# Generating Affective Natural Language for Parents of Neonatal Infants

**Saad Mahamood and Ehud Reiter**
Department of Computing Science
University of Aberdeen
Scotland, United Kingdom
{s.mahamood, e.reiter}@abdn.ac.uk

## Abstract

This paper presents several affective NLG strategies for generating medical texts for parents of pre-term neonates. Initially, these were meant to be personalised according to a model of the recipient's level of stress. However, our evaluation showed that all recipients preferred texts generated with the affective strategies, regardless of predicted stress level.

## 1 Introduction

In recent years there has been a great interest in building NLG systems that do not only inform but also take into consideration the recipients emotional state. The need to take such additional factors into account arises from the fact that end users in various contextual circumstances can have more than informational needs to meet. This is particularly apparent for parents of babies that are being looked after in a Neonatal Intensive Care Unit (NICU). This is an environment that has many challenges in receiving and understanding information. It is also an environment in which parents have to come to terms with information that they are not familiar with and have to deal with the emotional impact of the information presented to them. However, whereas medical staff can express affect through voice-tone and body language, the affect-limited nature of text means that information given to recipients by computers must carry the appropriate affective tone in the way that words are expressed to the recipient. The use of empathy to recognise and express emotions to efficiently convey the affective tone of information could allow computers to influence the mood of their users (Picard, 1997). NLG technology has made it possible to produce data-to-text information summaries for human recipients. However, very few NLG systems have any form of strategies that take into consideration the recipients emotional state when communicating information.

In this paper, we discuss our effort to do this in the context of the BabyTalk (Gatt et al., 2009) project. In particular, this paper will focus on the affective approaches used by the BabyTalk-Family system, which was designed to communicate medical information summaries for parents of pre-term neonatal infants. We describe the design, construction, and evaluation of this system in the hope to stimulate discussion on how best to incorporate human emotions as a component of communication between computers and humans.

## 2 Background

In the United Kingdom 12% of newly born babies need specialist medical care in a NICU or in a Special Care Baby Unit (SCBU). The length of stay for such infants can range from a few days to several months. Inside these units, critical life support, physiological monitoring, and medical attention are provided twenty-four hours a day. The babies that are cared for may have complex and serious medical problems. The environment of neonatal care is one of "high technology" in which babies are looked after in incubators surrounded by monitors, wires, and tubes.

For parents of children in NICU, the need for information that is tailored to emotional and informational needs is very much evident. The birth

of a child that requires neonatal care is a particular circumstance that has the potential to cause a considerable amount of stress and anxiety for the parents. The sequence of events in NICU can be akin to a roller coaster ride, with many unexpected ups, downs, and turns of events. Parents rarely feel safe from the fear and uncertainty of the problems that can occur whilst the child is in care (McGrath, 2001). In addition, the stress and shock of having a sick child in neonatal intensive care might also mean that parents will not be able to process large amounts of information (Brazy et al., 2001). However, the provision of information is important in giving parents a sense of hope and a feeling of involvement in their child's care (Charchuk and Simpson, 2005).

## 2.1 NLG and e-Health systems

NLG systems have been increasingly used for the creation of e-Health systems (Hüske-Kraus, 2003), such as generating information for smoking cessation patients (Reiter et al., 2000), breast surgery patients (DiMarco et al., 2007), and so forth. Within healthcare, increasing amounts of patient data are being stored within computerised health databases. This information is being stored in patient records and is combined with drug databases and knowledge bases of medical terminology. Besides helping to provide information support to clinicians, NLG is playing a greater role in providing patients with access to information in a personal form. One prime example is the HealthDoc project that aimed to customise patient information at an individual level based upon their medical condition, demographic, personality profile, and other relevant factors (DiMarco et al., 2007). Such personalisation compares favourably when compared to traditional patient literature which is often limited in its effectiveness by having to address a wide audience (DiMarco et al., 1995).

## 2.2 Affective NLG

Recently, the NLG community has explored the use of emotion as a way of adapting information to the recipient. This development has led to the rise of 'Affective' NLG (ANLG), which has been defined as "NLG that relates to, arises from, or deliberately influences emotions or other non-strictly rational aspects of the hearer" (de Rosis and Grasso, 2000). In

other words, it is a form of NLG that outputs text from a non-linguistic source, but unlike most NLG systems it also takes into account the emotional aspects of the recipient and modifies its textual output for the intended recipient. ANLG attempts to redefine NLG methods and knowledge sources to produce more affective texts (de Rosis and Grasso, 2000). One approach proposed by de Rosis and Grasso (2000) was to introduce models at the sentence planning stage that adapts the message for the intended recipient's communicative goal and also employs rule-based heuristics for the usage of empathy in the resultant text. Other ANLG systems have used emotional or physiological models to define the type of affective text generation. For example, the PERSONAGE system, whilst strictly not dealing with emotion, has shown that by using the 'Big Five' personality traits model it is possible to generate tailored output for particular personality traits (Mairesse and Walker, 2007).

