Using Spatial Reference Frames to Generate Grounded Textual Summaries of Georeferenced Data

Ross Turner, Somayajulu Sripada and Ehud Reiter

Dept of Computing Science,
University of Aberdeen, UK
`{rturner,ssripada,ereiter}@csd.abdn.ac.uk`

Ian P Davy

Aerospace and Marine Intl.,
Banchory, Aberdeenshire, UK
`idavy@weather3000.com`

Abstract

Summarising georeferenced (can be identified according to it's location) data in natural language is challenging because it requires linking events describing its non-geographic attributes to their underlying geography. This mapping is not straightforward as often the only explicit geographic information such data contains is latitude and longitude. In this paper we present an approach to generating textual summaries of georeferenced data based on spatial reference frames. This approach has been implemented in a data-to-text system we have deployed in the weather forecasting domain.

1 Introduction

Data-to-text systems are NLG systems that generate texts from raw input data. Many examples of such systems have been reported in the literature, which have been applied in a number of domains and to different types of input. For example, BabyTalk (Portet et al., 2007) generates medical reports from sensors monitoring a baby in a Neonatal Intensive Care Unit, while (Hallett and Scott, 2005) describe a system for generating reports from events in medical records. SumTime (Reiter et al., 2005), (Coch, 1998) and Fog (Goldberg et al., 1994) generate weather forecasts from the output of weather computer simulation models, while (Iordanskaja et al., 1992) and (Rösner, 1987) both generate summaries from employment statistics.

As the above examples show most work in data-to-text up to now has concentrated almost exclusively on time series data. Work on generating text from spatial data has been reported in Coral (Dale et al., 2005), which generates route descriptions of a path constructed from Geographical Information Systems (GIS) datasets. Unlike the input to Coral however, most georeferenced data contains only limited spatial information(in many cases, only latitude and longitude).

As (Roy and Reiter, 2005) point out, connecting language to the non-linguistic world is an important issue in Cognitive Science and Aritificial Intelligence; moreover, geographic data is becoming increasingly ubiquitous as the availability of low cost locational devices such as GPS increases, and GIS become more user friendly. Therefore, we believe exploring the issue of generating textual reports grounded in real world geographical data is an important challenge. On a more practical level, it is also a natural next step in the application of data-to-text technology to apply it to geographically referenced data.

In the RoadSafe project described in the following section, we have been investigating this issue in a data-to-text system that generates road ice weather forecasts. The subsequent focus of this paper is the adaption of NLG techniques to the task of summarising georeferenced data. In particular, the incorporation of spatial reference frames to generate grounded (from external GIS data sources) spatial references.

2 Background

Weather forecasting has been one of the most successful and widely researched application domains for NLG systems. The main novel aspect that sets Road-Safe apart from other weather forecast generators and indeed, other data-to-text systems, is it's application to spatio-temporal data. The input to Road-Safe is generated by a road ice simulation model, which outputs a large (in order of Megabytes) multivariate data set, shown in Figure 1.

The output of the model contains predicted measurements of 9 meteorological parameters for 1000's of points across a road network, each measured at 20 minute intervals during a 24 hour forecast period. A map of such a network, belonging to a local council in the UK, is shown in Figure 2. This model forms the basis of a road ice forecasting ser-

ID	Latitude	Longitude
UKHA001000000	50.7695	-1.6433
UKHA001000001	50.7591	-1.6258
UKHA001000002	50.7491	-1.5545
....	....	....

