Supplementary Materials: Street Name Data as a Reflection of Migration and Settlement History

Sarah J. Berkemer [1,2,3]* and Peter F. Stadler [1,2,4,5,6,7,8]*

1. Data Extraction and Cleaning

Europe	#Packages	North America	#Packages	Africa	#Packages
France	22	Canada	13	Africa	53
Germany	28	Mexico	1		
Great Britain	49	US	51		
Poland	15				
Netherlands	10				
Italy	5				
Others	21				
Packages total	152	Packages total	65	Packages total	53
Countries total	27	Countries total	3	Countries total	53

Table S1. Table listing regional packages that have been downloaded from geofabrik.de in order to extract street name data. European countries with a large amount of data are further divided into several regional packages. Canada and the US are divided into packages according to their states.

Geographic data has been downloaded from `geofabrik` (download.geofabrik.de, Feburary 2020: Europe and North America, September 2020: Africa). In order to deal with the amount of data, geofabrik provides smaller packages defined by state or country border which results in 152 parts for Europe and 65 parts for North America, as listed in Table S1. The data has been extracted and converted using osmium tool (osmcode.org). Street names have been extracted filtering the data by the tag *highway*, which defines the category of all types of streets in OSM. Together with the street names, Table S2 lists additional data that has been extracted for each street if available within OSM. If no data is available, fields are filled with NA. Each street entry in OSM is represented by a set of nodes with unique IDs and geographical coordinates (latitude and longitude). In order to obtain one pair of coordinates for each street entry, geographic coordinates of all nodes that form a street entry are collected and their geographic midpoints are calculated. Each street entry has now a pair of latitude and longitude coordinates that is associated to its geographic midpoint and will be used in further steps of the analysis.

Extracted Data	Description
Street ID	unique ID for each street entry
Street type	e.g. residential, primary, secondary, tertiary
Maximal allowed speed	given in mph or km/h
One way street	if the street is a one way street
Street reference	other names for that street, e.g. national routes
Street lit	if street lights are available
Number of nodes	number of nodes in OSM a street is composed of
Node IDs	unique IDs of OSM nodes
Latitude	latitude coordinates for each node
Longitude	longitude coordinates for each node
Latitude midpoint	geographic center of all nodes' latitudes
Longitude midpoint	geographic center of all nodes' longitudes

Table S2. Additional data that has been extracted from OSM together with the street name for each street entry.

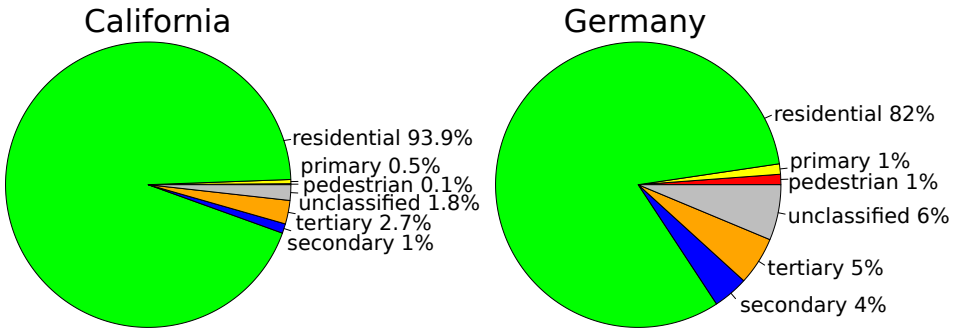

Figure S1. Percentages of street types in California (left) and Germany (right). Street types are part of the street entries in the OpenStreetMap data base.

Street entries in OSM are assigned a street type. In this analysis, we only include streets with types primary, secondary, tertiary, residential, pedestrian and unclassified. In this way, highways and motorways, which are typically named just by numbers, are excluded. Street types are used in the OSM database to classify streets by size and importance (wiki.openstreetmap.org/wiki/Key:highway). Primary, secondary and tertiary roads are named by importance level, connecting towns and villages. Residential roads are lined by houses and serve as an access to housing but do not connect towns. Pedestrian streets are used exclusively for pedestrians. The street type 'unclassified' does not indicate an unknown classification but defines the least important road type in a country, usually serving as access to properties.

Figure S1 shows percentages of street types for Germany and California. The higher amount of residential streets in California compared to Germany might be caused by the higher amount of cities and population density in Germany where residential areas are also located close to larger streets. Additionally, city centers typically have several pedestrian streets in most European cities.

2. Density of street names in Europe and Africa

Calculation of street densities is described in the main text. Figures S2 and S3 show plots for streets in Europe and Africa.