A review of past work in ANLG by Belz (2003) concluded that the research in ANLG has not yet been successful in making the connection between emotion and NLG. Empirical testing of ANLG systems can also pose many challenges as well, with very few past systems being tested. Work by van der Sluis and Mellish (2009) on measuring the emotions of recipients when given positively slanted texts has recently shown it is possible to measure the emotional effect. However, the overall lack of empirical testing from past ANLG systems makes it hard to determine the effectiveness of previous ANLG implementations and the relative importance of their individual techniques.

## 2.3 The BabyTalk Project

The goal of the BabyTalk project (Gatt et al., 2009) is to develop software that generates English summaries of medical data about babies in a NICU. The babies that are cared for may have complex and serious medical problems, and could require critical life support, physiological monitoring, and medical attention twenty-four hours a day. Large quantities of data (a megabyte per day or more) are generated from the real-time monitoring of the baby's physiological condition (e.g., heart rate, blood pressure) and discrete medical events (e.g., equipment settings, drug administration, parent interactions) are

also logged. This large, diverse array of information is stored by modern NICUs in an Electronic Medical Record (EMR). Typically, these EMRs are accessed by medical practitioners through a computer beside the baby's cot.

The main aim for all of the BabyTalk systems is to generate English summaries of EMR data for a variety of readers and purposes. They use signal analysis and medical data interpretation techniques to identify key events and inter-relationships, and NLG to express these events and relationships as a textual narrative. The overriding philosophy of BabyTalk is to use information only within the medical record and not to rely on any additional data input from its recipients. So as not to inconvenience clinicians, nurses, and parents with additional demands. Three BabyTalk (BT) systems have been built. BT-45 (Portet et al., 2007) generates summaries for medical professionals, to assist in real-time decision making; BT-Nurse (Hunter et al., 2011) generates summaries for nurses, to assist in shift handover; and BT-Family generates summaries for parents, to keep them informed about the condition of their child.

Our focus in this paper is on the BT-Family system. This is a system where it is essential that the texts generated be comprehensible to people who are not medical professionals, that the texts do not cause unnecessary stress and anxiety, and most importantly, the texts communicate the information that parents want to know.

## 3 Stress Modelling in BabyTalk-Family

To develop a more effective approach to communicating information to parents, BT-Family must take into account the possible state of mind of the intended recipient and the context or climate that the message would be received in (Berry, 2004). In neonatal care, one of the most predominate emotions that parents face is one of stress. The level of distress experienced by parents can be significant especially if their child is critically ill (Shields-Poë and Pinelli, 1997). Parents can also become distressed by noticing colour changes such as Jaundice, or witnessing episodes of Apnea or respiratory distress (Miles and Holditch-Davis, 1997; Bass, 1991). The small, fragile and undeveloped appearance of an infant in NICU whilst being surrounded by med-

ical apparatus such as respirators, intravenous fluid lines, and monitoring equipment can be very stressful for parents (Miles et al., 1991; Holditch-Davis and Miles, 2000). Parents can find the experience of having their child looked after in a technological environment considerably distressing and oppressive (Jämsä and Jämsä, 1998). The baby's appearance can have such an impact that even at one month of age, mothers of very low birthweight infants show a higher degree of stress compared to mothers of full-term infants (Jackson et al., 2003).

Since neonatal care is a dynamic environment, the sources and levels of stress for parents can change over time and therefore it is important to obtain repeated stress measurements to obtain an accurate assessment of parental stress (Reid et al., 2007). In BT-Family this was done through a stress prediction model called PNSS (Predictive Neonatal Stress Score). Unlike traditional stress self-questionnaire instruments for parents of pre-term neonatal infants, the focus of this model was to have a repeatable non-invasive way of calculating the recipient's level of stress. A detailed explanation of this model's implementation is beyond the scope of this paper, but in essence the model focused on utilising the baby's EMR data to generate a stress score on a three point Likert scale. The higher the score, the more likely the parent could potentially be stressed. The PNSS model composed of thirteen elements. These elements were derived from a partial subset of the Parental Stress Scale (PSS): NICU (Miles et al., 1991) and the Neonatal Unit Parental Stress (Reid et al., 2007) questionnaire instruments. As one of the main factors of parental stress is the physiological health of the child (Shields-Poë and Pinelli, 1997; Seideman et al., 1997), most of the elements in the PNSS model focus on this particular aspect. Another reason to focus on the physiological aspect was due to the fact that most data contained with the baby's EMR focused on the physiological state or medical treatments of the patient. Information about the parents is sparsely recorded or not recorded at all. Therefore, any attempt to simply use all the elements within existing stress questionnaire instruments is not possible.

