Atlas.txt: Exploring Linguistic Grounding Techniques for Communicating Spatial Information to Blind Users

Kavita E. Thomas

Department of Computing Science, University of Aberdeen

E-mail: kavita.e.thomas@gmail.com

Somayajulu Sripada

Department of Computing Science, University of Aberdeen

Tel.: +44 -1224-620

E-mail: yaji.sripada@abdn.ac.uk

Matthijs L. Noordzij

Department of Cognitive Psychology and Ergonomics, University of Twente

Tel: +31-53-489-2589

E-mail: m.l.noordzij@utwente.nl

Category: Long Paper

Abstract This paper describes exploratory research into automatically describing geo-referenced information to blind people. The goal is to produce texts giving an overview of the spatial layout, and a central concern of such texts is that they employ an appropriate linguistic reference frame which enables blind hearers to ground the information. The research presented in this paper was based on two hypotheses: (1) directly perceivable reference frames are easier to ground, and (2) spatial descriptions drawn from composite reference frame systems composed of more than one reference frame are easier to ground. An experiment exploring text comprehension on a range of texts employing different reference frame systems is presented. The main results indicate that the second hypothesis is supported. A prototype of a natural language generation system which generates texts describing geo-referenced information from data is described.

Keywords *blind users · geo-referenced spatial information · natural language generation· reference frame preference*

1 Introduction

Geo-referenced information like census data is distributed over a geographic area and is often conveyed via shaded thematic *choropleth* maps, which are inaccessible to blind users, as is shown in Figure 1. Although many thematic maps provided by official authorities provide charts along with the maps to make this information accessible, blind users explore charts via a screenreader and spreadsheet software, which restricts them to the mathematical functions on columns and rows like averaging, summation, max and min. This information crucially excludes any sort of overview about global structures or trends in the data which sighted users easily pick up from the corresponding thematic maps. Likewise, while embossed maps can be printed on special paper, haptic exploration does not enable a quick overview of the data [20] since unlike visual perception, which enables Gestalt forms to rapidly emerge from the scene, it proceeds by extensive low-level exploration before any global forms can be detected. Textual information labelling maps which can be read out via a screenreader can bridge this gap and provide an overview of how spatial information is laid out, communicating global forms and general trends which might be very hard to otherwise detect. The Atlas.txt project investigates how geo-referenced information like census data is best communicated to blind users, with the goal to proved information which is easy to comprehend and ground, and the user is able to mentally visualise the described distribution of the data. Census data typically involves describing the locations of significant features of a geo-referenced variable or statistic like crime rates in a given area. Typically, the information communicated includes the locations of the maxima and minima of the variable and general trends of the variable across the given region. The goal of texts describing this data is to enable easy mental visualisation of the layout of significant features of the geo-referenced data. The Atlas.txt project involves the development of a prototype data-to-text natural language generation (NLG) system which aims to process geo-referenced data stored in tables accompanying census maps, and communicate salient information describing the data via easily understood texts. Although the ideal solution for presenting such data is via multimodal interfaces, so that blind users can hear textual overviews and explore low-level features of the data via exploration in haptic and other modalities, In the

context of Atlas.txt, which is an exploratory investigation into automatically communicating geo-referenced information to blind users, the study of the role of non-linguistic modalities has been left for future work.

Given the goal of producing easily understood and grounded texts describing spatial data, the issue of which reference frame to choose becomes central. A linguistic reference frame identifies the location of an object via a spatial coordinate system. These are typically either absolute and involve arbitrary bearings (e.g., North), relative to another object (e.g., “next to the house”), or intrinsic to an object's perspective (e.g., “in front of you”), where the object can (for example) be the hearer or speaker. When describing spatial information to blind people the choice of which reference frame to select becomes central, as visual input to verify the layout of spatial information is unavailable and texts need to be easy to interpret and ground.

One hypothesis argued for here is related to the idea that some reference frames might be more familiar to blind hearers, and are therefore more readily grounded. For example, describing crime rates as highest “in Northeastern areas” might not be as easy to visualise as hearing that crime rates are highest “at 2 to 3 o'clock” for a blind user who has never seen a map but regularly uses a tactile clock. This paper argues that directly perceivable reference frames like clockface or body-centred directions are easier to ground than indirectly perceivable reference frames like cardinal directions.

A second hypothesis hinges on the nature of the survey perspective of space; unlike in a route perspective, (non-tactile) maps of large areas are designed with visual perception in mind.

Their advantage over text is that they enable viewers to perceive global forms and patterns easily, and to selectively focus on specific sub-regions. Unlike descriptions in the route perspective, which typically employ egocentric reference frames (i.e., with respect to the person following the route), survey descriptions (which are the focus of this work) tend to avoid egocentric reference frames and often employ relative or absolute reference frames, where objects or regions are either described with respect to other objects or areas in the map, or with respect to an arbitrary frame of reference like the cardinal directions. This leads to a range of possible reference frames which can be used to describe areas or objects on maps in the survey perspective, and the hypothesis put forward here is that

employing compound reference frames composed of two (single) reference frames provides additional conceptual perspectives on the described area which communicates more visualisation clues than are communicated by single reference frames. That is, hearing that, for example, crime rates are highest “at 2 to 3 o'clock, at the top-right of the map” might be easier to understand than hearing that they are highest “at 2 to 3 o'clock”.

