



University of Brighton

ITRI-04-05 **Be Polite or Be Summarized**

**Norton Trevisan Roman, Paul Piwek and
Ariadne Maria Brito Rizzoni Carvalho**

March, 2004

CNPq, Capes and EC Project NECA IST-2000-28580.

Information Technology Research Institute Technical Report Series

ITRI, Univ. of Brighton, Lewes Road, Brighton BN2 4GJ, UK
TEL: +44 1273 642900 EMAIL: firstname.lastname@itri.brighton.ac.uk
FAX: +44 1273 642908 NET: <http://www.itri.brighton.ac.uk>

Be Polite or Be Summarized*

Norton Trevisan Roman¹, Paul Piwek², and Ariadne Maria Brito Rizzoni
Carvalho¹

¹ Institute of Computing, Unicamp, Brazil.

norton@ic.unicamp.br

ariadne@ic.unicamp.br

² Information Technology Research Institute, University of Brighton, UK.

Paul.Piwek@itri.brighton.ac.uk

Abstract. Our main goal is to determine if human summarizers include the politeness degree of dialogues in their summaries. We describe two empirical pilot studies on the role of politeness in dialogue summarization. We intend to use the results to build an automatic dialogue summarization system. Amongst other things, we found that the politeness degree is included more often in summaries of dialogues which deviate from what would be considered normal or unmarked. The extent to which politeness degree is included does, however, seem to be affected by constraints on the summary length. We found that the point of view of the summarizer does not influence the extent to which politeness is reported in the summaries.

Key words: Automatic Dialogue Summarization, Automatic Summarization, Natural Language Processing, Politeness.

1 Introduction

Work on both language interpretation and generation has for a long time concentrated mainly on interpretation to and generation from truth-conditional representations of content, with some early exceptions such as [1]. The underlying assumption has been that the content of a natural language utterance can be captured by representing the conditions under which it is true. When considering dialogues, the limitations of a strictly truth-conditional approach are apparent. Many types of dialogue acts do not yield to a purely truth-conditional analysis (greetings, acknowledgments, closings, etc.) and the way a dialogue proceeds is often affected by the emotional states of the interlocutors [2–5].

Recent research on the influence of emotions on interpretation and generation includes work on recognizing the user’s emotional state, e.g., anger [6], so that a

* We would like to thank CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico – and CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – for sponsoring the research. Part of it was also supported by the EC Project NECA IST-2000-28580.

computer dialogue system can adapt its own behavior to prevent such feeling; the use of humor to facilitate interaction with users [7]; and strategies to establish a social and personal relationship with the user by means of “small talk” [8, 9]. An overview of the literature on affect/emotion in natural language generation can be found in [10].

Although there is a body of work on emotion and, more specifically, politeness, in dialogue generation and interpretation, to our knowledge there is not yet any systematic work on the role of politeness in *dialogue summarization*. In this paper, we address the question of whether people consider politeness features of the dialogue important enough to mention them in dialogue summaries. We describe two pilot studies carried out at the State University of Campinas – Brazil, in order to collect data to (partially) resolve the aforementioned question.

In the studies, our subjects were instructed to summarize a carefully selected set of automatically generated dialogues. The resulting collection of human-authored summaries constitutes our data. Our two studies were designed to complement each other. Thus, the first study intends to answer the question mentioned above in situations where summarizers are free to choose the summary’s characteristics. These characteristics are, amongst other things, the summary length, whether they will extract parts of the dialogue or reproduce them etc. The second study complements the first. In this study, summarizers were given a maximum summary size. Its purpose was to verify whether the results obtained in the first study change when the summaries have a limited length. Our longer term aim is to use the insights gained from studying these data in the construction of an automatic dialogue summarizer, which will be integrated with the NECA system [11].