To evaluate the accuracy of the PNSS model, it was validated against a set of PSS: NICU scores that were obtained from eight mothers who had a

child actively receiving care in a neonatal unit. Statistical analysis of the results obtained showed that the PNSS score had no statistically significant non-parametric correlation with the PSS: NICU score ($p=0.204$, $r_s=-0.504$). These forms of discrepancy could possibly be attributed to the lack of elements that describe the parental role in the PNSS model. This is primarily because this information is not available in the EMR. Whilst the philosophy of BabyTalk is to avoid asking external information from users (such as clinicians, nurses, and parents), in the case of stress scores such input would be required from parents. Further work is required in this area to produce a more accurate predictive stress score.

## 4 Implementing ANLG in BT-Family

The ANLG architecture (illustrated in Figure 1) used in BT-Family is an extension of the NLG data-to-text architecture that was proposed by Reiter (2007), in which natural language text is generated from a non-linguistic data source. Most of the core parts of this system are based upon the BT-Nurse system, in which there are six core components: *Badger EMR Database*, *BabyTalk Ontology*, *Signal Analysis*, *Data Interpretation*, *Document Planner*, and *Microplanner & Realisation*. A detailed explanation of how these core modules function can be found in Gatt et al., (2009) and Mahamood (2010).

The system presented in this section is built upon the BT-Nurse system, but there are crucial differences between these two systems that make both unique from each other. BT-Family contains additional affective extensions to produce textual output that takes into consideration the emotional status of the recipient. The modifications and innovations in the BT-Family system rest in three key areas:

1. Implementation of a stress model within a traditional NLG architecture.

2. The development of a selective document planner that reacts to parental level of distress.

3. The application of multiple affective strategies that attempt to mitigate emotional affect.

Figures 2 & 3 shows the difference between the 12-hour BT-Nurse and 24-hour BT-Family reports produced from the same EMR record.

The PNSS model implemented in BT-Family is based on the work described in the previous section. The implementation uses the BabyTalk ontology to query for the existence of particular factual details about the baby's records to help determine the score for each of the thirteen separate stress factors. This score is stored and made accessible to all other components of the BT-Family system. However, the PNSS score could be calculated by other means or even directly entered by a parent or medical staff; this would not affect the rest of the system.

**BT-Nurse – Patient: 100299, Shift Ending: 2004-02-16 20:00**
**Background**
The baby was born at 24 weeks weighing 755 g. He is 7 weeks old, with corrected gestational age of 30 weeks and 4 days, and weighs 1113 g. He is in an intensive care nursery.

**Current problems**

- Oxygen or ventilator requirement at 28 days of age (since 31/01/2004).
- Hyponatraemia (since 02/02/2004).
- PDA (since 07/02/2004).
- Thrombocytopaenia (since 09/02/2004).
- Confirmed bacterial sepsis (since 10/02/2004).

**Respiratory Support**
Currently, the baby is on CPAP in 27 % O2. CPAP pressure is 4.4 cms H2O.
SaO2 is variable within the acceptable range and there have been some desaturations.
The most recent blood gas was taken at around 11:45. There is fully compensated respiratory acidosis or secondary compensation of metabolic acidosis. pH is 7.32. CO2 is 9.52 kPa. BE is 9.7 mmol/L. The last oral suction was done at about 16:30.

**Events During the Shift**
Between 09:00 and 11:30, RR decreased from 81 to 38.
At around 10:00, the baby was given caffeine.

**Current Status**
Currently, HR is stable within the acceptable range although there have been some bradycardias. At about 19:45, it decreased from 157 bpm to 141 bpm. T1 is variable within the acceptable range.

Figure 2: A partial BT-Nurse report example.