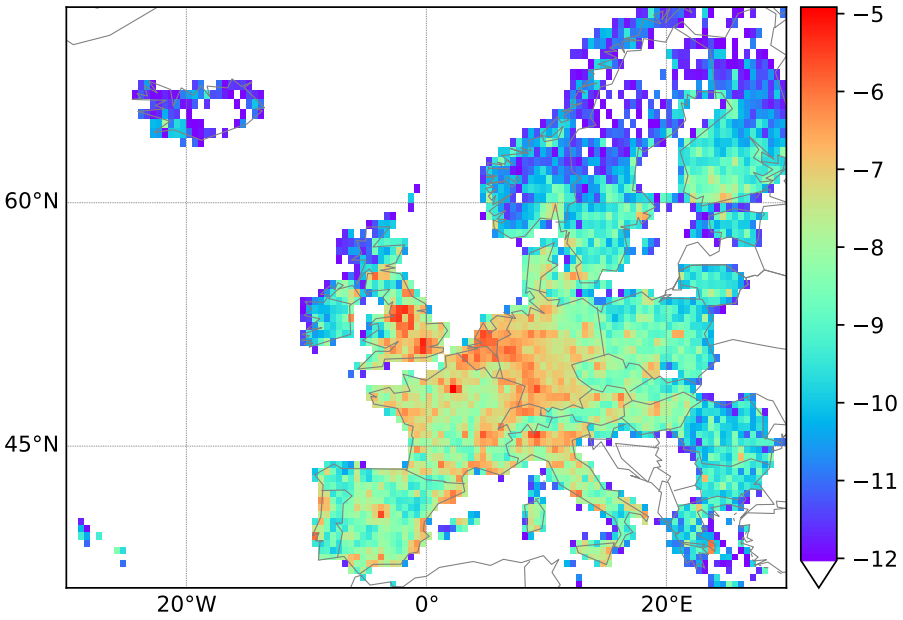

Figure S2. Heatmap of Europe showing the density of streets. Colors indicate the density of streets in this area, thus the logarithmic normalized number of streets per cell of the grid. Higher values and red colors show a higher density of streets.

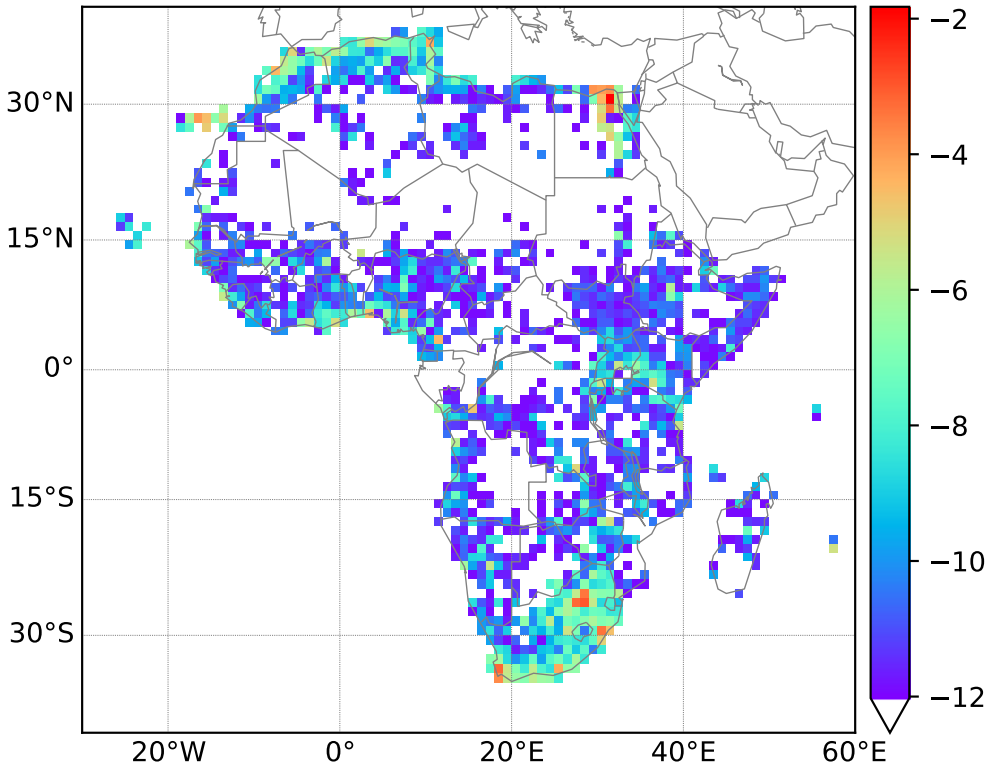

Figure S3. Heatmap of Africa showing the density of streets. Colors indicate the density of streets in this area, thus the logarithmic value of the number of streets normalized by all streets in the cell of the grid.

3. Extraction of street terms from street names

In many countries, street names are composed of two parts, a *specific* term that makes the street name unique in a certain area and a *generic* term, referring to the kind of street, such as road, way or boulevard [1]. In numerous cases, a third term is added which denotes the direction or orientation of the street such as east and west or upper and lower. Numbered streets are excluded from the analysis in order to be able to extract street names of European origin.

Figure S4 depicts the differences in term occurences for several Austrian terms where street kinds (generics) occur most frequently. The number of occurences is normalized by the number of all term occurences in the country. The second plot shows positionwise term occurences. With 'complete term', terms are counted that could be split from a street name by splitting at space (" ") or dash ("-") sign. Prefix, infix and suffix refer to parts of a single word as described above. The German language usually composes street names by concatenating all terms into one word, but there are also street names where terms are separated by a space or dash. Here, we can see again that street kinds mostly occur as suffixes or complete terms whereas the name 'josef' is either a prefix or a complete term (names are often separated by dash signs). The term 'hauptstraße' (translates to 'main street') is usually used as a single street name, however, the plot shows that it is also used as a street name suffix. Several terms such as 'dorf' (village), 'wald' (forest) or 'alt' (old) can be used in many ways and occur in all positions. Corresponding plots for street term positioning in English (Ireland) and French (Luxembourg) can be found in Figures S5 and S6 and show differences in language-specific composition of street names.

