

# Designing an Algorithm for Generating Named Spatial References

Rodrigo de Oliveira, Yaji Sripada, Ehud Reiter

Department of Computing Science  
University of Aberdeen, Scotland

{rodrigodeoliveira,yaji.sripada,e.reiter}@abdn.ac.uk

## Abstract

We describe an initial version of an algorithm for generating named references to locations of geographic scale. We base the algorithm design on evidence from corpora and experiments, which show that named entity usage is extremely frequent, even in less obvious scenes, and that names are normally used as the first focus on a global region. The current algorithm normally selects the Frames of Reference that humans also select, but it needs improvement to mix frames via a mereological mechanism.

## 1 Introduction

Geospatial data of public interest such as weather prediction data and river level data are increasingly made publicly available, e.g. DataPoint from the Met Office in the UK, River Level data from SEPA in Scotland and Global Forecast system data from NOAA in the US.

We are interested in developing computational techniques for expressing the information content extracted from these datasets in natural language using data-to-text natural language generation (Reiter et al., 2005) techniques. For example, from precipitation prediction data corresponding to several locations across Scotland, we are developing techniques to automatically generate the statement *Heavy rain likely to fall as snow on higher ground in the northeast of Scotland*.

An important subtask here is to automatically generate the spatial referring expression (SRE) *higher ground in the northeast of Scotland* to linguistically express the location of the snowing event found in the precipitation prediction data. This paper presents corpus analysis and experimental studies to guide the design of an algorithm for SRE generation. Studies of human written

SREs (Turner et al., 2010) show a broad range of descriptors such as *north*, *east*, *coastal*, *inland*, *urban*, and *rural* to specify locations. Descriptors belong to one of many perspectives on the scene, or *Frames of Reference* (Levinson, 2003) or FoR for short, such as direction, coastal proximity, population density and altitude.

Our own corpus studies (Section 2) show that geographic names are the dominant descriptors in weather forecast texts, route descriptions and river level forecast reports. Our experiment to empirically understand the extent of usage of geographical names in SREs (Section 3) also shows that names are the most used descriptors, as well as the FoR that sets the first focus on a region. Using this empirical knowledge we propose an initial version of an algorithm (Section 4) that automatically generates SREs using names as well as other descriptors.

## 2 Corpus Analysis

The first stab at the problem was a corpus analysis study. We gathered a total of 36 texts in 3 domains (route descriptions, weather forecasts, river forecasts), in 3 languages (English, Portuguese and Spanish), for 3 target audiences (general public, fishing enthusiasts, kayaking enthusiasts).

We define an SRE as an adverbial (*inland*) or a noun phrase (*the north*), which ties non-spatial information to one location. Only sentences that contained at least 1 SRE were included in the corpus. For each SRE at least 1 FoR was annotated. Below are 3 examples from the corpus, all originally in English, where SREs are underlined:

1. (a) The Red River is slowly rising (b) from Emerson (c) downstream to Winnipeg.
2. (a) From the north (b) the A1 (c) and M1 link (d) to the A14 dual carriageway (e) straight to the city.

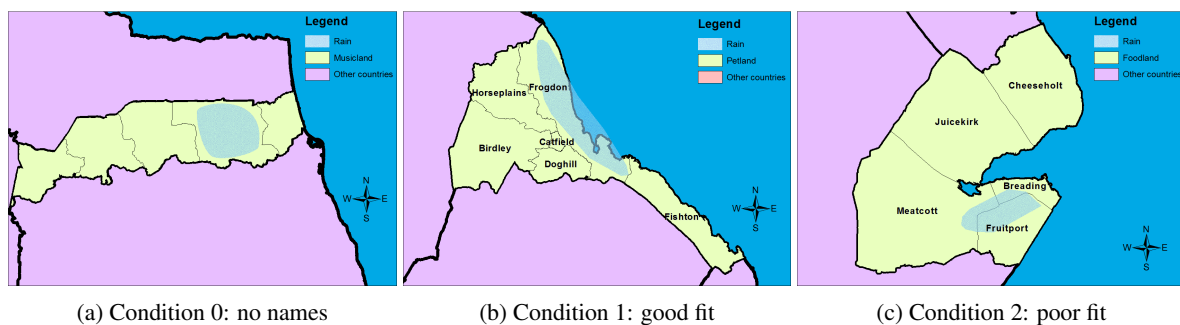


Figure 1: Stimuli in the pilot experiment.

