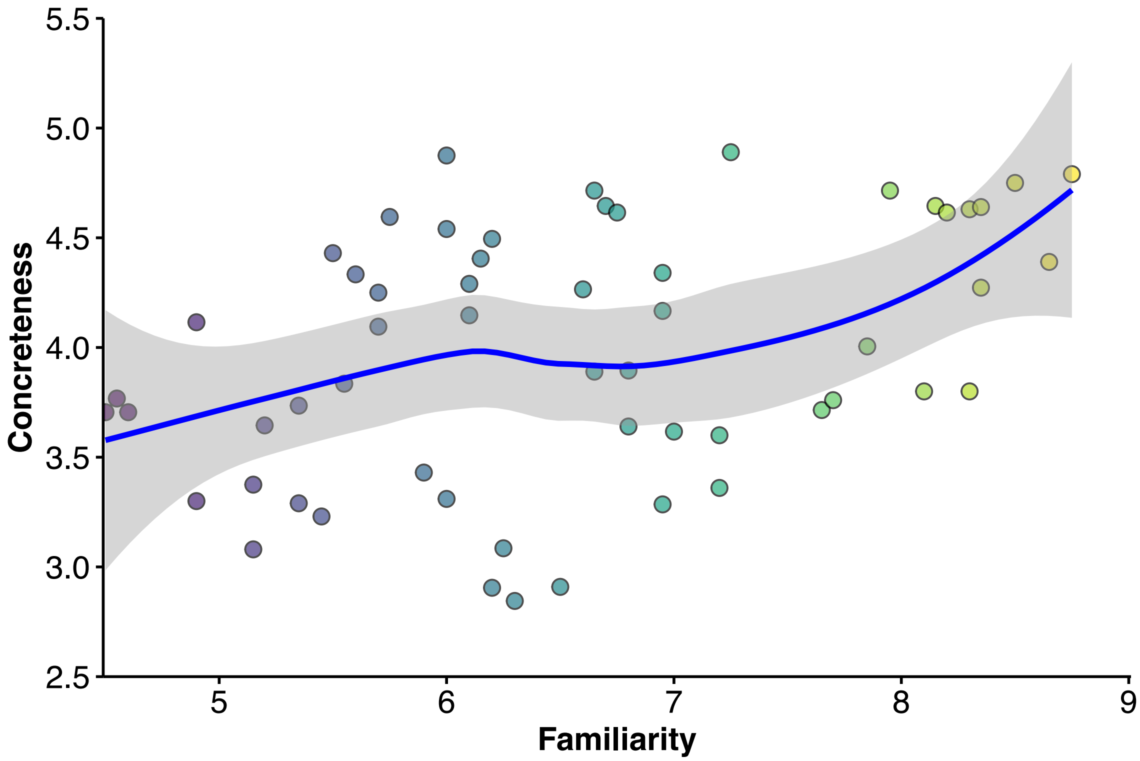
Concreteness ratings were obtained from Brysbaert et al. (2014) and Muraki et al. (2023). All items were rated on a scale from 1 to 5, where 1 indicated very abstract and 5 indicated very concrete. Existing concreteness ratings were used when possible. Averages were taken for collocated items that were not present in either dataset by obtaining the concreteness ratings for each word and then dividing across the number of words in the concept.

2.1.3.4 Familiarity and Concreteness.

The relationship between familiarity and concreteness was investigated using Spearman's rank-order correlation. There was a moderate, statistically significant positive correlation between familiarity and concreteness, $r_s(55) = .40$, $p = .002$ (see Figure 4). Items that were rated as more concrete were rated as more familiar.

Figure 4

Correlation Between Familiarity and Concreteness Ratings



Note: Each point represents a concept. The data were fit with a LOESS curve using locally weighted regression. The grey area represents the confidence interval.

2.2 Procedure

2.2.1 Property Listing Task

The property listing task was designed to systematically obtain an empirical set of features for different types of settlement concepts. This task has been used to study natural language use in the organization of categories and concepts (Buchanan et al., 2019; Kivisaari et al., 2023; McRae et al., 2005; Rosch et al., 1976) and to understand the content of semantic memory (Collins & Loftus, 1975; Collins & Quillian, 1969; Murphy et al., 2012; Taylor et al., 2007). The current task was adapted from methodology used by McRae et al.(2005).

Participants were recruited from Prolific, where they met initial screening criteria based on age and residency. Upon agreeing to participate in the study, they were provided a link to the Qualtrics survey where they were asked to read the Letter of Information and provide informed consent. After providing their Prolific ID, participants advanced to the property listing task (see Appendix C for the task instructions) where they were asked to generate a maximum of 10 features for each of the 30 concepts they saw. The 30 concepts were randomly selected from the set of 57 and presented one at a time (see Figure 5 for an example of the property listing task for one concept). Upon completion of the demographics questions, participants were redirected to Prolific to submit their completion code, which triggered their compensation. All procedures were approved by the Research Ethics Board at the University of Western Ontario (see

Appendix E). 105 participants generated features for 30 concepts, and 17 participants generated features for between 10 and 29 concepts, possibly due to the task timing out or because a concept was skipped. On average, features were obtained from *M* = 62.75 (range: 59-69) participants per concept (see Figure 6). The average time to complete the property listing task and demographics questionnaire was 00:56:18.

Figure 5

Example of the Property Listing Task and Responses for American City

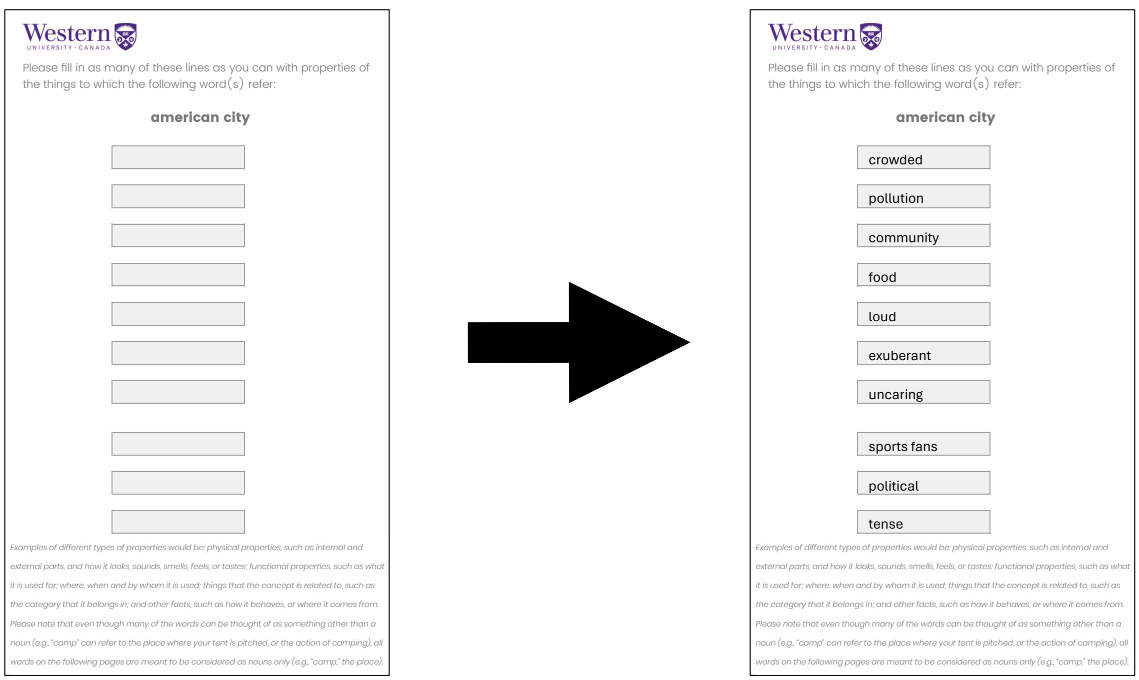
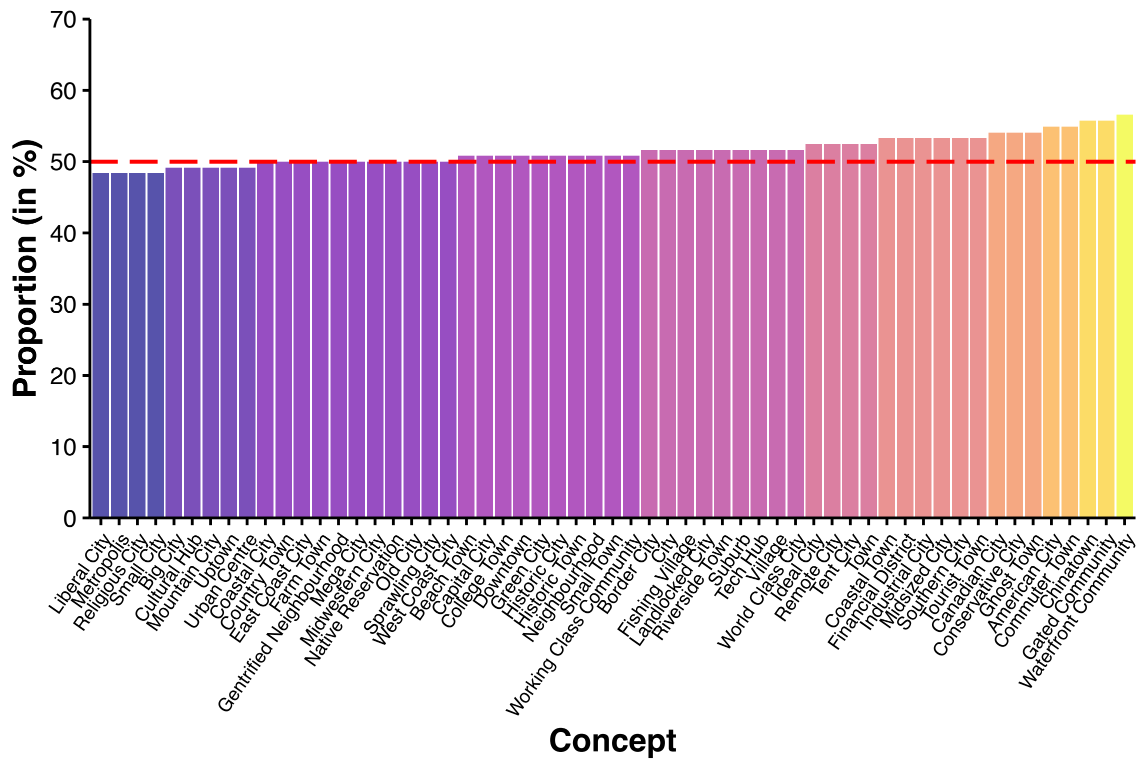


Figure 6

Proportion of Participants per Concept



Note: The red dashed line at the 50% mark provides a reference point for assessing which concepts had features generated by at least half of the participants. On average, each concept received features from approximately M = 62.75 (range = 59 - 69) participants (SD = 2.44).

Chapter 3

3 Analysis and Results

3.1 Data Cleaning and Preprocessing

Raw features were manually coded by a research team of four undergraduate students at the University of Western Ontario. Feature coding occurred in two phases.

In the first phase, the research team (a) determined the validity of each feature, (b) coded features synonymously when the meaning was the same, (c) if applicable, noted the type of change made to the original feature, and (d) identified rows containing multiple features. Valid responses provided semantic content that was not a cue repetition (e.g., listing <a city in the united states> for *American city*), property repetition (e.g., listing <may have homeless people> and <homelessness> for the same concept), or either a metacognitive or off-task comment (e.g., listing <this survey has lost my attention here because it is too long> for a concept). In the second phase, the 1,117 multi-feature rows identified in the first phase were split and standardized where necessary, and features with similar meanings that were missed in phase 1 were standardized. Features were processed through Canadian spell-check and formatted to lowercase. All analyses were performed using R Statistical Software (v4.2.1; R Core Team, 2022).

3.2 Production Frequencies

A custom R function counted the feature frequencies for each concept and stored them as a list. The individual lists were then merged into a single data frame using packages from the R Tidyverse. The average number of features generated for each concept was $M_{NOF\ Total}$ = 1236.40 ($M_{NOF\ Unique}$ = 306.26; see Appendix F for a complete summary of statistics for the concepts). Appendix G shows the top 10 feature production frequencies for the 57 concepts.

3.3 Cluster Analyses

Agglomerative hierarchical clustering (Maechler et al., 2023) was selected to explore the relational structure of the settlement concepts. The algorithm begins by treating each exemplar object as its own individual cluster. Next, comparisons are made between sets of two exemplars. If the exemplars are similar, they are merged. The pairs of clusters are then successively merged with each other to eventually form one large cluster. The data processing steps required for the hierarchical clustering algorithm are discussed below.

3.3.1 Full Dataset

The production frequencies data frame was reshaped and formatted as a matrix of feature production frequencies, with each row corresponding to one of 57 concepts and each column to one of 8474 features. Cells corresponding to features that were not mentioned for a concept contained a zero value.

A cosine similarity matrix was derived from the production frequency matrix by transposing the data and then obtaining the cosine similarity using the Latent Semantic Analysis package (lsa; Wild, 2022). The data were transposed such that each column represented a concept vector with rows representing the features. Taking the cosine similarity between the vectors produced a concept (57) x concept (57) matrix where each cell contained the cosine similarity for each pairwise comparison. See Appendix J for a schematic of how the cosine of the angle between two concept vectors was derived.

The cosine similarity between any two concepts was calculated using a function that divided the dot product of their vectors by the product of their magnitudes. Each of the 8474 concept vector components corresponded to the production frequency of a feature. Those components were multiplied, and the resulting products were summed to obtain the dot product.

The magnitude for each concept vector was calculated by taking the square root of the sum squared components of each concept vector and then multiplying them. Magnitude was used to normalize the dot product such that the cosine of the angle between the two vectors could be measured independent of their length. For example, American City had 1230 total features; and Canadian City had 1240. Normalizing vector length allowed meaningful comparisons between the concepts that were not influenced by the overall length of the total features, i.e., concepts could be compared regardless of the number of features generated. If represented visually, this process would involve plotting the vector for each concept in an 8474-dimensional feature space and taking the cosine of the angle (θ) between two concept vectors, giving a cosine similarity value between zero and one. For simplicity, the depiction in Appendix J uses a 3-dimensional feature space.

A cosine similarity value of one (i.e., $\cos(0°)$) indicates concepts are maximally similar and within the same position in the multidimensional space. Conversely, a value of zero (i.e., $\cos(90°)$) indicates concepts are orthogonal and do not share any features. Since the raw frequency values used to compute the matrices are not negative, the cosine values for the concept x concept matrix range from zero to one.