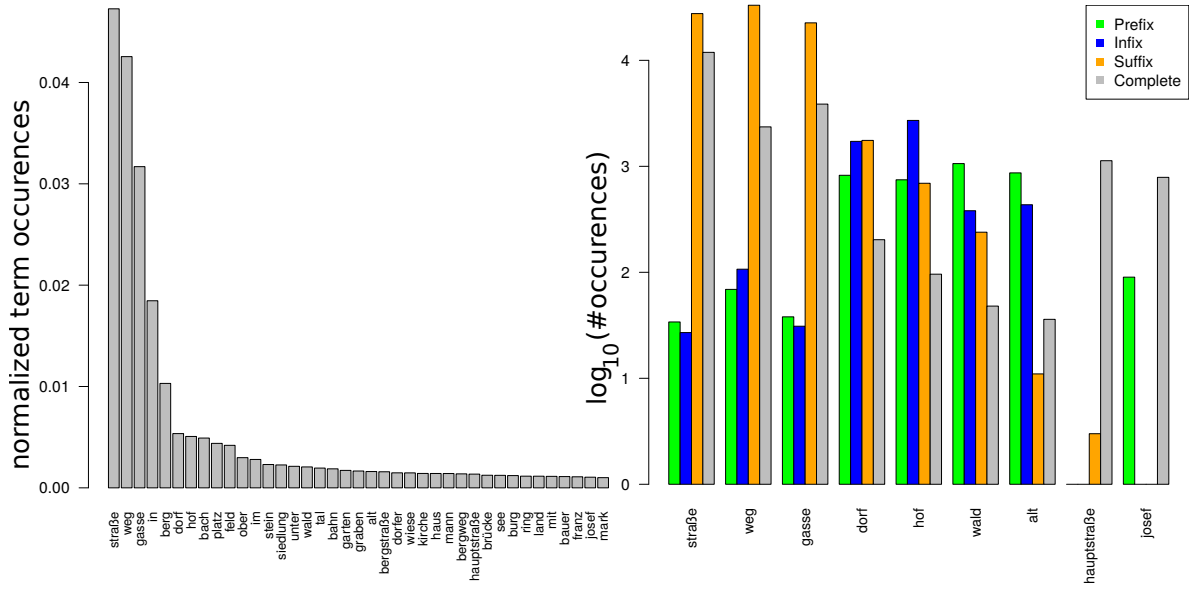

Figure S4. Left: Term occurences for selected street terms in Austria normalized by the complete number of terms. Right: Positionwise occurences of selected street terms in Austria. Prefix, infix and suffix refer to parts of one word whereas complete terms refers to the term having been separated from the street name by splitting the street name at space or dash symbol.

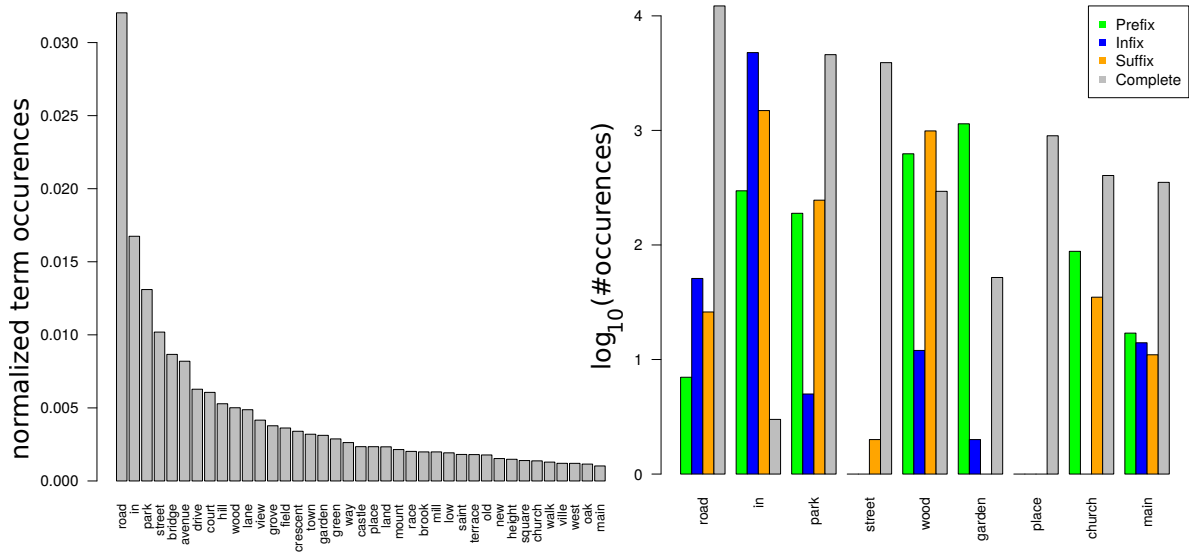

Figure S5. Left: Term occurences for selected street terms in Ireland normalized by the complete number of terms. Right: Positionwise occurences of selected street terms in Ireland. Prefix, infix and suffix refer to parts of one word whereas complete terms refers to the term having been separated from the street name by splitting the street name at space or dash sign.