ID	Parameter	1200	1220	1240	1300	1320	1340	1400	.....
UKHA001000000	T	4.7	4.8	4.9	5	5.1	5.2	5.2	.....
	W	-2.5	-2.5	-2.6	-2.7	-2.7	-2.8	-2.9	.....
	R	9.5	9.4	9.3	9.2	9.1	9	8.9	.....
	C	0	0	0	0	0	0	0	.....
	D	354	354	355	355	355	355	355	.....
	V	19	19	19	19	19	19	18	.....
	G	28	28	28.1	28.1	28.1	28	26.5	.....
	S	90.4	91.6	91.7	90.9	89.5	87.7	86	.....
	P	0	0	0	0	0	0	0	.....
UKHA001000001	T	4.7	4.9	5	5.1	5.2	5.2	5.3	.....
	W	2.4	-2.5	-2.5	-2.6	-2.7	-2.7	-2.8	.....
	R	9.2	9.1	9	8.9	8.7	8.6	8.5	.....
	C	0	0	0	0	0	0	0	.....
	D	354	354	355	354	355	355	355	.....
	V	19	19	18	18	18	18	18	.....
	G	28.7	28.7	27.2	27.2	27.2	27.2	27.2	.....
	S	90.4	91.6	91.7	90.9	89.5	87.7	86	.....
	P	0	0	0	0	0	0	0	.....
UKHA001000002	T	5	5.1	5.2	5.3	5.4	5.4	5.4	.....
	W	1.9	-1.9	-2	-2.1	-2.2	-2.3	-2.4	.....
	R	9.9	9.8	9.7	9.7	9.6	9.5	9.3	.....
	C	61	61	61	61	61	61	61	.....
	D	354	354	354	354	354	354	354	.....
	V	20	19	19	19	19	19	19	.....
	G	29.8	28.4	28.4	28.4	28.4	28.4	28.3	.....
	S	91.2	92	92.5	93	93.4	93.7	94.2	.....
	P	0	0	0	0	0.1	0.1	0.1	.....
....	....	.....	.....	.....	.....	.....	.....	.....	.....

Figure 1: Part of a RoadSafe input data set showing corresponding spatial and non-spatial attribute tables; T=Air Temperature (Deg C), W=Dew Point (Deg C), R=Road Surface Temperature (Deg C), C=Weather Code, D=Wind direction (Degrees), V=Mean wind Speed (knots), G=Wind Gust (knots),S=Sky Cover (%), P=Precipitation Water Equivalent (mm).

Figure 2: Road Ice Model Data Points Map

vice provided by Aerospace and Marine International (AMI), which is delivered to local councils via an online Road Weather Information System (RWIS). This service provides road engineers with up to the minute weather information using graphs, graphics and textual reports that allows them to base their road maintenance operations on during the winter months. In RoadSafe we have been working on generating the textual reports, such as the one shown in Figure 3, automatically from the model data.

The communicative goal of the textual reports is to complement detailed tabular and graphical presentations of the model data with a more general overview of the weather conditions. In the context of our work this presents a number of challenges:

1. The input data has to be analysed, this is non-trivial due to the complexity and size of the input data.

2. Our system is required to achieve a huge data/text compression ratio (Human authored texts are short and concise summaries). Therefore, content selection is a serious issue for our system.

3. Describing the effect of the underlying geography on weather conditions, such as 'possible gale force gusts on higher ground', is an integral part of the communicative goal of the text. Information containing such relationships is not explicit in the input data and therefore must be grounded.

'Another night with all routes quickly falling below zero this evening. Only isolated urban spots in the south will only drop to around zero. Freezing fog patches will become more widespread during the night but thin a little tomorrow morning especially in the south.'

Figure 3: Example Human Authored Corpus Text

3 Architecture

As noted in the previous section, the input data to our system contains only limited spatial information: a point identifier that ties the measurement site to a particular route and a latitude longitude coordinate. Therefore it is necessary for our system to perform additional spatial reasoning to characterise the input in terms of its underlying geography. The architecture of our system shown in Figure 4, extends
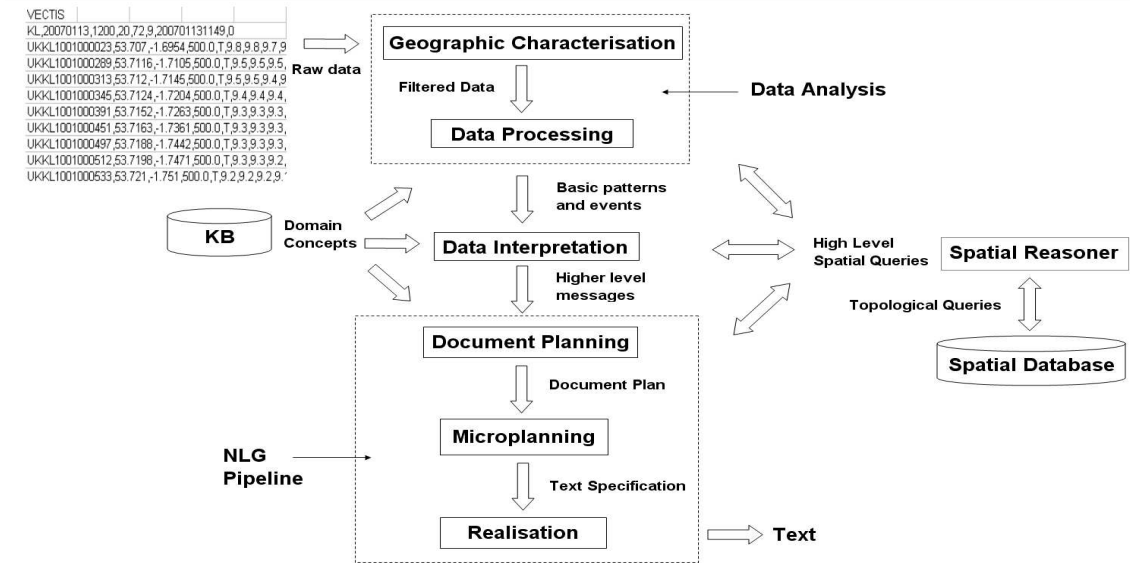
Figure 4: RoadSafe System Architecture