The BT-Family system has a document planner that generates a text structure in a more accessible narrative format for parents rather than producing technical diagnostic texts for nurses. Research findings from past knowledge acquisition phases with parents of neonatal infants were used to create additional subject matter in the generated reports that were of particular interest to parents, but not considered to be clinically relevant (Mahamood et al.,
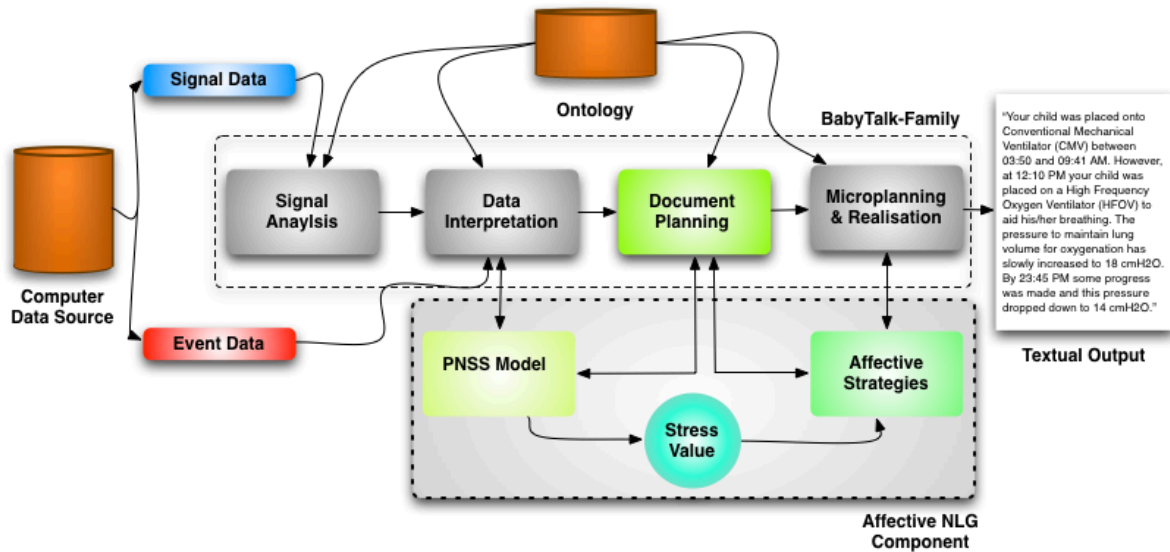
Figure 1: Diagram illustrating the overall ANLG architecture

**BT-Family – Baby 100299, 16/02/2004 08:00 to 17/02/2004 08:00**

John was in intensive care. Your child was stable during the day and night. Since last week, his weight increased from 860 grams (1 lb 14 oz) to 1113 grams (2 lb 7 oz). He was nursed in an incubator.

Yesterday, John was on a ventilator. The mode of ventilation is Bilevel Positive Airway Pressure (BiPAP) Ventilation. This machine helps to provide the support that enables him to breathe more comfortably. Since last week, his inspired Oxygen (FiO2) was lowered from 56% to 21% (which is the same as normal air). This is a positive development for your child.

During the day, Nurse Johnson looked after your baby. Nurse Stevens cared for your baby during the night.

Since last week Milk feeds have increased from 3.0 mls per every hour to 7.0 mls per every hour. This is a reassuring development for your baby.

Baby John had Mummy & Daddy provide some care to him yesterday. John had a gastric milk feed. Also, baby John had some visitors who came to visit him yesterday.

Figure 3: An equivalent BT-Family report.

2008; Moncur et al., 2009). This included matters such as addressing the nursing staff details, listing the parental based care given to child, a reminder of information leaflets given to the parents, and even details on whether the baby had slept well during the night. Some of the content topics are only addressed if the relevant medical details are present within the baby's medical record. Likewise,

BT-Nurse also contains detailed clinical information that is not present in BT-Family, such as blood gas test results (pH, CO2, and Base Excess (BE) levels).

The initial core of the BT-Family document planner, however, relies on addressing four main physiological topics in its textual reports:

1. Details of the baby's weight.

2. Ventilation and inspired oxygen details.

3. Baby's feeding details.

4. Arterial and IV tube insertion/removal details.

What makes these four subject matters different from BT-Nurse is how they are handled in the BT-Family document planner. Instead of just simply reporting the factual details, additional information is also generated to accompany these subject matters. Each of the four main subject matters uses one or more of these additional affective information types, which can consist of:

1. Explanatory Justifications / Details

2. Positive Trend Descriptions

3. Reassurance statements

The main fundamental difference between BT-Family and BT-Nurse resides within the document planner module of the two systems. In BT-Nurse, the document planner utilises several algorithms that specify the maximum length of the document and

16

the minimum importance an event must have to be mentioned. Key events are specifically identified by the BT-Nurse document planner whose importance exceeds a preset threshold and these events are placed at the head of a paragraph, with each paragraph being ordered by the time of occurrence of the key events (Gatt et al., 2009). This implementation differs substantially from BT-Family, where the document planner is based upon fixed categorical topics and clinical events are described in more general terms rather than having the specificity found in BT-Nurse.