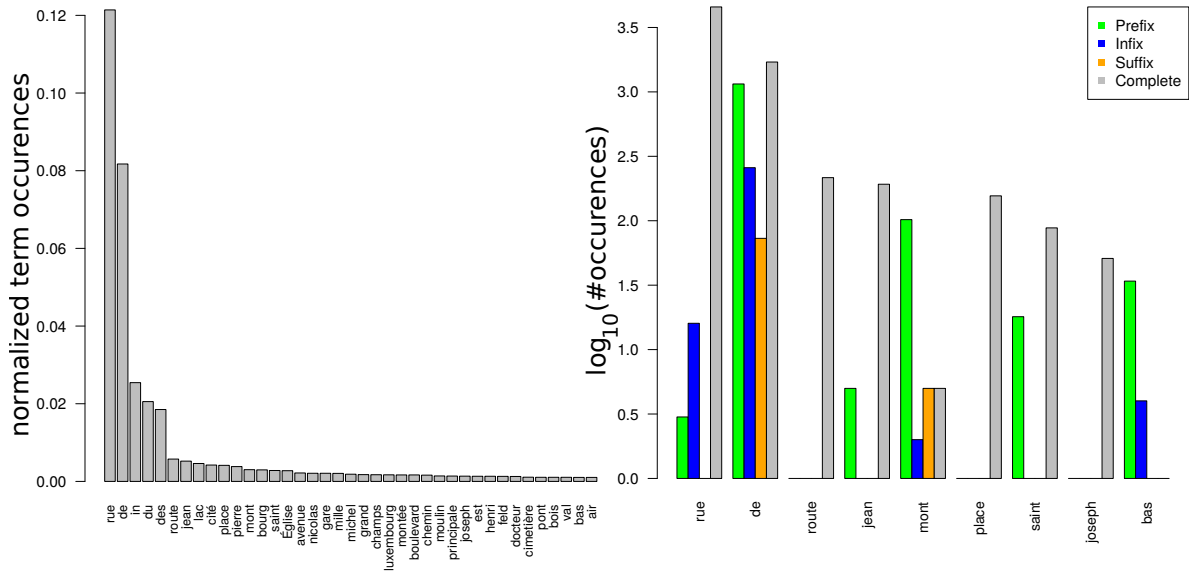

Figure S6. Left: Term occurences for selected street terms in Luxembourg normalized by the complete number of terms. Right: Positionwise occurences of selected street terms in Luxembourg. Prefix, infix and suffix refer to parts of one word whereas complete terms refers to the term having been separated from the street name by splitting the street name at space or dash sign.

4. Classification of street terms and names

In order to verify assigned countries of origin and classify terms that have not yet been classified, we apply the `fasttext` language identifier [2,3] that has been trained on Wikipedia data. For each street term, we extract the three most probable languages and corresponding countries, i.e. if the language is German, possible countries are Germany, Austria and Switzerland. For most of the street terms, it is possible to correctly verify the country of origin using the language identification tool. The results agreed to more than 80% with the results obtained by matching street terms of American countries to street terms in Europe.

Tables S4 and S5 show corresponding numbers for street specifics in North America that have been matched to European street specifics as well as numbers for correct verification of the country of origin with the language classifier. Tables S6 and S7 list the numbers for street generics.

5. Oldest street names can be found in rural areas - Example Wisconsin

The density of German street names is relatively high in Wisconsin (WI). Analysing those regions in more detail, one finds street names with German origin mostly in the rural areas outside of larger cities. One example are former German settlements in Jefferson County in Wisconsin. Table S3 shows numbers of inhabitants with German origin in Wisconsin and Jefferson County based on the US census data from 2014 (US census data: https://data.census.gov/cedsci/table?tid=ACSDP1Y2014.DP02&q=DP02&g=0400000US55_0500000US55055).

	Wisconsin State (WI)	Jefferson County, WI
Total Population	5,757,564	84,395 (+/- 22)
German Origin	2,333,895 (+/- 23,137)	42,080 (+/- 2,398)
German Origin in %	40.5	49.9

Table S3. Numbers of the US american census data of 2014 for the state of Wisconsin (WI) and Jefferson County in Wisconsin for the total population and inhabitants of German origin thereof.

6. The case of Yucatán in Mexico

Interestingly, the density of Spanish odonyms is expectionally low in the Mexican state of Yucatán, presumably reflecting the common use of Mayan languages in this region. This is shown in Figure S7 which shows a self-mapping from Mexican street specifics onto street names in Mexico. Thus, higher density values here show regions with street names occuring in large amounts all over the country. The region of Yucatán can be clearly distinguished in the plot and confirms that common Mexican street terms only match to a smaller extent to street names in Yucatán.