the architecture for data-to-text systems proposed in (Reiter, 2007) to include this additional processing. In Section 3.1 we explain some of the rationale behind these design decisions based on observations from our knowledge acquisition(KA) Studies. In Sections 3.2 and 3.3 we explain the additional modules we have introduced in more detail.

3.1 Observations from Knowledge Acquisition Studies

We have been working closely with experts at AMI for a number of winters now in the development of RoadSafe. During this time we have found that two interrelated aspects in particular have influenced the architecture of our system, which we describe next.

Spatial Reference Frames Frames of reference in this context are a particular perspective in which the domain can be observed. More precisely, they are sets of related geographical features (such as elevated areas) which partition the domain into meaningful sub areas for descriptive purposes. In Levinson's terminology (Levinson, 2003), they are absolute reference systems as they employ fixed bearings. In the RoadSafe domain we have identified 4 main spatial frames of reference used by experts in our corpus described in (Turner et al., 2008):

1. Altitude e.g. 'rain turning to snow on higher ground'.

2. Absolute Direction e.g. 'some heavier bursts in the north'.

3. Coastal Proximity e.g. 'strong winds along the coast'.

4. Population e.g. e.g. 'Roads through the Urban areas holding just above freezing'.

Communicative Purpose of Spatial Descriptions From our studies we have found that experts generally follow 4 steps when writing road ice forecasts:

1. Build frames of reference to geographical features that may affect general weather conditions.

2. Build an overview of the general weather pattern.

3. Select important features to communicate from the pattern.

4. Communicate the summary.

Building frames of reference to geographical features is important for a human forecaster to be able to take into account how the geography of the region influences the general weather conditions. Understanding the weathers interaction with the terrain enables them to make reliable meteorological inferences. For example a spatial description such as 'rain turning to snow in rural areas' may be geographically accurate, but does not make sense meteorologically as it is purely by chance that this event is occurring at that location.

From a NLG system perspective it is important to take into account the communicative purpose of spatial descriptions in this context, which are expressing causality (the effect of geographical features on weather conditions) rather than being purely locative. For example, changes in precipitation type are more commonly seen in higher elevation areas where the air temperature is generally lower, so a spatial description describing such an event should make use of a reference frame that reflects this interaction. Similarly, road surface temperatures are generally higher in urban areas where there is a general population effect. For a referring expression generation (REG) strategy this means that this requires not only adequate spatial representation and reasoning capabilities about an objects location, but also additional information about an objects function in space. This is a problem which has been acknowledged in the psycholinguistic literature e.g. (Coventry and Garrod, 2004).

3.2 Geographic Characterisation

Geographic Characterisation is responsible for grounding the location of the data by making the relationship between it's underlying geography explicit. As the first stage of data analysis it assigns additional spatial properties to each data point by intersecting the point with external GIS data sources representing the frames of reference we have identified. For example after characterisation, the first point in the spatial attribute table shown in Figure 1 is assigned values [0m,SSW,Urban,Coastal] to represent elevation, absolute compass direction, population density of its immediate area and its proximity to the coast respectively. This process is more commonly known as a form of data enrichment in the Spatial Data Mining community (Miller and Han, 2001). In the scope of our work it is important for two reasons: most importantly, it provides a set of properties that are used by the REG module to generate spatial descriptions; secondly, these properties can be taken into account by our analysis method during the initial segmentation of the data.