The agglomerative hierarchical clustering algorithm required a distance matrix to construct the dendrogram. Therefore, the matrix of cosine similarity values was converted into a dissimilarity matrix. The agnes function from the R cluster package (Maechler et al., 2023) was then applied to the dissimilarity matrix to construct the dendrogram from the bottom up. Each concept was initially treated as its own cluster (i.e., the dendrogram began with 57 clusters). Pairs of clusters with the lowest dissimilarity score were then identified and merged to form a new cluster. The dissimilarity matrix was updated to reflect this merging such that the dissimilarity between the new cluster and the other clusters was the average dissimilarity of the original clusters to the other clusters. Merging the closest clusters and updating the dissimilarity matrix continued iteratively until all items were merged into a single cluster at the top of the dendrogram.

3.3.1.1 Prototypes.

Prototypes were identified for each cluster to determine the central featural tendencies of each group. To extract the prototypical features, concepts within a cluster were identified along with the corresponding features. The mean of each feature (see Table 1) was calculated by taking the production frequency of that feature for the cluster and then dividing it by the total number of features within the cluster. The means were sorted in descending order to identify the features most common to a given cluster. A higher mean frequency value indicated a feature was more typical.

At the highest level two main groupings emerged (Clusters A and B in Figure 7). Each of these main clusters was further subdivided into two subclusters. Table 1 contains the featural prototypes extracted for each cluster.

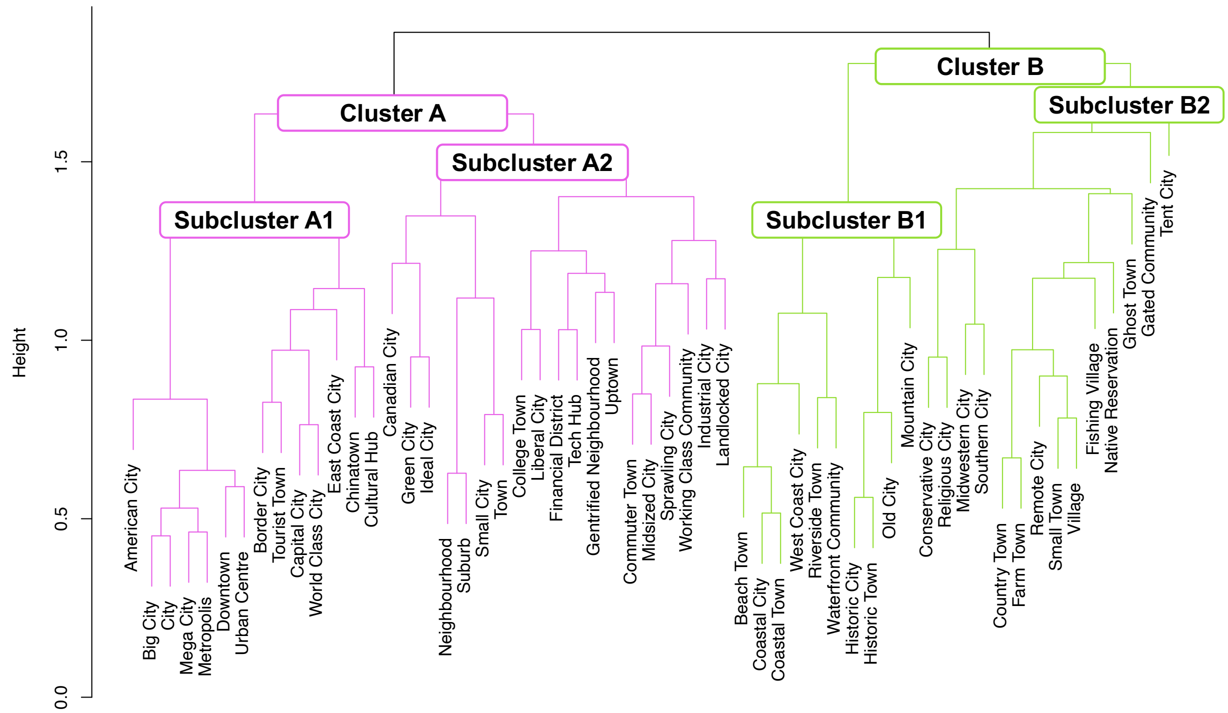
Cluster A reflected the general features of concepts that one might expect of a large city, such as <busy>, <public transportation>, <tourism>, and <expensive>. To understand which subclusters were most highly associated with the features, prototypes were also obtained for the subclusters. Subcluster A1 appeared to represent large and developed settlement concepts, largely accounting for features such as <busy>, <tourism>, <public transportation>, and <expensive>. Subcluster A2 appeared to reflect a more goal-oriented grouping and was characterized by features such as <public transportation>, <busy>, <cars>, <parks>, and <quiet>. The concept closest to the featural prototype for Cluster A was *City* (M = 0.408).

The second main grouping, Cluster B, generally emphasized <tourism> in <small>, <quiet>, <expensive>, <nature> areas with a <low population>. Subcluster B1 largely accounted for the <tourism> feature, characterizing concepts with <beaches> near the <ocean> that tend to be <expensive>. Subcluster B2 possessed a different prototype structure, with more emphasis on <small>, <quiet>, <rural>, <traditional> concepts. The concept closest to the featural prototype for Cluster B was *Historic Town* (M = 0.294).

The overall clustering of the concepts and the features most typical of each cluster provided a broad overview of these data. What these results alone cannot reveal is how such concepts might differ as a function of region. To explore how regional variation influenced the structure of settlement concepts, the next set of analyses segmented the feature data by participants’ geographic location.

Figure 7

Agglomerative Hierarchical Clustering Using Cosine Similarity (Overall Sample)



Note: The agglomerative coefficient (*AC* = 0.56) is a measure of overall clustering.

Table 1

Top 10 Prototypical Features for Clusters A and B and Subclusters in the Overall Sample

Cluster A		**Subcluster A1**		**Subcluster A2**	
Feature	M_F	**Feature**	M_F	**Feature**	M_F
busy	9.79	busy	15.29	public transportation	6.53
public transportation	8.09	tourism	10.50	busy	5.74
tourism	5.52	public transportation	10.21	cars	5.11
expensive	4.64	expensive	6.14	parks	4.37
cars	4.33	diverse	6.07	quiet	4.05
diverse	4.18	restaurants	6.00	green space	3.95
restaurants	3.82	crowded	5.14	access to education	3.84
parks	3.30	skyscrapers	4.71	expensive	3.53
people	3.18	loud	4.64	walkable	3.47
moderate traffic	3.15	moderate traffic	4.50	clean	3.37
Cluster B		**Subcluster B1**		**Subcluster B2**	
Feature	M_F	**Feature**	M_F	**Feature**	M_F
tourism	10.58	tourism	21.10	small	7.29
small	5.75	beaches	7.40	quiet	6.93
quiet	5.38	ocean	6.80	rural	5.50
expensive	4.29	expensive	6.50	traditional	4.93
nature	4.21	boats	6.10	unpopulated	4.93
unpopulated	3.62	water	5.90	poverty	4.79
rural	3.54	nature	4.20	conservative	4.36
traditional	3.46	fishing	3.90	nature	4.21
beaches	3.21	old	3.80	friendly	3.93
boats	3.17	small	3.60	community	3.43

Note. *N* = 122 participants were in the overall sample. Mean values (M_F) represent the average frequency of the feature relative to the number of concepts in the cluster. See Figure 7 for the accompanying dendrogram.

3.3.2 Subset Data: Canadian and American Participants

The subsequent analysis examined whether regional context influences the features people generate for settlements. To address any potential confounds related to participants' residential mobility, the analysis considered the duration of residence within specific regions instead of simply dividing the total dataset by region. To filter the dataset, a demographic question asked participants to state whether they had lived in a particular region throughout their entire lives, selecting only participants residing in Canada or the United States. This approach yielded a subset of 44 individuals, evenly split with 22 participants in each region. The preprocessing steps for the subset data, including the computation of their respective matrices, remained consistent with the general analysis framework. For each regional group (Canada and the United States), the prototypical features of the resulting clusters and subclusters are presented and compared.

3.3.2.1 Canadian Prototypes.

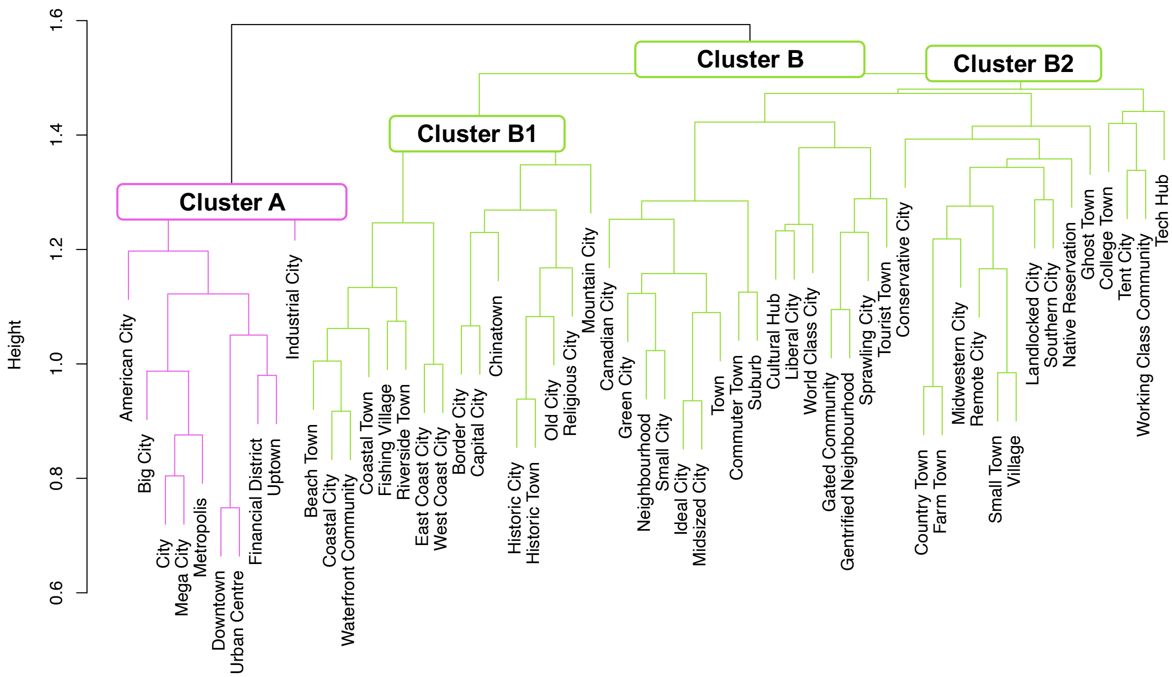
At the highest level, the dendrogram split into two clusters labelled A and B (Figure 8; Table 2). Similar to the arrangement in the Overall sample, Cluster A consisted of settlement concepts that are <busy>, have <public transportation>, <moderate traffic>, and are <expensive>. However, unlike the Overall sample, Cluster A formed one main cluster. The concept closest to the featural prototype for Cluster A was *Metropolis* (*M* = 0.373).

Concepts formed distinct subclusters in Cluster B. At the highest level, Cluster B emphasized <tourism> in <friendly>, <small>, <expensive>, <quiet> areas with <green space> and <nature>. Subcluster B1 appeared to largely account for the <tourism>

feature, characterizing concepts that are close to the <water>, with access to <beaches> and <boats>, and tend to be <expensive>. Subcluster B2 possessed a different prototype structure, with more emphasis on <friendly>, <public transportation>, <green space>, <quiet> features. The concept closest to the featural prototype for Cluster B was *Village* (*M* = 0.162).

Figure 8

Agglomerative Hierarchical Clustering Using Cosine Similarity (Canada; n = 22)



Note: The agglomerative coefficient (AC_{CA} = 0.31) is a measure of overall clustering.

Table 2

Top 10 Prototypical Features for Clusters in the Canada Sample

Cluster A		Cluster B		Cluster B1		Cluster B2	
Feature	M_F	**Feature**	M_F	**Feature**	M_F	**Feature**	M_F
busy	3.30	tourism	1.40	tourism	3.44	friendly	1.13
public transportation	2.40	friendly	0.96	water	1.44	public transportation	1.03
moderate traffic	1.70	small	0.94	beaches	1.12	green space	1.00
expensive	1.60	expensive	0.89	boats	1.00	quiet	1.00
businesses	1.50	quiet	0.87	expensive	1.00	small	0.90
people	1.50	green space	0.79	small	1.00	community	0.84
pollution	1.40	nature	0.72	busy	0.69	expensive	0.84
noisy	1.30	public transportation	0.72	fishing	0.69	nature	0.81
restaurants	1.30	community	0.66	ocean	0.69	walkable	0.71
cars	1.20	walkable	0.60	sunny	0.69	cars	0.68

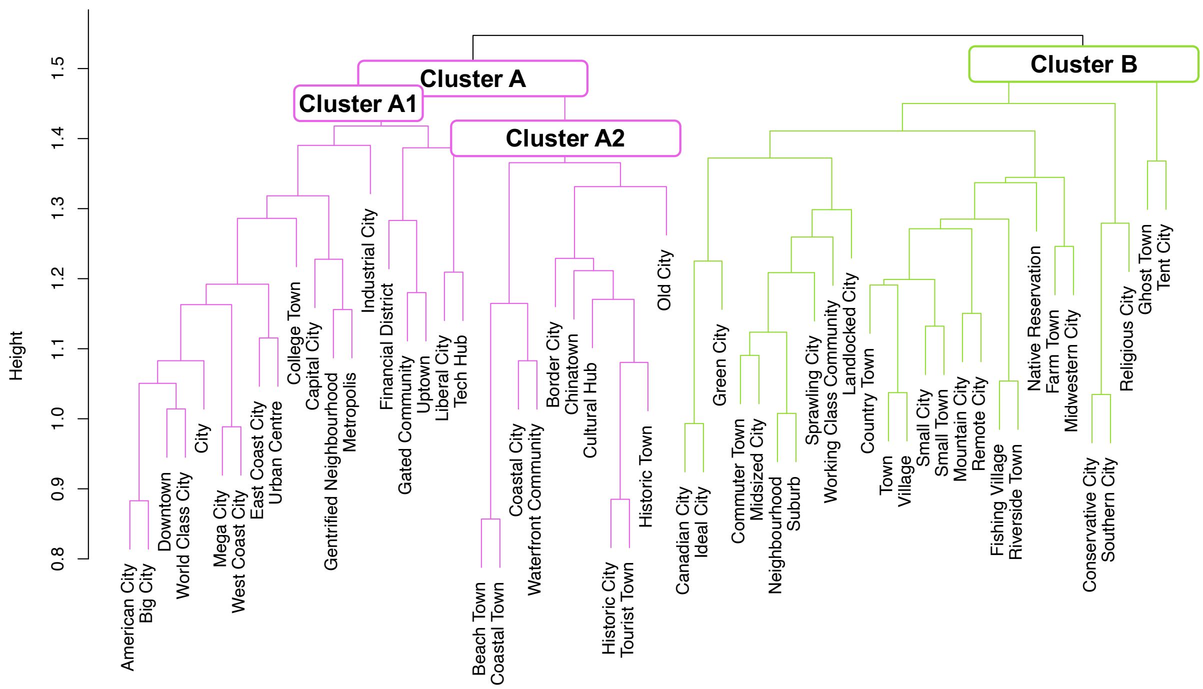
Note. n = 22 participants were in the Canada sample. Mean values (M_F) represent the average frequency of the feature relative to the number of concepts in the cluster. See Figure 8 for the accompanying dendrogram.

