# A Comparison of Hedged and Non-hedged NLG Texts

**Saad Mahamood, Ehud Reiter, and Chris Mellish**
Department of Computing Science
University of Aberdeen, Untied Kingdom
{smahamoo, ereiter, cmellish}@csd.abdn.ac.uk

## Abstract

We assess the use of hedge phrases in "affective" NLG texts. A simple experiment suggests non-native speakers prefer texts that contain hedge phrases, but native speakers prefer texts that do not contain hedge phrases.

## 1 Introduction

In recent years there has been growing interest in "affective" Natural Language Generation (NLG). Affective NLG 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, 2001).

One open issue in affective NLG is how texts that communicate emotionally sensitive information should be worded. In this paper we focus on the usage of 'hedge phrases' which communicate empathetic information, such as "unfortunately." An experiment in the domain of communicating exam results to students suggests that such emphathetic hedge phrases are appreciated by non-native English speakers, but disliked by native English speakers.

## 2 Motivation

Some NLG systems produce texts that communicate emotionally sensitive information. In particular, the BT-Parent system, which is part of the BabyTalk project (Portet et al., 2007; Reiter, 2007), produces texts that summarise the condition of a baby in a neonatal intensive care unit, for the baby's parents. Such texts must be worded in a way which minimises emotional distress, while of course still being truthful. The work here is an initial attempt to explore one aspect of how sensitive information should be communicated. Because of ethical considerations, we have conducted this initial experiment with a different group, students who are being told about exam results.

## 3 Hedge Phrases

Hedging can be described as a strategy by which speakers mitigate and soften the force of their utterances (Nikula, 1997). The term was first coined by Lakoff (1973), but the original definition has been widened in part due to the observation that certain verbs and syntactic constructions convey hedged performatives (Markkanen and Schröder, 1997a). Verbs such as *suppose*, *guess*, and *think* are examples of these hedged performatives. However these verbs have also been defined as hedges that affect the amount of commitment which a speaker may have to the truth-value of a particular sentence (Markkanen and Schröder, 1997a). Prince et al. (1982) in their analysis on hedging found that hedges can be be classified into two distinct groups with respect to their effect on the truth-value.The first are *approximators* which do modify the truth-condition of the proposition (e.g. I *suppose* the sky is blue). The second is *shields* which unlike *approximators* do not affect the truth-conditon of the proposition, but instead show the amount of commitment that the speaker/writer has to the proposition (e.g. I *think* his feet are blue) (Markkanen and Schröder, 1997a).

In the hedging literature distinctions are not always made between the strategies to be applied and the modifying devices used to achieve the hedging (Clemen, 1997). The five distinct strategies defined by Clemen (1997) are: *Politeness*, *Indirectness*, *Mitigation*, *Vagueness*, and *Understatement*. It can be argued that hedging is determined by a combination of factors, namely the type of context (discourse type), the colloquial situation, the speaker's/writer's intention and knowledge of the background dialogue/conversation (Clemen, 1997).

The aspects of hedging that were of interest to us was the usage of evaluative adverbs (Bonami and Godard, 2006) as a modifying device (such as *unfortunately*, *sadly*, etc.) combined with aspects of

the *Indirectness* and *Mitigation* hedging strategies defined by Clemen (1997).

## 4 Empathy in NLG

Reiter et al. (2000) asked two doctors to rewrite a computer-generated smoking cessation letter, with one doctor asked to enhance the empathy of the letter and another doctor asked to enhance the readability of the letter, thus producing two differing letters. They then asked 20 smokers which version they preferred, and why. They found strong individual differences, with 8 smokers preferring one version and 9 smokers preferring the other (and 3 expressing no preference). In terms of qualitative comments, some of the participants finding the empathic version too "patronising", whilst others found it to be "encouraging".

Perhaps the most important lesson from this experiment is that there seem to be major individual differences in how diffent people react to "empathetic" texts. Hence it would be very helpful if we could create rules which predicted which texts people prefer, based on their characteristics.

## 5 System Implementation

We created a simple NLG system which generated short texts which summarised the exam results of a student; this used the simplenlg API (Reiter, 2007). The system can insert evaluative adverbs (words and phrases) into generated sentences. These adverbs were classified into two distinct groups: Negative hedges and Positive hedges. Negative hedges were adverbs that were used when a given proposition was negative in content. While the positive hedges were used when dealing with a positive content. The example below illustrates these two types of hedges:

(1) ***Unfortunately**, you got CAS 9 in CS5038*

(2) ***Happily**, the CAS result for CS5035 was 19*

The Common Assessment Scale (CAS) score is a scoring mechanism used by the University of Aberdeen. The scale comprises of 21 discrete points with 0 being the lowest and 20 being the highest score. For the NLG system, a student's score was used to determine whether a given proposition was positive or negative.