3. (a) *Dry with sunny spells on Saturday and Sunday these mainly inland* (b) *with Aberdeenshire coast becoming cloudy.*

Sentence 1 was extracted from a river level report for Manitoba, Canada, which seems to be aimed at the general public. In the instance, we identified 3 SREs, all of which using named entities as FoR. Sentence 2 is a route description for drivers to reach Cambridge, England, so it is also aimed at the general public. 2a uses a cardinal direction as FoR, 2e uses the entity’s type, while 2b-c use named entities. Sentence 3 also seems to be intended for the general public; it is extracted from a weather forecast report for Aberdeenshire, Scotland. Both SREs use coastal proximity as FoR, while 3b also includes a named entity.

In total the corpus yielded 556 SREs, out of which 318 (57%) use named entities, either in isolation or combined with other FoR. It is important to remember that another 7 FoR appear in the corpus – cardinal direction, coastal proximity, population density, type, motion sequence, river segment and size – which means that names account for more than half of a total of 8 choices.

With the corpus in place, it became clear that names do not compete with other FoR in a balanced manner. Because of this expressive imbalance, we were lead to the suspicion that humans choose to refer to geographic regions by their names using a different strategy than when choosing other FoR. We suspect people may be more precise when they use FoR such as cardinal direction or coastal proximity, but they can be very imprecise when using names. This suspicion lead us to our first hypothesis:

**Hypothesis 1: People mostly use named entities to refer to locations of geographic scale, even if the fit between the named location and the located entity/event is poor.**

By the above hypothesis we mean that named entities are used as spatial references also in situations where using a name as reference is not so obvious. For instance, if the named location only covers a small portion of a located entity/event, or if the located entity/event is much smaller than the named location, we suspect that most people still use the named location as reference, hence the high frequency of named entities in the corpus.

### 3 Experiment

Even though the corpus analysis returned fruitful insights, we remained with a major shortfall to design a computational algorithm for an NLG system. We expect such an algorithm to be used in data-to-text systems – i.e. systems that write text from information stored in data bases – so a data-and-text parallel corpus is more suitable to inform us what our SREG algorithm must consider. Thus we resorted to experiments with human participants to collect spatial expressions, while having full access to the data underlying the text.

#### 3.1 Pilot

To test hypothesis 1, we designed a pilot experiment (see Figure 1), where we showed 3 different maps (conditions) of fictitious countries to 14 human participants and asked them to describe where on those countries they could see a patch of rain. Both the no-name condition and the good-fit condition placed the rain patch very neatly on one specific region of the country, with the difference that the no-name condition did not have any names for the regions and the good-fit condition did. In the poor-fit condition, named regions were also present but the patch covered only a small portion of several regions. Participants were split into balanced groups and each group saw maps in a different order. The rationale behind the no-name con-

Condition	SRE Type						$\Sigma$
	name-only	other-only	name-1st	name-2nd	both-1st	none	
no-name	2	86	0	0	0	5	93
good-fit	43	3	41	5	1	0	93
poor-fit	12	2	72	5	2	0	93
$\Sigma$	57	91	113	10	3	5	279

Table 1: Experimental results showing types of SREs per condition. SREs can contain only names, only other FoR, none, or mix names with other FoR. When mixed, names can be the first or second focus, or both types can be first focus.

dition is to certify that people resort to other FoR when names are not available.

Curiously, names were not as dominant in the pilot experiment as they are in the corpus. The FoR used by all participants were names, cardinal direction (north, south, etc.) and some proximity (coast, border, etc.). In the vast majority of responses (94%), people used multiple FoR to refer to the location of the rain patch, which we believe helped balance the usage of FoR across responses. Names were used in 79% of responses in the good-fit condition – proximity 86% and direction 50% – and in the poor-fit condition names were used in 64% of responses – direction 79% and proximity 57%.