3.3.2.1 American Prototypes.

At the highest level, the dendrogram split into two clusters labelled A and B (Figure 9; Table 3). Cluster A consisted of settlement concepts with <tourism>, <expensive>, <busy>, and <crowded> features. Cluster A formed two main Subclusters. Subcluster A1 largely accounted for <expensive>, while also having <busy>, <public transportation>, and <skyscrapers> features. Subcluster A2 accounted for <tourism>, while also being characterized by <beaches>, and tending to be <expensive>. The concept closest to the featural prototype for Cluster A was *Mega City* (M = 0.187). Cluster B emphasized <cars> in <quiet>, <small>, <traditional>, <rural> areas with <green space> and <nature>. The concept closest to the featural prototype for Cluster B was *Town* (M = 0.154).

Figure 9

Agglomerative Hierarchical Clustering Using Cosine Similarity (United States; n = 22)



Note: The agglomerative coefficient (AC_{US} = 0.27) is a measure of overall clustering.

Table 3

Top 10 Prototypical Features for Clusters in the United States Sample

Cluster A		Cluster A1		Cluster A2		Cluster B	
Feature	M_F	**Feature**	M_F	**Feature**	M_F	**Feature**	M_F
tourism	1.70	expensive	1.89	tourism	3.64	cars	1.15
expensive	1.60	busy	1.74	beaches	1.09	quiet	1.07
busy	1.37	public transportation	1.21	expensive	1.09	small	0.89
crowded	0.90	skyscrapers	1.16	old	1.00	traditional	0.81
public transportation	0.87	crowded	1.05	ocean	0.91	rural	0.78
skyscrapers	0.73	dirty	0.79	arts scene	0.82	green space	0.67
diverse	0.70	loud	0.68	museums	0.82	nature	0.67
dirty	0.60	buildings	0.68	busy	0.73	community	0.59
beaches	0.57	diverse	0.68	culture	0.73	unpopulated	0.59
restaurants	0.57	homelessness	0.68	good food	0.73	access to education	0.56

Note. $n = 22$ participants were in the United States sample. Mean values (M_F) represent the average frequency of the feature relative to the number of concepts in the cluster. See Figure 9 for the accompanying dendrogram.

3.3.2.2 Comparing Canadian and American Representations.

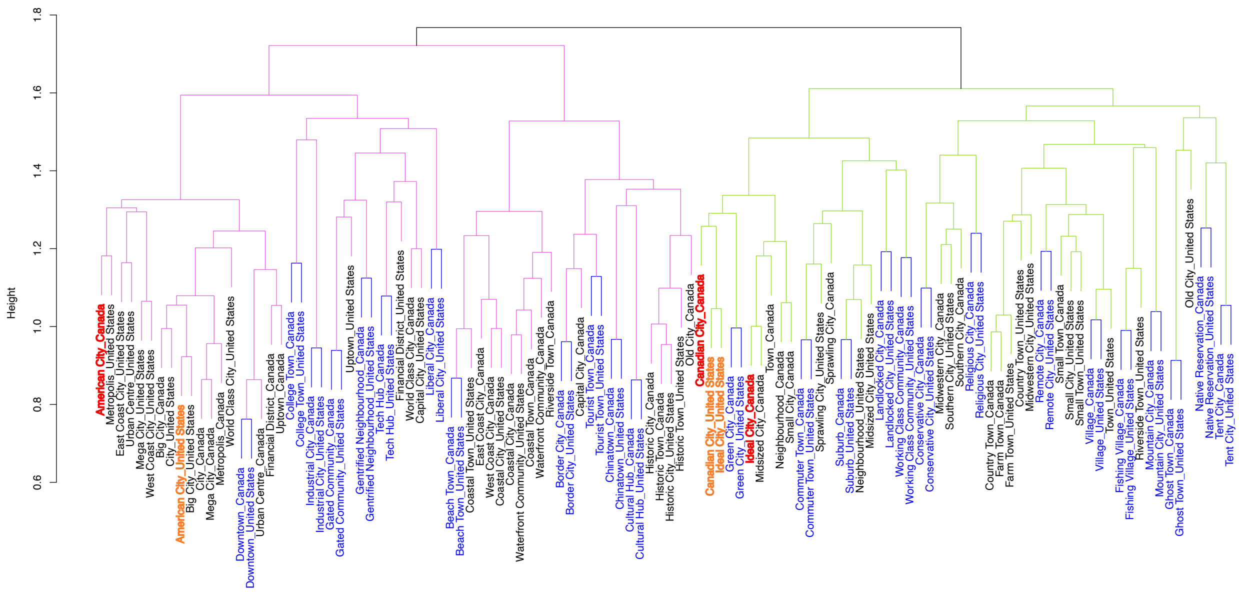
The cluster analyses for the Canada and United States samples revealed similarities and differences in where the concepts grouped together. Overall, the two groupings formed at the highest level across datasets showed a cluster for large developed city-type concepts and small rural town-like concepts. However, the groupings of concepts within the two largest clusters differed depending on region, as did the concepts most typical of the featural prototypes extracted for each cluster. In the overall sample, *City* was most typical of Cluster A, *Metropolis* was most typical in the Canadian subset, and *Mega City* was most typical in the American subset. For Cluster B, *Historic Town* was most typical for the overall sample, *Village* was most typical in the Canadian subset, and *Town* was most typical in the American subset.

The next set of analyses plotted the features split by region together (see Figure 10) to visualize the similarity structure (note: prototypes were not extracted for clusters in Figure 10). Bifolious concepts (i.e., two-leaved clades that emerged from a single node, denoted in blue) were maximally similar to each other relative to the other concepts within their cluster and associated subclusters and accounted for 45.61% of the concept groupings. The remaining 54.39% were within different subclusters (43.86%), trifolious (i.e., three-leaved clades that emerged from a single node; 7.02%), or entirely different clusters (either A or B; 3.51%), reflecting a graded similarity structure between the two groups.

Although the preceding sections detail the overall structure of settlement concepts and representational differences between regional groups, they do not explicitly probe the impact of lived experience on the features generated for these concepts. Nor do they examine how ideals might diverge among participants from different regions where participants had lived experience or how their ideals might differ. The final set of analyses aimed to address these questions by comparing representations of *Canadian City*, *American City*, and *Ideal City* across the entire participant pool, as well as between the Canadian and American subsets.

Figure 10

Agglomerative Hierarchical Clustering Using Cosine Similarity (Combined Canadian and United States Subsets; n = 44)



Note: The agglomerative coefficient ($AC_{CA\&US}$ = 0.39) is a measure of overall clustering. The two main clusters are highlighted in pink and green. Bifoliate concepts are denoted in blue. *Canadian City*, *American City* and *Ideal City* are distinguished with bold typeface, and Canada and United States subset concepts are presented in red and orange, respectively.

3.4 Cross-Concept Feature Comparisons

The structure of the concepts differed in relation to each other when the data were parsed by region. Prototypes extracted from the cluster analyses showed the arrangements differed due to the features generated by participants in the property listing task. While some concepts maintained consistent arrangements across the subsets, parsing the data by region showed that only 45.61% of the concepts formed maximally similar groupings (indicated in blue text in Figure 10). The remaining 54.39% fell within a different arrangement that varied by region. Among those concepts were *Canadian City*, *American City*, and *Ideal City* (indicated in bold typeface in Figure 10 with red text for Canada and orange for United States). Features for the three concepts of interest were compared across the overall dataset and the subset data. Table 4, Table 5, and Table 6 present the top 10 features and their frequencies expressed as a proportion relative to the sample size in each group. See Appendix H and Appendix I for the top 10 feature production frequencies of the concepts for each region.

3.4.1 Canadian City

One feature was shared across the Overall sample and the Canada and United States subsets top 10 features of *Canadian City*, <cold>. The relative proportion of <cold> differed depending on region, with the United States sample being comparable to the Overall sample, and with Canada listing this feature less frequently. Features that only appeared within the top 10 for the Overall sample were <busy>, <hockey>, <cold during winter>, and <tourism>. The remaining top 10 features were shared with one of the Subset samples, or unique to each Subset.

Shared features among the top 10 listed for participants in the Overall sample and Canada subset for *Canadian City* were <diverse>, <multicultural>, and <polite>. Proportions in the Canada subset were higher for <diverse> by 5.44%, <multicultural> by 6.26%, and <polite> by 3.35%. The top 10 features unique to the Canada subset were <trees>, <nature>, <rural>, <safe>, <activities>, and <bad smells>.

The top 10 features shared between the Overall sample and the United States subset for *Canadian City* were <friendly> and <french speaking>. The proportion for <friendly> was larger within the Overall sample with a difference of 1.57%, and the proportion for <french speaking> was larger in the United States subset with a difference of 1.71%. The top 10 features unique to the United States subset were <clean>, <clean air>, <healthcare>, <access to education>, <active>, <basement houses>, and <beautiful>.

Table 4

Top 10 Features Listed for Canadian City

Overall		**Subset$_{CA}$**		**Subset$_{US}$**	
Feature	**%**	**Feature**	**%**	**Feature**	**%**
cold	13.93	diverse	13.64	cold	13.64
friendly	10.66	multicultural	13.64	clean	9.09
diverse	8.20	trees	13.64	clean air	9.09
busy	7.38	cold	9.09	french speaking	9.09
french speaking	7.38	nature	9.09	friendly	9.09
multicultural	7.38	polite	9.09	healthcare	9.09
hockey	6.56	rural	9.09	access to education	4.55
cold during winter	5.74	safe	9.09	active	4.55
polite	5.74	activities	4.55	basement houses	4.55
tourism	5.74	bad smells	4.55	beautiful	4.55

Note. N = 122 participants were in the Overall sample, while n = 22 were in each of the Canada and United States subsets. Frequencies of the top 10 features are expressed as a proportion relative to the sample size in each group.

3.4.2 American City

One feature was shared across the Overall sample and the Canada and United States subsets top 10 features of *American City*, <diverse>. The relative proportion of <diverse> differed depending on region, with the United States sample being comparable to the Overall sample, and with Canada listing <diverse> more frequently. Features that only appeared within the top 10 for the Overall sample were <busy>, <cars>, <dangerous>, and <patriotic>. The remaining top 10 features were shared with one of the Subset samples, or unique to each Subset.

Shared features among the top 10 listed for participants in the Overall sample and Canada subset for *American City* were <guns>, <loud>, and <homelessness>. Proportions in the Canada subset were higher for <guns> by 8.34%, <loud> by 0.07%, and <homelessness> by 3.35%. The top 10 features unique to the Canada subset were <business>, <festive>, <heavy traffic>, <proud>, <skyscrapers>, and <suburbs>.

The top 10 features shared between the Overall sample and the United States subset for *American City* were <fast food> and <moderate traffic>. The proportion for <fast food> was larger within the United States subset with a difference of 3.35%, and the proportion for <moderate traffic> was larger in the United States subset with a difference of 8.72%. The top 10 features unique to the United States subset were <crowded>, <apartment buildings>, <pollution>, <public transportation>, <skyscrapers>, <smog>, and <access to education>.

Table 5

Top 10 Features Listed for American City

Overall		**Subset$_{CA}$**		**Subset$_{US}$**	
Feature	**%**	**Feature**	**%**	**Feature**	**%**
busy	9.84	guns	18.18	crowded	13.64
guns	9.84	diverse	13.64	moderate traffic	13.64
loud	9.02	business	9.09	apartment buildings	9.09
diverse	8.20	festive	9.09	diverse	9.09
cars	6.56	heavy traffic	9.09	fast food	9.09
dangerous	6.56	homelessness	9.09	pollution	9.09
fast food	5.74	loud	9.09	public transportation	9.09
homelessness	5.74	proud	9.09	skyscrapers	9.09
moderate traffic	4.92	skyscrapers	9.09	smog	9.09
patriotic	4.92	suburbs	9.09	access to education	4.55

Note. N = 122 participants were in the Overall sample, while n = 22 were in each of the Canada and United States subsets. Frequencies of the top 10 features are expressed as a proportion relative to the sample size in each group.

3.4.3 Ideal City

Two features listed for *Ideal City*, <green space> and <healthcare>, were common across all three groups. However, their relative proportions differed. Within the Overall sample and the Canada subset, <green space> was one of the most frequently occurring features. This feature was listed less frequently for the United States sample. On the other hand, <healthcare> was listed less frequently for the Overall and Canada sample relative to the United States. Features that only appeared within the Overall sample were <safe> and <sustainable>. The remaining top 10 features were either shared with the Overall sample and one of the Subset samples, or unique to each Subset.

Among the top 10 features shared between the Canada subset and the Overall sample were <public transportation>, <walkable>, <parks>, and <friendly>. Relative to the Overall sample, the proportions were higher in the Canada sample for <public transportation> by 5.96%, <walkable> by 1.34%, <parks> by 4.62%, and <friendly> by 10.80%. The top 10 features unique to the Canada subset were <fun>, <accepting>, <arts scene>, and <cheap>.

The overlapping features shared between the United States subset and the Overall sample were <clean> and <access to education>. Relative to the Overall sample, the proportions were higher in the United States sample for <clean> by 1.34% and <access to education> by 9.16%. The top features unique to the United States subset were <mountains>, <affordable housing>, <beautiful>, <inexpensive>, <low crime rate>, and <nice>.

Table 6

Top 10 Features Listed for Ideal City

Overall		**Subset$_{CA}$**		**Subset$_{US}$**	
Feature	**%**	**Feature**	**%**	**Feature**	**%**
public transportation	21.31	green space	27.27	access to education	18.18
green space	13.93	public transportation	27.27	clean	13.64
clean	12.30	friendly	18.18	healthcare	13.64
safe	12.30	fun	13.64	mountains	13.64
walkable	12.30	parks	13.64	affordable housing	9.09
access to education	9.02	walkable	13.64	beautiful	9.09
parks	9.02	accepting	9.09	green space	9.09
friendly	7.38	arts scene	9.09	inexpensive	9.09
healthcare	7.38	cheap	9.09	low crime rate	9.09
sustainable	7.38	healthcare	9.09	nice	9.09

Note. N = 122 participants were in the Overall sample, while n = 22 were in each of the Canada and United States subsets. Frequencies of the top 10 features are expressed as a proportion relative to the sample size in each group.