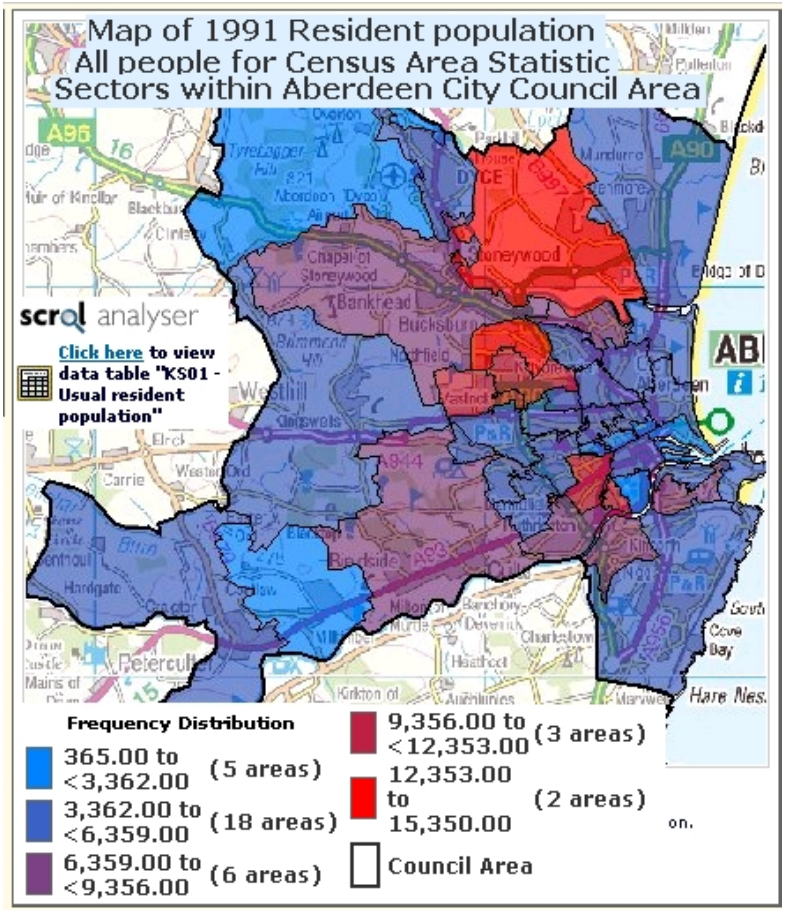

Fig. 1 A typical map showing geo-referenced data

This paper presents an experiment which investigates comprehension of census texts employing different reference frame strategies in order to determine whether these hypotheses hold. The main findings from the experiment support at least the second hypothesis, as composite descriptions employing cardinal and body-centred, clockface and cardinal, and cardinal and clockface reference frames obtained the highest scores. The first hypothesis seems to be partially supported, as the highest scoring reference frame systems all involved at least one directly perceivable reference frame. The paper starts by describing related work and then presents the experiment and finishes by addressing how the findings from the

experiment can be integrated into a natural language generation system producing summaries of spatial data.

1.1 Related Work

The idea of generating textual descriptions of data to make it accessible to the blind is not new; several projects have focused on generating textual descriptions of graphs, ([4]; [8]). The iSonic project ([21]) has focused on communicating census data, the same domain addressed here, but in this project the output modalities were sonification and haptic rather than text. Haptic exploration starts by discovering low-level structures and makes detection of global structures difficult [20]. Sonification also faces similar constraints, as this modality also needs to sweep across the entire data set in order to learn global structures. Aside from the iSonic project, which describes census maps, several projects related to navigational assistance have focused on describing routes to the visually impaired. The work reported in this paper differs from all of these because it aims to (1) effectively describe spatial data via natural language to blind people, and (2) focuses on a domain in which spatial descriptions aim to broadly sketch how data is distributed, rather than a route through space.
One question which one might ask in order to automate spatial descriptions of census data for blind people is whether blind people differ from sighted people with respect to spatial learning, and in particular whether they learn best from the same spatial descriptions which are appropriate for sighted people. In terms of spatial learning, Millar [10] argues that efficient and fast coding of spatial relations between objects without considering the position of the body almost certainly needs vision, or at least memories of visual experiences. Millar raises the possibility that vision is only important during a critical period in life, after which the ability for certain spatial processing mechanisms, such as coding within an allocentric reference frame (i.e., one which refers to locations with reference to a given object, e.g., “in front of the house”), is functional and no longer dependent on vision. Ungar et al [20] have shown that blind people learning a spatial layout were more successful when they tended to explore relative locations of objects and their locations (1) with respect to the external frame, and (2) using two or more exploratory strategies. Several researchers note that certain spatial coding strategies employed by blind people might differ from those employed by sighted

people with respect to the preferred reference frame ([10]; [11]). Millar [10] argues that blind people tend to code spatial information (especially of large spaces) in the form of a local, sequential representation based on routes, whereas sighted people mostly code spatial information in the form of a more global, externally based representation. Noordzij et al [11] found that this holds for spatial descriptions as well; blind people perform better after listening to route descriptions in contrast with sighted people, however they can build up mental models from both survey and route modes. Turning to learning from spatial descriptions, in [11] it was also showed that early and late blind people can construct spatial mental models on the basis of verbal descriptions, and it was argued that visual experience does not seem necessary in order to be able to form some spatial representations. Additionally, it was argued that spatial mental model construction may be affected by the way in which navigational skills are learned by blind people, since blind participants performed better in spatial tasks given route descriptions than survey descriptions. They refer to work showing that blind people have different coding strategies [10] or behavioral strategies [16] than sighted people. Brambring [2] elicited spatial descriptions from both blind and sighted participants and found that, while sighted persons gave environment-oriented descriptions, blind participants tended to use descriptions related to their own position. Brambring concluded that sighted persons seem to give global, externally based descriptions and blind people tend to give local, internally based descriptions.

There has also been research on whether the mental models built by sighted people given spatial descriptions indicates that descriptions in both route and survey modes probably produce the same mental models. Taylor and Tversky [14] conducted a study comparing mental models which sighted participants derived from survey vs. route descriptions in which they read texts in either perspective and then answered verbatim or inference questions from both perspectives and drew maps of the environments. In all studies, Taylor and Tversky observed that participants were faster and more accurate answering verbatim rather than to inference questions, suggesting that verbatim questions are verified against a representation of the text of the descriptions. They also noted that participants were as fast and accurate answering inference questions from the read perspective as from the new perspective, suggesting that inference questions

are verified against a representation of the situation described by the text. Map drawings were very accurate for both description types. They surmised that readers form the same spatial mental models capturing the spatial relations between landmarks from both survey and route descriptions, and from maps. They refer to work by Thorndyke and Hayes-Roth [17] which argues that route and survey perspectives induce different spatial representations, as they found that subjects who learned an environment by walking through it gave more accurate estimates of local distances, whereas subjects who had studied maps gave more accurate estimates of survey measurements.

However, none of this research evaluates specific linguistic reference frames from absolute, relative and intrinsic modes in terms of how they affect comprehension and grounding of spatial information. In addition, none of these approaches explores whether compound reference frames improve comprehension. The effect of multiple or compound reference frames on spatial description comprehension has not been explored before, but Tenbrink [15] investigates how the principle of redundant verbalisation applies to spatial descriptions in spatially complex tasks involving object identification. Tenbrink found that in situations involving particularly complex scenes, or where different kinds of description strategies were available and equally appropriate, participants tended to use several spatial descriptions where one spatial description would have sufficed to identify the given object. This violation of the Gricean maxim of brevity [5] was preferred in a web-study eliciting speaker descriptions. This paper also takes into account additional information manifested as a redundant spatial description, and investigates whether this redundancy aids comprehension of spatial information, where the redundant information in the spatial descriptions considered here makes use of different description strategies, that is, different reference frame choices.