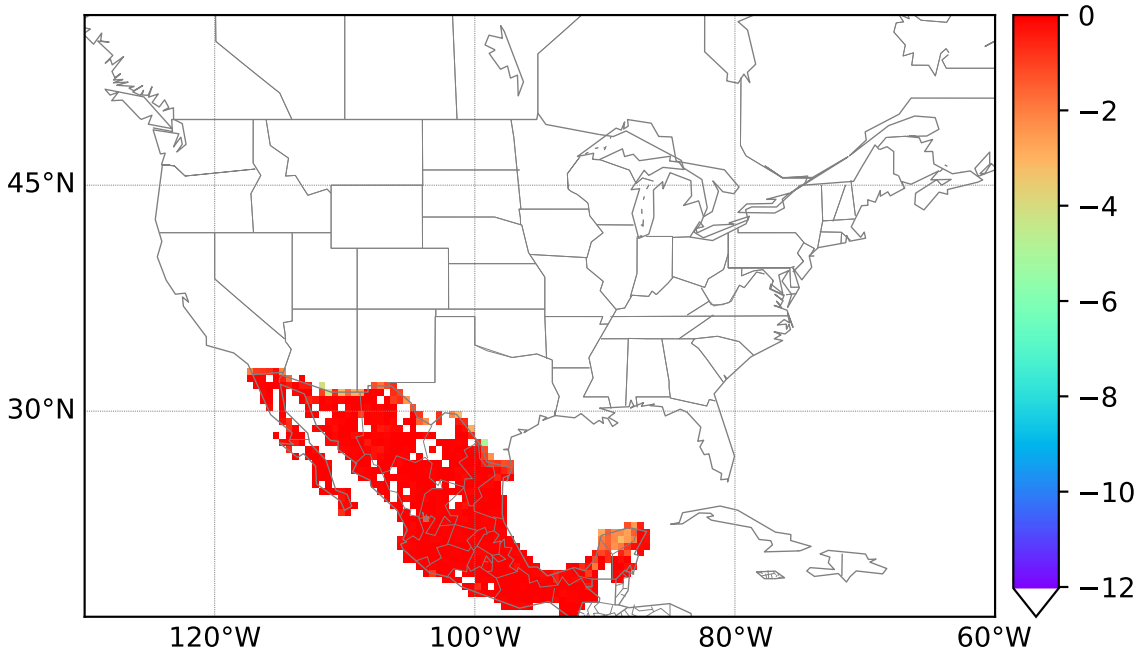

Figure S7. Heatmap of North America showing the distribution and density of Mexican street specifics over Mexican street names. Colors indicate the density of streets in this area, thus the logarithmic value of the number of streets normalized by all streets in the cell of the grid.

7. Additional Tables

1. Badariotti, D. Les noms de rue en géographie. Plaidoyer pour une recherche sur les odonymes/Street names, an argument for a geographic research. Annales de géographie. JSTOR, 2002, pp. 285–302.
2. Joulin, A.; Grave, E.; Bojanowski, P.; Mikolov, T. Bag of Tricks for Efficient Text Classification. *arXiv preprint arXiv:1607.01759* **2016**.
3. Joulin, A.; Grave, E.; Bojanowski, P.; Douze, M.; Jégou, H.; Mikolov, T. FastText.zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651* **2016**.

Country	Specifics (S)	S	S	S	S	S EU	S EU	S	S
	total	EU	EU(%)	non-EU	non-EU(%)	correct	correct(%)	NA	NA(%)
Alabama	8441	4856	0.575	3585	0.425	4233	0.872	35	0.007
Alaska	2696	1885	0.699	811	0.301	1691	0.897	18	0.009
Arizona	7978	4705	0.590	3273	0.410	4006	0.851	36	0.007
Arkansas	7713	4734	0.614	2979	0.386	4064	0.858	40	0.008
California	20329	10097	0.497	10232	0.503	8380	0.830	85	0.008
Colorado	6631	4004	0.604	2627	0.396	3484	0.870	36	0.009
Connecticut	4646	3341	0.719	1305	0.281	3046	0.912	21	0.006
Delaware	2440	1913	0.784	527	0.216	1746	0.913	12	0.006
Florida	14400	7282	0.506	7118	0.494	6138	0.843	60	0.008
Georgia	11648	6108	0.524	5540	0.476	5266	0.862	46	0.007
Hawaii	1679	438	0.261	1241	0.739	291	0.664	17	0.037
Idaho	4730	3106	0.657	1624	0.343	2712	0.873	25	0.008
Illinois	9662	5373	0.556	4289	0.444	4651	0.866	39	0.007
Indiana	8531	5003	0.586	3528	0.414	4402	0.880	29	0.006
Iowa	4604	2841	0.617	1763	0.383	2544	0.895	20	0.007
Kansas	4577	2750	0.601	1827	0.399	2457	0.893	22	0.008
Kentucky	9212	5245	0.569	3967	0.431	4542	0.866	39	0.007
Louisiana	8848	5515	0.623	3333	0.377	4693	0.851	48	0.009
Maine	4405	3060	0.695	1345	0.305	2754	0.900	19	0.006
Maryland	9016	5639	0.625	3377	0.375	4932	0.875	39	0.007
Massachusetts	7673	5037	0.656	2636	0.344	4485	0.890	31	0.006
Michigan	10914	6176	0.566	4738	0.434	5263	0.852	54	0.009
Minnesota	6506	3671	0.564	2835	0.436	3177	0.865	28	0.007
Mississippi	7140	4422	0.619	2718	0.381	3844	0.869	34	0.007
Missouri	8800	5121	0.582	3679	0.418	4433	0.866	45	0.009
Montana	3678	2504	0.681	1174	0.319	2191	0.875	20	0.008
Nebraska	3171	1942	0.612	1229	0.388	1753	0.903	13	0.006
Nevada	5915	4048	0.684	1867	0.316	3507	0.866	32	0.008
New Hampshire	3470	2591	0.747	879	0.253	2370	0.915	17	0.006
New Jersey	7938	5188	0.654	2750	0.346	4565	0.880	34	0.006
New Mexico	5742	3818	0.665	1924	0.335	3273	0.857	40	0.010
New York	10945	6337	0.579	4608	0.421	5491	0.866	50	0.008
North Carolina	14004	7117	0.508	6887	0.492	6023	0.846	67	0.009
North Dakota	1104	696	0.630	408	0.370	646	0.928	6	0.007
Ohio	11716	6516	0.556	5200	0.444	5650	0.867	42	0.006
Oklahoma	6575	3241	0.493	3334	0.507	2871	0.886	21	0.006
Oregon	6533	4088	0.626	2445	0.374	3571	0.874	27	0.006
Pennsylvania	11814	6497	0.550	5317	0.450	5579	0.859	42	0.007
Puerto-rico	2892	2226	0.770	666	0.230	1805	0.811	25	0.011
Rhode Island	2557	2028	0.793	529	0.207	1857	0.916	13	0.006
South Carolina	10559	6032	0.571	4527	0.429	5194	0.861	47	0.008
South Dakota	2440	1440	0.590	1000	0.410	1310	0.910	16	0.010
Tennessee	10695	5806	0.543	4889	0.457	5048	0.869	42	0.007
Texas	17990	8422	0.468	9568	0.532	7078	0.840	67	0.008
Utah	3547	2516	0.709	1031	0.291	2255	0.896	24	0.009
Vermont	2538	1963	0.773	575	0.227	1802	0.918	10	0.005
Virginia	11482	6559	0.571	4923	0.429	5708	0.870	49	0.007
Washington	6867	4034	0.587	2833	0.413	3504	0.869	28	0.007
West-virginia	4985	3368	0.676	1617	0.324	2986	0.887	24	0.007
Wisconsin	8017	4859	0.606	3158	0.394	4139	0.852	36	0.007
Wyoming	2031	1454	0.716	577	0.284	1307	0.899	14	0.009