### 4.1 Explanatory justification

Explanatory justification is an affective technique that aims to provide additional explanatory textual information to parents for why particular medical actions have occurred by stressing the positive effects for the baby. For example, if a child is moved from one ventilation equipment type to another, that could be viewed by the parents as a negative reflection on the child's well being. But with an explanatory justification statement, the positive benefits are stated for the parents to offset any possible negative perceptions. These statements were implemented as fixed statements that are inserted into the document plan automatically for the last three main topics, besides the baby's weight. This strategy is similar to one employed by Haimowitz (1991) and de Rosis et al. (1999) in which empathy is used by "stressing favourable information while downplaying or offsetting unfavourable information" (Haimowitz, 1991).

> *"This machine helps to provide the support that enables him to breathe more comfortably."*

Figure 4: Explanatory justification example.

### 4.2 Positive Trend Descriptions

Trend descriptions, on the other hand, present trend information over the previous twenty-four hours or week weight, inspired oxygen, and feeding quantities. This is BT-Family's second affective technique. Unlike other strategies BT-Family has discretion when reporting trends. Only those trends that could be considered positive or stable by the parents are reported as such. If no positive or stable trends could be identified, then BT-Family will always present the current value by itself without any

trend description. The trend analysis module tries to determine a positive or stable trend by analysing previous medical data to see if a relative decline or increase has occurred in the given timeframe. Twenty-four hour trends are only computed by the system when the infant has been hospitalised for less than seven days. For inspired oxygen, IV feeds, and nitric oxide, a positive trend was one that has declined over time, whereas for milk feeds and weight, a positive trend was defined as a trend that increased over a specific time period. Table 1 details the trend expectations of parents for each of the different data sources. If no positive or stable trend could be identified, then no additional trend statement was created at all and the factual statement, such as the babys weight, would be presented on it's own.

It was hoped that the use of trend statements would give parents a better understanding of the medical situation that their child faces. The provision of such extra information could prevent parents from feeling that they are not being told every detail and thus would lose hope or assume the worst (Charchuk and Simpson, 2005).

> *"Since last week, his inspired Oxygen (FiO2) was lowered from 56% to 21% (which is the same as normal air)."*

Figure 5: Trend description example.

### 4.3 Reassurance Statements

Finally, within BT-Family there are two forms of reassurance statements: Positive Assurance and Supportive Reassurance statements. If no positive or stable trend from the parents perspective was identified by the system, then a supportive reassurance sentence would be used instead. This would help to reassure the parent that whilst no additional progress has been made, the medical staff nevertheless will continue to support the baby. Such a strategy is also used to help parents cope with the distress of seeing their child having a temporary downturn by helping to reassure parents that the baby has made significant progress over the long term: *"We'll continue to monitor your baby's condition and provide all the support he needs."*

For Positive Assurance, statements like *"Your baby has made good progress today"* are used by the

| Data Source | Trend Expectation | Trend Explanation |
|---|---|---|
| Babys Weight | Positive | *Increases in the Babys weight in grams over a 24-hour / 7-day period.* |
| IV Feed Fluid | Negative | *Decreases in the amount of Dextrose (mls) over a 24-hour / 7-day period.* |
| Milk Feed | Positive | *Increases in the amount of Milk (mls) over a 24-hour / 7-day period.* |
| Inspired Oxygen | Negative | *Decreases in the amount of Inspired Oxygen (%) over a 24-hour / 7-day period.* |
| Nitric Oxide | Negative | *Decreases in the amount of Nitric Oxide (%) over a 24-hour / 7-day period.* |

Table 1: Trend expectations listing for each data source

system to assure parents that positive physiological progress is being made by the child.

### 4.4 Affective and non-Affective Texts

The PNSS model was simply used as a means of switching the affective strategies on or off. In practice this resulted in two forms of text: Affective and non-Affective texts for a given situation. Statements such as trend descriptions and reassurance statements were used as part of an Affective text if the strategies were turned on, but otherwise remained absent from their equivalent non-affective text.The affective strategies would be activated at key junctures of the document planner that dealt with any of the four core physiological topics listed above. Other information types such as explanatory justifications can differ between affective and non-affective texts as shown in Figure 6. Both explanatory statements try to communicate the same concept to the parent. However the affective version is far more concise than the non-affective version so that parents would be prevented from being overwhelmed with information.