NECA (*Net Environment for Embodied Emotional Conversational Agents*) is a conversational agent platform in which the user can create characters by specifying their roles, personalities and interests. On the basis of these settings, the system automatically generates dialogues between the characters. The result of the generation process is a script which can be performed by two or more embodied agents [12, 13]. The system has been developed for two domains: *eShowRoom*, from which we have taken the dialogues for our study (first version of October 2002); and which concerns car sales, where one agent is the vendor while the other is the customer. In the second domain, *Socialite*, the characters are inhabitants of a student district in Vienna. We used NECA’s scripted dialogues as a test corpus because the NECA system allowed us to systematically change the agents’ politeness.

2 Study 1

2.1 Methodology

The subjects for this study were students and former students, from the State University of Campinas. Table 1 show details about the subjects, according to their educational level, sex and working area. We gave them the dialogues, either

printed out, or by e-mail, and asked them to summarize those dialogues and return the summaries to us, either by e-mail or in paper, typed or handwritten.

<i>Degree</i>		<i>Sex</i>		<i>Course</i>	
Graduated	1	Masc.	18	Physics	1
MPhil Student	8	Fem.	12	Biology	1
PhD Student	20			Civil Eng.	1
PhD	1			Computer Sci.	27

Table 1. Subjects' data.

The study was carried out over a period of two months – July and August, 2003 – in the State University of Campinas, Brazil. First, we generated four dialogues with NECA system [11]. The dialogues take place in a car sales scenario where Ritchie, the vendor, tries to sell a car to Tina, the customer. The dialogues can be found in Appendix I.

For the purpose of our studies we varied the politeness of both interlocutors in the dialogues. This was possible because NECA allows us to control some dialogue parameters such as, for example, the interlocutors' interests, which influence the size of the dialogue to the extent that if the customer is interested in many features of the car, the dialogue will be longer; their agreeableness, which also influences the dialogue length, for it defines how easily the customer can be persuaded to buy the car, or how easily the vendor will give up the sale; and politeness, which determines how polite the interlocutors will be (currently the system allows only two politeness degrees: polite and impolite).

The dialogues were, then, printed and presented to the subjects in the following order:

- Dialogue D_1 shows an interaction in which both vendor and customer were polite and do not use language that is likely to cause offence. The dialogue begins with the vendor attending to the customer, and ends with the customer buying the car;
- D_2 shows an interaction in which the vendor is polite and the customer is impolite. The interaction begins with the customer not finding the vendor in the showroom (i.e., the vendor was apparently absent for a moment) and ends with the customer not buying the car;
- D_3 shows an interaction in which the vendor is impolite and the customer is polite. Again, the dialogue begins with the vendor being apparently absent for a moment, and ends with the customer not buying the car;
- D_4 shows a very short interaction in which, again, both vendor and customer are polite. Like the first dialogue, it begins with the vendor introducing himself to the customer, and ends with the customer buying the car.

The subjects were told to summarize the dialogues as if they had been present during the dialogue and were telling a friend what happened. This included 10

subjects who were told to summarize the dialogue from the point of view of the customer, i.e., they were asked to pretend that they had been the customer, a further 10 who were told to summarize the dialogues as if they were the vendor and, finally, 10 who were told to summarize the dialogues as if they had overheard the dialogue without directly participating in it. The subjects were given no other constraint, apart from the point of view they should take. So, it was up to them to decide how long the summary would be, as well as whether they would quote parts of the dialogue or rephrase them. Although the dialogues are in English, the summarizers were instructed to summarize them either in Portuguese or in English, depending on their preference.

The produced summaries were manually annotated by one of the researchers, sentence by sentence, for politeness content. We classified summaries based on whether the subjects mentioned the interlocutors’ politeness. Thus, a summary is classified as + *politeness* if it has some remark concerning the politeness, such as “I asked a very disgusting vendor...”, “A loser was here...”, “I was extremely attentive...”, “I was in a bad mood...”, and “Ritch, showing lack of respect and politeness...”; and – *politeness* otherwise.

2.2 Results and Analysis

Table 2 summarizes the study’s results. In this table, the rows differentiate among the users who have only taken politeness into account (+ *politeness (only)*), ignoring technical information; the users who have taken the politeness degree of the interaction into account, as well as technical information (+ *politeness* in the tables), when producing the summary; and those who have not taken the politeness into account at all (– *politeness*), i.e., those who have only included the exchanged technical information in the summary. The main columns indicate the number of subjects for each dialogue. Each dialogue column, in turn, is subdivided according to the three points of view adopted by subjects, i. e., the customer’s (**C**), the vendor’s (**V**) and an observer’s (**O**) point of view.