Chapter 4

4 General Discussion

Although the role of settlement context is a constant and fundamental aspect of human experience that situates people within the environment, many questions remain about the nature of such representations. The present study had several goals, largely aimed at supporting future efforts related to this project. The first goal was to empirically derive semantic feature norms for a set of settlement concepts to understand broadly how people think about the different types of urban environments humans encounter. As expected, concepts tended to cluster together based on the types of features shared and their relative production frequency. Moreover, the concepts within a given cluster showed a graded relationship to the overall family resemblance prototype extracted for each cluster, which is evidence of a prototype effect. The second goal was to understand how individual context and priors affect the information people choose to attend to, and what people deem to be ideal or most important. Participants who self-reported living in Canada or the United States for their entire lives showed differences in their conceptual structure of settlement such that the clustered concepts formed different configurations, and the prototypes derived for each cluster differed depending on region. This finding is consistent with prototype effects found in cross-linguistic studies of conceptual structure (Pavlenko & Malt, 2011), but does not explain why differences emerged within the same language for my study. Other potential drivers of these differences in prototypes are likely accounted for by region and culture.

4.1 Regional Differences

The differences in conceptual structure of settlement found by comparing feature norms for long-term residents of Canada with those residing in the United States can be explained by synthesizing Gibson's (1979) theory of *affordances*, the *embodied simulation hypothesis* (Gallese & Lakoff, 2005), and *dynamical systems theory* (Spivey & Dale, 2004). Affordances are opportunities for action that objects, environments, or situations provide to an individual (Gibson, 1979). Within this frame, the same environmental feature can offer various affordances to different people, and even to the same person at different times. Embodied cognition supports the theory of affordances because cognitive processes are deeply rooted in the body's interactions with the world. Different usage contexts seemed to evoke different embodied simulations during the property listing task. The results of the cluster analysis, therefore, highlight the features that imply the kinds of interactions and behaviours that participants believe such environments afford.

Prototypical features overlapped across participant groups, though mean frequencies of the features differed. At the highest level of the cluster analyses, prototypical features for the Overall sample in Cluster A, such as <busy>, <public transportation>, and <expensive>, appeared within Cluster A across participant groups. The consistent presence of this family resemblance suggests the overall groupings at the highest level possessed some shared meaning across participants, regardless of region. The overall family resemblance at the level of Cluster B accounted for slightly more variability when participants were split by region, and is at least partially attributed to differences in ideals and experiences. This finding suggests elements of a common conceptual structure among participants about what constitutes the essence of Clusters A and B, possibly reflecting a shared cultural or cognitive schema about certain urban characteristics.

Despite sharing a family resemblance at the level of Cluster A and B across participant groups, the mean feature frequencies differed. This was due to a few factors. For one, the sample sizes in the subset data were smaller, resulting in a smaller chance of consensus across listed features. Further, the structure of the clusters differed in the number of concepts that fell within a particular grouping, and the concept most similar to the featural prototypes for each cluster also differed between groups. Mean frequency differences across groups also hint at regional variations in how these features were prioritized or perceived, likely influenced by local environments and the specific dynamics of urban life in their regions. For example, <expensive> might be a more salient feature in cities with higher costs of living. Similarities and differences between regions were expected based on what has been found for landscape concepts (van Putten et al., 2020), the general observation made by UN-Habitat (UN-Habitat, 2018), and the literature on context-dependent conceptual activation (Colston & Rasse, 2021; Yee & Thompson-Schill, 2016). While there is evidence of a clear family resemblance across participant groups at the highest level of the cluster analyses, local contexts and differences in ideals may partially account for the variability observed in feature emphasis and cluster structure.

The specific structure of concepts hypothesized to most likely be influenced by experience and goals were *Canadian City*, *American City*, and *Ideal City*. Findings for each of these concepts showed a smaller family resemblance of features across participant groups, and a graded feature typicality structure. Participants seemed to agree less depending on whether the concept was congruent with their region. For example, Canadian participants' concept of *Canadian City* shared one top 10 feature with American participants, <cold>, and American participants' concept of *American City* shared one top 10 feature with American participants, <diverse>. *Ideal City* shared two top 10 features across the groups, <green space> and <healthcare>. Although the features were shared across groups, the relative proportions of feature frequency differed, which like the above analyses implies regional variations in how features were prioritized or perceived by participants. For example, the proportion for <healthcare> in the American participants subset was higher than the Overall and Canadian subset. Such differences in relative proportions likely reflect the features most associated with a concept, and in the case of *Ideal City*, what participants deem to be important.

Relative proportions of features could also reflect regional norms, stereotypes and biases, and the impact of current events. For example, American participants might generalize healthcare access as a feature of most Canadian Cities. One reason healthcare might not be within the top 10 features for Canadians is because it has become such a consistent part of the Canadian concept that Canadians do not feel the need to mention it. However, Canadians, especially those living in rural regions, report having lower quality or limited access to healthcare (Wilson et al., 2020), demonstrating that this feature does not necessarily generalize in every case. Similarly, Canadian participants might generalize gun culture as being ubiquitous across American cities when this is not the experience for Americans. Canadians' perception of American gun culture may be less influenced by personal experience and more by media and public discourse (Hill, 2023). Furthermore, the current temporal context in relation to recent events might influence the salience of features associated with a region (Hill, 2023). Temporal context is a facet of the dynamic nature of settlement concepts that must therefore be considered to accurately capture their representation – even if this means the behavioural feature norms collected here are a mere snapshot in time.

4.2 Limitations and Future Directions

The main limitations of this work relate to sampling. The participants, all from Canada or the United States and communicating in English, largely represent a WEIRD (Western, Educated, Industrialized, Rich, and Democratic) demographic. Heinrich et al. (2010) highlight the inherent issues with such a sample, suggesting that findings may not extend beyond this demographic – let alone the English-speaking community as a whole. Participants were also recruited from Prolific, which excludes individuals without internet or computer access. Further, 63.93% of participants self-reported having a university-level education. Education and the associated life experiences that come with this privilege are closely linked to socioeconomic status (SES; Gelbgiser, 2021), an additional variable that was not collected for participants. Consequently, sampling methods likely affected generalizability of the behavioural norms to the spectrum of educational attainment levels and SES and the experiences afforded by these contexts and are a limitation of the present study.

Findings from Colson and Rasse (2021) further suggest variances in the semantic understanding of English across different geographic locations could affect the generalization of the results. One difference I might observe if I were to compare non-English representations from non-WEIRD populations is the number of features referring to context versus objects, based on findings from Miyamoto et al. (2006). Future studies should code features based on whether they are elements of overall context or objects to better understand cross-cultural and regional similarities and differences in global versus local processing of settlements.

There was also potential for interference across concepts. All 57 concepts were from the same superordinate category of settlement. Steps were not taken to avoid participants listing features for semantically similar concepts within the settlement category. This likely led to explicit comparisons and possibly biased the salience of the features listed through proactive and retroactive interference, making it challenging to say whether the first-listed features truly represent the most salient aspect of the concept. However, only one concept was presented at a time on each page.

Additionally, participant task completion varied. 105 participants generated features for 30 concepts, while the remaining 17 varied between 10 and 29 concepts. McRae et al. (2005) had 30 participants list features for each concept. As a result, the number of participants who generated features for each settlement concept varied. Ideally, having each concept receive features from the same number of participants would ensure comparability across concepts. However, the varying number of participants likely caused some concepts to be underrepresented, leading to skewed production frequencies. The uneven number of participants who completed the task could affect the reliability of the results, as the variability attributed to the number of features listed could be confounded with differences between participants. Variability in task completion might reflect different levels of engagement or understanding of the task requirement. Future studies could include a question at the end of the survey that obtains participant feedback to better understand such effects. The property listing task necessitates collecting a substantial amount of data to capture a comprehensive array of features for each concept. Such variability within and between participants therefore makes it important that each concept is seen an equal number of times.

The study's reliance on linguistic-based feature norms may also limit the scope of the findings, as some information about settlement concepts may not be clearly conveyed through language or easily verbalized. Future studies could employ Virtual Reality paradigms to observe behaviours in a settlement context directly. Task designs that simulate experiences within a given settlement context could capture elements that are not easily verbalized and yet still part of a person's concept.

One element that was not considered in the survey design was participant strategy. It was not immediately clear from the responses whether participants thought of specific instances of the concepts (unless they listed a specific instance as a feature) or if they generated features based on an abstracted prototype. Including such a question would have allowed distinguishing between exemplar-based and prototype-based processing in feature generation. As a future direction, it would be beneficial to explicitly ask participants post-task about their approach. In addition to complementing the prototypes extracted from the clusters, knowing about participant strategy would enhance understanding of the cognitive mechanisms underlying feature listing.

It is also possible that the task design might have biased participant responses toward features that distinguish concepts. Given that the concepts were all part of the same superordinate category, participants might have assumed that listing distinguishing features was what the task wanted of them. It could have also been the case that contrastive features just happened to be the most salient because they are distinguishing.

The feature coding process might have also introduced limitations. The data received a single overall pass and, therefore, likely missed other semantically similar features. This could have resulted in smaller production frequencies in the subset data. Additionally, though care was taken to eliminate potential coder biases, coding decisions implicitly influenced by bias could still impact the decisions made when coding the concepts that might have been different from one coder to the next.

While the study provides valuable information into how English-speaking participants from North America conceptualize settlement concepts, the findings for regional comparisons should be interpreted with caution due to the potential for constrained generalizability and potential biases introduced by the study design and participant pool. Nonetheless, the behavioural feature norms derived for the settlement concepts offer a unique dataset of semantic content within the context of the behavioural norming literature.

The set of behavioural feature norms derived from this work can be used to design experiments testing theories of cognition in relation to the built environment. Additionally, future studies can apply this general framework to obtain behavioural feature norms for settlement concepts in other regions of the world and across languages. Settlement norms from other regions and languages can then be compared with the current North American English representations of settlements to identify similarities and differences in how humans collectively think about the categories people use to describe the various places people inhabit on a global scale.

References

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *11*(4), 629–654. https://doi.org/10.1037/0278-7393.11.1-4.629

Berlin, B., & Kay, P. (1969). *Basic color terms: their universality and evolution*. University of California Press.

Brown, R. (1958). How shall a thing be called. *Psychological Review*, *65*(1), 14–21. https://doi.org/10.1037/h0041727

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Buchanan, E. M., Valentine, K. D., & Maxwell, N. P. (2019). English semantic feature production norms: An extended database of 4436 concepts. *Behavior Research Methods*, *51*(4), 1849–1863. https://doi.org/10.3758/s13428-019-01243-z

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240–247. https://doi.org/10.1016/S0022-5371(69)80069-1

Colston, H. L., & Rasse, C. (2021). Embodiment across Englishes: The comprehension of metaphors in popular song lyrics in Canadian, Austrian, and American English.

Metaphor in Language and Culture across World Englishes, 129–152. https://doi.org/10.5040/9781350157569.ch-007

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 1 billion+ words, 1990-present*. [Website]. https://www.english-corpora.org/coca/

Devereux, B. J., Tyler, L. K., Geertzen, J., & Randall, B. (2014). The Centre for Speech, Language and the Brain (CSLB) concept property norms. *Behavior Research Methods*, *46*(4), 1119–1127. https://doi.org/10.3758/s13428-013-0420-4

Dieciuc, M. A., & Folstein, J. R. (2019). Typicality: Stable structures and flexible functions. *Psychonomic Bulletin & Review*, *26*(2), 491–505. https://doi.org/10.3758/s13423-018-1546-2

Diveica, V., Pexman, P. M., & Binney, R. J. (2023). Quantifying social semantics: An inclusive definition of socialness and ratings for 8388 English words. *Behavior Research Methods*, *55*(2), 461–473. https://doi.org/10.3758/s13428-022-01810-x

Gallese, V., & Lakoff, G. (2005). The Brain’s concepts: the role of the Sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*(3), 455–479. https://doi.org/10.1080/02643290442000310

Garrard, P., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. (2001). Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology*, *18*(2), 125–174. https://doi.org/10.1080/02643290042000053

Gelbgiser, D. (2021). Socioeconomic Segregation, Campus Social Context, and Disparities in Bachelor’s Degree Attainment. *Demography*, *58*(3), 1039–1064. https://read.dukeupress.edu/demography/article-abstract/58/3/1039/173293

Gelman, S. A., & Davidson, N. S. (2013). Conceptual influences on category-based induction. *Cognitive Psychology*, *66*(3), 327–353. https://doi.org/10.1016/j.cogpsych.2013.02.001

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*(3), 183–209. https://doi.org/10.1016/0010-0277(86)90034-x

Gibson, J. J. (1979). The theory of affordances. In *The Ecological Approach to Visual Perception*.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*, 585–612. https://doi.org/10.1146/annurev.psych.49.1.585

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews. Neuroscience*, *15*(8), 536–548. https://doi.org/10.1038/nrn3747

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2–3), 61–83; discussion 83-135. https://doi.org/10.1017/S0140525X0999152X

Hill, M. (2023, August 16). *Canada's gun control debate has been Americanised and the consequences are real*. The Hub. https://thehub.ca/2023/08/16/canadas-gun-control-debate-has-been-americanised-and-the-consequences-are-real/

Hou, Y., Qu, Y., Zhao, Z., Shen, J., & Wen, Y. (2021). Residents' Spatial Image Perception of Urban Green Space through Cognitive Mapping: The Case of Beijing, China. *Forests, Trees and Livelihoods*, *12*(12), 1614. https://doi.org/10.3390/f12121614

Jang, K. M., & Kim, Y. (2019). Crowd-sourced cognitive mapping: A new way of displaying people's cognitive perception of urban space. *PloS One*, *14*(6), e0218590. https://doi.org/10.1371/journal.pone.0218590

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: making the connection. *Cognitive Psychology*, *16*(2), 243–275. https://doi.org/10.1016/0010-0285(84)90009-4

Kivisaari, S. L., Hultén, A., van Vliet, M., Lindh-Knuutila, T., & Salmelin, R. (2023). Semantic feature norms: a cross-method and cross-language comparison. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02311-1

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2023). *cluster: Cluster Analysis Basics and Extensions* (2.1.4) [Computer software]. Comprehensive R Archive Network (CRAN). https://CRAN.R-project.org/package=cluster

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: a cognitive neuropsychological perspective. *Annual Review of Psychology*, *60*(1), 27–51. https://doi.org/10.1146/annurev.psych.60.110707.163532

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, *16*(1), 1–27. https://doi.org/10.1016/0010-0285(84)90002-1

McCloskey, M., & Glucksberg, S. (1979). Decision processes in verifying category membership statements: Implications for models of semantic memory. *Cognitive Psychology*, *11*(1), 1–37. https://doi.org/10.1016/0010-0285(79)90002-1

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547–559. https://doi.org/10.3758/bf03192726

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. https://doi.org/10.1037/0033-295x.85.3.207

Mervis, C. B., & Crisafi, M. A. (1982). Order of acquisition of subordinate-, basic-, and superordinate-level categories. *Child Development*, *53*(1), 258–266. https://doi.org/10.1111/j.1467-8624.1982.tb01315.x

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *27*(3), 775–799. https://doi.org/10.1037/0278-7393.27.3.775

Minda, J. P., & Smith, J. D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *28*(2), 275–292. https://doi.org/10.1037//0278-7393.28.2.275

Minda, J. P., & Smith, J. D. (2011). Prototype models of categorization: basic formulation, predictions, and limitations. In Pothos, E M & Willis, A J (Ed.), *Formal Approaches in Categorization* (pp. 40–64). Cambridge University Press.