3.3 Spatial Reasoner and Spatial Database

The spatial database provides a repository of geographic information. Frames of reference are stored as thematic layers from various GIS data sources consisting of sets of boundary objects. For example, altitude is represented as sets of polygons representing altitude contours at a given resolution and population is a set of town boundary polygons. The spatial reasoning module provides a high level interface between the spatial database and the rest of the system.

It is responsible for performing geographic characterisation and providing spatial query functionality to the rest of the system.

4 Text Generation

In Section 2 we outlined 3 main challenges that our system must address. Our approach to the first, analysis of the input data, is described in (Turner et al., 2007). In the following Sections 4.1 and 4.2, we describe the approach taken by our text generator to the former two: content selection and generating spatial references.

4.1 Content Selection

The input to the document planning module of our system is a series of meteorological events (such as rises in temperature) describing each parameter over specific periods of time and locations. The basic events are generated by data analysis which are then abstracted into higher level concepts by data interpretation. As it is impossible to include all these events in such a short summary our system also generates a table as well as text shown in Figure 5.

In our KA studies we have found experts use a qualitative overview of weather conditions when writing forecasts to perform this task, confirming similar observations reported in (Sripada et al., 2001). We take the same approach as experts in our system by including the internal information of the table (generated by the data analysis module) as input to document planning. This serves as the overview for content selection and allows construction of an initial document plan consisting of overview event leaf nodes. An example of this structure for the system output shown in Figure 5 is given in Figure 6. Each overview event corresponds to a column (or columns in the case of snow and rain) in the table if the column indicates a significant threshold for the parameter it describes (i.e. yes for ice).
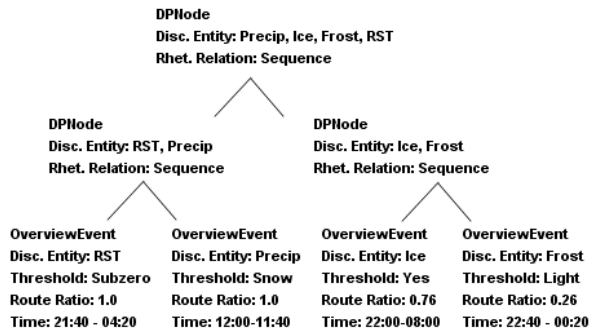

Figure 6: Overview event tree for the text output in Figure 5

Issued by Aerospace & Marine International:- Valid from noon on 11/03/2008

					24 Hour Forecast for Kirklees				
Confidence/Comments		This initial analysis is carried out by the advanced RoadSafe system and is computer generated.							
All Routes	**Min RST**	**Time <= 0c**	**Ice**	**Hoar Frost**	**Snow**	**Fog**	**MaxGusts**	**Rain**	**TS**
Worst/Best	-2.4 /0	21:40/ 04:20	YES /NO	LIGHT /NO	HEAVY/HEAVY	NO/NO	82/69	HEAVY/MODERATE	NO
Overview	Road surface temperatures will fall below zero on all routes during the late evening and tonight except in areas below 100M.								
Wind (mph)	SW 10-25 gusts 45-50 this afternoon in southwestern areas, veering WSW and increasing 15-35 after midnight, gusts 55-60 during the evening and tonight except in areas above 500M, increasing 20-45 then veering W by early morning, gusts 70-75 tomorrow morning in most southern and central places.								
Weather	Wintry precipitation will affect most routes at first, falling as snow flurries in some places above 300M at first. Snow spreading throughout the forecast period to all areas and persisting in some places above 300M until end of period. Ice will affect most routes from the late evening until early morning except in areas below 100M. Hoar frost will affect some routes during the late evening until around midnight. Road surface temperatures will fall slowly during the mid afternoon until tonight, reaching zero in some places above 500M by 22:00.								
Route	All routes summary worst/best								
1	-0.3 /0.5	00:40 /NA	YES /NO	NO/NO	HEAVY/HEAVY	NO/NO	71/65	MODERATE/MODERATE	NO
2	0.0 /1.1	04:00 /NA	NO /NO	NO/NO	HEAVY/HEAVY	NO/NO	70/57	MODERATE/MODERATE	NO
3	-0.2 /0.8	01:00 /NA	YES /NO	NO/NO	HEAVY/HEAVY	NO/NO	71/53	MODERATE/MODERATE	NO
4	-0.2 /0.8	01:00 /NA	YES /NO	NO/NO	HEAVY/HEAVY	NO/NO	70/64	MODERATE/MODERATE	NO
5	-0.1 /0.7	01:20 /NA	YES /NO	NO/NO	HEAVY/HEAVY	NO/NO	70/52	MODERATE/MODERATE	NO
6	0.0 /1.1	01:40 /NA	NO /NO	NO/NO	HEAVY/HEAVY	NO/NO	70/63	MODERATE/MODERATE	NO
7	0.0 /0.9	04:00 /NA	NO /NO	NO/NO	HEAVY/HEAVY	NO/NO	70/53	MODERATE/MODERATE	NO
8	0.0 /0.8	03:40 /NA	NO /NO	NO/NO	HEAVY/HEAVY	NO/NO	70 /49	MODERATE/MODERATE	NO
9	0.0 /1.0	04:00 /NA	NO /NO	NO/NO	HEAVY/HEAVY	NO/NO	69/53	MODERATE/MODERATE	NO
10	0.0 /1.0	02:00 /NA	NO /NO	NO/NO	HEAVY/HEAVY	NO/NO	69 /49	MODERATE/MODERATE	NO
11	-0.3 /0.9	00:40 /NA	YES /NO	NO/NO	HEAVY/HEAVY	NO/NO	71/59	MODERATE/MODERATE	NO
12	-1.4 /0.3	23:20 /NA	YES /NO	LIGHT /NO	HEAVY/HEAVY	NO/NO	81/64	MODERATE/LIGHT	NO
......	......	......	......	......	....	......	......	....	......