2 Experiment

2.1 Hypotheses

1. Blind people will prefer (i.e., exhibit better comprehension and state an explicit subjective preference for) reference frames that are directly perceivable (e.g., body-centred directions) over reference frames that are

not perceptually grounded (e.g., cardinal directions). That is, the hypothesis is formulated that grounded representations work better with blind people for comprehending spatial descriptions. This follows the findings of Noordzij et al [11], showing that experience with the interface is key. Spatial imagery requires experience and may work best if the input channel for perceiving this imagery is currently open, as Noordzij et al. argue.

2. Blind people will prefer spatial descriptions from compound reference frames (i.e., spatial location descriptions from more than one reference frame system), as this correlates to employing multiple conceptual perspectives to facilitate comprehension.

2.2 Design

In order to explore these hypotheses, the experiment reported here focuses on text comprehension as a measure of the goodness of the reference frame strategy employed. Blind participants were present with texts which were for practical purposes identical, except for the spatial descriptions they use. These spatial descriptions refer to a range of comparable geometric configurations and differ in the reference frames they employ. Reading times of the texts were recorded. Participants were presented with four comprehension questions following each text, identical across texts except that they apply to the given text which they follow. Answering time was measured, as well as the answers to the comprehension questions.

The experiment was conducted online in order to recruit a large number of blind participants, who were asked to browse through the web-pages with the help of their usual screenreaders. Participants were recruited through messages posted onto blind interest mailing lists and discussion forums, and were informed about the goal of the project. Participation was unpaid and on a voluntary basis. Participants were told they could participate whenever it was most convenient, and the only constraint was that they needed to do the whole experiment in one session without taking breaks, which would take them about half an hour. A total of 40 people participated, of which 36 completed the experiment. The downside of online web experiments for measuring reading or response time is that the measurements themselves will vary slightly between machine and internet

connection; however, if one assumes that this machine and internet variation is constant for all interactions in the space of the half-hour experiment, then this added delay is constant across all of a given participant's responses to the various texts. Given that the experiment focused on within-subjects difference between stimuli (i.e., text type), it is argued that this constant delay factor does not affect the results unduly. Another difficulty with online experiments is that participants may take breaks, etc., which will affect reading and response time measurements. However, the participants were warned against this several times, and participated voluntarily in order to assist the project, so it is assumed that the vast majority of the users participated in good faith. Reading and response time results were also checked for outliers, and it was found that there were less than 10-15% of cases with values larger or smaller than the mean +/- two standard deviations. The replicability of the results in a lab based experiment is beyond the scope of this work, but there is good evidence that web-based psycholinguistic experiments yield similar results as their lab based counterparts [6]. Additionally, answers to comprehension questions were also evaluated, which are an equally valid form of measurement under online and laboratory conditions, assuming that participants are in good faith.

The experiment began with criteria for participation (participants needed to be blind or severely visually-impaired and be native speakers of English), information about the duration of the experiment, and general instructions about not taking breaks or closing the browser window. Participants were then given a detailed questionnaire about demographic data, including:

- Education level, i.e., finished high school, finished university, worked or taken a course in geography or statistics.
- Fluency in English: this question aimed to check that participants were native speakers; also dialect of English spoken (e.g., North American, British, Australian)
- Whether they had taken the experiment before; this question was intended to check that they did not take the experiment more than once
- Sense of geography in large spaces (a self-rating was elicited)
- Frequency of use of a tactile clock or watch and when they last used a tactile watch; the objective was to find out whether they were familiar with the clockface reference frame

- Frequency of situations in which they need to understand the layout of a large area
- Severity of their visual impairment (i.e., totally blind, cannot see screen, can partly see screen) and onset of visual impairment (i.e., born blind, onset between 0 to 3 years, between 3 to 8 years, between 8 to 20 years or over 20 years)
- Screenreader used; this was asked in order to make sure that the used screenreader was supported in the experiment design (in terms of accessibility of web pages).

The questionnaire was followed by more detailed instructions about the specific task and an example text and questions. Specifically, they were told that the reading time was being measured, and that they needed to read each text as fast as they could for comprehension, as they would then need to answer questions about the text they just read, but that they could read each text as many times as they needed to visualise the data. As a subsequent step, they were told to press the "Next" button at the bottom of the page as soon as they were sure they understood the data in order to navigate to the next page. The web site contained a series of 10 texts, each followed by a page with comprehension questions. The 10 texts included spatial descriptions from the 10 different simple and combined reference frames considered here, as follows:

- Simple reference frames:
  - Cardinal directions (absolute reference frame), e.g., "in the North"
  - Map-based directions (relative reference frame), e.g., "at the top of the map"
  - Clockface directions (absolute reference frame), e.g., "at 12 o'clock"
  - Body-centred directions (egocentric reference frame), e.g., "in front of you"
- Compound reference frames:
  - Cardinal and body-centred, e.g., "to the North in front of you"
  - Cardinal and clockface, e.g., "to the North at 12 o'clock"
  - Cardinal and map-based, e.g., "to the North at the top of the map"

- Clockface and map-based, e.g., “at 12 o'clock at the top of the map”
- Clockface and body-centred, e.g., “at 12 o'clock in front of you”
- Map-based and body-centred, e.g., “in front of you at the top of the map” .

Familiarity of participants with the clockface directions is investigated via their response to the preliminary questionnaire which asked about the frequency with which they use tactile clocks, as Noordzij et al. [12] indicated that familiarity with tactile clocks might play a role in the results. They investigated the role of spatial imagery in a task which asked early-blind, late-blind and sighted participants to estimate angles between the hands of a clock, and found that early blind subjects performed worse than late-blind participants in this task; one of the reasons for this might be related to the possibility that the participants did not frequently use tactile clocks.

A sample text (using cardinal directions) follows:

> *Crime levels in the region are on the whole quite low. However there's a rough trend for crime to increase towards the Northeast. Crime is highest in the Northeast with the exception of the far Northeast. In addition, crime is high in Central areas. Crime is lowest in the South with the exception of some areas in the far South. In addition crime is low in the Northwest.*

The texts vary with respect to the reference frames used for the spatial descriptions, the geo-referenced variable considered (e.g., crime, property prices, health levels, etc.) in order to avoid participants correlating texts, and the geometric configuration. That is, the locations of maxima and minima differ in each text, but all configurations are mirror-images of each other, so that differences in difficulty between geometric configurations do not affect comprehension. After each text, participants were reminded to click on the next button. The next page contained comprehension questions about the text they just read. For example, for the text reported above, the following questions were presented:

1. Question 1: Roughly how many times did you read the preceding text? Please choose one of the 5 choices below:

- once
- twice
- thrice
- four times
- more than four times

2. Question 2: Imagine there's a flag in the centre of the region. Where is crime high? Please choose one of the 4 choices below:
    - Northeast of the flag (partly correct)
    - Northeast of the flag and areas near the flag (correct)
    - Northwest of the flag and South of the flag (wrong)
    - don't know
3. Question 3: For which area or areas below is crime low? Please choose one of the 4 choices below:
    - far North (wrong)
    - Northwest and far North (partly correct)
    - Northwest (correct)
    - don't know
4. Question 4: Where else are crime rates low? Fill in the blank text box below. (correct answer: South)

The locations of the answer choices (i.e., correct, partly correct, wrong and don't know) did not vary across texts, nor did the order of questions differ. However, the order of texts themselves differed randomly across participants. Participants were reminded to click on the next button at the bottom of the page to go to the next text after answering the questions. Question 4 differs from the others in that it elicits a response which is designed to indicate both (1) comprehension (based on whether it is correct, partly correct, wrong or a “don't know”), and (2) test which reference frame participants choose to frame their answer in: the same one as in the text, one that is partly similar (this only applies to compound reference frames where participants might choose one of the two reference frames making up the descriptions in the text), or drawn from a different reference frame.