Even though names were not dominant, people still used names in most cases, even in a scenario where using a name was not so obvious (the poor-fit condition), speaking in favour of hypothesis 1. After results from the pilot experiment, we could see that most responses use a *first focus frame* (of reference) and a *a second focus frame*. Take the SRE *coastal areas of Frogdon* for instance. *Frogdon* (a name) indicates the first focus area, while *coastal areas* (proximity) sets a second focus on one particular portion of the first focus area. We suspect that most first-focus areas are named regions, which leads us to a second hypothesis:

**Hypothesis 2: When mixing named entities with other FoR, people use named entities mostly for first-focus areas and other FoR for second-focus.**

### 3.2 The main experiment

The above results were not formally verified with statistical tests because we believe our sample of 14 participants was not representative. In order to test our hypotheses with more statistical vigor, we ran a slightly modified version of the pilot experiment on Amazon Mechanical Turk, where 93

participants successfully completed the task. The difference from the pilot is that, instead of having only 1 image per condition, we prepared 2 images for the control condition, 4 for the good-fit condition and 4 for the poor-fit condition. As in the pilot, participants saw only 1 image per condition, so the system randomly chose a single map to display in each condition. With multiple images available in the experiment, we reduced the level of specificity between stimuli and responses.

Responses varied from single clauses to sentences containing 2 or more clauses, or even full paragraphs containing 2 or more sentences. We considered the entire response as 1 SRE, but since we are now interested in names versus non-names, we combined all other FoR that are not named entities. We marked SREs with 1 of the 5 following annotations:

**name-only** If the SRE only contained named entities as spatial references: *It will rain in Doghill.*

**other-only** If the SRE only contained non-named-entities as spatial references: *Rain can be expected in the south-most region of Musicland<sup>1</sup>. There is no other chance of rain.*

**name-2nd** If both names and other FoR were used, but named entities were used as second focus: *Throughout the far south of Foodland, going through Meatcott and Fruitport, rain is to be expected.*<sup>2</sup>

<sup>1</sup>Since we are interested in knowing *where in Musicland it is raining*, the descriptor *Musicland* is tautological, thus not counted as chosen descriptor to locate the event. Therefore, named entities are also ignored as contributing FoR in this description.

<sup>2</sup>Here it is not a lexical item that informs us that the named entities are used as second focus. It is the fact that the a direction (*the far south*) is used to select a larger sub-region of the global region (the country), within which the named regions (*Meatcott and Fruitport*) exist.

**name-1st** If both names and other FoR were used, but named entities were used as first focus.

**both-1st** If names and other FoR don't compete for first focus, but remain on the same level, so the resulting subregion is a union of multiple sub-regions. For example: *northwestern Fruitport... southwest of Breeding... eastern part of Meacott... not in the far northeast or southeast*. Fruitport, Breeding and Meacott are named regions but far north-east and south-east are directions. None is a part of the other, so the named areas and *not far north-east and south-east* complement each other at the same focus level.

**none** If no FoR, but only vague descriptors were used.

Finally we counted all possible combinations of FoR usage and aligned those with experimental conditions, as displayed in Table 1. The first intriguing observation is that 5 responses did not use any FoR, according to our annotation. 2 of them used only a quantifier (*much, most*), 2 only the name of the country (*Musicland*), and 1 used both (*some parts of Musicland*). Using only the name of the country does not successfully complete the task, because it does not answer the question “where *in the country* will it rain?”. Quantifiers were also not annotated as other FoR because they are extremely vague. We were aiming at FoR that help a hearer more precisely identify referenced locations.

Even more interesting, 2 SREs created named entities in the no-name condition, i.e. where no name was available as per task. One participant decided to name an unnamed subregion of Musicland as *Drum County* and referred to it ‘by its name’. Although odd, this suggests how people strongly feel the necessity for named entities when describing geographies. This is very similar to another response in the pilot experiment, where the participant described one unnamed subregion as *the penultimate state before reaching the coast*, and later stated in the comments that names should be on the map.

Hypothesis 1 states that people use names with a high frequency in any condition where names are available. If we exclude the no-name condition from the count, this hypothesis is supported with 97% (90/93) of name usage in the good-fit

condition and 98% (91/93) in the poor-fit condition. We did *not* observe a significant difference in name usage between good-fit and poor-fit conditions,  $\chi^2(1, N=186) = 0.21, p = .65$ .

Hypothesis 2 was also supported, again excluding the no-name condition. People very often (113/126 or 90%) use names as the *first-focus area* and other FoR as the *second focus-area*.

After testing the above hypotheses, we observed the same phenomenon as identified by Turner and colleagues (2010): that people resort to other FoR more often when the fit between (rain) patch and region is poorer. In the good-fit condition 54% (50/93) of responses used other FoR, while 87% (81/93) of poor-fit responses contain other FoR. This means that there is a *significant* need for other FoR when moving from a good-fit to a poor-fit scenario,  $\chi^2(1, N=186) = 26.18, p < .001$ .