	D_1			D_2			D_3			D_4		
	C	V	O	C	V	O	C	V	O	C	V	O
+ <i>politeness (only)</i>	0	0	0	0	0	0	5	2	1	0	0	0
+ <i>politeness</i>	2	0	0	6	10	9	4	7	9	3	1	1
– <i>politeness</i>	8	10	10	4	0	1	1	1	0	7	9	9

Table 2. Data for the first study, where summaries have no length constraint.

From Table 2, we can see that although for D_1 and D_4 there was a small percentage of subjects who included the interaction politeness in the summary

(respectively, 7% for D₁ and 17% for D₄), this percentage increases remarkably when we consider D₂ and D₃ (respectively, 83% for D₂ and 93% for D₃³).

This difference might be explained by the fact that D₂ and D₃ show unusual situations for car sales dialogues. In these dialogues either the customer or the vendor shows a rude behaviour which, apparently, is worth mentioning in the summary. D₁ and D₄, on the other hand, are more usual, or “neutral”, dialogues. If we group D₁ and D₄, and D₂ and D₃ together, we come up with approximately 12% of the subjects mentioning politeness in D₁ and D₄, and 88% mentioning politeness in D₂ and D₃. A $2 \times 2 \chi^2$ analysis revealed that this is a significant difference, $\chi^2(1, N=120) = 70.53$, at the significance level of $p < 0.001$. So, apparently, whether subjects include politeness into the summary depends on the type of the dialogue.

Another point that is worth mentioning is the increasing number of subjects who reported the politeness in D₄ when compared to D₁ (from 2 subjects in D₁ to 5 in D₄). This could be owed to the fact that D₂ and D₃ present such an unusual interaction that it was worth noticing when the interaction goes back to its “normal” form. The difference between D₁ and D₄ in this respect is, however, not statistically significant.

Furthermore, if we count the total number of summaries mentioning politeness and those that do not, and relate them to the viewpoint, we have 20 summaries reporting politeness and 20 not reporting it at all, for each point of view. There is no statistically significant relation between reporting politeness and the point of view.

3 Study 2

3.1 Methodology

This study was also carried out over a period of two months – November and December, 2003 – at the State University of Campinas, Brazil. In this second study we used the same set of dialogues as in the first study, because we intended to compare the results. We followed the same procedure and used same materials, i.e., we presented the same set of dialogues, in the same order, to the same subjects (see Section 2 for details). Subjects received the dialogues either in electronic form or printed out, and returned the summaries in one of the two forms as well.

Once again, subjects were asked to produce a summary of each dialogue according to the same point of view they assumed in the first study. The only difference is that the summary length was restricted to 10% of the number of words in the dialogue. For annotation of the summaries we used the procedure as described in Section 2.

³ This percentage is the sum of the users who have taken politeness into account and those who have *only* taken politeness of the interaction into account.

3.2 Results and Analysis

Table 3 summarizes the study’s results. In this table, like in Table 2, the rows differentiate among the users who have taken the politeness degree of the interaction and the technical information into account (+ *politeness*); the users who have only included the technical information (- *politeness*); and those who have only taken politeness into account (+ *politeness (only)*).

	D_1			D_2			D_3			D_4		
	C	V	O	C	V	O	C	V	O	C	V	O
+ <i>politeness (only)</i>	1	0	0	0	1	0	1	4	1	0	0	0
+ <i>politeness</i>	0	0	0	5	3	5	8	5	8	0	1	0
- <i>politeness</i>	9	10	10	5	6	5	1	1	1	10	9	10

Table 3. Data for the second study, where summaries have a restricted size.