There were also two different types of hedges deployed; Front hedges and back hedges. Front hedges, as illustrated above, are hedges in which an adverb is injected in front of a given proposition, whereas back hedges are phrases that are inserted after a proposition:

(3) ***Sadly**, you got CAS 9 in CS5035 and CS5037. **Hopefully, this won't effect your degree result by much**.*

Our system randomly chooses which hedge phrase to use; however, it will never insert more than one front hedge or more than one back hedge into a sentence.

The system differs from the ADVISOR expert system (Elhadad, 1991) because of its focus on a specific collection of adverbial words that act as hedges rather than using adjectives for argumentative usage.

## 6 Experiment

### 6.1 Aims

This initial experiment compared the individual preference of participants when presented with hedged and non-hedged texts from our system. The texts gave results for mythical students, not for the experiment participants.

### 6.2 Participants

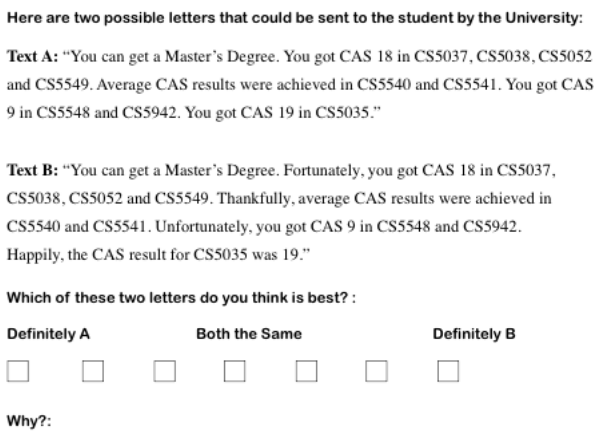
The trial for this experiment was organised as follows. A 1-page, double-sided questionnaire was distributed to a total of 37 Masters students (9 females and 28 males). A total of 5 questionnaires were rejected as incomplete. In the experiment 14 participants were native and fluent in English (Native speakers), whilst 18 participants were not native to the English language. The 18 non-native English speakers were asked if they were fluent in English: 7 said they were (Fluent speakers), and 11 said they were not (Non-fluent speakers).

### 6.3 Procedure

The questionnaire in the experiment showed to each participant an artificially generated exam test result for two differing scenarios. The first was in a positive context in which a hypothetical student has achieved a set of high results. The second in a negative context in which a set of low results was presented. For each scenario participants were shown two texts summarising the exam results: one without hedges (Text A) and one with hedges (Text B). The participants were asked to state which of the two texts they would prefer in the context of having their results delivered to them.

### 6.4 Evaluation Criteria

The questionnaire asked each participant for a set of personal details; Age, English fluency level, and gender. The main questionnaire then consisted of two sections as described in the previous section.


Here are two possible letters that could be sent to the student by the University:

**Text A:** "You can get a Master's Degree. You got CAS 18 in CS5037, CS5038, CS5052 and CS5549. Average CAS results were achieved in CS5540 and CS5541. You got CAS 9 in CS5548 and CS5942. You got CAS 19 in CS5035."

**Text B:** "You can get a Master's Degree. Fortunately, you got CAS 18 in CS5037, CS5038, CS5052 and CS5549. Thankfully, average CAS results were achieved in CS5540 and CS5541. Unfortunately, you got CAS 9 in CS5548 and CS5942. Happily, the CAS result for CS5035 was 19."

Which of these two letters do you think is best? :

Definitely A | Both the Same | Definitely B

Why?:

Figure 1: A section of the questionnaire given to participants

Subjects were asked to state a preference which ran from -3 for Text A to + 3 for Text B. If both texts were considered by a participant to be the same then a score of 0 was given. The participants were asked to provide free text comments on why they made their particular choice of text.

### 6.5 Results

The overall results in this experiment were calculated by a mean average of the participants selection of hedging preference in sections 1 and 2 of the questionnaire. These results are presented in table 1. It is quite clear that on average most participants favoured texts without any form of hedging. However, if we break the results down by language fluency a different picture emerges.

A one-way ANOVA statistical analysis shows a significant effect by group ($p = 0.01$). Further analysis using Post-Hoc Tukey HSD shows a significant difference ($p = 0.013$) between native and fluent speaker groups. This significant difference is also present between native and non-fluent ($p = 0.002$) speaker groups. There was no significant difference ($p = 0.487$) between the fluent and non-fluent speaker groups. It can be concluded from this statistical analysis that there is a correlation between native/non-native status and the amount of preference for hedged and non-hedged texts.