Figure 5: Example system output with text and partial table

The next stage is to construct messages from the leaf nodes of the document plan. This is done in a top down fashion by further annotating the tree with events from the input list. Additional events are selected by using the information from the overview events to retrieve them from the list. This has the benefit of keeping the content of both text and table consistent. The final tree comprises the input to the microplanner where messages are realised as sentences in the final text and typically contain two events per message (as observed in our corpus). For example the overview event describing Precip in Figure 6 is realised as two sentences in Figure 5: Wintry precipitation will affect most routes throughout the forecast period at first [overview event], falling as snow flurries in some places above 300M at first [event]. Snow spreading throughout the forecast period to all areas [event] and persisting in some places above 300M until end of period [event].

4.2 Generating Spatial References to Geographic Areas

Approaches to REG to date have concentrated on distinguishing descriptions (e.g. (Gatt and van Deemter, 2007),(van Deemter, 2006),(Horacek, 2006),(Krahmer et al., 2003),(Dale and Reiter, 1995); more specifically that is given a domain, they look to generate a description of a target object that uniquely distinguishes it from all other objects within that domain. In a large geographic environment such as a road network consisting of 1000's of points,

where the task is to refer to an event occurring at a small subset of those points, it is impractical (generated descriptions may be long and complex) and prohibitively expensive (large numbers of spatial relations between objects may have to be computed) to take this approach. A more practical approach is to generate spatial descriptions in terms of regions that are not strictly distinguishing (i.e. urban areas, high ground) rather than in terms of the points contained within that region. Indeed, this is the strategy employed by human authors in our corpus. Therefore, in a description such as 'road surface temperatures will fall below zero in some places in the south west', distractors can be defined as the set of points within the south western boundary that do not satisfy this premise.

The relaxation of the requirement to generate a distinguishing description simplifies the REG task in this context as a single referring expression may be deemed acceptable to refer to a wide range of situations. For example, 'in some places in the south west' could be used to refer to a large number of possible subsets of points that fall within the south western boundary of the network. A simple REG strategy is to find the set of properties to use in a description that introduce the least number of distractors. However, as mentioned previously in Section 3.1, an added constraint is that a spatial description should use an appropriate frame of reference in the context of the event it is describing. For example, describing a change in precipitation type using population as

a frame of reference (i.e 'rain turning snow in some rural places') is not a sound meteorological inference because population density does not affect precipitation. This could cause a reader to infer false implicatures (Grice, 1975), and consequently lead to unnecessary treatment of part of the road network so should be avoided. To account for this, following (Dale and Reiter, 1995) we include a preference set of reference frames for each type of event that must be described. Absence from the set signifies that the specified frame of reference should not be used in that context.