Lastly, participants filled in a questionnaire asking them to provide their opinion on the difficulty of the experiment, and rank the 10 reference frames used on a scale of 1 to 10 to indicate which descriptions they preferred and found

easiest to understand. They were then debriefed on the experiment purpose and tasks. The experiment was piloted by a blind computer training expert at Grampian Society for the Blind and a blind computing science student formerly involved with the project who tested it with Jaws, Supernova and Hal for accessibility in the design and to check that the content was at an appropriate level of difficulty.

2.3 Results

The results of the experiment include reading times, answering times, comprehension questions, reference frame elicitation, and preference ranking of reference frames. Each of these aspects is discussed in turn below, as along with some information on the demographics of the participants.

2.3.1 Demographics

A total of 36 participants completed the web experiment. Of these 21 described themselves as blind. Ten of these were born blind, two lost their sight before the age of three, two lost their sight between the ages of three and four, and seven lost their sight after the age of 20. Eight participants described themselves as being unable to see the screen, six of which experienced visual impairment between the ages of three and eight, and two of which had visual impairment after the age of 20. One participant described himself or herself as partly able to see the screen. Several participants did not answer these and other demographics questions, so information on the remaining participants is not available.
Self-described map ability was fairly widespread but leaning toward good map ability rather than poor, with two people describing themselves as being very bad at navigating with maps, while six listed themselves as being very good at using maps. There were nine people who described themselves as average at maps, six as poor map users and 13 as good map users.

Frequency of use of tactile clocks varied to a high degree, but leaned towards less frequent use of tactile clocks, with 10 people using them daily, two monthly, and 17 using them less than yearly. Of those who used clocks daily, three preferred one or more reference frames which used clockface directions, which, if normalised by the number of users is 0.3. For the monthly tactile clock users, one preferred a clockface reference frame, giving a normalised preference of 0.5. Of

the people who used tactile clocks less than yearly, seven people preferred clockface reference frames, giving a normalised preference of 0.41. These scores indicate that frequency of use of tactile clocks may not be the only factor involved in preference for clockface directions, and in the future it might be usful to investigate the participants' usage of tactile clocks earlier in life and also age of onset of blindness. Seven of the 17 participants who used tactile clocks less than yearly lost their sight after the age of 20, and these participants would likely have had experience seeing clocks and telling the time before then. Three of those seven people preferred reference frames using clockface directions. Leaving out those three, one can count 10 participants as having very little experience with either tactile or visually-perceived clockfaces, of which 5 preferred clockface directions, giving a normalised score of 0.5. This indicates that familiarity does not seem to be strongly correlated with the preference for clockface directions. Of course, this does not account for blind participants who lost their sight early but used tactile clocks in their childhood, which may well have been the case. Also since several participants did not enter the age of onset or severity of their visual impairment, these scores are only rough indicators, and a study which accounts for this information would need to be run in order to firmly establish connections between familiarity and preference of reference frames.

2.3.2 Reading Time

Reading time is a standard measure of text comprehension, and this study compares reading times within-subjects in order to detect whether reference frame choices in the 10 different texts affected comprehension. All reading time measures were normalised by the number of words in the respective texts. Reading times were then filtered to exclude cases with reading times outside the range of two standard deviations more or less than the mean. A repeated measures ANOVA was then performed. Normalised reading times showed a significant difference within subjects ($p=.011$), $F(9,270)=2.436$. Subjects were grouped by onset of blindness, but this factor was not significant ($p=.762$), $F(27,270)=.792$. The means and standard deviations for reading times can be seen in Table 1. Notice that texts with compound reference frames have lower normalised reading

times than texts with single reference frames. This finding supports the compound reference frame hypothesis.

Another question which arises is whether familiarity with a tactile clock affects comprehension of texts using the clockface reference frame. Noordzij et al [12] indicated that the poorer performance of blind users on interpreting angles formed by clock hands might be due to a lack of experience with clocks (tactile or ordinary clocks via visual experience). In order to investigate this issue, participants were grouped into three groups, where the highest score (1) was given to people who used tactile clocks on a daily basis, a score of 0.5 was given to those who used them on a monthly basis, and a score of 0 was given to those who used them less often than that (which in all participants amounted to using tactile clocks less than once a year). An ANOVA on these groups for clockface reading time was not significant ($p=.495$), $F(2,25)=.724$, though the between-subjects measurement on web-based reading times is itself not too reliable due to variation in participants' average reading speed. For future work, reading times for this particular issue should be tested in the laboratory.

2.3.3 Answering Time

Time taken to answer recall questions is another standard measure of text comprehension, and again, what is compared here is within-subjects times on the different texts. The answering times were converted to a logarithmic scale (base 10) as the times varied a lot in size. Answering times outside the range specified by two standard deviations above or below the mean were excluded, leaving 17 participants within two standard deviations of the mean, and a repeated measures ANOVA was run. It was found that answering times did not differ significantly ($p=.286$), $F(9,117)=1.226$. There was likewise no significance between subjects based on age of onset of blindness ($p=.373$), $F(27,117)=1.082$. The means and standard deviations are shown in Table 2.

2.3.4 Comprehension Questions

Recall questions are another way to test text comprehension and do not involve the confidence issues raised with reading and answering times measured online. In this case, the questions asked for direct recall of information given in the texts, though in one case the answers were only possible in a different

reference frame from the stimulus, and in one case they involved elicitation of the answer in a reference frame of the participant's choice (through textbox entry). Here, the cumulative correctness of these three questions is considered. The questions were scored as wrong, correct and partly correct (if some information given was correctly but other information necessary to answer the question was either missing or incorrect). Scores for each reference frame type were summed into three groups (number correct, partly correct and wrong), and separate Friedman tests were run on each group (e.g., correct) across text types. The differences between text types in terms of correct answers was significant (p=.009, Chi-Square=22.004, N=40). The two top scoring reference frames are both compound. However this is also true of the two lowest scoring reference frames.