For this study, the percentage of subjects who included politeness in the summary, for D_1 and D_4 , is 3%. This percentage increases to 47% and 90% for D_2 and D_3 , respectively. Again we notice an increase in the number of subjects who report the dialogue’s politeness when the dialogue becomes more extreme or unusual. Interestingly, in such extreme cases where there is also an upperbound on the summary length, summarizers who report politeness might be forced to leave out technical information which they could have included otherwise.

If we group D_1 and D_4 , and D_2 and D_3 , we have approximately 3% of the subjects mentioning politeness in D_1 and D_4 , and 68% mentioning politeness in D_2 and D_3 . A $2 \times 2 \chi^2$ analysis revealed that this was a significant difference, $\chi^2(1, N=120) = 55.14$, at the significance level of $p < 0.001$. Thus, we have evidence that there may be a relationship between the politeness of the dialogue and subjects reporting it in the summary, even with the restriction on the summary length.

Notice that the number of subjects who included politeness in the summaries for D_1 and D_4 is exactly the same, i.e., apparently, the fact that D_4 was presented after two unusual dialogues did not affect the subjects, given that they were restricted to a small summary length.

Counting again the total number of summaries mentioning politeness and those not mentioning it at all, and relating them to the assumed viewpoint, we have 15 summaries mentioning politeness and 25 not mentioning it, for the customer’s point of view; 14 and 26, respectively, for the vendor’s; and 14 and 26 for the observer’s. This presents no statistically significant relation between summaries noticing politeness and the point of view. In the next section we make some comparisons between both experiments.

4 Comparison

From both studies (Tables 2 and 3), we see that subjects reported the politeness in the summary mainly for the more impolite dialogues (D_2 and D_3). This fact apparently does not depend on the summary size. So, even if we restrict the summary size to a value as low as 10% of the number of words present in the dialogue, subjects will still include remarks about the emotional state of the interlocutors, related to their politeness, when the dialogues show more extreme negative behaviours.

If, however, we count the summaries in which subjects reported politeness (+ p) and those in which they do not ($- p$), and group them according to the study they belong to, we have evidence that the overall number of summaries reporting politeness depends on the study, i.e., it depends on whether they are restricted to a 10% summary size or not restricted at all. A $2 \times 2 \chi^2$ analysis revealed that this is a significant result, $\chi^2 (1, N=240) = 4.92$, at the significance level of $p < 0.05$. So, although in both studies we have evidence that people do report politeness when the dialogues are more extreme, i.e., the fact that politeness is worth mentioning for the more extreme dialogues does not depend on the study, the overall number of summaries reporting politeness decreases when the summaries have a 10% restriction in their length.

Table 4 summarizes the results. In this table, columns show the number of summaries separated by dialogue, while its rows distinguish between summaries of both studies.

	<i>Total</i>		<i>D₁</i>		<i>D₂</i>		<i>D₃</i>		<i>D₄</i>	
	+ p	$- p$	+ p	$- p$	+ p	$- p$	+ p	$- p$	+ p	$- p$
<i>Study 1</i>	60	60	2	28	25	5	28	2	5	25
<i>Study 2</i>	43	77	1	29	14	16	27	3	1	29

Table 4. Summaries mentioning politeness for both studies.

When we consider each dialogue separately, we find almost no statistically significant results, except for D_2 . We have non-statistically significant χ^2 values for D_1 , D_3 and D_4 , indicating that most probably the results do not depend on the study, for these three dialogues. However, if we apply the same test to D_2 , we have $\chi^2 (1, N=60) = 8.86$, which is significant at the significance level of $p < 0.01$. This indicates that, for D_2 , whether subjects report politeness may depend on the study. Actually, this high level of significance for D_2 may have polluted the data for the whole experiment, to the extent that, if we remove D_2 from consideration, we come up with the conclusion that reporting politeness did not depend on the study ($\chi^2 (1, N=90) = 0.87$, $p < 0.5$).

Thus, apparently, the restriction in summary size affected only D_2 . This result could be explained, perhaps, on the basis of pre-conceived expectations about the dialogue. Because in a customer-vendor interaction the vendor is expected to be polite and not the customer, subjects have dropped politeness when facing

a restriction on the summary size for D_2 . Another possible explanation for this phenomenon may be that D_2 is actually less impolite than D_3 , i.e., it may be not so “extreme”. This could make it more difficult for summarizers to decide how important it was to mention politeness in the summary. This is a matter that deserves further investigation.