### 3.3 Preliminary conclusions

To date this project has shown evidence that:

- Humans use several FoR when referring to geographical locations.
- Regardless of scenario, named entities are almost always used.
- Named areas mostly function as a first focus area, wherein a descriptor of a second FoR can still be selected.

## 4 Algorithm

We used the knowledge described above to inform an algorithm that selects Frames of Reference. The procedure is basically the ContentSelector algorithm of the RoadSafe project (Turner, 2009), which looks at an event that takes place in a geography and selects one or more frames out of an array of frames. The input to the algorithm, as for many geographic information systems, is a set of points with latitude-longitude coordinates and some other value denoting the status of the point in some event. In Turner's sense, a Frame of Reference is a set of descriptors, and a descriptor is a non-overlapping partition of a geographic region where each descriptor can be used to refer to a specific partition. The frame contains all points of the dataset, but each descriptor encompasses a particular subset of points.

For instance, take the US as our global geography, which contains several thousands of points.

The Frame of Reference StateNames contains 50 descriptors, one for each US state, so each descriptor contains a couple of hundreds of points. Altogether StateNames contains all points that form the US. Another frame could be CoastalProximity, which is composed of only 2 descriptors, Coastal and Inland, where most points belong to the Inland descriptor and the rest to Coastal. Note that in this example, all points that belong to the descriptor Kansas of the frame StateNames also belong to the descriptor Inland of the frame CoastalProximity, but such overlaps are not always true. Out of the points that form the descriptor Texas, some belong to Inland and others to Coastal.

Following the US example, the high-level goal of the algorithm is to select one or more descriptors that *best* locate a target subset of all the points in the US. For instance if our dataset contains a binary variable for “rain” for each point, and we are interested in describing the location of the “raining points” – or simply answering the question “where in the US is it raining?” – the algorithm’s task is to return a set of descriptors that encompasses the majority of points with rain=true values. If the result is {Colorado, Coast}, the NLG system where the algorithm lives should be able to produce the sentence “it will rain on the Coast and in Colorado”.

Turner describes the ContentSelection algorithm in detail (p. 122), so below we highlight its main steps:

1. Take as input a set of points representing an event, along with meta-data for Frames of Reference.
2. Count the density of target points for each descriptor of each frame.
3. Remove a frame if all its descriptors have non-zero densities.
4. Of the remaining frames, rank them by a pre-defined preference order.
5. Use the first frame with non-zero densities.
6. Try adding each subsequent frame, if this reduces the number of false positives.
7. Use the descriptors with non-zero densities of the chosen frames.

We take the algorithm and include, first of all, a NamedAreas frame. This however is currently

done in the same fashion as all other frames in the RoadSafe project. The true conceptual modification to the original algorithm was the threshold of density (step 3). RoadSafe fixes this value at 0, which means that if all descriptors of a Frame of Reference have at least 1 target point, then this frame cannot be chosen. We suspect that humans are more lenient when computing density. We believe that humans can choose frames where all descriptors have non-zero densities, by focussing on descriptors with high densities and ignoring descriptors with low (yet non-zero) densities. Therefore our version of the algorithm selects a descriptor as candidate if it reaches a density threshold, and it ignores a FoR if all its descriptors are candidates.

#### 4.1 A small-scaled quantitative evaluation

To test how the algorithm currently performs, we ran it using 7 weather forecast datasets provided by the UK’s meteorology agency: MetOffice. The data contained numerical predictions for a region in the UK (Grampian), and each dataset also accompanies a textual summary, against which we used to compare our algorithm. We chose DICE to evaluate how comparable each output was. This metrics has been widely used by the Referring Expression community (Gatt et al., 2008; Belz and Gatt, 2008). The results are displayed in Table 2.

To compare MetOffice’s FoR choices with those by our algorithm, we ran it using 6 different density thresholds: 0.0, 0.2, 0.4, 0.6, 0.8 and 1.0. A density threshold is in this sense the minimum event density a descriptor can have to be accepted as a candidate. If you recall the explanation of the algorithm above, a Frame of Reference is rejected if all its descriptors are rejected, but equally if all its descriptors cannot be rejected. For example, it only makes sense to select Inland as a descriptor if Coastal is not a candidate; if both Inland and Coastal are equally valid, then we can say the event (e.g. rain) is taking place in the entire region, as far as coastal proximity is concerned. As explained above, the fixed density threshold in the original algorithm was 0.0, which means that 1 single point was enough to make a descriptor invalid. By running the algorithm with different density thresholds, we are able to have an idea of some optimal threshold, where non-zero-density descriptors still get rejected.