---

**Affective Version**
*The mode of ventilation is Conventional Mechanical Ventilation (CMV). This machine helps to provide the support that enables him to breathe more comfortably.*

**Non-Affective Version**
*The mode of ventilation is Conventional Mechanical Ventilation (CMV). This kind of ventilation helps your babys breathing by inflating his lungs, oxygenating the blood, and removing carbon dioxide so that he breathe a lot more easily.*

---

Figure 6: A comparison of explanatory justification statements

Figure 7 shows the difference between the non-affective and affective texts when describing inspired oxygen information for a neonate. From the version it is apparent that the affective version

attempts to reassure the parent far more than the non-affective version. On the other hand, the non-affective version attempts to only communicate the factual status of the baby. The affective version not only addresses the parents information need like the non-affective version but also goes beyond and attempts to reassure as well.

---

**Non-Affective Text Version:**
*"Since yesterday, he was in air (which is the same as 21% oxygen)."*

*"In the evening, John was fed on specialised milk at 56 mls. John was able to take his milk feeds well yesterday. ."*

**Affective Text Version:**
*"Since last week, his inspired Oxygen (FiO2) was lowered from 56% to 21% (which is the same as normal air). This is a positive development for your child."*

*"Milk feeds have increased since last week from 53.0 mls per every three hours to 56.0 mls per every three hours. This is a reassuring development for your baby. John was able to take his milk feeds well yesterday."*

---

Figure 7: A comparison between non-affective and affective versions of inspired oxygen information and milk feeds.

## 5 Evaluation

BT-Family was evaluated with parents that previously had a child in neonatal care (intensive care, high dependency care, or special care) to see the effectiveness of several aspects of the generated texts. A total of thirteen parents were recruited for this study. Recruited participants were mostly socio-economically affluent and well-educated. Parents with babies currently in NICU were not used for this study due to ethical constraints.

**Methodology:**

The thirteen participants were asked to evaluate two different types of text that communicated the same information over ten different medical scenarios, making a total of twenty texts in total. These two types of texts were the computer generated BT-Family reports that were presented as "affective" and "non-affective" variants to the participants. These texts were presented to parents randomly labelled as either text 'A' or 'B', with no indication given of which text was which. In order to reduce the time required for the study, parents were shown only half of a BT-Family report instead of a complete report.

One of the main objectives of this evaluation was to understand the parent's preference for either an "affective" or "non-affective" text for a given medical scenario. In particular, did the participating parents share the same preferences for affective texts in complex medical scenarios and for non-affective texts in simpler medical situations? As subjects could not be made stressed for ethical reasons, they were instead given medical details for each scenario and were asked to imagine that they were the parent of the baby. The PNSS score was used as a way of indirectly identifying between complex and simple medical scenarios as the majority of the score's indicators dealt with the baby's physiological health. Ten different medical scenarios were chosen with five medical scenarios having a high PNSS score and the other five scenarios having a low PNSS score.

To present the two texts for each scenario, the researcher gave a verbal description describing the medical condition and circumstances of the baby in the scenario. This description helped the participant to mentally familiarise themselves with the situation of the baby in the given scenario without prejudicing their preferences. After presenting the participants with a verbal description of the medical circumstance of the baby, they were asked to examine the two texts generated by BT-Family for the given scenario. The participants were informed that the system can produce two types of texts that can express the same information but were not informed about the explicit reason why the texts differ, so as not to prejudice their choice. Following which the researcher asked the participants the following sets of questions:

1. Text style preference: Whether the participants preferred either text A or B for the given scenario.

2. The level of understandability for text A and B (Likert scale of 1 to 5).

3. The helpfulness for both text A and B (Likert scale of 1 to 5).

4. The level of which both text A and B appropriately considers the parents' emotional state in the given scenario (Likert scale of 1 to 5).

5. Participant's comments about the two texts.

**Results:** An overwhelming number of parents (80%) preferred the affective text version than compared to the non-affective text version (20%) in the first five high PNSS score scenarios. Contrary to expectations, in the low PNSS score scenarios, the non-affective text version was disapprovingly looked upon by the parents (13%) compared to the affective text version (87%). It seems that for both cases, parents overwhelmingly prefer the affective text version compared to the non-affective version representing the same information regardless of the baby's scenario. On average, all of the understandability, helpfulness, and emotional appropriateness ratings were weighted in favour of the affective texts across all scenarios.

Several reasons were given by participants for their overwhelming preference for affective texts across all scenarios. For the high PNSS scenarios, the affective texts were favoured due to the fact that the non-affective texts were viewed as "too technical" for some of the parents or that they contain "too much information". Secondly, for low PNSS scenarios, the opposite reaction occurred. Parents stated that the non-affective texts contained less information compared to the affective version, as they contained additional trend and reassurance statements that were not present in the non-affective text. The use of positive reassurance statements in the affective texts were well received by the parents and also were perceived as producing "more friendly text". However, one parent in particular did find the language used by some of the affective texts "a bit patronising". Additionally, the presence of trend statements for the baby's weight, inspired oxygen, and feeds were positively welcomed by the parents. The combination of these factors led most parents to prefer the affective text version in all scenarios regard-

less of the emotional and situational circumstances of the scenario.