Recall from Section 3.2 that properties in this case relate directly to sets of boundary objects within a frame of reference. Our content selection module takes as input a series of individual proportions describing the spatial distribution of each parameter within each frame of reference at a particular time point. A score is calculated for each set of properties by averaging over the sum of proportions for each frame of reference. An appropriate frame of reference is then selected by choosing the one with the highest score from the preference set for the given event. An example[1] of the input for the generated description 'falling as snow flurries in some places above 300M at first' in Figure 5 is shown in Figure 7.

5 Evaluation

The system presented in this paper is in its first incarnation, RoadSafe is still actively under development in preparation for a full scale user evaluation. We have been evaluating the quality of the output of the current system using post edit techniques and feedback from expert meteorologists at AMI. Our prototype has been installed at AMI since the start of the year and is being used to generate draft road ice forecasts for one of their local council clients. One forecast is generated per day which is then post-edited by an on duty forecaster before it is sent to the client. While common in Machine Translation post-edit evaluations are still relatively rare in NLG. The only large scale post-edit evaluation of an NLG system to our knowledge has been reported in (Sripada et al., 2005).

Our current evaluation is small in comparison to that evaluation; SumTime-Mousam, the system being evaluated in that work was generating 150 draft forecasts per day. However, it does try to address some of the problems the authors encountered during that evaluation. The main issue outlined by (Sripada

---

[1]N.B. this example is taken from route network that is land locked and therefore coastal proximity is not taken into account in this case.

Parameter:	Snow	
Class:	Flurries	
Time point:	12:00 {	
Reference Frame	Boundary	Proportion
Altitude		
	0m:	0.0
	100m:	0.0
	200m:	0.0
	300m:	0.07
	400m:	1.0
	500m:	1.0
Direction		
	CentralNE:	0.0
	CentralNW:	0.0
	CentralSE:	0.0
	CentralSW:	0.0
	EastNorthEast:	0.0
	EastSouthEast:	0.0
	SouthSouthEast:	0.0
	SouthSouthWest:	0.18
	TrueNorthEast:	0.0
	TrueSouthEast:	0.0
	TrueSouthWest:	0.56
	WestSouthWest:	0.23
Population		
	Rural:	0.02
	Urban:	0.0
}		

Figure 7: Example input to content selection for REG. Proportions are number of points affected by snow within given boundary at the specified time point. Scores by Reference Frame: Altitude = 0.35, Direction = 0.07, Population = 0.01

et al., 2005) was that their analysis was post-hoc and therefore not supported by authors or by an editing tool, which made it difficult to analyse why post-edits were made. We have accounted for this by including post-editing as part of our development process and making use of a simple online interface that allows the editor to select check boxes as they edit and insert any general comments they may have. Check boxes record edit reasons at a high level, for example content, sentence order, spatial description used etc. This is because it is not reasonable to expect a time-constrained forecaster to spend time recording every edit he makes.

Another important lesson pointed out by (Sripada et al., 2005) is the need for a pilot study to analyse the post-edit behaviour of individual authors to account for noisy data. This is certainly worthwhile, but is difficult to carry out in our domain where forecasters work in variable shift patterns and on variable forecasting tasks at different times. Instead, we

have used feedback forms as a way to gain qualitative data on both the general quality of the texts and the post-editing process. We present our results in Section 5.1. In Section 5.2 we provide some discussion of the results and describe future work.

5.1 Results

Our post-edit corpus currently consists of 112 texts, 2 texts(1 generated,1 edited) for 56 forecast days. Of the 56 generated texts 54 have been edited before being released to the user. As a general evaluation criterion, our generated texts are generally too long with a mean word length of 72 (standard deviation of 21) compared to a mean word length of 53 (standard deviation of 17). The mean word count difference per forecast is 21 (standard deviation of 15). In general analysis of the corpus is difficult, as in some cases (18) texts have been basically rewritten. This is not reflecting the quality of the text as such, but the fact that the author has access to other information sources such as satellite maps, which can lead him to draw different inferences to those in the raw model data available to the system. Furthermore, (Hopwood, 2004) acknowledge as ice prediction models have become increasingly advanced, the primary added value provided by weather forecasters is to function as quality control and error mitigation for the model, using their insight and experience to make amendments particularly on marginal nights (where the road surface temperature may or may not fall below zero). Such cases can only be considered as noise for analysis purposes, and the fact that our system cannot account for this without the additional information has been acknowledged by all forecasters in their editing comments and feedback forms.