Miyamoto, Y., Nisbett, R. E., & Masuda, T. (2006). Culture and the physical environment. Holistic versus analytic perceptual affordances. *Psychological Science*, *17*(2), 113–119. https://doi.org/10.1111/j.1467-9280.2006.01673.x

Muraki, E. J., Abdalla, S., Brysbaert, M., & Pexman, P. M. (2023). Concreteness ratings for 62,000 English multiword expressions. *Behavior Research Methods*, *55*(5), 2522–2531. https://doi.org/10.3758/s13428-022-01912-6

Murphy, G. L., Hampton, J. A., & Milovanovic, G. S. (2012). Semantic memory redux: An experimental test of hierarchical category representation. *Journal of Memory and Language*, *67*(4), 521–539. https://doi.org/10.1016/j.jml.2012.07.005

Murphy, G. L., & Lassaline, M. (1997). Hierarchical structure in concepts and the basic level of categorization. *Knowledge Concepts and Categories*, 93–131. https://doi.org/10.4324/9780203765418-4/hierarchical-structure-concepts-basic-level-categorization-gregory-murphy-mary-lassaline

Murphy, G. L., & Wisniewski, E. J. (1989). Categorizing objects in isolation and in scenes: What a superordinate is good for. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *15*(4), 572–586. https://doi.org/10.1037/0278-7393.15.4.572

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, *36*(3), 402–407. https://doi.org/10.3758/bf03195588

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology. General*, *115*(1), 39–61. https://doi.org/10.1037//0096-3445.115.1.39

Pavlenko, A., & Malt, B. C. (2011). Kitchen Russian: Cross-linguistic differences and first-language object naming by Russian–English bilinguals*. *Bilingualism:*

Language and Cognition, *14*(1), 19–45. https://doi.org/10.1017/S136672891000026X

Petkov, G., & Petrova, Y. (2019). Relation-Based Categorization and Category Learning as a Result From Structural Alignment. The RoleMap Model. *Frontiers in Psychology*, *10*, 563. https://doi.org/10.3389/fpsyg.2019.00563

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3, Pt.1), 353–363. https://doi.org/10.1037/h0025953

Purves, R. S., Striedl, P., Kong, I., & Majid, A. (2023). Conceptualizing Landscapes Through Language: The Role of Native Language and Expertise in the Representation of Waterbody Related Terms. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12652

R Core Team. (2022). *R: A language and environment for statistical computing* (v4.2.1) [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Rosch, E. H., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rosch, E. H., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-x

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology*, *25*(2), 231–280. https://doi.org/10.1006/cogp.1993.1006

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Harvard University Press. https://doi.org/10.4159/harvard.9780674866270

Spivey, M. J., & Dale, R. (2004). On the continuity of mind: Toward a dynamical account of cognition. *The Psychology of Learning and Motivation: Advances in Research and Theory, Vol 45.*, *45*(297), 87–142. https://psycnet.apa.org/fulltext/2005-01906-003.pdf

Stanton, R. D., Nosofsky, R. M., & Zaki, S. R. (2002). Comparisons between exemplar similarity and mixed prototype models using a linearly separable category structure. *Memory & Cognition*, *30*(6), 934–944. https://doi.org/10.3758/bf03195778

Striedl, P., Majid, A., & Purves, R. S. (2024). Sensory, motor, and emotion associations for landscape concepts differ across neighbouring speech communities. *Landscape Research*, 1–18. https://doi.org/10.1080/01426397.2024.2334055

Taylor, K. I., Moss, H. E., & Tyler, L. K. (2007). The conceptual structure account: A cognitive model of semantic memory and its neural instantiation. In J. Hart & M. A. Kraut (Eds.), *Neural Basis of Semantic Memory* (pp. 265–301). Cambridge University Press. https://doi.org/10.1017/cbo9780511544965.012

UN-Habitat. (2018). *What is a city?* United Nations Human Settlements Programme. https://unhabitat.org/sites/default/files/2020/06/city_definition_what_is_a_city.pdf

Valério, D., Hussain, A., & Almeida, J. (2023). Semantic feature production norms for manipulable objects. *Cognitive Neuropsychology*, *40*(3–4), 167–185. https://doi.org/10.1080/02643294.2023.2279185 van Putten, S., O'Meara, C., Wartmann, F., Yager, J., Villette, J., Mazzuca, C., Bieling, C., Burenhult, N., Purves, R., & Majid, A. (2020). Conceptualisations of landscape differ across European languages. *PloS One*, *15*(10), e0239858. https://doi.org/10.1371/journal.pone.0239858

Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, *40*(1), 183–190. https://doi.org/10.3758/brm.40.1.183

Voorspoels, W., Storms, G., & Vanpaemel, W. (2011). Representation at different levels in a conceptual hierarchy. *Acta Psychologica*, *138*(1), 11–18. https://doi.org/10.1016/j.actpsy.2011.04.007

Wild, F. (2022). *lsa: Latent Semantic Analysis* (0.73.3) [Computer software]. Comprehensive R Archive Network (CRAN). https://cran.r-project.org/package=lsa

Wilson, C. R., Rourke, J., Oandasan, I. F., & Bosco, C. (2020). Progress made on access to rural health care in Canada. *Canadian Family Physician / Le Médecin de Famille Canadien*, *66*(1), 31–36. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7012120

Wittgenstein, L. (2009). *Philosophical Investigations* (P. M. S. Hacker & J. Schulte, Eds.; G. E. M. Anscombe, P. M. S. Hacker, & J. Schulte, Trans.; 4th ed.). Blackwell Publishing Ltd. https://play.google.com/store/books/details?id=n4hKDwAAQBAJ

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review*, *23*(4), 1015–1027. https://doi.org/10.3758/s13423-015-0948-7

Zhuang, T., & Lingnau, A. (2022). The characterization of actions at the superordinate, basic and subordinate level. *Psychological Research*, *86*(6), 1871–1891. https://doi.org/10.1007/s00426-021-01624-0

Appendix A

Demographic Characteristics of Participants

Characteristic	Full sample	
	n	%
Gender		
Woman	57	46.72
Man	57	46.72
Non-binary	4	3.28
Genderfluid	2	1.64
Prefer not to answer	2	1.64
Country of Residence		
Canada	80	65.57
United States	42	34.43
Country of Birth		
Canada	54	44.26
United States	39	31.97
Nigeria	3	2.46
China	3	2.46
Bangladesh	2	1.64
Philippines	2	1.64
South Africa	2	1.64
Vietnam	2	1.64
Algeria	1	0.82
Argentina	1	0.82
Guatemala	1	0.82
Guyana	1	0.82
India	1	0.82
Japan	1	0.82
Mexico	1	0.82
Morocco	1	0.82

Nepal	1	0.82
Pakistan	1	0.82
Poland	1	0.82
Puerto Rico	1	0.82
Russian Federation	1	0.82
United Kingdom	1	0.82
Prefer not to answer	1	0.82
First Language		
English	105	86.07
Vietnamese	3	2.46
French	2	1.64
Polish	2	1.64
Spanish	2	1.64
Cantonese	1	0.82
Japanese	1	0.82
Korean	1	0.82
Tagalog (Filipino)	1	0.82
Prefer not to answer	4	3.28
Ethnicity		
Canadian	38	31.15
English	14	11.48
Irish	5	4.10
African American	4	3.28
Chinese	4	3.28
Vietnamese	4	3.28
Indian	3	2.46
Jewish	3	2.46
Polish	3	2.46
Afro-Canadian	2	1.64
Americo-Liberian	2	1.64
Anglo Afro-Caribbean	2	1.64

Arab	2	1.64
Bengali	2	1.64
Russian	2	1.64
Anglo-Canadian	1	0.82
Anglo-Irish	1	0.82
Bahun	1	0.82
Bazigar	1	0.82
Caribbean	1	0.82
Caucasian	1	0.82
Croat	1	0.82
Dutch	1	0.82
Efik	1	0.82
European	1	0.82
Filipino	1	0.82
Finn	1	0.82
German	1	0.82
Hispanic	1	0.82
Italian	1	0.82
Khmer	1	0.82
Korean	1	0.82
Sindhi	1	0.82
Slovak	1	0.82
South Asian	1	0.82
Ukrainian	1	0.82
Prefer not to answer	11	9.02
Hometown		
City	97	79.51
Rural	23	18.85
Prefer not to answer	2	1.64
Mode of transportation		
Car	82	67.21
Public transit	23	18.85

Walk	5	4.10
Bicycle	4	3.28
Uber or Taxi	4	3.28
Motorcycle	2	1.64
Scooter	1	0.82
Prefer not to answer	1	0.82
Current		
City	102	83.61
Rural	18	14.75
Prefer not to answer	2	1.64
Mode of transportation		
Car	83	68.03
Public transit	23	18.85
Walk	8	6.56
Uber or Taxi	5	4.10
Bicycle	2	1.64
Prefer not to answer	1	0.82
Driver’s license		
Yes	108	88.52
No	12	9.84
Expired	1	0.82
Prefer not to answer	1	0.82
Education		
Doctorate	3	2.46
Master’s degree	16	13.11
Bachelor’s degree	59	48.36
Vocational school	5	4.10
High school degree or equivalent	36	29.51
Prefer not to answer	3	2.46
Employment		
Full-time job	64	52.46
Full-time student, full-time job	2	1.64

Full-time student, part-time job	3	2.46
Full-time student, unemployed	8	6.56
Part-time job	18	14.75
Part-time student, part-time job	1	0.82
Retired	5	4.10
Unemployed	12	9.84
Prefer not to answer	9	7.38

Appendix B

Familiarity Rating Task

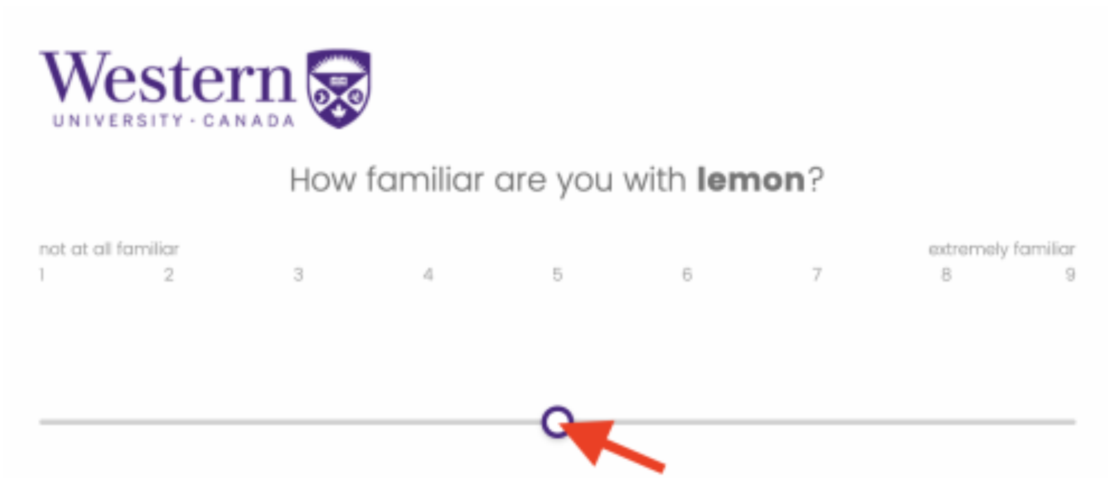
Instructions

On the next screen, you'll be asked how well you know the thing or idea that a given word points to.

Note that this isn't about the word itself but about the **thing or idea** it represents.

Think about your own life - your experiences, conversations, films or TV programs you've watched, or things you've read. Use these to decide how familiar you are with the thing or idea that the word refers to.

Here is an example of what you will see:



The rating scale ranges from 1 (not at all familiar) to 9 (extremely familiar) To indicate your familiarity, click on the slider and drag it to the position that best represents your level of familiarity with that item. Click the purple navigation arrow at the bottom of the screen to continue

Familiarity Rating Task

How familiar are you with **american city**?

not at all familiar extremely familiar

1 2 3 4 5 6 7 8 9

Appendix C

Property Listing Task

Task Instructions

On the next screen, there are words that each denote a concept, with each being followed by 10 blank lines. Please fill in as many of these lines as you can with properties of the concept to which the word refers.

Examples of different types of properties would be: physical properties, such as internal and external parts, and how it looks, sounds, smells, feels, or tastes; functional properties, such as what it is used for; where, when and by whom it is used; things that the concept is related to, such as the category that it belongs in; and other facts, such as how it behaves, or where it comes from.

Please note that even though many of the words can be thought of as something other than a noun (e.g., "camp" can refer to the place where your tent is pitched, or the action of camping), all words on the following pages are meant to be considered as nouns only (e.g., "camp," the place).

Below, we have provided 3 examples to give you an idea of what might be considered a property description of a concept.

Duck	Cucumber	Stove
is a bird		
is an animal	is a vegetable	is an appliance
waddles	has green skin	produces heat
flies	has a white inside	has elements
migrates	is cylindrical	has an oven
lays eggs	is long	made of metal
quacks	grows in gardens	is hot
swims	grows on vines	is electrical
has wings	is edible	runs on wood
has a beak	is crunchy	runs on gas
has webbed feet	used for making pickles	found in kitchens
has feathers	eaten in salads	used for baking
lives in ponds		used for cooking food
lives in water		
hunted by people		
is edible		

Here is an example of what you will see:

You may be able to think of more and/or different types of properties for these concepts, but these examples should give you an idea of what is requested of you. Please do not languish an extraordinary amount of time on each word, but also please take a bit of time to consider the relevant properties of each concept. In other words, complete this questionnaire reasonably quickly, but keep the relevant types of properties in mind.

Thank you very much for completing this questionnaire.

Click the purple navigation arrow at the bottom of the screen to continue

Property Listing Task

Please fill in as many of these lines as you can with properties of the things to which the following word(s) refer:

american city

____
____
____
____
____
____
____
____
____
____

Examples of different types of properties would be: physical properties, such as internal and external parts, and how it looks, sounds, smells, feels, or tastes; functional properties, such as what it is used for; where, when and by whom it is used; things that the concept is *related to, such as the category that it belongs in; and other facts, such as how it behaves, or where it comes from. Please note that even though many of the words can be thought of as something other than a noun (e.g., “camp” can refer to the place where your tent is pitched, or the action of camping), all words on the following pages are meant to be considered as nouns only (e.g., “camp,” the place).*

Appendix D

Demographics Questionnaire

Welcome to the **final phase** of the study.
Here, you will be asked to fill out a brief **demographics questionnaire**. We kindly request that you respond to all questions truthfully and endeavour to complete the remaining questions in **one uninterrupted session**.