A last point we want to report about this study is that the difference between the number of summaries including politeness in D_1 and D_4 for the first study disappears completely in the second study. So, if the fact that D_4 was presented after two unusual dialogues could have affected the summaries in the first study, it had no effect when the summary size was much restricted.

5 Conclusion

In this paper we described two exploratory studies designed to gather information about how people summarize dialogues. More specifically, we investigate how people deal with politeness in dialogues. Our eventual goal is to use the results to inform the construction of an automatic dialogue summarization system capable of taking dialogue politeness into account when summarizing. We hope that the resulting summaries will be considered more “natural”, since they should be closer to human produced summaries.

In both studies we observed that the percentage of summaries mentioning politeness increased as the dialogues became more impolite. This behaviour was independent of the point of view. The extent to which politeness is reported was influenced by whether subjects were restricted to a 10% summary or not for one of the four dialogues we presented to subjects (dialogue D_2).

Interestingly, for D_2 we observed that whether subjects report politeness depends on the permitted summary length. For some reason, when the summary length was restricted for D_2 , subjects were less likely to include politeness. This effect was not observed for the other impolite dialogue, D_3 . This suggests that there may be a difference in the politeness degree of D_2 and D_3 . We also noticed that the order in which the dialogues were presented may have had little influence on whether people included politeness in the summary. We found no statistically significant differences with respect to the reporting of politeness between the neutral dialogue before the two impolite dialogues (D_1), and the one after them (D_4).

Our main direction for future work is to investigate in detail *how* people report politeness in dialogue summaries; our final goal is to build an automatic dialogue summarizer. We would also like to determine whether people bias their summaries according to the point of view they are summarizing under.

The currently used dialogues are automatically generated. In further studies we would like to test the obtained results from these preliminary studies with more complex dialogues, like those produced by human writers for TV, theatre, etc. Finally, so far we have only dealt with politeness, neglecting other possible social rules. In the future, we would like to investigate dialogues in which social rules other than politeness are broken, generalizing the current results.

References

1. Hovy, E.: *Generating Natural Language Under Pragmatic Constraints*. Lawrence Erlbaum, Hillsdale, New Jersey (1988)
2. Craggs, R., Wood, M.M.: Annotating emotion in dialogue. In: *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue*, Sapporo, Japan (2003)
3. Fischer, K.: Annotating emotional language data. Technical Report 236, Verbmobil Project (1999)
4. Schmitz, B., Quantz, J.: Dialogue acts in automatic dialogue interpreting. Technical Report 173, Verbmobil Project (1996)
5. Reithinger, N., Kipp, M., Engel, R., Alexandersson, J.: Summarizing multilingual spoken negotiation dialogues. In: *Proceedings of the 38th Meeting of the Association for Computational Linguistics (ACL – 2000)*. (2000)
6. Huber, R., Batliner, A., Buckow, J., Nöth, E., Warnke, V., Niemann, H.: Recognition of emotion in a realistic dialogue scenario. In: *Proceedings of the International Conference on Spoken Language Processing*. Volume 1., Beijing, China (2000) 665–668
7. Nijholt, A.: *Humor and embodied conversational agents*. CTIT Technical Report series 03-03, University of Twente (2003)
8. Bickmore, T., Cassell, J.: How about this weather? social dialogue with embodied conversational agents. In: *Proceedings of the AAAI Fall Symposium on Socially Intelligent Agents*. (2000)
9. Bickmore, T.: Social dialogue is serious business. In: *Proceedings of the CHI 2002 Workshop on Socially Adept Technologies*. (2002)
10. Piwek, P.: An annotated bibliography of affective natural language generation. Technical Report ITRI-02-02, ITRI – University of Brighton (2003)
11. Krenn, B., Pirker, H., Grice, M., Baumann, S., Piwek, P., van Deemter, K., Schroeder, M., Klesen, M., Gstrein, E.: Generation of multimodal dialogue for net environments. In Busemann, S., ed.: *KONVENS 2002*, Deutsches Forschungszentrum fuer Kuenstliche Intelligenz (DFKI), Saarbruecken, Germany (2002) 91–98
12. Piwek, P., Krenn, B., Schröder, M., Grice, M., Baumann, S., Pirker, H.: RRL: A rich representation language for the description of agent behaviour in NECA. In: *Proceedings of the Workshop “Embodied conversational agents - let’s specify and evaluate them!”*, held in conjunction with AAMAS-02, Bologna, Italy (2002)
13. Piwek, P., van Deemter, K.: Dialogue as discourse: Controlling global properties of scripted dialogue. In: *AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*, Menlo Park, California, AAAI Press (2003) Technical Report SS-03-06.