From this initial evaluation, we could verify

Dataset	MO	BL	D	DT=.0	D	DT=.2	D	DT=.4	D	DT=.6	D	DT=.8	D	DT=1	D
May 21	nam, cst	nam	0.7	*	0	dir	0	nam	0.7	nam	0.7	nam	0.7	nam	0.7
May 25		nam	0	nam	0	nam	0	-	1	-	1	-	1	-	1
May 27	nam	nam	1	*	0	*	0	dir	0	dir	0	nam, dir	0.7	nam	1
May 28	nam	nam	1	*	0	*	0	nam	1	nam, dir	0.7	-	0	-	0
Jun 01	nam, dir	nam	0.7	*	0	dir	0.7	nam, dir	1	nam	0.7	nam	0.7	nam	0.7
Jun 02	nam	nam	1	nam	1	nam	1	nam	1	nam, dir	0.7	dir	0	-	0
Jun 04	dir	nam	0	nam	0	nam, dir	0.7	nam	0	nam, dir	0.7	-	0	-	0
Average			0.6		0.1		0.3		0.7		0.6		0.4		0.5

Table 2: Comparison of 1st-focus FoR choice between MetOffice texts and the algorithm running with different density thresholds. Assigning 2 (or more) 1st-focus FoR to a dataset is very similar to assigning “both-1st” to experimental responses. Please refer to Section 3.2 for a more detailed discussion on multiple 1st-focus FoR. Abbreviations: nam = NamedArea; dir = Directions; cst = CoastalProximity; MO = MetOffice; BL = Baseline; DT = Density Threshold; D = DICE score; \* = all descriptors reach the threshold, so no FoR is discriminative enough to be chosen; - = no descriptor reaches the threshold, so no FoR qualifies as candidate to be chosen.

that, at its current state, the algorithm is performing relatively well in choosing the ‘favourite’ frame, which is NamedAreas. Another important observation is that the algorithm reached, at this relatively small evaluation, its optimal density threshold at 0.4, as indicated by the DICE value of 0.7, which is higher than the baseline of 0.6. The baseline is simply the most common FoR in the dataset, which is named entities. Surely a more substantial evaluation with a larger dataset will be required before we are safe to make stronger claims about thresholds and performance.

It is important to highlight how we annotated our corpus texts. Frames were considered chosen if they were the first-focus FoR in the description (see 3.1 for a discussion on first vs. second-focus FoR). For instance, if “in Aberdeen and in the west” was the expression, both names and direction were annotated as first-focus frames; if “in western Aberdeen” was the case, then only name was considered first-focus, with direction annotated as second-focus and therefore outside the comparison with the algorithm. This is necessary because, although we gained valuable knowledge about first and second-focus with previous studies, the functionality for focus is not yet present in the algorithm, thus we are not yet ready to evaluate it for this mechanism.

## 4.2 An example

Below we provide an example of how the algorithm decides for Frames of Reference and descriptors. We take a dataset used in the evaluation

exercise, which contains rain forecast data for the Grampian region, in Scotland. The region has a coastal line at the North Sea and is composed of 3 authority areas, namely: Aberdeen, Aberdeenshire and Moray.

As explained above, the data is provided by MetOffice, who also provides textual summaries for the data. From an analysis of the summaries we identified 3 Frames of Reference used with a frequency higher than 5% to describe rain events. These frames, their descriptors and frequencies are:

**NamedAreas (83%):** Aberdeen, Aberdeenshire and Moray.

**Directions (33%):** NorthEast, SouthEast, SouthWest, NorthWest.

**CoastalProximity (17%):** Coastal, Inland.

In the Directions frame, we coded only the inter-cardinal directions as descriptors. This is necessary because the algorithm needs to compute each descriptor as a non-overlapping atomic partition. A North descriptor would overlap with an East descriptor, forming exactly the partition North-East. For this reason, a description such as “the North” is achieved if the algorithm selects the descriptors North-West and North-East, but not South-West and South-East.