Partly correct answers across text types were not significant (p=.290, Chi-Square= 10.789). Wrong answers across text types were significant (p=.019, Chi-Square=19.761). In this case, the highest scorer is a single reference frame text (cardinal directions), and the two texts with the lowest number of wrong answers are compound, corroborating the results for correct answers, which indicate that compound reference frames lead to better comprehension.

Considering whether participants' responses differ based on age of onset of blindness, it emerged that those who became blind before the age of three years had a nearly significant difference across text types for correct answers (p=.063, Chi-Square=15.843, N=13). These congenital and early blind participants had the most correct answers with a compound reference frame and the least number of correct answers with a single reference frame. This seems to indicate that, at least for this group, comprehension was better with compound rather than single reference frames.

Late blind participants who became blind after the age of 20 did not differ significantly across text types (p=.327, Chi-Square=10.293, N=10), and this group scored highest with single reference frames and worst with compound reference frames.

One of the comprehension questions involved a textbox answer where the participants had to type the answer in their own terms. This elicited reference frame selection, and the responses were scored according to whether they used the same (a score of 1), mixed (0.5) or different (0) reference frames from the text

they had previously read. This choice of elicited reference frame (i.e., same, different or mixed) was significant across text types (p=.001, Chi-Square=25.601, N=10) in a Friedman test. Texts using cardinal directions elicited the most responses in the same reference frame, followed by the majority of compound reference frames. This is somewhat odd, as compound reference frames seem more cumbersome and one might expect participants to be driven by the Gricean maxim of brevity in their responses, and therefore answer briefly, favouring a single reference frame texts. Map directions elicited the fewest responses in the same reference frame, followed by body-centred directions.

Table 1 Normalised Reading Times (mean;SD)

Groups	NrCases	Card	Clock	Body	Map	CardBody	ClockBody	CardClock	CardMap	ClockMap	BodyMap
Cong	11	.56;.27	.67;.47	.65;.46	.58;.30	.56;36	.47;.31	.56;.44	.50;.31	.52;.27	.40;.22
0-3	2	.64;.26	.54;.10	.17;.24	.34;.18	.36;.04	.39;.56	.38;.22	.65;.15	.39;.08	.39;.18
3-8	11	1.19;1.06	.89;.78	.94;.71	.79;.65	.72;.42	.78;.44	.75;28	.74;.60	.68;.38	.62;.54
>20	10	.71;.38	.66;.35	.56;.31	.48;.32	.53;.32	.49;.15	.38;.22	.30;.17	.32;.10	.31;.23
Total	34	.81;.69	.73;.55	.69;.54	.61;.45	.59;.36	.57;.36	.56;.35	.53;.42	.51;.30	.45;.37

Table 2 Answering Times (log10, mean;SD)

Groups	NrCases	Card	Clock	Body	Map	CardBody	ClockBody	CardClock	CardMap	ClockMap	BodyMap
Cong	5	1.89;.145	2.03;.22	1.94;.23	1.98;.27	1.79;.29	1.77;.22	1.87;.26	1.80;.29	1.94;.35	1.89;.32
0-3	1	2.21;-	2.22;-	2.14;-	2.24;-	2.14;-	2.12;-	2.04;-	2.04;-	2.15;-	1.89;-
3-8	5	1.81;.10	1.88;.23	2.01;.19	2.01;.17	2.01;.19	1.90;.17	1.97;.11	1.88;.24	2.05;.19	1.99;.29
>20	6	1.83;.14	1.91;.16	1.92;.23	1.92;.14	1.83;.16	1.90;.25	1.88;.15	1.84;.11	1.84;.09	1.89;.32
Total	17	1.86;.15	1.95;.21	1.92;.20	1.98;.19	1.89;.23	1.87;.22	1.91;.17	1.85;.21	1.95;.23	1.92;.23

Table 3 Mean Rank for Correct Answers (decreasing left to right)

CardBody	CardClock	Card	Body	Map	ClockBody	Clock	BodyMap	CardMap	ClockMap
6.71	5.96	5.90	5.88	5.44	5.23	5.16	5.08	5.01	4.64

Table 4 Mean Rank for Partly Correct Answers (decreasing left to right)

Body	CardMap	BodyMap	Map	CardClock	ClockBody	ClockMap	CardBody	Clock	Card
6.14	5.93	5.74	5.45	5.41	5.39	5.39	5.38	5.23	4.96

Table 5 Mean Rank for Wrong Answers (decreasing left to right)

Card	CardMap	BodyMap	Map	ClockMap	Body	CardBody	Clock	ClockBody	CardClock
6.53	6.22	5.85	5.82	5.79	5.31	5.06	5.03	4.84	4.54

Table 6 Mean Rank for Correct Answers for those Blind before 3 yrs (decreasing L to R)

CardBody	Card	BodyMap	CardClock	Body	Map	ClockBody	CardMap	ClockMap	Clock
7.92	5.96	5.88	5.85	5.38	5.00	4.92	4.85	4.69	4.54

Table 7 Mean Rank for Correct Answers for those Blind after 20 yrs (decreasing L to R)

Body	Clock	CardClock	CardBody	Card	ClockBody	Map	BodyMap	CardMap	ClockMap
6.50	6.40	6.05	6.00	5.85	5.45	5.45	5.00	4.65	3.65

Table 8 Mean Rank for Reference Frame Selection (decreasing L to R)

Card	CardClock	BodyMap	CardBody	Clock	ClockBody	ClockMap	CardMap	Body	Map
7.15	6.95	6.80	6.55	6.55	5.20	4.55	4.35	3.60	3.30

Table 9 Mean Rank for Reference Frame Preference (decreasing L to R)

ClockBody	CardClock	BodyMap	CardMap	ClockMap	CardBody	Body	Map	Clock	Card
6.68	6.39	5.80	5.79	5.66	5.36	5.34	5.29	4.50	4.20

2.3.5 Preferences

At the end of the experiment, participants were asked to rank the different reference frame types based on which ones were easier to mentally visualise on a scale of 1 to 10. A Friedman test was run on these rankings and this difference was significant (p=.038, Chi-Square=17.644, N=28), with participants clearly preferring the compound reference frames over the single ones.
The next section addresses incorporating these findings into a prototype system which describes census data.