From the ratings results gathered, a two-tailed Pearson cross-correlation statistical test was calculated. The null hypothesis was that there should be no correlations between any of the affective and non-affective ratings in all of the three categories. This proved to be incorrect, as three significant results were identified. The first result shows a statistically significant positive correlation between the emotional appropriateness rating and the helpfulness rating for affective texts ($p=0.049$, $r=0.633$). Additionally, the emotional appropriateness rating had a second statistically significant correlation with understanding ratings for affective texts ($p=0.001$, $r=0.885$). What these two results seem to indicate is that there is a relationship between the emotional appropriateness rating for affective texts and the rating scores given by parents for the levels of helpfulness and understandability for the affective texts. The non-affective texts showed only one significant result, a positive correlation between the level of emotional appropriateness and the level of helpfulness ($p=0.006$, $r=0.793$).

## 6 Current Work

BT-Family is still work in progress. Work on BT-Family is preparing for on-ward evaluations with parents, scheduled for late 2011. Using the feedback from parents in the previous section and comments from clinicians, refinements have been made in the textual output of the system and additional topics have been added that were not covered previously, such as drug medication, blood sugar levels, stool and urine output, details of the baby's cot location, and more. Ideas such as the PNSS model and non-affective text output have been removed as ultimately they have proved to be unsuccessful when evaluated.

Ethical permission was sought and granted for two on-ward evaluations with parents of neonatal infants. The first evaluation will focus on refining the quality of the texts from a content and readability perspective with parents providing direct feedback on texts generated for their own baby. The second evaluation will focus on evaluating the usefulness of the texts by seeing how frequently parents access

the texts through a web based portal. We will also conduct post-discharge interviews with participating parents to see if they thought BT-Family texts were communicating appropriate and understandable information. These two evaluations will ultimately help to assess the usefulness of generating such reports for parents.

## 7 Conclusion

We have presented in this paper several affective strategies for communicating medical information for parents of neonatal infants. Initially, we tried to personalise this for the recipient's level of stress. We found that all recipients preferred texts generated with our affective strategies. The key finding is that the use of such affective strategies may be appropriate whenever an NLG system is communicating emotional sensitive information to an non-expert recipient.

## References

Linda S. Bass. 1991. What do parents need when their infant is a patient in the NICU? *Neonatal Network*, 10(4):25–33, December.

Anja Belz. 2003. And now with feeling: Developments in emotional language generation. Report itri-03-21, Information Technology Research Institute, University of Brighton.

Dianne C. Berry. 2004. *Risk, Communication and Health Psychology*. Open University Press.

Jane E. Brazy, Barbara M. H. Anderson, Patricia T. Becker, and Marion Becker. 2001. How parents of premature infants gather information and obtain support. *Neonatal Network*, 20(2):41–48, March.

Margo Charchuk and Christy Simpson. 2005. Hope, Disclosure, and Control in the Neonatal Intensive Care Unit. *Health Communication*, 17(2):191–203.

Fiorella de Rosis and Floriana Grasso. 2000. Affective natural language generation. *Affective Interactions, Springer Lecture Notes in AI*, pages 204 – 218.

Fiorella de Rosis, Floriana Grasso, and Dianne C. Berry. 1999. Refining instructional text generation after evaluation. *Artificial Intelligence in Medicine*, 17(1):1–36.

Chrysanne DiMarco, Graeme Hirst, Leo Wanner, and John Wilkson. 1995. Healthdoc: Customizing patient information and health eduction by medical condition and personal characteristics. In *Workshop on Artificial Intelligence in Patient Education,*, Glassgow, Scotland, UK.

Chrysanne DiMarco, H. Dominic Covvey, P. Bray, D. Cowan, V. DiCiccio, E. Hovy, and D. Mulholland J. Lipa. 2007. The Development of a Natural Language Generation System For Personalized e-Health Information. In *12th International Health (Medical) Informatics Congress*, Brisbane, Australia. Medinfo 2007.

Albert Gatt, François Portet, Ehud Reiter, Jim Hunter, Saad Mahamood, Wendy Moncur, and Somayajulu Sripada. 2009. From Data to Text in the Neonatal Intensive Care Unit: Using NLG Technology for Decision Support and Information Management. *AI Communications*.

