# Identifying Patterns in Geospatial Natural Language

Dr Kristin Stock, Massey University, Auckland, New Zealand
k.stock@massey.ac.nz

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

**1. side of the (1/136); the side of (24/46)**

$C1$ = the side of the [gf: river, road, Greet, main road, main lane, main thoroughfare, main street, main highway, creek, chancel, church, square, town, station, highway, stream, River Trent, car park]

$C2$ = the [dir adj: south, east, west, north, eastern, western, northern, southern, north west] side of the [gf: as for c1]

$C3$ = the [s deictic: other, same, opposite, far, near, left, right, left hand, right hand] side of the [gf: as for c1]

$C4$ = [enum: either, each] side of the [gf: as for c1]

$C5$ = [quant: one] side of the [gf: as for c1]

$C6$ = the [gf: platform] side of the [gf: as for c1]

$C7$ = [s prep: by, on, along, around, at, down] the side of the [gf: as for c1]

$C8$ = [s prep: on] [enum: each] side of the [gf: as for c1]

$C9$ = [s prep: on, along] the [dir adj: south, north, left, right, west, right hand, left hand] side of the [gf: as for c1]

$C10$ = [s prep: on, to] the [s deictic: other, far, opposite, same] side of the [gf: as for c1]

$C11$ = [s prep: on] the [gf: platform] side of the [gf: as for c1]

$C12$ = [lmv: lies, park, stands, is] [s prep: on, by] the side of the [gf: as for c1]

$C13$ = [gf: as for c1] [s prep: by] the side of the [gf: as for c1]

$C14$ = [gf: as for c1] [s prep: by] the side of where [det: a] [gf: as for c1] is.

$C15$ = [gf: as for c1] [s prep: on] the [dir adj: south] side of the [gf: as for c1]

$C16$ = [gf: as for c1] [s prep: on] the [dir adj: south] side of the [gf: as for c1]

$C17$ = [gf: as for c1] [s prep: on] the [gf: platform] side of the [gf: as for c1]

$C18$ = a [adj: good] [gf: as for c1] [s prep: by] the side of the [gf: as for c1]

Overlaps common, treated together

**General constructions for 'side of the':**

$$Cg1 = \text{gf lmv s prep} \begin{bmatrix} \textit{the dir adj} \\ \textit{the s deictic} \\ \textit{one} \\ \textit{enum} \end{bmatrix} \text{side of the gf}$$

**General construction for spatial language:**

$$[\textit{div noun of}] \begin{bmatrix} \text{det gf} \\ \text{pn} \\ \text{det lmn} \end{bmatrix} \textit{lmv s adv} \text{ s prep } [\textit{det div noun of}] \begin{bmatrix} \text{det gf} \\ \text{pn} \\ \text{det lmn} \end{bmatrix}$$

## Why?

As the first step in interpreting geospatial natural language for:

- harvesting of location from natural language documents
- identifying locations from verbal descriptions (+ speech recognition);
- natural language spatial querying.

Most previous work has focussed on prepositions [6-9], and particular aspects of structure [10-14].

## How?

### Background to the Method

- In the cognitive linguistics tradition (in contrast to generative linguistics a la Chomsky) [1,2].
- Grammar (form) cannot be separated from meaning.
- View language in terms of lexical phrases (i.e. groups of fixed words), combined with members of grammatical categories with which they commonly appear [3-5].

words + slots = constructions

### Method

- We tried to automate identification of constructions with cluster analysis and systematic wildcard discovery, but neither were successful.
- Defined constructions manually, starting with trigrams.
- Ground up from trigrams, searched with concordancer and identified patterns.

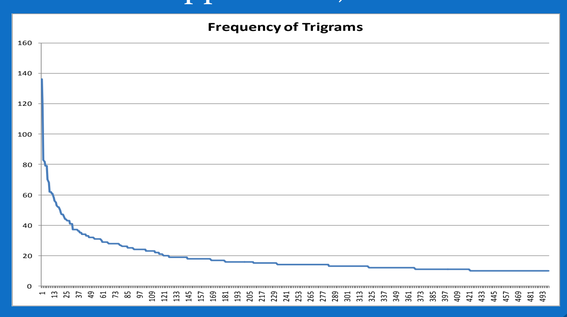
Let the data speak for itself

## What?

- Analysed content of Nottingham Corpus of Geospatial Language (NCGL) (10k spatial expressions)

| rank | freq | trigram |
|---|---|---|
| 1 | 136 | side of the |
| 3 | 83 | part of the |
| 7 | 70 | the top of |
| 8 | 68 | from London to |
| 11 | 61 | end of the |
| 13 | 58 | in the north |
| 14 | 56 | next to the |
| 15 | 55 | to the south |
| 16 | 53 | to the right |
| 19 | 52 | to the north |
| 20 | 50 | in the south |
| 21 | 48 | the heart of |
| 22 | 47 | of the city |
| 23 | 47 | the end of |
| 24 | 46 | the side of |

- Studied 15 most frequently occurring trigrams.
- Trigrams contained the largest proportion of expressions that were spatial in nature (38 out of 50) than shorter and longer ngrams.
- Ensures common expressions not missed (as in preposition focussed approaches).