Focusing on 74 real post-edits (not attributed to model data) recorded in our corpus, they can be classified into the following broad error categories: content edits - 65% and microplanning edits 35%. One major problem we have identified with the current generated text is the way in which overview events described in 4.1 are realised. Deletions of whole sentences describing overview events such as the one highlighted in bold in Figure 8 constitute over half (52%) of content edits, which may help to explain the large descrepency in word counts. Essentially forecasters believe they can often communicate similar information as subsequent statements about the same parameter making the texts repetitive at times. Therefore they suggest they should either be omitted or be realised as more interpretative statements, such as 'A marginal night for most routes' for the omitted statement in Figure 8. Forecasters also often delete subsequent statements following overview

---

Generated Text:

'**Road surface temperatures will reach near critical levels on some routes from the late evening until tomorrow morning.** Rain will affect all routes during the afternoon and evening. Road surface temperatures will fall slowly during the mid afternoon and evening, reaching near critical levels in areas above 500M by 21:00.'

Post-edited Text:

'Rain will affect all routes during the afternoon and evening. Road surface temperatures will fall slowly during the mid afternoon and evening, reaching near critical levels in areas above 500M by 21:00.'

Figure 8: Content selection post-edit example (road surface temperature overview information removed)

Generated Text:

'Road surface temperatures will reach near critical levels on some routes after midnight until tomorrow morning. Rain will affect all routes **throughout** the forecast period, **falling as snow in some places above 500M by 08:00. Snow clearing by 08:00.** Road surface temperatures will fall slowly during the late evening and tonight, reaching near critical levels in areas above 500M by 03:00.'

Post-edited Text:

'Road surface temperatures will reach near critical levels on some routes after midnight until tomorrow morning. Rain will affect all routes **during** the forecast period, **this may fall as sleet later on highest ground before dying out.** Road surface temperatures will fall slowly during the late evening and tonight, reaching near critical levels in areas above 500M by 03:00.'

Figure 9: Microplanning post-edit example (lexicalisation and aggregation)

sentences when they describe an event (such as rain turning heavy) occuring only at a small number of locations. So the spatial extent of an event and not only its meteorological importance should be considered during content selection. RoadSafe does not currently include much domain reasoning at the document planning level to be able to do this.

Microplanning edits, as highlighted in bold in Figure 9, are due to individual lexical choice or aggregation issues. In all questionnaires experts have commented that the generated texts are grammatically sound but could flow better. Aggregation is done in a fairly basic fashion in our system at present as is lexicalisation. There have been no edits to the frame of reference used in the generated spatial descriptions, which we have taken as indication that our REG strategy works well.

5.2 Discussion

The general feedback to our system has been encouraging. In terms of the exploitability of the system in its current form it has received mixed reviews from 4 forecasters: 1 forecaster rated the system as good for content and very poor on fluency; 1 rated it as ok for both; 1 forecaster rated it as poor for content and ok for fluency; 1 forecaster rated it as poor for both. Generally all forecasters believe the generated texts should tell a more fluent story about weather conditions with more causal linking between events. In terms of the techniques and approach outlined in this paper they have worked well, although as acknowledged in the previous section more sophisticated domain reasoning and aggregation techniques are required if the text is to function as a concise summary, and indeed reach the standard of human authored texts.

Making the required improvements highlighted in the previous section is the focus of current work. After these improvements have been made we plan to carry out an evaluation with users of the forecasts. We hope to also extend the functionality of the system by generating individual route forecasts, which can be accessed interactively through the table.

6 Conclusions

We have presented an approach to generating geographically grounded summaries of georeferenced data using spatial reference frames. This approach has been implemented in a data-to-text system for generating road ice forecasts. An important task in summarising georeferenced data is to describe the data in terms of its underlying geography it references. This presents an interesting challenge for conventional REG approaches as finding a distinguishing description for large numbers of objects in geographic space is not practical. We have found characterising the geography in terms of spatial reference frames provides a good solution as it provides a flexible representation to describe set of objects in terms of geographic areas.