The frequencies become the weights of each frame in the algorithm, and the decision for a descriptor is based on the utility score of a descriptor. Utility is computed by multiplying the

event density within a descriptor and its Frame of Reference weight. The event density is the percentage of points of a given descriptor that are also within the event. For example, if the descriptor NorthEast has 32 points in total and 18 are marked with  $\langle \text{rain}, \text{true} \rangle$ , while 14 are marked with  $\langle \text{rain}, \text{false} \rangle$ , the rain-event density of NorthEast is 0.44.

As discussed above, the algorithm was tested with different density thresholds, which set the minimum density value for a descriptor to be considered as candidate. In table 3, we can see why Aberdeen (of the NamedAreas frame) was selected for a setting where density threshold was set to 0.4.

Frame of Reference	Descriptor	Point Count		Density
		Event	Frame	
CoastalProximity	Coastal	27	44	0.61
	Inland	17	83	0.20
Directions	NorthWest	7	27	0.26
	SouthEast	18	27	0.67
	SoutWest	1	41	0.02
	NorthEast	18	32	0.56
NamedAreas	Aberdeen	9	9	1.00
	Aberdeenshire	27	88	0.31
	Moray	8	30	0.27

Table 3: Event densities of a dataset used in the evaluations.

Following the description of the algorithm (in Section 4), the algorithm receives the set of points that ‘are raining’ as well as what descriptors can be assigned to each point. It counts the event density of each descriptor and attempts to reject any descriptor whose density is lower than the threshold. When the density threshold is set to 0, no descriptor is rejected so no frame can be selected. However, when we set the threshold to 0.4, Inland, NorthWest, SouthWest, Aberdeenshire and Moray get rejected. Because each frame now contains a rejected descriptor, all frames are good candidates as SREs. To break the tie, the algorithm resorts to frame weights and densities (i.e. utility). It computes that the utility score of Aberdeen is higher than that of the other non-rejected descriptors, NorthEast, SouthEast, and Coastal, so it selects the descriptor Aberdeen (and the NamedAreas Frame of Reference).

## 5 Conclusions and future work

In this paper we described an initial version of an algorithm that is able to select one or more Frames of Reference – and appropriate descriptors

thereof – to describe an event taking place at a geographic scene. The current state of the algorithm seems promising insofar that it prefers the frame that humans also prefer: NamedAreas. This preference was better observed when the event-density threshold of the algorithm was set to 0.4. However this performance is only verified for first-focus frames, those that are used to reduce the global region to a smaller sub-region.

To enable the algorithm to compute second-focus frames, the key aspect will be *mereology*. A Frame of Reference mix is, at the current state of the algorithm, the geometrical union of two or more descriptors, which in turn share the same global region. Take for instance Texas and North; they belong to different Frames of Reference – StateNames and Directions respectively – but, in isolation, assume the same global area: the US. Although this may be a good mechanism to mix frames in some cases, our corpora are abundant of examples where one descriptor assumes another descriptor as its global region. Take the expression “northern Texas” for instance. It is not the case that the expression refers to the union of Texas and the north of the US. While “Texas” has the entire US as its global region, “northern” refers to the sub-area within Texas. In experiment 1 (see Section 3.2) we showed how names are very frequently the first meorological level when frames are mixed meorologically. We believe that a systematic approach to compute meorological Frames of Reference will substantially improve the performance of the algorithm. Based on evidence found, we also believe that named areas will play a particularly important role in meorological operations.

## 6 Related Work

The subtask of generating referring expressions such as *the green plastic chair* and *the tall bearded man* has been extensively studied by the NLG research community (Dale and Reiter, 1995; Van Deemter, 2002; Krahmer and Van Deemter, 2012). However, relatively fewer studies have been reported on SREs. A notable work is that of Turner and colleagues (2010), which implements the notion of FoR to generate approximate descriptions of geographical regions. As such Turner’s algorithm seem to be too domain specific, as it covers only a subset of FoR that exist.

The algorithm we propose aims to **not** be domain specific but it may be constrained to generat-

ing expression that refer to locations of geographical scale such as regions of a country. Initially we are not concerned with describing the position of small-scale scenes such as a cup on a table. Below we describe how these *spaces* can be significantly different for our task. We also review the *backbone* concept for the algorithm, that of FoR, and we finally list some existing implementations for generating spatial referring expressions.