What is your **age** (in years)?
For example, if you are 18, you would write "18" in the space below.

How would you best describe your **gender** identity?

- Woman
- Man
- Genderfluid
- Non binary
- Two spirit
- Trans woman
- Trans man
- Gender category/identity not listed (please specify below)
- Prefer not to say

What **languages** can you speak? Select **all** that apply.
If the language is not listed, please specify (up to three languages) in the space(s) provided.

- Arabic
- Cantonese
- Cree
- English
- French
- German
- Greek
- Hindi
- Hungarian
- Inuktitut
- Italian
- Japanese
- Korean
- Mandarin
- Michif
- Ojibwemowin
- Persian (Farsi)
- Polish
- Portugese
- Punjabi

Russian
Spanish
Tagalog (Filipino)
Tamil
Urdu
Vietnamese
Not listed 1 (please specify below)
Not listed 2 (please specify below)
Not listed 3 (please specify below)

Of the language(s) you selected, which is your **first language**?

▼ Arabic ... Not listed 3 (please specify below)

Rate your **level of proficiency** in the language(s) you listed.

The rating scale ranges from 0 (no proficiency) to 6 (native or bilingual proficiency).

Indicate your response by clicking on the slider and dragging it to the rating that best represents your level of proficiency with that language.

*NOTE: If you are <u>unsure how to respond</u>, see the **response guide** at the bottom of the current survey window.You are <u>not</u> required to read the response guide.*

No proficiency | Elementary proficiency | Limited working proficiency | Professional working proficiency | Full professional proficiency | Native or bilingual proficiency

0 1 2 3 4 5 6

<u>Response Guide:</u>
No proficiency
Your spoken language is limited to occasional, isolated words, and you can ask questions or make statements accurately only with memorized phrases. Your reading ability is limited to numbers, isolated words and phrases, and familiar names or signs. Understanding is restricted to isolated words or memorized phrases for immediate needs. You might be able to write using basic symbols in an alphabetic or syllabic system or 50 common characters.
Elementary proficiency
You can meet basic travel needs and behave politely. You can ask and answer simple questions within a limited range of experience, though native speakers may need to rely on context and put in extra effort to understand you. Your comprehension of basic speech can be improved with aids like slower speech or repetition. Your vocabulary is just broad enough to communicate the most fundamental needs. Your writing consists of simple sentences or fragments, often with continuous spelling and grammar errors. Most tasks at this level include basic functions like buying goods, telling time, ordering

simple meals, and asking for basic directions.
Limited working proficiency
You can meet routine social demands and handle limited work requirements with confidence. You can navigate basic social situations, including introductions and casual conversations about current events, work, and personal information. Although you may need help with complex situations, you can understand the gist of non-technical conversations and respond simply, even if you occasionally need to talk around a word. While your accent might not be perfect, it's understandable. You can usually use basic grammar accurately, although you may not have full confidence or control over it.
Professional working proficiency
You can speak the language effectively in most practical, social, and professional conversations, even discussing your interests and special areas of expertise with ease. Your understanding is comprehensive at a normal speed of speech, and your vocabulary is wide enough that you rarely need to search for a word. Although your accent might be noticeably non-native, your strong grasp of grammar and minimal errors do not affect your understanding or confuse native speakers. You can use the language in professional settings, reliably extract information and opinions from native speakers, answer objections, clarify points, state and defend policies, conduct meetings, and read almost all types of prose, including news reports, routine correspondence, and technical material in your fields of expertise, whether the topics are familiar or not.
Full professional proficiency
At this level, you can fluently and accurately use the language across professional contexts. You can understand and engage in conversations within your experience, with precise vocabulary, even in unfamiliar situations. Although you may not always pass as a native speaker due to occasional lapses in idioms or cultural references, you can interpret the language informally and effectively participate in various verbal exchanges, including conferences and debates. Your understanding extends to intricate details and implications of concepts differing from your native language.
Native or bilingual proficiency
At this level, you speak as fluently as a well-educated native speaker. You have mastered the language completely, and your speech is easily understood and accepted by native speakers in every aspect. This includes using a wide range of vocabulary and common sayings, as well as references that make sense in the culture of the language you're using.

Which ethnic group(s) do you identify with (select up to 3)?
NOTE: Ethnicity refers to groups (such as Irish, Fijian, Sioux, etc.) that share a collective identity rooted in common ancestry, language, or culture.

Ethnicity (primary)

▼ !Kung ... Balkar

Ethnicity (secondary)

▼ !Kung ... Balkar

Ethnicity (tertiary)

▼ !Kung ... Balkar

If you **prefer not to answer**, scroll all the way to the bottom of the dropdown menu and select 'prefer not to answer.'

You indicated that your ethnicity was not listed.
Which ethnic group(s) do you identify with (provide up to 3)?

The next questions ask about your **hometown**. This should be the place **where you grew up or where you've spent most of your life, <u>whichever is the longest duration</u>**. You will be asked to indicate the following information with respect to your hometown:
Name of the country
Name of the province, state, territory, or equivalent
Name of the city, town, village

Where did you grow up?
Home country

▼ Afghanistan ... Liechtenstein

If home state, province, or territory **does not apply** to your country, scroll all the way to the bottom of the dropdown menu below and select 'does not apply.'

If your home state, province, or territory is **not listed**, scroll all the way to the bottom of the dropdown menu below and select 'not listed.'

If you **prefer not to answer**, scroll all the way to the bottom of the dropdown menu below and select 'prefer not to answer.'

Home state, province, or territory

▼ Alabama ... Connecticut

You indicated that your home state, province, or territory was **not listed**. What is your home state, province, or territory?

hometown What is the **name of your hometown**? (E.g., if you grew up in Toronto, you would write Toronto in the space below)

Would you consider your hometown a **city** or a **rural** area?

City
Rural

How long have you lived in your hometown, either in the past or currently?
I have been living in my hometown for __ years.
I lived in my hometown for __ years before moving to my current location.
I have lived in my hometown my entire life.

Do you **currently** live in a city or rural area?
City
Rural

What was your primary mode of transportation in your hometown?
Public transit
Uber or Taxi
Car
Bicycle
Walk
Other (please specify)

What mode of transportation do you **primarily** use at the moment?
Public transit
Uber or Taxi
Car
Bicycle
Walk
Other (please specify)

Do you have a driver's license?
Yes
No
I did, but it's expired

How old were you when you started driving?

Based on where you **currently live**, how important is it for you to be able to **drive a car**?

not at all important extremely important

1 2 3 4 5 6 7 8 9

Based on where you **grew up/spent the most time** (i.e., your hometown), how important was it for you to be able to **drive a car**?

not at all important extremely important

1 2 3 4 5 6 7 8 9

Is **public transit** *available* where you **currently live**?
  Yes
  No

Based on where you **currently live**, how important is it for you to have access to **public transit**?

not at all important extremely important

1 2 3 4 5 6 7 8 9

Was **public transit** *available* in your **hometown**?
  Yes
  No

Based on where you **grew up/spent the most time** (i.e., your **hometown**), how important was it for you to have access to **public transit**?

not at all important extremely important

1 2 3 4 5 6 7 8 9

What is the **highest degree** or **level of school** you have completed?
  High school degree or equivalent
  Bachelor's degree (e.g., BA, BSc)
  Master's degree (e.g. MA, MSc, MEd)
  Doctorate (e.g. PhD, EdD)
  Other (please specify)
  Prefer not to say

What is your current **employment** status?
  Full-time job
  Part-time job
  Unemployed
  Retired
  Full-time student, unemployed
  Full-time student, part-time job
  Full-time student, full-time job
  Part-time student, unemployed
  Part-time student, part-time job
  Part-time student, full-time job
  Prefer not to say

Appendix E

NMREB Continuing Ethics Review Approval Letter

Date: 16 April 2024

To: Dr. John Paul Minda

Project ID: 115056

Study Title: A Study Exploring the Conceptual Structure of the City

Application Type: Continuing Ethics Review (CER) Form

Review Type: Delegated

Date Approval Issued: 16/Apr/2024 12:08

REB Approval Expiry Date: 13/May/2025

Dear Dr. John Paul Minda,

The Western University Non-Medical Research Ethics Board has reviewed this application. This study, including all currently approved documents, has been re-approved until the expiry date noted above.

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Electronically signed by:

Mr. Joshua Hatherley, Ethics Coordinator on behalf of Dr. Isha DeCoito, NMREB Chair 16/Apr/2024 12:08

Reason: I am approving this document

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Appendix F

Concept Summary Statistics

Concept	Mean_Concr	SD_Concr	Pronunciation	COCA_FPM	COCA_log_freq_mil	Familiarity	Length_Letters	NOFs_uniq	NOFs_total
american city	4.645	0.549	əˈmɛrɪkən ˈsɪti	0.678	0.225	6.70	12	315	1230
beach town	4.715	0.775	biːʧ taʊn	0.194	0.077	7.95	9	313	1235
big city	4.273	0.786	bɪg sɪti	2.587	0.555	8.35	7	239	1232
border city	4.540	0.845	ˈbɔrdər ˈsɪti	0.144	0.058	6.00	10	287	1233
canadian city	4.645	0.549	kəˈneɪdiən ˈsɪti	0.059	0.025	8.15	12	295	1240
capital city	3.800	1.398	ˈkæpətəl ˈsɪti	1.813	0.449	8.30	11	325	1233
city	4.790	0.570	sɪti	377.998	2.579	8.75	4	313	1253
coastal city	4.340	0.930	ˈkoʊstəl ˈsɪti	0.267	0.103	6.95	11	326	1232
coastal town	4.265	1.035	ˈkoʊstəl taʊn	0.311	0.118	6.60	11	325	1241
college town	4.630	0.800	ˈkɑlɪʤ taʊn	0.523	0.183	8.30	11	293	1244
commuter town	3.835	1.025	kəmˈjutər taʊn	0.008	0.003	5.55	12	327	1229
conservative city	3.375	0.880	kənˈsɜrvətɪv ˈsɪti	0.036	0.015	5.15	16	300	1231
country town	4.405	0.980	ˈkʌntri taʊn	0.151	0.061	6.15	11	274	1232
cultural hub	2.845	1.030	ˈkʌlʧərəl hʌb	0.036	0.015	6.30	11	260	1236
downtown	4.390	0.880	ˈdaʊnˈtaʊn	33.097	1.533	8.65	8	244	1241
east coast city	4.495	0.744	iːst koʊst ˈsɪti	0.019	0.008	6.20	13	318	1235
farm town	4.615	0.855	fɑrm taʊn	0.122	0.050	6.75	8	316	1240
financial district	2.905	1.240	fəˈnænʃəl ˈdɪstrɪkt	0.586	0.200	6.20	17	306	1242
fishing village	4.595	0.660	ˈfɪʃɪŋ ˈvɪləʤ	0.435	0.157	5.75	14	295	1243
gated community	4.167	0.718	ˈgeɪtɪd kəmˈjunəti	0.469	0.167	6.95	14	245	1233
gentrified neighbourh	3.735	0.915	ˈʤɛntrɪˌfaɪd ˈneɪbərˌ	0.000	0.000	5.35	23	359	1239
ghost town	2.909	1.514	goʊst taʊn	0.893	0.277	6.50	9	318	1228
green city	4.430	0.895	griːn ˈsɪti	0.056	0.024	5.50	9	291	1247
historic city	3.360	0.890	hɪˈstɔrɪk ˈsɪti	0.120	0.049	7.20	12	306	1237
historic town	3.285	0.995	hɪˈstɔrɪk taʊn	0.111	0.046	6.95	12	287	1237
ideal city	3.080	0.620	aɪˈdiːl ˈsɪti	0.037	0.016	5.15	9	324	1267
industrial city	3.895	1.055	ɪnˈdʌstriəl ˈsɪti	0.237	0.092	6.80	14	337	1241
land locked city	4.290	0.995	ˈlænˌdlɑkt ˈsɪti	0.003	0.001	6.10	14	302	1228
liberal city	3.230	0.800	ˈlɪbərəl ˈsɪti	0.052	0.022	5.45	11	331	1244
mega city	3.430	0.925	ˈmɛgə ˈsɪti	0.029	0.013	5.90	8	300	1245
metropolis	3.890	1.340	məˈtrɑpələs	2.602	0.557	6.65	10	310	1240
midsized city	3.715	0.963	ˈmɪdˌsaɪzd ˈsɪti	0.009	0.004	7.65	12	300	1232
midwestern city	3.768	0.963	mɪˈdwɛstərn ˈsɪti	0.116	0.048	4.55	14	310	1239
mountain city	4.875	0.385	ˈmaʊntən ˈsɪti	0.052	0.022	6.00	12	248	1229
native reservation	3.085	1.490	ˈneɪtɪv ˌrɛzərˈveɪʃən	0.006	0.003	6.25	17	356	1245
china town	4.715	0.640	ˈʧaɪnəˌtaʊn	0.043	0.018	6.65	9	292	1233
neighbourhood	4.750	0.440	ˈneɪbərˌhʊd	1.477	0.394	8.50	13	268	1233
old city	3.600	1.578	oʊld ˈsɪti	1.211	0.345	7.20	7	351	1230
religious city	3.645	1.015	rɪˈlɪʤəs ˈsɪti	0.011	0.005	5.20	13	307	1234
remote city	4.115	1.025	rɪˈmoʊt ˈsɪti	0.021	0.009	4.90	10	319	1231
riverside town	4.250	1.010	ˈrɪvərˌsaɪd taʊn	0.014	0.006	5.70	13	339	1233
small city	4.005	1.120	smɔːl ˈsɪti	0.716	0.235	7.85	9	316	1231
small town	3.800	1.056	smɔːl taʊn	7.000	0.903	8.10	9	340	1233
southern city	3.705	0.930	ˈsʌðɚn ˈsɪti	0.386	0.142	4.60	12	328	1240
sprawling city	3.705	0.935	ˈsprɔːlɪŋ ˈsɪti	0.080	0.033	4.50	13	292	1243
suburb	3.760	1.270	sʌbɜːrb	5.234	0.795	7.70	6	291	1246
tech hub	3.310	1.320	tɛk hʌb	0.033	0.014	6.00	7	343	1248
tent city	3.300	1.494	tɛnt ˈsɪti	0.335	0.126	4.90	8	269	1235
tourist town	4.615	0.765	ˈtʊrəst taʊn	0.117	0.048	8.20	11	277	1230
town	4.640	0.780	taʊn	157.958	2.201	8.35	4	294	1227
uptown	3.290	1.300	ʌptaʊn	2.636	0.561	5.35	6	309	1229
urban centre	3.640	1.275	ˈɜrbən ˈsɛntər	0.016	0.007	6.80	11	285	1231
village	4.890	0.310	ˈvɪləʤ	49.553	1.704	7.25	7	308	1228
waterfront community	4.095	1.050	ˈwɔtərˌfrʌnt kəmˈjunə	0.021	0.009	5.70	19	350	1227
west coast city	4.147	1.143	wɛst koʊst ˈsɪti	0.015	0.007	6.10	13	276	1232
working class commu	3.617	1.310	ˈwɜrkɪŋ klæs kəmˈju	0.014	0.006	7.00	21	370	1229
world class city	4.333	0.920	wɜrld klæs ˈsɪti	0.015	0.007	5.60	14	338	1239