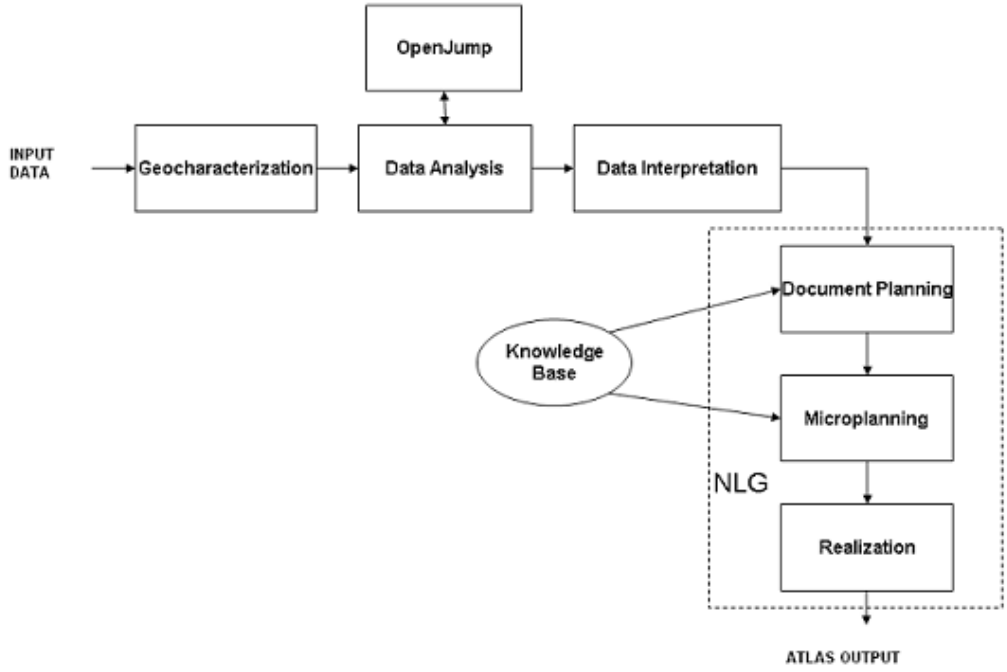

Fig. 2 Architecture of the Atlas.txt Prototype

3 The Atlas.txt System

Atlas.txt is a prototype Natural Language Generation (NLG) system that produces textual summaries of UK 2001 Census data, such as unemployment statistics and population density. These textual summaries describe how values of a census variable are geographically distributed. The experiment described in the previous section shows that compound reference frames could be more useful than single reference frame for visually impaired users in comprehending spatial descriptions. The current version of Atlas.txt has been designed to generate spatial descriptions using several combinations of single reference frames. It works as a

plugin for the opensource lightweight GIS (Geographic Information System) called OpenJump [1]. The architecture of the Atlas.txt system is shown in Figure 2. It has been adapted from the RoadSafe system's architecture [17]. This section describes the main modules of the system. The microplanner (described in section 3.7) has been designed based on the results from the experiment described in the previous section.

3.1 Input

The main input to the Atlas.txt system is derived from UK census data files. An extract of census data is shown in Table 10 below. Census values for several geographic areas (e.g., Census Area Statistics or CAS Wards) making up a higher level geography (e.g., Council Area) are listed in the table. In the data table, each lower level geographic area is referred to by its toponym, and serves as the geo-referenced information anchoring the census data. In order to analyse the census data spatially, geo-referenced information such as longitude and latitude values for the geographic boundary of the area is needed. This digital boundary data forms the second input to the Atlas.txt system.

3.2 Output

> *Some wards in the Eastern and Central parts of the city (to your right and where you are) have a high percentage of unemployed people aged 16-74 (i.e., above 03.25%) and several areas scattered over the city have a low percentage of unemployed people aged 16-74 (i.e., below 03.25%).*

The above text is an example of the output generated by the system for the data shown in Table 10 below. In its current form, the output text describes how the census variable (e.g., percentage of unemployment) is geographically distributed, and helps the reader to gain an overview of the underlying data set. In future work it is planned to include additional details interactively based on user requirements. The output text uses compound reference frames in spatial referring expressions. For example, the spatial referring expression “in the Eastern and Central parts of the city (to your right and where you are)” uses two reference frames. The two reference frames chosen here are cardinal directions and body-centred, since these two reference frames obtained low mean reading and

answering times, and also had the highest mean rank for correct answers in the experiment discussed earlier.

Table 10 Extract of Input Data Showing Percentage Unemployment Values for CAS Wards in Aberdeen City Council Area

AreaName	Percentage Unemployment
Pitmedden	2.047
Bankhead/ Stoneywood	1.747
Danestone	1.155
Jesmond	2.038
Oldmachar	1.663

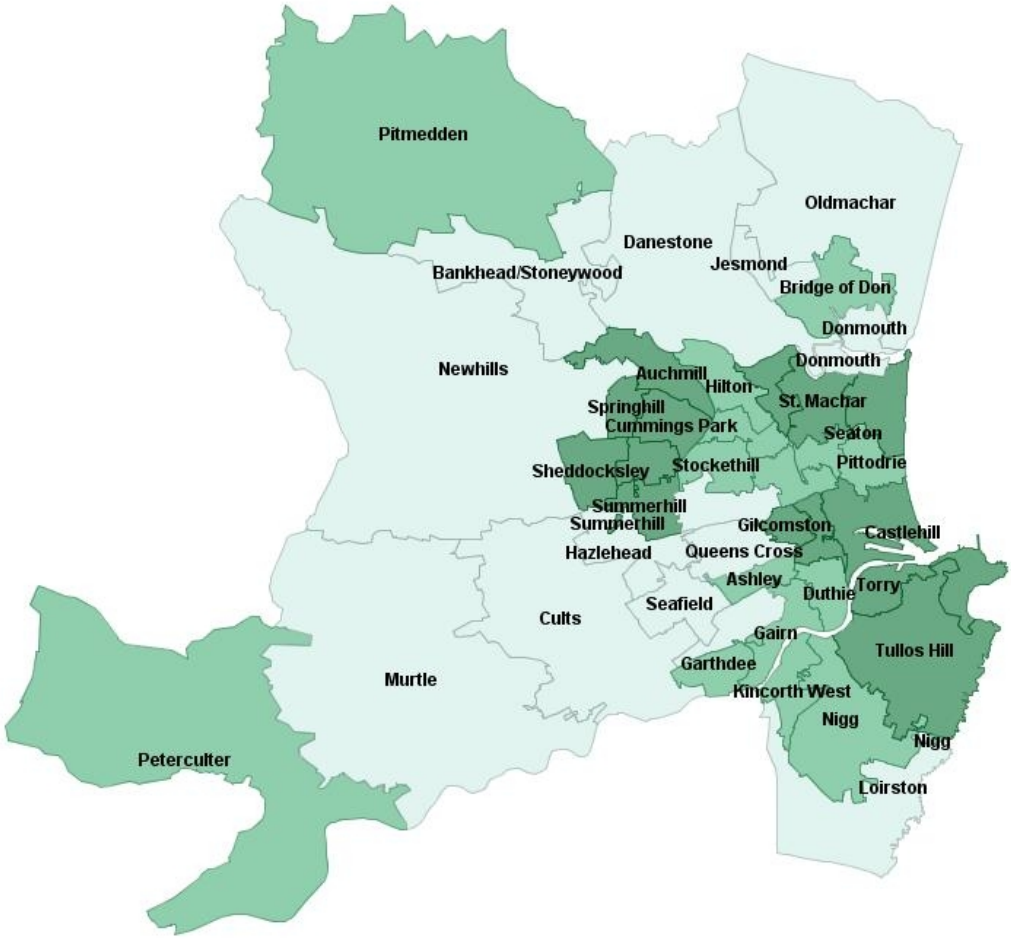

Fig. 3 Aberdeen City Map showing Percentage Unemployment Information (higher values have darker colouring on the map)