## What Next?

- Calculation of measures of the degree to which a slot attracts a particular lexeme in corpus text.
- Investigation of interactions and correlations in slot values (co-varying collexemes), and their semantic coherence.
- Study of most suitable level of generalisation of constructions for the automated interpretation task.
- Ongoing research using the most general construction.

## References

1. Evans, V. and Green, M.: Cognitive Linguistics: An Introduction. Edinburgh: Edinburgh University Press (2006).
2. Chomsky, Noam: Three models for the description of language. IRE Transactions on Information Theory 2 (3), 113–124 (1956).
3. Goldberg, A.: Constructions at Work: The Nature of Generalization in Language. Oxford: Oxford University Press (2006).
4. Hunston, S. and Francis, G.: Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English. Amsterdam: John Benjamins Publishing Company (2000).
5. Langacker, Ronald.: Foundations of Cognitive Grammar. 2 vols. Stanford: Stanford University Press (1987, 1991).
6. Coventry, K.R. and Garrod, S.C.: Saying, Seeing and Acting: The Psychological Semantics of Spatial Prepositions. East Sussex: Psychology Press (2004).
7. Herskovits, A.: Language and Spatial Cognition: An interdisciplinary study of the prepositions in English. Cambridge: Cambridge University Press (1986).
8. Talmy, L.: Toward a Cognitive Semantics: Volume 1 Concept Structuring Systems. Cambridge, Massachusetts: MIT Press (2000).
9. Tyler, A. and Evans, V.: The Semantics of English Prepositions. Cambridge: Cambridge University Press (2003).
10. Klippel, A., Xu, S., Li, R., & Yang, J.: Spatial event language across domains. 2nd Workshop on Computational models of Spatial Language Interpretation (CosLI-2) (2011).
11. Xu, S., Klippel, A., MacEachren, A., Mitra, P., Turton, I., Zhang, X., & Jaiswal, A.: Exploring regional variation in spatial language - a case study on spatial orientation with spatially stratified web-sampled documents. In Spatial Cognition Conference – Poster Session, Mt. Hood/Portland Oregon (2010).
12. Wang, X., Matsakis, P., Trick, L., Nonnecke, B. and Veltman, M. A.: A study on how humans describe relative positions of image objects. Proceedings of 13th Int. Symposium on Spatial Data Handling, Springer Publications, 1-18 (2008).
13. Blaylock, N., Swain, B. and Allen, J.: Mining Geospatial Path Data from Natural Language Descriptions. In ACM QUeST'09, November 3, 2009, Seattle, WA, USA (2009).
14. Tomai, E. and Kavouras, M.: Where the city sits? Revealing Geospatial Semantics in Text Descriptions. 7th AGILE Conference on Geographic Information Science, Heraklion, Greece, 189-194 (2004).

| Code | Meaning |
|---|---|
| Spatial Categories (defined for the purposes of this work) | |
| gf | Geographic feature type, indicating a category of geographic features, rather than a specific one. E.g. library, building, city, mountain. |
| pn | Place name, official or vernacular, name by which a location is known, referring to a specific geographic feature. E.g. Trent River |
| lmn | Location and movement noun. A noun that is relevant to location or movement, and may also be considered an abstract geographic feature type. E.g. journey, route. |
| div noun | Division noun, indicating some component of a larger object. E.g. part, corner, centre, side, quadrant, edge. |
| lsgf | Location specified geographic feature, being a particular subset of geographic features, specified by location. E.g. any Dutch station, major cities in the south of France. |
| ll | Informal location, referring to a specific location rather than a geographic feature type, usually vague in nature so without a definitive place name. E.g the north, downtown. |
| s prep | Spatial preposition (also including other forms of adposition, but in English most are prepositions). E.g. in, near, on, from. |
| s deictic | Spatial deictic, describing location relative to the speaker or some other object, often classed as determiners, but in this work, they are given a special class because they are important from the point of view of interpreting spatial location. E.g. here, there, other. |
| lmv | Location and movement verb, being a verb indicating either location or movement. E.g. passes, is located. |
| s adv | Spatial adverb, referencing an lmv. Included here if they have a spatially specific sense. E.g. right (next to), just (north of)). |
| dir adj | Direction adjective, indicating direction in absolute, not relative terms (that is, relative to the earth), including compass directions and gravity related directions. E.g. north, northern, northerly, top. Sometimes the reference noun is implicit. |
| div adj | Division adjective indicating some division of an object. A superclass of dir adj. E.g. central, main. |
| s adj | General spatial adjective (indicating size, shape, direction, etc.) not already included in more specific categories. E.g. small, sharp. |
| mo | Moving object, often used as reference in route descriptions. E.g. passengers, you. |
| veh | Moving vehicle, often used as a reference in route descriptions. E.g. car, train. |
| s unit | spatial unit - distance, area, etc. E.g. miles, minutes, square miles. |
| General Grammatical Categories (from conventional grammar) | |
| spec | Specifier, referring to determiners that are not covered by indefinite and definite articles, quantifiers or enumerators. E.g. another, other, same, when used non spatially. |
| enum | Enumerators and indicators of alternatives. E.g. each, any, every, either, neither (technically a type of determiner, but included here because of their spatial importance. |
| quant | Quantifier, indicator of quantity. E.g. numbers, little, few, most. |
| adv | Adverb, describing a verb. |
| time period | A period of time. E.g. day |
| det | Determiner, referring only to definite and indefinite articles. E.g. this, your. |
| v | Verb. |