Combined
Brysbaert 2014
Muraki 2022

Appendix G

Top 10 Feature Production Frequencies (All Participants; *N* = 122)

concept	feature_1	n_1	feature_2	n_2	feature_3	n_3	feature_4	n_4	feature_5	n_5	feature_6	n_6	feature_7	n_7	feature_8	n_8	feature_9	n_9	feature_10	n_10
American City	busy	12	guns	12	loud	11	diverse	10	cars	8	dangerous	8	fast food	7	homelessness	7	moderate traffic	6	patriotic	6
Beach Town	tourism	30	sand	26	sunny	15	ocean	11	beaches	8	restaurants	8	expensive	7	relaxing	7	warm	7	water	6
Big City	busy	20	public transportation	18	dense population	16	diverse	11	skyscrapers	11	moderate traffic	9	crime	8	expensive	8	pollution	8	cars	7
Border City	tourism	23	police	8	busy	7	crime	7	diverse	7	travel	7	security	6	immigration	5	multilingual	5	people	5
Canadian City	cold	17	friendly	13	diverse	10	busy	9	french speaking	9	multicultural	9	hockey	8	cold during winter	7	polite	7	tourism	7
Capital City	tourism	17	busy	15	government	13	expensive	8	moderate traffic	6	ottawa	6	public transportation	6	businesses	5	crowded	5	expensive cost of living	5
Chinatown	busy	16	tourism	16	chinese food	11	restaurants	11	food	10	crowded	9	culture	8	good food	8	asian people	7	chinese people	7
City	public transportation	22	busy	15	buildings	12	people	11	moderate traffic	10	noisy	10	pollution	9	big	8	diverse	8	urban	8
Coastal City	tourism	23	ocean	16	water	16	beaches	15	boats	9	expensive	7	sand	7	busy	6	seafood	5	warm	5
Coastal Town	tourism	23	beaches	22	ocean	21	boats	11	fishing	9	water	8	breezy	5	elderly population	5	ports	5	swimming	5
College Town	young population	20	bars	18	students	14	parties	12	busy	8	loud	8	public transportation	7	walkable	7	access to education	6	college	6
Commuter Town	cars	19	public transportation	15	quiet	10	suburban	9	residential	6	trains	6	bike friendly	5	boring	5	highways	5	safe	5
Conservative City	traditional	16	religious	14	racism	10	close minded	9	conservative	9	wealthy	9	churches	8	family oriented	7	guns	7	political	6
Country Town	farms	14	rural	12	low population	11	nature	11	friendly	10	quiet	9	farmland	8	horses	8	small	8	farming	7
Cultural Hub	arts scene	13	diverse	13	museums	11	busy	10	community	10	culture	9	tourism	8	art galleries	7	creative	7	food	7
Downtown	busy	31	restaurants	25	skyscrapers	16	centre	14	public transportation	13	businesses	11	heavy traffic	11	expensive	9	shops	9	cars	8
East Coast City	busy	13	beaches	8	cold	8	tourism	7	ocean	6	public transportation	6	seafood	6	water	6	boats	5	friendly	5
Farm Town	tractors	13	animals	12	rural	10	crops	8	quiet	8	conservative	7	cows	7	farmers markets	7	green space	7	livestock	7
Financial District	banks	20	busy	17	businesses	16	wealthy	16	suits	13	expensive	11	skyscrapers	11	money	10	people in suits	8	fast paced	7
Fishing Village	small	19	boats	15	water	11	quiet	9	elderly population	8	fish	8	fishing	8	unpopulated	8	fish smell	7	coastal	6
Gated Community	safe	26	expensive	23	exclusive	19	wealthy	17	rich	15	private	14	security	11	home owners association	10	quiet	10	pools	9
Gentrified Neighbourhood	expensive	8	wealth disparity	6	white people	6	trendy	5	affluent	4	cars	4	diverse	4	low crime rate	4	money	4	poverty	4
Ghost Town	abandoned	24	empty	17	quiet	15	unpopulated	10	dusty	8	old	8	abandoned buildings	7	tourism	7	no people	6	boring	5
Green City	bike friendly	15	trees	15	recycling	14	green space	12	nature	12	parks	12	public transportation	12	sustainable	11	walkable	10	clean	9
Historic City	tourism	33	old	16	museums	13	monuments	7	history	6	architecture	5	clean	5	arts scene	4	brick	4	crowded	4
Historic Town	tourism	24	old	18	museums	14	small	9	history	8	unpopulated	7	elderly population	6	monuments	6	old buildings	6	quaint	6
Ideal City	public transportation	26	green space	17	clean	15	safe	15	walkable	15	access to education	11	parks	11	friendly	9	healthcare	9	sustainable	9
Industrial City	pollution	23	factories	15	busy	12	noisy	10	warehouses	7	jobs	6	loud	6	moderate traffic	6	public transportation	6	dirty	5
Landlocked City	distant from water	10	cars	8	car dependent	6	surrounded by land	6	isolated	5	roads	5	busy	4	dry	4	highways	4	land	4
Liberal City	young population	10	diverse	9	accepting	7	progressive	7	busy	5	lgbtq	5	democratic	4	green space	4	parks	4	public transportation	4
Mega City	public transportation	22	busy	17	expensive	11	crowded	10	tourism	10	moderate traffic	9	skyscrapers	9	people	8	populated	8	heavy traffic	7
Metropolis	busy	21	public transportation	11	skyscrapers	10	large	9	populated	9	heavy traffic	8	densely populated	7	urban	7	diverse	6	moderate traffic	6
Midsized City	public transportation	11	suburban	9	access to education	8	moderate traffic	8	cars	5	diverse	5	parks	5	affordable housing	4	amenities	4	average	4
Midwestern City	agriculture	6	farming	6	flat	6	friendly	6	boring	5	cold	5	farmland	5	farms	5	hot	5	community	4
Mountain City	tourism	20	cold	15	skiing	13	snow	12	clean air	9	expensive	8	hiking	8	nature	7	fresh air	6	hills	6
Native Reservation	poverty	17	nature	11	traditional	11	isolated	7	poor public healthcare	7	casinos	6	rural	6	alcoholism	5	community	5	green space	5
Neighbourhood	community	16	houses	15	friendly	14	access to education	13	families	11	quiet	10	parks	9	safe	9	walkable	9	cars	8
Old City	tourism	14	historic	10	elderly population	8	walkable	7	run down	6	historical	5	landmarks	5	history	4	small	4	traditional	4
Religious City	churches	17	conservative	10	religious	9	tourism	8	close minded	6	religious events	5	strict	5	traditional	5	holy	4	quiet	4
Remote City	isolated	18	quiet	11	rural	10	small	9	far away	8	poor public transportation	8	unpopulated	8	distant	6	nature	6	poor public healthcare	6
Riverside Town	fishing	19	tourism	16	nature	15	boats	13	water	10	quiet	8	small	7	beautiful	6	river	6	bridges	5
Small City	quiet	8	friendly	6	unpopulated	6	community	5	businesses	4	cars	4	everyone knows each other	4	poor public transportation	4	quaint	4	small	4
Small Town	rural	14	quiet	11	unpopulated	11	lack of education	10	conservative	9	local businesses	6	cars	5	community	5	friendly	5	nature	5
Southern City	hot	16	conservative	10	humid	8	friendly	7	warm	7	racism	6	traditional	6	charming	5	historic	5	small	5
Sprawling City	cars	13	public transportation	10	busy	9	expanding	9	large	9	poor public transportation	9	suburban	7	car dependent	6	diverse	6	expansive	6
Suburb	quiet	17	access to education	15	families	14	parks	13	cars	12	houses	9	safe	9	children	8	residential	8	green space	7
Tech Hub	busy	9	computers	9	expensive	9	modern	9	technology	9	innovation	8	money	7	advanced	6	young population	6	clean	5
Tent City	homelessness	37	poverty	21	drugs	15	dirty	12	tents	10	crime	9	unsafe	9	community	5	dangerous	5	nature	5
Tourist Town	tourism	22	expensive	16	tourism based economy	13	souvenir shops	12	restaurants	11	attractions	10	busy	10	hotels	10	tourist attractions	8	sightseeing	7
Town	small	19	people	11	unpopulated	9	community	8	friendly	6	parks	6	public transportation	6	restaurants	6	businesses	5	cars	5
Uptown	busy	11	restaurants	9	rich	8	affluent	7	expensive	7	quiet	6	expensive housing	5	families	5	apartment buildings	4	big houses	4
Urban Centre	busy	13	public transportation	12	diverse	10	crime	8	businesses	7	dense	7	restaurants	7	homelessness	6	loud	6	people	6
Village	small	35	unpopulated	20	friendly	11	quiet	11	rural	11	nature	10	traditional	9	community	8	peaceful	7	quaint	6
Waterfront Community	boats	21	expensive	16	tourism	14	beaches	11	water	11	water sports	9	wealthy	9	nature	7	beautiful	6	fishing	6
West Coast City	beaches	17	ocean	15	tourism	14	expensive	13	liberal	9	homelessness	8	sunny	7	drugs	6	mountains	6	busy	5
Working Class Community	blue collar	13	busy	8	hard working	8	public transportation	7	families	6	lack of education	6	cars	5	factories	5	middle class	5	access to education	4
World Class City	tourism	19	public transportation	15	busy	14	expensive	10	crowded	7	wealthy	7	clean	6	safe	6	skyscrapers	6	diverse population	5

Appendix H

Top 10 Feature Production Frequencies (Long-Term Residents of Canada; *n* = 22)