## 6.1 Spatial frames of reference

When choosing how to represent space with words, we need to select not only spatial entities but a spatial relation between them. Choosing a spatial relation depends largely on the perspective with which one looks at (or imagines) a scene. In cognitive sciences, people have used the term *Frames of Reference* (FoR) to refer to such perspectives. Levinson (2003) classifies cognitive FoR into 3 types:

**Intrinsic** Objects *have* spatial parts such as front or top.

**Relative** The 3rd object position is taken into account.

**Absolute** Fixed bearings such as latitude longitude coordinates.

In this work, we take the same position as (Turner et al., 2010), which perceives the absolute FoR as the one employed by humans when surveying geographical spaces.

## 6.2 Generation of spatial referring expressions

The first systems to use an SREG module date back to the 1990s. FOG (Goldberg, 1995) was the first large scale commercial application of NLG and it generated weather forecasts in English and French.

Similar to FOG, many other systems focus on generating descriptions for weather data (Coch, 1998; Reiter et al., 2005; Bohnet et al., 2007). We can expect the spatial language in the output of such systems to employ the absolute FoR, given the geo-referenced input data. The other type of systems normally use SREG modules to describe a medium-scale (e.g. street) or a small-scale (e.g. room) space (Ebert et al., 1996; Dale et al., 2005; Kelleher and Kruijff, 2006). In such systems, we can expect intrinsic and relative frames.

RoadSafe (Turner et al., 2010), is to the best of our knowledge the most recent system to implement an SREG module. Output spatial language employs absolute FoR and geo-referenced data is processed using DE-9IM (Clementini et al., 1993). RoadSafe implements the most sophisticated SREG module to describe geographical scenes using non-named FoR. We need to enable NLG systems to generate named spatial references as well.

## References

- Anja Belz and Albert Gatt. 2008. Intrinsic vs. extrinsic evaluation measures for referring expression generation. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, pages 197–200. Association for Computational Linguistics.
- Bernd Bohnet, François Lareau, Leo Wanner, et al. 2007. Automatic production of multilingual environmental information. In *Proceedings of the 21st Conference on Informatics for Environmental Protection (EnviroInfo-07)*, Warsaw, Poland.
- Eliseo Clementini, Paolino Di Felice, and Peter van Oosterom. 1993. A small set of formal topological relationships suitable for end-user interaction. In *Advances in Spatial Databases*, pages 277–295. Springer.
- Jose Coch. 1998. Multimeteo: multilingual production of weather forecasts. *ELRA Newsletter*, 3(2).
- Robert Dale and Ehud Reiter. 1995. Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive science*, 19(2):233–263.
- Robert Dale, Sabine Geldof, and Jean-philippe Prost. 2005. Using Natural Language Generation in Automatic Route Description. *Journal of Research & Practice in Information Technology*, 37(1):89–105.
- Christian Ebert, Ralf Meyer-klabunde, Daniel Glatz, Martin Jansche, and Robert Porzel. 1996. From Conceptualization to Formulation in Generating Spatial Descriptions. In *Proceedings fo the 5th European Conference on Cognitive Modelling*, pages 235–241.
- Albert Gatt, Anja Belz, and Eric Kow. 2008. The tuna challenge 2008: Overview and evaluation results. In *Proceedings of the Fifth International Natural Language Generation Conference*, pages 198–206. Association for Computational Linguistics.
- Adele E Goldberg. 1995. *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.



- John Kelleher and Geert-Jan Kruijff. 2006. Incremental generation of spatial referring expressions in situated dialog. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 1041–1048. Association for Computational Linguistics.
- Emiel Krahmer and Kees Van Deemter. 2012. Computational generation of referring expressions: A survey. *Computational Linguistics*, 38(1):173–218.
- Stephen C Levinson. 2003. Frames of reference. In *Space in Language and Cognition: Explorations in Cognitive Diversity*, chapter 2, pages 24–61. Cambridge University Press.
- Ehud Reiter, Somayajulu Sripada, Jim Hunter, Jin Yu, and Ian Davy. 2005. Choosing words in computer-generated weather forecasts. *Artificial Intelligence*, 167(1):137–169.
- Ross Turner, Somayajulu Sripada, and Ehud Reiter. 2010. Generating approximate geographic descriptions. In *Empirical methods in natural language generation*, pages 121–140. Springer.
- Ross Turner. 2009. *Georeferenced data-to-text: techniques and application*. Ph.D. thesis, University of Aberdeen.
- Kees Van Deemter. 2002. Generating referring expressions: Boolean extensions of the incremental algorithm. *Computational Linguistics*, 28(1):37–52.