3.3 Geocharacterization

An area defined by a digital boundary (specified by a vector of longitude and latitude values) can be geo-referenced in several different ways. For example, an area can be specified by its postcode and the cardinal direction (such as North and East) of its location. Each of these geo-reference specifications is known as a frame of reference. In other words, each frame of reference offers a set of values (such as North, East, West and South) that can be associated to a spatial object to specify its location. Geo-referenced data sets never come with these alternative reference frames explicitly specified. In Table 10, the toponym of a CAS ward provides its geo-reference. In geocharacterization, publicly available resources are used that define additional frames of reference for the same toponyms, such as rural-urban, coastal-inland and cardinal directions.

3.4 Data Analysis

The data analysis module is responsible for performing the low level spatial data analysis, which involves computing partitions from the input data. Each partition consists of a list of toponyms. Each class can have more than one partition, because a partition merges only those areas that belong to the same class of census values which are also neighbours.

3.5 Data Interpretation

The data interpretation module analyses the collection of partitions obtained from the previous module. A feature-value vector made up of the results of all these analyses is the main output of this module. For the input data shown in Figure 3, the following feature-value vector is computed:

AREAWISE DOMINATING CLASS=2,
DISTRIBUTION=UNEVEN,
PARTITIONWISE DOMINATING CLASS=2,
CLASS OF DOMINATING PARTITION=1

In the above feature-value vector, AREAWISE DOMINATING CLASS is set to the class that has the maximum number of areas, which in this case is Class 2.

Because there are no partitions (areas) in Class 0, the distribution of areas is uneven among classes. Therefore, DISTRIBUTION is set to UNEVEN. PARTITIONWISE DOMINATING CLASS is set to the class with the maximum number of partitions. For this example this class is Class 2 again. CLASS OF DOMINATING PARTITION is set to the class that has the maximum number of areas in one single partition. For this example this class is Class 1.

3.6 Document Planning

The document planning module is responsible for deciding which messages to include in the output summary text and also for determining the document structure. In the current version of the system, document schemas are used for document planning. Each schema is associated with a set of feature values that identify it. The selection of a schema is driven by matching the feature-value vector computed in the previous module with the feature values for available schemas. For the input data shown in Figure 3 this module selects a schema that describes Class 3 partitions, followed by partitions belonging to Classes 1 and 2. This is because the feature-value vector shows that both Classes 1 and 2 are important in this data set.

3.7 Microplanning

The microplanner is responsible for generating spatial and non-spatial referring expressions, and also for lexicalization of the document plan by using the lexicalization mappings defined in the knowledge base. Spatial referring expressions are generated using the method described in [16]. This method is based on mapping the spatial layout of partitions computed by the data analysis module onto an n-by-n matrix, where n represents the granularity of geo-referenced data considered. In the developed prototype, a 3-by-3 matrix is used representing the cardinal and inter-cardinal directions as shown below:

```
[NW N NE]
[ W  C  E]
[SW  S SE]
```

Depending on the location of a partition with respect to the 3-by-3 spatial grid overlaid on the map, the matrix is populated with the corresponding values. For the running example, there are four partitions in Class 3 as shown below:

Partition 1 [Summerhill CummingsPark Auchmill Springhill Mastrick Sheddocksley]
Partition 2 [TullosHill Torry]
Partition 3 [Castlehill Langstane Gilcomston]
Partition 4 [St.Machar Woodside Seaton]

Using a 3-by-3 grid index scheme for the input map, these partitions can be mapped onto the matrix as shown below:

[
[0 0 0]
[0 1 1]
[0 0 1]

Using this matrix, the microplanner computes that these partitions are located in Eastern and Central parts of the map. The algorithm does not compute an exact specification for the locations. For example, the matrix has a zero in the top right corner and yet the algorithm produces “Eastern and Central parts”, which leads to approximate spatial descriptions. To summarise, the microplanner uses the matrix to map the spatial layout of the partitions to cardinal and inter-cardinal directions. This spatial reference information is then used to generate the final spatial referring expression.

As described in section 3.3, there are different frames of reference to describe the spatial layout of the partitions. The reported experiment showed that visually impaired users prefer using compound frames of reference such as cardinal directions and clockface directions. Because cardinal directions are more frequently used to describe map data in survey perspective, cardinal directions are adopted as one of the compound frames of reference. Using the matrix obtained above, the microplanner generates the spatial referring expression “in the Eastern and Central parts of the city”. The microplanner then uses a pairing of reference frames to compound with cardinal directions. Since the compound reference frame composed of cardinal and body-centred directions had low mean reading and answering times, and also had the highest mean rank for correct answers in the experiment, this compound reference frame is generated by the microplanner. Generating body-centred directions proceeds as for cardinal directions, where the 3-by-3 spatial partitioning of the data is mapped onto the 9 body-centred

directions. This pairs a commonly used abstract reference system with one which is directly perceivable and does not involve familiarity issues, as in the case of the clockface system, since everyone is familiar with body-centred directions. However, this particular compound reference frame is only one of the many possible reference frames which can be generated. The microplanner can easily generate other reference frames by simply mapping the partitioned data onto other reference frames where directions are specified in a 3-by-3 grid. This means that in a more interactive version of the system, users would be able to select the reference frames of their choice to be included in the output text, thereby creating individually-customised spatial referencing in their summaries.

3.8 Realization

The realization module is responsible for producing grammatically well-formed sentences and clauses in the final surface text using knowledge of English syntax and morphology. In the current version, the *simplenlg* package [11] is used for text realization. The *simplenlg* package contains Java classes for programmatically specifying inputs for realization. For example, *simplenlg* generates the correct verb form that agrees in number with the subject in the sentence based on the specification of the root verb form, the tense and the mood.