Table S4. Table listing numbers for street specifics. For each US state, total number of street specifics, number of specifics classified by exact matches to European street specifics and the ones that could not be matched to a European street term are given. The second part shows how many of the North American street specifics were correctly identified with the language identifier. The last two columns show the total number and percentage of street specifics that could neither be matched to a European term nor be identified by the language identifier.

Country	Specifics (S)	S	S	S	S	S EU	S EU	S	S
	total	EU	EU(%)	non-EU	non-EU(%)	correct	correct(%)	NA	NA(%)
Alberta	3498	1858	0.531	1640	0.469	1686	0.907	10	0.005
British Columbia	6526	4315	0.661	2211	0.339	3787	0.878	35	0.008
Manitoba	1491	1224	0.821	267	0.179	1130	0.923	5	0.003
New Brunswick	2277	1893	0.831	384	0.169	1735	0.917	11	0.005
Newfoundl./Lab.	1395	1043	0.748	352	0.252	975	0.935	6	0.005
NW Territories	52	38	0.731	14	0.269	36	0.947	0	0.000
Nova Scotia	2438	1965	0.806	473	0.194	1803	0.918	11	0.005
Nunavut	4	0	0.000	4	1.000	0	0	0	0
Ontario	8943	5810	0.650	3133	0.350	5054	0.870	34	0.006
Pr. Edward Isl.	594	527	0.887	67	0.113	496	0.941	3	0.004
Quebec	7856	5343	0.680	2513	0.320	4568	0.855	30	0.005
Saskatchewan	1392	1099	0.790	293	0.210	1022	0.930	2	0.001
Yukon	105	87	0.829	18	0.171	79	0.908	0	0.000
Mexico	16233	6856	0.422	9377	0.578	5323	0.776	81	0.012

Table S5. Table listing numbers for street specifics. For each Canadian state and Mexico, total number of street specifics, number of specifics classified by exact matches to European street specifics and the ones that could not be matched to a European street term are given. The second part shows how many of the North American street specifics were correctly identified with the language identifier. The last two columns show the total number and percentage of street specifics that could neither be matched to a European term nor be identified by the language identifier.

Country	Generics (G)	G	G	G	G
	total	correct	correct(%)	NA	NA (%)
Alberta	45	36	0.800	2	0.044
British Columbia	72	59	0.819	7	0.097
Manitoba	42	37	0.881	4	0.095
New Brunswick	41	39	0.951	1	0.024
Newfoundland/Labrador	40	36	0.900	2	0.050
Northwest Territories	26	25	0.962	1	0.038
Nova Scotia	37	34	0.919	2	0.054
Nunavut	17	15	0.882	1	0.059
Ontario	59	50	0.847	5	0.085
Prince Edward Isl.	27	25	0.926	1	0.037
Quebec	55	45	0.818	6	0.109
Saskatchewan	36	32	0.889	3	0.083
Yukon	33	32	0.970	0	0.00
Mexico	87	72	0.828	7	0.080

Table S6. Table listing numbers for street generics in Canada and Mexico. For each state, number of street generics that have been correctly classified by the language identifier are listed. The last column shows the total number of street generics that could neither be matched to a European term nor be identified by the language identifier.