In comparing the number of participants that preferred hedged text in sections 1 & 2 it seems that 16 people preferred the hedged text in section 1 compared to 8 people who preferred the opposite tendency in section 2. However, there was no statistical significant factor found which determined the preference of either hedged text when a univariate analysis of variance was conducted.

## 7 Discussion

From the results above we can see that there is a statistical correlation between the overall preference for these types of hedges and whether the participant is a native speaker. This correlation between being a native speaker and preference for hedging may be partially due to the differing cultural expectations in text of non-native speakers compared to those that have English as their primary language. This is consistent with research conducted by Crismore et al. (1993) which has shown that there is a difference in the amount of hedges used in written text between Finnish and American University students.

The scenario in which the hedges were inserted may play a role in defining the preference for hedging. One of the typical comments which was repeated by several native speakers was that the hedges were less favourable since they were inappropriate for the textual content type. In general the hedges were perceived to be adding opinion to text or were just "wishy-washy". The quote below is from a native speaker and is a good overall representative of the comments received:

> "[Text] B is over-personal, yet in this context it could be constructed as condescending as well. Although [Text] A is a bit more abrupt, it's more formal and suited to University letters."

Such sentiments were not as pervasive in the other speaker groups, which instead commented on how the hedged text was more "humanised" and comfortable. An example of this from a non-fluent speaker is shown below:

> "Good result, so 'wonderful' can make the student happy. Humanised."

It is possible that cultural expectations may play a big role in the wide disparity in perception between the native and non-native speaker groups. However, it is also possible that the preference for such hedges

| Speaker Group | Total Mean | Section 1 | Section 2 |
|---|---|---|---|
| Native Speakers | -1.75 | -1.42 | -2.07 |
| Fluent Speakers | 0.27 | 0.09 | 0.45 |
| Non-Fluent Speakers | 1.14 | 2 | 0.28 |
| **All Groups** | **-0.48** | **-0.14** | **-0.82** |

Table 1: Overall average participant preferences for sections 1 & 2 by language fluency

just reflects the varying degrees of English understanding by the different groups. Such "hedges" may act as "emotional navigators" for those who aren't fluent with English language conventions. On the other hand it is quite possible that the preference for hedging is based upon personal preference rather than any singular principle.

## 8 Future Work

There are at least two areas that need further investigation. The first is to explore the impact that cultural background may have on a particular individual's preference for hedging. It would be interesting to see the types of hedges preferred by some cultures and the ones that aren't. It is possible that this line of research could in turn allow a NLG system to adapt its textual output depending on the cultural background of the recipient.

The second area would look into defining a model for hedging. Currently there is no model for the placement and frequency of hedges within NLG text. One approach for defining such a model would be to look into the statistical frequency of hedges within different corpora to see whether there are any particular defining grammatical or contextual features that are shared between hedges.

## References


Olivier Bonami and Danièle Godard. 2006. Lexical semantics and pragmatics of evaluative adverbs.

Gudrun Clemen. 1997. In Markkanen and Schröder (Markkanen and Schröder, 1997b), chapter The Concept of Hedging: Orgins, Approaches and Definitions, pages 235–248.

Avon Crismore, Raija Markkanen, and Margaret S. Steffenson. 1993. Metadiscourse in persuasive writing: A study of texts written by american and finnish unversity students. *Written Communication*, 10(1):39–71, January.

Fiorella de Rosis. 2001. Towards adaptation of interaction to affective factors. *User Modelling and User-Adapted Interaction 11*.

M. Elhadad. 1991. Generating adjectives to express the speaker's argumentative intent. In *Proceedings of the 9th National Conference on Artificial Intelligence (AAAI 91)*, pages 98–104.

George Lakoff. 1973. Hedges: A study in the meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic 2*.

Raija Markkanen and Hartmut Schröder. 1997a. In Markkanen and Schröder (Markkanen and Schröder, 1997b), chapter Hedging: A Challenge for Pragmatics and Discourse Analysis, pages 3–18.

Raija Markkanen and Hartmut Schröder, editors. 1997b. *Hedging and Discourse: Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts*. Walter de Gruyter.

Tarja Nikula. 1997. In Markkanen and Schröder (Markkanen and Schröder, 1997b), chapter Interlanguage View of Hedging, pages 188–207.

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

Ellen F. Prince, Joel Frader, and Charles Bosk, 1982. *Linguistics and the Proffessions*, chapter On hedging in physician-physician discourse, pages 83–97. Ablex.

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.