concept	feature_1	n_1	feature_2	n_2	feature_3	n_3	feature_4	n_4	feature_5	n_5	feature_6	n_6	feature_7	n_7	feature_8	n_8	feature_9	n_9	feature_10	n_10
American City	guns	4	diverse	3	business	2	festive	2	heavy traffic	2	homelessness	2	loud	2	proud	2	skyscrapers	2	suburbs	2
Beach Town	sand	6	sunny	4	tourism	4	seafood	3	beaches	2	boats	2	expensive	2	friendly	2	tourist attractions	2	walkable	2
Big City	big	2	buildings	2	busy	2	congestion	2	dense population	2	expensive	2	public transportation	2	skyscrapers	2	tourism	2	apartment buildings	1
Border City	tourism	6	concrete	2	people	2	police	2	security	2	violence	2	active	1	angry	1	annoying	1	border control	1
Canadian City	diverse	3	multicultural	3	trees	3	cold	2	nature	2	polite	2	rural	2	safe	2	activities	1	bad smells	1
Capital City	tourism	4	government	3	businesses	2	large	2	skyscrapers	2	administrative	1	apartment buildings	1	attraction	1	big	1	big buildings	1
Chinatown	bakeries	3	chinese food	3	tourism	3	busy	2	crowded	2	cultural centre	2	good food	2	trinkets	2	asian	1	asian people	1
City	public transportation	5	moderate traffic	3	pollution	3	airport	2	buildings	2	businesses	2	dense population	2	diverse	2	people	2	populated	2
Coastal City	water	8	beaches	4	tourism	4	sand	3	boats	2	expensive	2	fisheries	2	nature	2	seafood	2	sun	2
Coastal Town	beaches	5	boats	4	small	3	water	3	breezy	2	fishing	2	laid back	2	ocean	2	relaxing	2	sandy	2
College Town	loud	3	access to education	2	alcohol	2	bars	2	public transportation	2	students	2	walkable	2	young people	2	activities	1	american	1
Commuter Town	cars	3	public transportation	3	bike friendly	2	quiet	2	suburban	2	trains	2	accessible	1	bars	1	blue collar	1	boring	1
Conservative City	religious	4	blue	3	wealthy	3	close minded	2	guns	2	political	2	public transportation	2	restrictive	2	traditional	2	accepting	1
Country Town	farmland	3	farms	3	horses	3	cowboys	2	cows	2	dirt roads	2	farming	2	fresh air	2	friendly	2	green space	2
Cultural Hub	community	5	arts scene	3	culture	3	good food	3	museums	3	music	3	art galleries	2	dancing	2	diverse culture	2	food	2
Downtown	restaurants	7	busy	5	heavy traffic	3	shops	3	bike friendly	2	cars	2	crime	2	crowded	2	expensive	2	food	2
East Coast City	busy	4	atlantic ocean	3	boats	3	urban	3	water	3	accents	2	beaches	2	cultural	2	ocean smell	2	party	2
Farm Town	quiet	3	cattle	2	chickens	2	community	2	farm animals	2	farmlands	2	farms	2	flat	2	green space	2	livestock	2
Financial District	businesses	4	expensive	4	busy	3	suits	3	wealthy	3	banks	2	fast paced	2	high rise	2	professionals	2	Pedestrians	1
Fishing Village	boats	4	exports	3	docks	2	elderly population	2	fish market	2	fish markets	2	fishing	2	friendly	2	quiet	2	rural	2
Gated Community	expensive	7	safe	6	exclusive	4	cars	3	clean	2	community	2	inaccessible to outsiders	2	posh	2	private	2	quiet	2
Gentrified Neighbourhood	expensive	3	cars	2	drugs	2	poverty	2	rich	2	affluent	1	angry	1	arts scene	1	attraction	1	boring	1
Ghost Town	abandoned	3	boring	3	deserted	3	quiet	3	unpopulated	3	abandoned buildings	2	creepy	2	dilapidated	2	dusty	2	empty	2
Green City	nature	4	green space	3	walkable	3	bike friendly	2	bushes	2	few cars	2	no pollution	2	public transportation	2	trees	2	young population	2
Historic City	tourism	4	old	3	arts scene	2	historic	2	important	2	knowledgeable	2	architecture	1	authentic	1	boring	1	brick	1
Historic Town	tourism	5	cobblestone	4	museums	4	old	3	small	3	educational	2	elderly population	2	famous	2	fascinating	2	historical	2
Ideal City	green space	6	public transportation	6	friendly	4	fun	3	parks	3	walkable	3	accepting	2	arts scene	2	cheap	2	healthcare	2
Industrial City	pollution	8	busy	4	factories	4	noisy	3	cars	2	concrete	2	gray	2	grey	2	high crime rate	2	poor public healthcare	2
Landlocked City	distant from water	3	immigrants	2	quiet	2	access to education	1	boring	1	car dependent	1	cars	1	developed	1	different cultures living together	1	difficult trading	1
Liberal City	accepting	3	diversity	2	progressive	2	arts scene	1	belief in equality	1	busy	1	calgary	1	community	1	convenience	1	developed	1
Mega City	public transportation	5	moderate traffic	3	big	2	busy	2	heavy traffic	2	huge	2	overpopulated	2	people	2	populated	2	stores	2
Metropolis	busy	7	moderate traffic	4	people	3	populated	3	public transportation	3	trains	3	buildings	2	cars	2	dense	2	densely populated	2
Midsized City	friendly	3	limited traffic	3	public transportation	3	walkable	3	access to education	2	accessible	2	green space	2	police	2	affordable	1	affordable cost of living	1
Midwestern City	cheap housing	2	desert	2	friendly	2	heat	2	hot	2	slow paced	2	accents	1	affordable	1	arenas	1	auctions	1
Mountain City	cold	3	remote	3	elevated	2	expensive	2	outdoor activities	2	scenic	2	skiing	2	snow	2	tourism	2	active	1
Native Reservation	community	2	cultural	2	historical points of inte	2	nature	2	traditional	2	alcoholism	1	animal fur	1	arts scene	1	authentic	1	cheap	1
Neighbourhood	friendly	5	quiet	3	clean	2	community	2	green space	2	houses	2	kids	2	parks	2	smells bad	2	strong sense of community	2
Old City	tourism	4	cultured	2	dusty	2	elderly population	2	green space	2	historical	2	run down	2	traditional	2	walkable	2	abandoned	1
Religious City	churches	4	religious	4	tourism	4	quiet	3	devout	2	libraries	2	long history	2	spiritual	2	strict	2	access to education	1
Remote City	isolated	2	nature	2	public transportation	2	rural	2	small	2	travel	2	accessible	1	boring	1	businesses	1	car dependent	1
Riverside Town	fishing	5	tourism	4	beautiful	2	nature	2	water	2	beaches	1	beautiful scenery	1	beautiful views	1	boating	1	built arund river	1
Small City	friendly	3	walkable	2	active	1	animals	1	better air	1	bike friendly	1	bored youth	1	brain drain	1	bricks	1	cheap	1
Small Town	conservative	3	friendly	3	no traffic	3	unpopulated	3	local businesses	2	peaceful	2	poverty	2	rural	2	agriculture	1	bars	1
Southern City	hot	3	cultural	2	historic	2	humid	2	small	2	accents	1	aquariums	1	authentic	1	beaches	1	big	1
Sprawling City	cars	4	car dependent	2	developed	2	expanding	2	expensive	2	has drugs	2	active	1	air	1	buses	1	congested	1
Suburb	quiet	5	bike friendly	3	access to education	2	cars	2	dense	2	expensive	2	families	2	family oriented	2	green space	2	outskirts of a city	2
Tech Hub	computers	6	learning	3	wires	3	advanced	2	busy	2	clean	2	internet	2	modern	2	access to education	1	active	1
Tent City	poverty	7	homelessness	4	crime	3	camping	2	community	2	alcohol	1	begging	1	black market economy	1	bonfire	1	campgrounds	1
Tourist Town	expensive	5	hotels	4	sightseeing	3	tourism	3	casinos	2	food	2	live music	2	pickpocketing	2	restaurants	2	souvenir shops	2
Town	cars	3	friendly	3	close knit	2	food	2	people	2	public transportation	2	shopping	2	small	2	stores	2	active	1
Uptown	expensive	3	rich	3	busy	2	events	2	gentrified	2	moderate traffic	2	noisy	2	people	2	public transportation	2	restaurants	2
Urban Centre	busy	6	businesses	4	active	3	expensive	3	noisy	3	public transportation	3	crowded	2	dense	2	lots of people	2	loud	2
Village	small	11	unpopulated	6	traditional	5	nature	4	quaint	3	quiet	3	rural	3	community	2	conservative	2	elderly population	2
Waterfront Community	expensive	4	nature	3	tourism	3	water	3	beaches	2	swimming	2	water sports	2	activities	1	aquatic	1	authentic	1
West Coast City	ocean	4	liberal	3	beaches	2	bike friendly	2	busy	2	cars	2	diverse population	2	drugs	2	expensive	2	fun	2
Working Class Community	hard working	3	access to education	2	blue collar	2	close knit	2	friendly	2	kids	2	lack of education	2	struggling	2	affordable housing	1	amazon hubs	1
World Class City	public transportation	3	expensive	2	modern	2	popular	2	accepting	1	access to education	1	accessible	1	activities	1	architecture	1	best job opportunities	1

Appendix I

Top 10 Feature Production Frequencies (Long-Term Residents of the United States; *n* = 22)

concept	feature_1	n_1	feature_2	n_2	feature_3	n_3	feature_4	n_4	feature_5	n_5	feature_6	n_6	feature_7	n_7	feature_8	n_8	feature_9	n_9	feature_10	n_10
American City	crowded	3	moderate traffic	3	apartment buildings	2	diverse	2	fast food	2	pollution	2	public transportation	2	skyscrapers	2	smog	2	access to education	1
Beach Town	sunny	4	tourism	4	restaurants	3	sand	3	beaches	2	hot	2	hotels	2	ocean	2	warm	2	bars	1
Big City	dense population	6	skyscrapers	5	moderate traffic	4	access to education	3	busy	3	crowded	3	loud	3	nightlife	3	arts scene	2	dirty	2
Border City	tourism	4	crime	3	diverse	3	immigration	2	multicultural	2	multilingual	2	police	2	travel	2	attractions	1	bad economy	1
Canadian City	cold	3	clean	2	clean air	2	french speaking	2	friendly	2	healthcare	2	access to education	1	active	1	basement houses	1	beautiful	1
Capital City	expensive	3	centre of economic activity and development	2	crowded	2	dangerous	2	embassy	2	houses the government and key institutions	2	public transportation	2	US congress	1	air quality	1	attractive	1
Chinatown	good food	4	chinese restaurants	3	tourism	3	asian culture	2	chinese food	2	chinese people	2	chinese writing	2	congested	2	family businesses	2	food	2
City	buildings	4	public transportation	3	arts scene	2	big	2	dense population	2	densely populated	2	employment	2	expensive	2	people	2	sidewalks	2
Coastal City	ocean	4	water	3	beaches	2	boats	2	busy	2	cars	2	cool weather	2	diverse	2	expensive	2	foot traffic	2
Coastal Town	beaches	5	tourism	5	ocean	3	coastal	2	historical	2	hot	2	surfing	2	wildlife	2	bars	1	based on trade	1
College Town	bars	2	busy	2	dirty	2	loud	2	parties	2	small	2	sports	2	access to education	1	active	1	alcohol	1
Commuter Town	bike friendly	3	public transportation	3	cars	2	pollution	2	quiet	2	requires a car	2	residential	2	suburban	2	traversable	2	access to education	1
Conservative City	traditional	6	churches	2	conservative	2	racism	2	religious	2	rural	2	unwelcoming	2	bad roads	1	bars	1	bumper stickers	1
Country Town	close knit	3	low population	3	rural	3	farmland	2	houses	2	slowed down	2	small	2	tractors	2	traditional	2	accents	1
Cultural Hub	arts scene	6	culture	3	community	2	cultural diversity	2	diverse	2	diverse culture	2	good food	2	history	2	languages	2	museums	2
Downtown	busy	5	businesses	4	restaurants	4	skyscrapers	4	centre	3	crowded	3	expensive	3	public transportation	3	business district	2	bustling	2
East Coast City	access to education	2	busy	2	loud	2	multicultural	2	police	2	airports	1	apartment buildings	1	bad city plannings	1	beaches	1	beautiful	1
Farm Town	animals	3	cows	3	crops	3	green space	3	horses	3	conservative	2	corn	2	farms	2	land	2	livestock	2
Financial District	banks	4	money	4	wealthy	4	businesses	2	bustling	2	capital flow	2	expensive	2	fancy restaurants	2	male dominated	2	people in suits	2
Fishing Village	small	5	fishing	4	fish	3	water	3	families	2	fish smell	2	fun	2	has fishing	2	intergenerational	2	nature	2
Gated Community	private	5	exclusive	4	expensive	4	security	4	big houses	3	home owners association	3	pools	3	safe	3	expensive cars	2	guarded entry	2
Gentrified Neighbourhood	expensive	4	diverse	2	wealth disparity	2	affluent	1	boston	1	capitalism	1	cars	1	changing	1	cleaned up graffiti	1	coffee	1
Ghost Town	abandoned	5	empty	5	old	3	dusty	2	haunted	2	sad	2	unpopulated	2	abandoned buildings	1	apathy	1	broken	1
Green City	sustainable	4	expensive	3	green space	3	public transportation	3	recycling	3	bike friendly	2	clean air	2	conservation	2	eco friendly	2	liberal	2
Historic City	tourism	8	museums	4	old	3	brick	2	monuments	2	old buildings	2	small	2	ancient ruins	1	archaeology	1	architectural heritage	1
Historic Town	old	5	culture	2	history	2	museums	2	tourism	2	architecture	1	artifacts	1	arts scene	1	attractions	1	beautiful	1
Ideal City	access to education	4	clean	3	healthcare	3	mountains	3	affordable housing	2	beautiful	2	green space	2	inexpensive	2	low crime rate	2	nice	2
Industrial City	pollution	6	warehouses	5	noisy	3	public transportation	3	blue collar	2	buildings	2	community	2	contribution to the national economy	2	distribution centres	2	factories	2
Landlocked City	cars	3	distant from water	3	civil unrest	2	isolated	2	lakes	2	affordable	1	airport	1	apartment buildings	1	better environments	1	big glass buildings	1
Liberal City	democratic	3	progressive	3	young population	3	political	2	violence	2	access to education	1	alive	1	annoying	1	arts scene	1	bad behavior	1
Mega City	expensive	3	homelessness	3	busy	2	crowded	2	diversity	2	loud	2	noisy	2	overpopulated	2	people	2	police	2
Metropolis	large	3	diverse	2	expensive	2	heavy traffic	2	ballets	1	buildings	1	bustling	1	busy	1	capital	1	corruption	1
Midsized City	cars	2	moderate traffic	2	not much to do	2	not special	2	parks	2	quiet	2	suburban	2	access to education	1	affordable housing	1	attraction	1
Midwestern City	lakes	2	moderate traffic	2	slow paced	2	agriculture	1	americas heartland	1	big lakes	1	blue skies	1	buffets	1	busy in summer	1	camping	1
Mountain City	cold	4	hills	4	snow	4	cabins	3	quiet	3	clean air	2	fresh air	2	hard to breathe	2	hiking trails	2	isolated	2
Native Reservation	poverty	4	nature	3	alcoholism	2	casinos	2	dirty	2	live off land	2	peaceful	2	quiet	2	trees	2	bartering	1
Neighbourhood	cars	5	houses	4	access to education	3	pets	3	families	2	moderate traffic	2	quiet	2	small	2	stores	2	streets	2
Old City	dirty	2	run down	2	small	2	tourism	2	alleys	1	architecture	1	authentic food	1	bad trash collection	1	big	1	big city	1
Religious City	religious	2	religious events	2	traditional	2	abbey	1	awful people	1	beirut	1	beliefs	1	berlin	1	bible	1	chapter house	1
Remote City	distant	3	isolated	3	poor public transportation	3	rural	3	boring	2	farmland	2	island	2	nature	2	private	2	quiet	2
Riverside Town	fishing	5	boats	3	beautiful	2	california	2	nature	2	quiet	2	river	2	scenic	2	water	2	adventure	1
Small City	cars	2	elderly population	2	no traffic	2	quaint	2	unpopulated	2	abandoned	1	accessible	1	agriculture	1	buildings	1	charming	1
Small Town	lack of education	4	cars	2	elderly population	2	limited shopping	2	quiet	2	rural	2	small stores	2	suburban	2	access to education	1	affordable housing	1
Southern City	friendly	4	hot	4	racism	3	traditional	3	arts scene	2	charming	2	comfort food	2	football	2	good food	2	high population	2
Sprawling City	cars	4	expansive	2	heavy traffic	2	industrial	2	low density	2	public transportation	2	stores	2	back roads	1	big	1	bike friendly	1
Suburb	accessible	3	cars	3	access to education	2	community	2	families	2	friendly	2	green space	2	homes	2	located on the outskirts of a city	2	parks	2
Tech Hub	busy	4	progressive	4	technology	4	innovation	3	startups	3	business	2	clean	2	computers	2	expensive	2	high population	2
Tent City	homelessness	6	drugs	3	tents	3	abandoned buildings	2	dirty	2	bad smells	1	begging	1	camping outside	1	community	1	crime	1
Tourist Town	tourism	8	tourism based economy	4	shopping	3	architecture	2	attractions	2	expensive	2	landmarks	2	money	2	noisy	2	accommodating	1
Town	small	5	cars	2	community	2	good food	2	main street	2	old	2	unpopulated	2	access to education	1	animals	1	beautiful	1
Uptown	affluent	4	new york	3	busy	2	expensive	2	expensive cars	2	expensive housing	2	families	2	fine dining	2	moderate traffic	2	modern	2
Urban Centre	crime	3	large population	3	buildings	2	busy	2	dirty	2	diverse	2	foot traffic	2	homelessness	2	public transportation	2	access to education	1
Village	small	5	community	3	dangerous	2	jungle	2	nature	2	safe	2	close knit	1	closed off	1	communal eating	1	countryside	1
Waterfront Community	boats	3	expensive	3	water	3	beaches	2	big	2	lake	2	water sports	2	apartment buildings	1	beach town	1	beautiful	1
West Coast City	beaches	4	expensive	4	homelessness	3	los angeles	3	busy	2	celebrities	2	dirty	2	high cost of living	2	laid back	2	large population	2
Working Class Community	blue collar	3	affordable cost of living	2	affordable housing	2	busy	2	cars	2	contributes to the local economy	2	families	2	industries and manufacturing	2	lack of education	2	low income	2
World Class City	public transportation	4	skyscrapers	4	tourism	4	crowded	3	attractions	2	busy	2	corporations	2	developed	2	diversity	2	expensive	2

Appendix J

------------------------------ *Features* ------------------------------

-- *Concepts* --

	cars	community	friendly	... N = 8474
City	6	1	1	
Village	1	8	11	
Town	5	8	6	
... N = 57				

$$similarity(City, Village) = cos(\theta) = \frac{City \cdot Village}{\| City \| \ \| Village \|} = \frac{\sum_{i=1}^{n} City_i \, Village_i}{\sqrt{\sum_{i=1}^{n} City_i^2 \, \sum_{i=1}^{n} Village_i^2}}$$