Country	Generics (G)	G	G	G	G
	total	correct	correct(%)	NA	NA (%)
Alabama	74	59	0.797	7	0.095
Alaska	53	44	0.830	4	0.075
Arizona	87	68	0.782	8	0.092
Arkansas	73	57	0.781	8	0.110
California	98	74	0.755	9	0.092
Colorado	77	63	0.818	7	0.091
Connecticut	43	37	0.860	4	0.093
Delaware	38	33	0.868	3	0.079
Florida	95	76	0.800	9	0.095
Georgia	86	68	0.791	8	0.093
Hawaii	33	23	0.697	7	0.212
Idaho	60	49	0.817	3	0.050
Illinois	80	65	0.812	8	0.100
Indiana	81	67	0.827	6	0.074
Iowa	56	49	0.875	2	0.036
Kansas	51	44	0.863	2	0.039
Kentucky	75	59	0.787	9	0.120
Louisiana	84	68	0.810	9	0.107
Maine	57	49	0.860	2	0.035
Maryland	60	51	0.850	4	0.067
Massachusetts	47	39	0.830	5	0.106
Michigan	89	71	0.798	11	0.124
Minnesota	68	53	0.779	5	0.074
Mississippi	58	52	0.897	2	0.034
Missouri	77	62	0.805	7	0.091
Montana	50	40	0.800	6	0.120
Nebraska	44	39	0.886	2	0.045
Nevada	58	49	0.845	3	0.052
New Hampshire	45	38	0.844	3	0.067
New Jersey	52	44	0.846	4	0.077
New Mexico	83	69	0.831	7	0.084
New York	79	62	0.785	10	0.127
North Carolina	89	70	0.787	11	0.124
North Dakota	36	31	0.861	2	0.056
Ohio	66	51	0.773	6	0.091
Oklahoma	64	56	0.875	2	0.031
Oregon	58	48	0.828	5	0.086
Pennsylvania	70	56	0.800	7	0.100
Puerto Rico	47	41	0.872	4	0.085
Rhode Island	26	22	0.846	3	0.115
South Carolina	80	65	0.812	9	0.113
South Dakota	40	34	0.850	1	0.025
Tennessee	78	62	0.795	10	0.128
Texas	93	74	0.796	9	0.097
Utah	51	44	0.863	2	0.039
Vermont	39	34	0.872	2	0.051
Virginia	74	57	0.770	10	0.135
Washington	83	67	0.807	6	0.072
West-virginia	65	52	0.800	7	0.108
Wisconsin	88	72	0.818	7	0.080
Wyoming	41	32	0.780	6	0.146

Table S7. Table listing numbers for street generics in the US. For each state, number of street generics that have been correctly classified by the language identifier are listed. The last column shows the total number of street generics that could neither be matched to a European term nor be identified by the language identifier.

Country	Specifics (S) total	S EU	S EU(%)	S non-EU	S non-EU(%)	S EU correct	S EU correct(%)	S NA	S NA(%)
Algeria	2238	490	0.219	1748	0.781	358	0.160	9	0.004
Angola	430	301	0.700	129	0.300	255	0.593	5	0.012
Benin	16	3	0.188	13	0.812	3	0.188	0	0.000
Botswana	63	23	0.365	40	0.635	22	0.349	1	0.016
Burkina Faso	83	45	0.542	38	0.458	33	0.398	2	0.024
Burundi	46	13	0.283	33	0.717	12	0.261	1	0.022
Cameroon	272	150	0.551	122	0.449	127	0.467	5	0.018
Canary Islands	3399	2547	0.749	852	0.251	2110	0.621	21	0.006
Cape Verde	96	86	0.896	10	0.104	77	0.802	2	0.021
Cent. African Rep.	3	0	0.000	3	1.000	0	0.000	0	0.000
Chad	96	14	0.146	82	0.854	5	0.052	2	0.021
Comores	10	0	0.000	10	1.000	0	0.000	1	0.100
Congo Brazzav.	207	106	0.512	101	0.488	90	0.435	3	0.014
DR Congo	1168	400	0.342	768	0.658	304	0.260	12	0.010
Djibouti	74	5	0.068	69	0.932	4	0.054	1	0.014
Egypt	6643	6	0.001	6637	0.999	5	0.001	1	0.000
Equatorial Guinea	24	3	0.125	21	0.875	2	0.083	0	0.000
Eritrea	33	1	0.030	32	0.970	0	0.000	1	0.030
Ethiopia	209	77	0.368	132	0.632	57	0.273	6	0.029
Gabun	1	0	0.000	1	1.000	0	0.000	0	0.000
Ghana	714	341	0.478	373	0.522	264	0.370	7	0.010
Guinea Bissau	1	0	0.000	1	1.000	0	0.000	0	0.000
Guinea	12	7	0.583	5	0.417	6	0.500	0	0.000
Ivory Coast	315	95	0.302	220	0.698	63	0.200	3	0.010
Kenya	942	344	0.365	598	0.635	282	0.299	8	0.008
Lesotho	23	5	0.217	18	0.783	5	0.217	1	0.043
Liberia	8	3	0.375	5	0.625	3	0.375	0	0.000
Libya	412	25	0.061	387	0.939	18	0.044	2	0.005
Madagascar	90	40	0.444	50	0.556	36	0.400	0	0.000
Malawi	9	1	0.111	8	0.889	1	0.111	1	0.111
Mali	46	20	0.435	26	0.565	17	0.370	1	0.022
Mauritania	138	24	0.174	114	0.826	16	0.116	2	0.014
Mauritius	380	289	0.761	91	0.239	258	0.679	4	0.011
Morocco	2763	773	0.280	1990	0.720	555	0.201	20	0.007
Mozambique	135	103	0.763	32	0.237	85	0.630	2	0.015
Namibia	490	308	0.629	182	0.371	270	0.551	5	0.010
Nigeria	1981	746	0.377	1235	0.623	552	0.279	27	0.014
Niger	58	33	0.569	25	0.431	21	0.362	0	0.000
Rwanda	13	8	0.615	5	0.385	7	0.538	0	0.000
Senegal/Gambia	254	137	0.539	117	0.461	102	0.402	3	0.012
Seychelles	2	0	0.000	2	1.000	0	0.000	0	0.000
Sierra Leone	378	261	0.690	117	0.310	243	0.643	3	0.008
Somalia	110	19	0.173	91	0.827	11	0.100	2	0.018
South Africa	12988	6391	0.492	6597	0.508	5140	0.396	66	0.005
South Sudan	63	7	0.111	56	0.889	4	0.063	1	0.016
Sudan	147	12	0.082	135	0.918	7	0.048	2	0.014
Swaziland	3	0	0.000	3	1.000	0	0.000	0	0.000
Tanzania	441	146	0.331	295	0.669	107	0.243	5	0.011
Togo	69	38	0.551	31	0.449	31	0.449	1	0.014
Tunisia	1474	100	0.068	1374	0.932	72	0.049	3	0.002
Uganda	380	172	0.453	208	0.547	145	0.382	5	0.013
Zambia	99	42	0.424	57	0.576	38	0.384	2	0.020
Zimbabwe	497	315	0.634	182	0.366	277	0.557	3	0.006