Ira J. Haimowitz. 1991. Modelling all dialogue system participants to generate empathetic responces. *Computer Methods and Programs in Biomedicine*, 35(4):321–330, August.

D. Holditch-Davis and M.S. Miles. 2000. Mothers' Stories about Their Experiences in the Neonatal Intensive Care Unit. *Neonatal Network*, 19(3):13–21.

James Hunter, Yvonne Freer, Albert Gatt, Ehud Reiter, Somayajulu Sripada, Cindy Sykes, and Dave Westwater. 2011. BT-Nurse: Computer Generation of Natural Language Shift Summaries from Complex Heterogeneous Medical Data. *Journal of the American Medical Informatics Association*, 18:621–624.

D. Hüske-Kraus. 2003. Text Generation in Clinical Medicine — a review. *Methods of Information in Medicine*, 1:51–60.

Karin Jackson, Britt-Marie Ternestedt, and Jens Schollin. 2003. From alienation to familarity: experiences of mothers and fathers of preterm infants. *Journal of Advanced Nursing*, 43(2):120–129.

Kaisa Jämsä and Timo Jämsä. 1998. Technology in neonatal intensive care — a study on parents' experiences. *Technology and Health Care*, 6(4):225–230, March.

Saad Mahamood, Ehud Reiter, and Chris Mellish. 2008. Neonatal Intesive Care Information for Parents – An Affective Approach. In *Proceedings of 21st IEEE International Symposium on Computer-Based Medical Systems*, Jyväskylä, Finland.

Saad Mahamood. 2010. *Generating Affective Natural Language for Parents of Neonatal Infants*. Ph.D. thesis, University of Aberdeen, Department of Computing Science, Aberdeen, Scotland, United Kingdom.

François Mairesse and Marilyn Walker. 2007. PERSONAGE: Personality generation for dialogue. In *Annual Meeting-Association For Computational Linguistics*, volume 45, pages 496–503.

Jacqueline M. McGrath. 2001. Building Relationships with Families in the NICU: Exploring the Guarded Alliance. *Journal of Perinatal & Neonatal Nursing*, 15(3):74–83.

Margaret Shandor Miles and Diane Holditch-Davis. 1997. Parenting the Prematurely Born Child: Pathways of Influence. *Seminars in Perinatology*, 21(3):254–266, June.

Margaret Shandor Miles, Sandra G. Funk, and Mary Ann Kasper. 1991. The Neonatal Intensive Care Unit Enviroment: Sources of Stress for Parents. *AACN Clinical Issues in Critical Care Nursing*, 2(2):346–354.

Wendy Moncur, Saad Mahamood, Ehud Reiter, and Yvonne Freer. 2009. Involving Healthcare Consumers in Knowledge Acquisition for Virtual Healthcare. Technical Report FS-09-07, AAAI Press, Menlo Park, California, USA, November.

Rosalind W. Picard. 1997. *Affective Computing*. MIT Press, Cambridge (Mass.).

François Portet, Ehud Reiter, Jim Hunter, and Somayajulu Sripada. 2007. Automatic generation of textual summaries from neonatal intensive care data. In *Proceedings of the 11th Conference on Artificial Intelligence in Medicine (AIME '07)*. LNCS, July.

Tilly Reid, Ros Bramwell, Nicola Booth, and A. M. Weindling. 2007. A new stressor scale for parents experiencing neonatal intensive care: the NUPS (Neonatal Unit Parental Stress) scale. *Journal of Reproductive and Infant Psychology*, 25(1):66–82, February.

Ehud Reiter, Roma Robertson, and Liesel Osman. 2000. Knowledge acquisition for natural language generation. In *First International Conference on Natural Language Generation (INLG-2000)*, pages 217–224.

Ehud Reiter. 2007. An architecture for data-to-text systems. In *Proceedings of the 11th European Workshop on Natural Language Generation (ENLG'07)*, June.

Ruth Young Seideman, Margaret A. Watson, Karen E. Corff, Phillip Odle, Joan Haase, and Jane L. Bowerman. 1997. Parent Stress and Coping in NICU and PICU. *Journal of Pediatric Nursing*, 12(3):169–177, June.

Donna Shields-Poë and Janet Pinelli. 1997. Variables Associated with Parental Stress in Neonatal Intensive Care Units. *Neonatal Network*, 16(1):29–37, February.

Ielka van der Sluis and Chris Mellish. 2009. Towards Empirical Evaluation of Affective Tactical NLG. In *Proceedings of the Twelfth European Natural Language Generation Conference (ENLG 2009)*.

21