4 Conclusions and Discussion/Future Work

This paper has presented a prototype system under development for communicating geo-referenced information in census data to blind people. One of the central issues involved in communicating geo-referenced information to blind people is how to provide spatial referring expressions describing salient features of the data in a way which is easily understood and grounded. To address this issue, this paper has presented an experiment evaluating comprehension and preference of texts employing various reference frame strategies with blind participants in order to evaluate two hypotheses: (1) that directly perceived reference frames are preferred, and (2) that compound reference frames are preferred. The results of this experiment inform the design of the prototype system, by enabling the implementation of the best possible grounding strategy for spatial texts. Although only inconclusive evidence was obtained to support the first hypothesis, the second hypothesis was supported in the results of the

experiment. Although the high scoring reference frame has differed slightly over all the tests considered, the factor in common for most of the high scoring reference frames is that they are compound, leading to the conclusion that these compound reference frames improve comprehension. Additionally, the reading time results are corroborated by the mean rank results for answers to comprehension questions, further lending support to the second hypothesis. Furthermore, all the high scoring compound reference frames involve a directly perceivable reference frame, lending tentative support also to the first hypothesis.

One line of future work should further explore the first hypothesis in conjunction with hypotheses about familiarity and onset of blindness for just single and just compound reference frames. The second hypothesis indicates that more information than is required to identify the approximate region described helps comprehension, in line with the findings of Tenbrink [15]. In this case it was unclear whether this preference for spatial descriptions using compound (i.e., redundant) reference frames arose due to the complexity of information described, and the effect of complexity on redundancy with this user group should be further explored. It would also be interesting to explore the effect of visual impairment on the preference for compound reference frames, i.e., do sighted persons presented with the same texts show the same preference, and do they also comprehend these texts better with compound reference frames?

The present findings have implications that go beyond the description of geo-referenced data like census data, as they can be generalised to situations involving geographic spatial location descriptions, where successful strategies for spatial reference are crucial for grounding locations for blind users. This information is also directly relevant for humans producing spatial descriptions in support of orientation and mobility training of blind people.

The findings can also be generalised to describing the location of elements on web-pages. Visually impaired users are confronted with a range of difficulties when they have to interact with graphical user interfaces. Harper et al [7] have described the various issues that visually impaired users encounter when interacting with complex hypermedia environments such as a web browser. For sighted users the general overview of a website and the selection of appropriate information are usually quickly accessible. However, this quick selection based on a global overview is not possible for blind people. Although these kinds of

environments provide numerous challenges to visually impaired users which fall beyond the scope of this work, the generation of easily grounded spatial location descriptions has the potential to facilitate the accessibility of global page layout. For example, when a given webpage is loaded the first piece of information that might be presented through spoken output is a spatial description of the general page layout (i.e., the placement of a number of pre-selected major features (e.g., frames, links, text). These spatial descriptions should be easily grounded, which is where specific findings for generating spatial location descriptions that work best for blind users comes in.

The developed system provides easy adaptability of the spatial referring expression algorithm, thus different reference frames can easily be selected by mapping spatial partitions of information onto the grid of location descriptions for a given reference system. This approach can support user-customisable spatial descriptions, where users can select the frame of reference which works best for them in order to hear the information with different spatial referring expressions. While there are clearly more open issues to be resolved when considering how best to generate spatial location descriptions for blind users, this work has presented both some findings with respect to which reference systems best facilitate comprehension of spatial descriptions and a prototype computational system for generating such spatial descriptions.

Acknowledgements We would particularly like to thank Charles Clark from the Grampian Society for the Blind for his help in preparing and piloting the experiment. Thanks also to Hussein Patwa for piloting the experiment and for helpful feedback. Thanks to Albert Gatt for his help with SPSS. Lastly we would like to thank EPSRC for funding the Atlas.txt project (EP/D052882/1).

References

1. Openjump. http://jump-pilot.sourceforge.net/.

2. M. Brambring. Language and Geographic Orientation for the Blind. Wiley, 1982.

3. M. Ester, A. Frommelt, H. Kriegel, and J. Sander. Algorithms for characterization and trend detection in spatial databases. 1998.

4. L. Ferres, A. Parush, S. Roberts, and G. Lindgaard. Helping people with visual impairments gain access to graphical information through natural language: The igraph system. Proceedings of ICCHP 2006, Lecture Notes in Computer Science, 4061, 2006.
5. H. Grice. Studies in the Way of Words (SWW). Harvard University Press, 1989.
6. S. Gunasekharan. Evaluation of Web Experiments. MSc Thesis, School of Informatics, University of Edinburgh, 2007.
7. S. Harper, C. Goble, and R. Stevens. Web mobility guidelines for visually impaired surfers. Journal of Research and Practice in Information Technology, 33(1), 2001.
8. K. McCoy, S. Carberry, T. Roper, and N. Green. Towards generating textual summaries of graphs. 2001.
9. A. Mehnert and P. Jackway. An improved seeded region growing algorithm. Pattern Recognition Letters, 18(10), 1997.
10. S. Millar. Understanding and Representing Space: Theory and Evidence from Studies with Blind Children. Oxford University Press, 1994.
11. M. Noordzij, S. Zuidhoek, and A. Postma. The influence of visual experience on the ability to form spatial mental models based on route and survey descriptions. Cognition, 100(2), 2006.
12. M. Noordzij, S. Zuidhoek, and A. Postma. The influence of visual experience on visual and spatial imagery. Perception, 36(1), 2007.
13. E. Reiter and A. Gatt. Simplenlg. http://www.csd.abdn.ac.uk/ ereiter/simplenlg/, 2008
14. H. Taylor and B. Tversky. Spatial mental models derived from survey and route descriptions. Journal of Memory and Language, 35, 1992.
15. T. Tenbrink. Space, Time, and the Use of Language: An Investigation of Relationships. Mouton de Gruyter, 2007.
16. C. Thinus-Blanc and F. Gaunet. Representation of space in blind persons: Vision as a spatial sense? Psychological Bulletin, 121, 1997.
17. P. Thorndyke and B. Hayes-Roth. Differences in spatial knowledge acquired from maps and navigation. Cognitive Psychology, 14, 1982.
18. R. Turner, S. Sripada, and E. Reiter. Generating approximate geographic descriptions. Proceedings of European Natural Language Generation, 2009.
19. R. Turner, S. Sripada, E. Reiter, and I. Davy. Using spatial reference frames to generate grounded textual summaries of georeferenced data. In Proceedings of International Natural Language Generation, 2008.
20. S. Ungar, A. Simpson, and M. Blades. Strategies for Organising Information While Learning a Map by Blind and Sighted People. Uiversidad Nacional de Educacion a Distancia, 2004.
21. H. Zhao, C. Plaisant, and B. Schneiderman. I hear the pattern: interactive sonification of geographical data patterns. 2005.