Table S8. Table listing numbers for street specifics in Africa. For each country total number of street specifics, number of specifics classified by exact matches to European street specifics and the ones that could not be matched to a European street term are given. The second part shows how many of the African street specifics were correctly identified with the language identifier. The last two columns show the total number and percentage of street specifics that could neither be matched to a European term nor be identified by the language identifier.

Country	Generics (G)	G	G	G	G
	total	correct	correct(%)	NA	NA (%)
Algeria	49	44	0.898	2	0.041
Angola	34	32	0.941	1	0.029
Benin	36	30	0.833	1	0.028
Botswana	25	23	0.920	0	0.000
Burkina Faso	23	19	0.826	4	0.174
Burundi	25	22	0.880	1	0.040
Cameroon	33	28	0.848	2	0.061
Canary Islands	32	30	0.938	1	0.031
Cape Verde	31	28	0.903	0	0.000
Cent. African Rep.	15	11	0.733	1	0.067
Chad	20	16	0.800	2	0.100
Comores	14	11	0.786	1	0.071
Congo Brazzav.	22	18	0.818	3	0.136
DR Congo	33	31	0.939	1	0.030
Djibouti	13	11	0.846	1	0.077
Egypt	6	6	1.000	0	0.000
Equatorial Guinea	20	17	0.850	1	0.050
Eritrea	5	5	1.000	0	0.000
Ethiopia	22	19	0.864	1	0.045
Gabun	0	0	0.000	0	0.000
Ghana	44	39	0.886	3	0.068
Guinea Bissau	10	9	0.900	0	0.000
Guinea	36	30	0.833	3	0.083
Ivory Coast	25	20	0.800	2	0.080
Kenya	41	32	0.780	3	0.073
Lesotho	16	15	0.938	0	0.000
Liberia	22	21	0.955	0	0.000
Libya	14	13	0.929	0	0.000
Madagascar	28	24	0.857	1	0.036
Malawi	29	24	0.828	1	0.034
Mali	18	17	0.944	1	0.056
Mauritania	20	15	0.750	2	0.100
Mauritius	35	32	0.914	1	0.029
Morocco	40	38	0.950	1	0.025
Mozambique	21	20	0.952	0	0.000
Namibia	34	29	0.853	1	0.029
Nigeria	49	41	0.837	4	0.082
Niger	31	21	0.677	2	0.065
Rwanda	14	10	0.714	1	0.071
Senegal/Gambia	48	32	0.667	6	0.125
Seychelles	10	9	0.900	1	0.100
Sierra Leone	19	15	0.789	1	0.053
Somalia	20	13	0.650	2	0.100
South Africa	92	75	0.815	7	0.076
South Sudan	18	14	0.778	0	0.000
Sudan	9	8	0.889	0	0.000
Swaziland	11	11	1.000	0	0.000
Tanzania	27	24	0.889	1	0.037
Togo	17	14	0.824	2	0.118
Tunisia	17	15	0.882	1	0.059
Uganda	26	24	0.923	0	0.000
Zambia	27	26	0.963	0	0.000
Zimbabwe	29	22	0.759	3	0.103

Table S9. Table listing numbers for street generics in Africa. For each state, number of street generics that have been correctly classified by the language identifier are listed. The last column shows the total number of street generics that could neither be matched to a European term nor be identified by the language identifier.