We have also implemented a simple top down content selection approach based on the idea of overview, taken from how we have observed experts commonly performing the summarisation task. While this approach works well for content selection, a post-edit evaluation with experts has highlighted that realising the overview in the text can make texts verbose and have the effect of making subsequent statements describing related events in the discourse sound repetitive. This is important as experts require a short concise summary of weather conditions.

References

J. Coch. 1998. Multimeteo: multilingual production of weather forecasts. *ELRA Newsletter*, 3(2).

K. R. Coventry and S. C. Garrod. 2004. *Saying, Seeing and Acting: The Psychological Semantics of Spatial Prepositions*. Psychology Press.

R. Dale and E. Reiter. 1995. Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive Science*, 19:233–263.

R. Dale, S Geldof, and J-P Prost. 2005. Using natural language generation in automatic route description. *Journal of Research and Practice in Information Technology*, 37(1):89–105.

A. Gatt and K. van Deemter. 2007. Lexical choice and conceptual perspective in the generation of plural referring expressions. *Journal of Logic, Language and Information*, 16:423–443.

E. Goldberg, N. Driedger, and R. Kittredge. 1994. Using natural-language processing to produce weather forecasts. *IEEE Expert*, 9(2):45–53.

H. Grice. 1975. Logic and conversation. In P. Cole and J. Morgan, editors, *Syntax and Semantics*, volume 3, Speech Acts, pages 43–58. Academic Press: New York.

C. Hallett and D. Scott. 2005. Structural variation in generated health reports. In *Proceedings of the 3rd International Workshop on Paraphrasing (IWP2005)*, pages 33–40, Jeju Island, Republic of Korea.

Philip Hopwood. 2004. Improvements in road forecasting techniques & their applications. In *12th International Road Weather Conference*, Bingen, Germany.

Helmut Horacek. 2006. Generating references to parts of recursively structured objects. In *Proceedings of the 4th International Conference on Natural Language Generation*, pages 47–54.

Lidija Iordanskaja, Richard Kittredge, Benoit Lavoie, and Alain Polguère. 1992. Generation of extended bilingual statistical reports. *COLING-92*, pages 1019–1023.

Emiel Krahmer, Sebastiaan van Erk, and Andr Verleg. 2003. Graph-based generation of referring expressions. *Computational Linguistics*, 29(1):53–72.

Stephen C. Levinson. 2003. *Space in language and cognition: explorations in cognitive diversity*. Cambridge University Press, Cambridge.

Harvey J. Miller and Jiawei Han. 2001. Geographic data mining and knowledge discovery: An overview. In *Geographic Data Mining and Knowledge Discovery*, chapter 1, pages 1–32. Taylor & Francis.

F. Portet, E. Reiter, J. Hunter, and S. Sripada. 2007. Automatic generation of textual summaries from neonatal intensive care data. In *11th Conference on Artificial Intelligence in Medicine (AIME 07)*, pages 227–236.

E. Reiter, S. Sripada, J. Hunter, J. Yu, and I. Davy. 2005. Choosing words in computer-generated weather forecasts. In *Artificial Intelligence*, volume 67, pages 137–169.

E. Reiter. 2007. An architecture for data-to-text systems. In *ENLG07*, pages 97–104.

D. Rösner. 1987. The automated news agency: Semtex: A text generator for german. In *Natural Language Generation: New Results in Artificial Intelligence, Psychology, and Linguistics*. Nijhoff.

D. Roy and E. Reiter. 2005. Connecting language to the world. *Artificial Intelligence*, 167:1–12.

S. Sripada, E. Reiter, J. Hunter, and Jin Yu. 2001. A two-stage model for content determination. In *ENLG2001*, pages 3–10.

S Sripada, E Reiter, and L Hawizy. 2005. Evaluation of an nlg system using post-edit data: Lessons learnt. In *10th European Workshop on Natural Language Generation*.

R. Turner, S. Sripada, E. Reiter, and I. Davy. 2007. Selecting the content of textual descriptions of geographically located events in spatio-temporal weather data. In *Applications and Innovations in Intelligent Systems XV*, pages 75–88.

R. Turner, S. Sripada, E. Reiter, and I. Davy. 2008. Building a parallel spatio-temporal data-text corpus for summary generation. In *Proceedings of the LREC2008 Workshop on Methodologies and Resources for Processing Spatial Language*, Marrakech, Morocco.

K van Deemter. 2006. Generating referring expressions that involve gradable properties. *Computational Linguistics*, 32:195–222.