Acknowledgments

We thank all the subjects for their volunteer work. We also thank Richard Power, Kees van Deemter, and Catalina Hallett for their valuable remarks on a draft for a previous version of this paper.

Appendix I

In this section we show the four dialogues used in the experiment, in the same order they were presented to the subjects.

Dialogue 1

Size: 99 words. Summary size: 10 words.

R: Hello! I am Ritchie.
T: Can you tell me something about this car?
R: It is very safe.
R: It has anti-lock brakes.
R: It has airbags.
T: How much does it consume?
R: It consumes 8 liters per 60 miles.
T: Interesting.
T: What kind of luggage compartment does this car have?
R: It has a spacious luggage compartment.
T: Excellent!
T: What kind of interior does this car have?
R: It has a spacious interior.
T: Excellent! (Tina)
T: How much does this car cost?
R: It costs 25 thousand Euros.
T: Well, well.
T: All in all this is a perfect car. It is a deal!
R: Really? I am sure you won't regret it.

Dialogue 2

Size: 182 words. Summary size: 18 words.

T: Anybody here? I haven't got all day.
R: I am sorry to have kept you waiting. I will be with you in a moment.
R: How can I help you?
T: I want you to tell me about this car.
R: It is very friendly for the environment.
R: It is made of recyclable materials.
R: It has a catalytic converter.
T: How much does it consume?
R: It consumes 8 liters per 60 miles.
T: It is not exactly very friendly for the environment.
R: Just consider this.
R: This car is very prestigious.
R: It has leather seats.
R: It has broad tires.
R: It drives 100 and 10 miles per hour.
R: It has 80 horse power.
T: What kind of luggage compartment does it have?
R: It has a spacious luggage compartment.
T: Well, well.
T: What kind of interior does this car have?
R: It has a spacious interior.
T: Well, well.
T: How much does this car cost?
R: It costs 25 thousand Euros.
T: It is not exactly very cheap.
R: Just consider this.
R: This car is very safe.
R: It has anti-lock brakes.
R: It has airbags.
T: Enough is enough! You keep your car!
R: Well, sorry that I cannot help you.

Dialogue 3

Size: 124 words. Summary size: 12 words.

T: Anybody in today?
R: Yeah yeah, keep your calm.
T: Can you tell me something about this car?
R: It is very safe.
R: It has anti-lock brakes.
R: It has airbags.
T: Does it have power windows?
R: What do you think? Of course not!
T: This car is not exactly very prestigious.
R: Just consider this.
R: This car has leather seats.
R: It has broad tires.
R: It drives 100 and 10 miles per hour.
R: It has 80 horse power.
T: What kind of luggage compartment does it have?
R: It has a spacious luggage compartment.
T: Well, well.
T: What kind of interior does this car have?
R: For god's sake, how should I know?
T: No problem.
T: Thank you for your help. I have to think a bit more about this.
R: Great! Thanks for wasting my time.

Dialogue 4

Size: 53 words. Summary size: 5 words.

R: Hello! My name is Richie.
R: What can I do for you?
T: Could you tell me something about this car?
R: This is a very sporty car.
R: It can drive 100 miles per hour.
T: How much does it consume?
R: It consumes 8 liters per 60 miles.
T: Not bad.
T: Ok! I'll take it.
R: Congratulations! Exquisite taste.