RC22690 (W0301-014) January 6, 2003
Computer Science

IBM Research Report

Controlled Generation for Speech-to-Speech MT Systems

Arendse Bernth
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Controlled Generation for Speech-to-Speech MT Systems

Arendse Bernth
IBM T.J. Watson Research Center
P.O. Box 218
Yorktown Heights
NY 10598
arendse@us.ibm.com

Abstract

In spoken dialog systems, a well-crafted prompt is important in order to get the user to respond with an expected type of utterance. We identify a new, important area for research in speech-to-speech translation, which focuses on the fact that the output of the MT system serves as the prompt for the user on each end. The MT engine used in speech-to-speech translation must pay special attention to its generation component, to such an extent that it makes sense to talk about controlled generation. Some rules for controlled generation are given.

1 Introduction

It is valuable to combine controlled language (CL) and machine translation (MT); see e.g. (Huijsen, 1998; Mitamura, 1999; Bernth and Gdaniec, 2001). Often text that is input to the MT system is constrained to conform to a specific CL that will make it easier for the MT system to perform well. In other words, CL is applied to the *input*. In this paper we propose a different scenario where CL is applied to the *output*. This seems of particular use for speech-to-speech MT systems, but other uses are not hard to imagine.

In spoken dialog systems (SDSs), a well-crafted prompt is important in order to get the user to respond with an utterance that the system can handle. In this paper we identify a new, important area for research in speech-to-speech translation, which focuses on the fact that the output of the MT system serves as the prompt for the user on each end. Since prompt design is extremely important in eliciting expected types of responses, the MT engine used in speech-to-speech translation must pay special attention to its generation component, to such an extent that it makes sense to talk about controlled generation. Thus this paper stands at the intersection of three important natural language technologies: Controlled language, machine translation, and speech processing. We will first, in Section 2, give some examples demonstrating the importance of a good prompt for SDSs. In Section 3 we relate MT output to dialog prompts, and in Section 4 we give some suggestions for CL rules in the context of speech-to-speech MT.

2 Importance of a Good Prompt

There are at least two reasons that a well-crafted prompt is important. One is to ensure that the system does not get a user response that it cannot handle, and the other is to ensure that the user does not get confused. Since the users of speech-to-speech MT typically come from different linguistic backgrounds with different linguistic conventions, it is particularly important to pay attention to the problem of potential confusion.

In this section we illustrate both cases with examples taken from the world of travel. For the target language parts, just the gloss in English is given.[1]

[1] Some of the examples are applicable to translations of other types of dialogs besides SDSs. For example text-to-text dialog, where the communication is by typing through a

(1) English-speaking airline ticket agent:
Do you have an e-ticket or a boarding pass?
MT translation in German:
Do you have an e-ticket ...
Traveller (barging in): *Yes.*
MT translation in English: *Yes.*
(Unless the system generates a non-recognition error message because the grammar is not expecting yes/no utterances)
English-speaking ticket agent:
So what do you have then, an e-ticket or a boarding pass??

Even with a perfect translation by the MT system, a disjunctive yes/no question is an ill-advised prompt in a speech system because of the mismatch that occurs between the system's expectations as expressed in its dialog management system, and the user's incomplete understanding of a prompt interrupted under barge-in. Additionally, the recipient may have trouble giving a suitable reply because of the ambiguity in a disjunctive question - as to whether the speaker is asking the hearer to name one of the disjuncts, or perhaps just to say whether the disjunction is true. An example of the latter might be: "Will you have a credit card or enough cash with you while you're on the trip, so that you can cover all the expenses?"

(2) English-speaking tourist:
Won't I need a reservation?
MT translation in Japanese:
Won't I need a reservation?
Japanese hotel receptionist: *Yes.*
MT translation in English: *Yes.*
English-speaking tourist:
How do I make a reservation?
MT translation in Japanese:
How do I make a reservation?'
Japanese hotel receptionist:
Sorry, I just said that you don't need a reservation; I don't understand.
Dialog manager: *I didn't get that; please repeat.*

The problem in this example is caused by the fact that the responses to negative yes/no questions are not semantically standard across languages. In English, the "yes" response means "Yes, you need a reservation." In other languages, e.g. Japanese or Chinese, there are "yes"-answer words that can express agreement with the negative proposition embedded in the question, thus creating the meaning "True, you do not need a reservation."[2] One possible solution to this problem would be for the MT system to convert a "yes" to a "no" (and vice versa) for language pairs that have different conventions. However, this solution does not adequately address the confusion that might occur if one of the speakers is aware of the difference and tries to take this into account when giving his reply, thus effectively misleading the MT system. This type of problem obviously interferes with a smooth conversation.

(3) First speaker's utterance, and its translation:
Do you know if there is an Italian restaurant nearby?
Second speaker's response, and its translation:
No.
First speaker:
You don't know, or there isn't?

Like Example 1, the example in (3) involves disjunction, but in this case the problem arises due to the ambiguity in replies to constructions involving a subset of what we might term "wh-raising verbs"; this subset includes verbs like "know", "decide" and "remember". If there is stress on the verb ("know"), the question is essentially disambiguated to be about that verb, but unstressed cases similar to this example, experienced by the author of this paper, show that humans can have trouble with this type of construction. Another example involving a wh-raising verb is given in (4).

(4) *Do you know when the next train to London leaves?*

The examples in (3) and (4) illustrate some problems caused by speech acts. Another example is given in (5):

(5) *Can you lift that suitcase for me?*

computer, as in online chat. Or human-computer interaction.

[2] Of course things are not always this simple, but this simple statement of the situation suffices to illustrate the point.

Here the issue is again an ambiguity between a yes/no reply and something else; in this case a request for some action. Speech acts in general seem to have potential for introducing problems and are worthwhile investigating in future work.

3 MT Output in the Role of Dialog Prompts

The examples in the previous section illustrate that even with perfect speech understanding and perfect MT, the interplay between the two can create a very imperfect speech-to-speech translation system. In dialog systems, a common technique for avoiding such problems is to design the system so that it has maximum control over the dialog. This is done by carefully designing the system prompts to constrain user input. Dialog Management, in particular how much freedom to give the user, is an important topic in Dialog Systems (Hochberg et al., 2002; Xu et al., 2002). As pointed out earlier, in a speech-to-speech translation system, the prompts stem directly from translations of user input, and hence the control is less, unless the system takes control over the output in a stricter manner than just rendering a faithful translation. So the system design for a speech-to-speech MT system has to pay careful attention to the generation component as it relates to the domain and even the individual state of the dialog. Otherwise, scenarios like those above will occur, or even worse.

A common measure of the quality of MT output is the extent to which it faithfully renders the source text in an idiomatic manner in the target language, without losing information or changing the style. For a speech-to-speech translation system, it may make more sense to state the ideal as a system that renders the information content perfectly in a smooth dialog, regardless of faithfulness to the exact way the source text presents it. Thus a speech-to-speech MT system may be better viewed as a sort of "multilingual information exchange system" than as a "traditional" MT system. This is in line with the idea of *translation objective* (Schmitz, 1997). "The translation objective specifies which aspects of a source-language utterance are to be rendered in a target-language utterance." In the Verbmobil scenario decribed in (Schmitz, 1997), the translation objective is modelled by dialog acts, examples of which are *Suggest a date*, *Claryifying answer*, and *Give reason*. As far as the author has been able to determine, the Verbmobil project did not specifically address the issue of MT output as dialog prompt.

So, instead of the usual idea of using CL to constrain the input text for MT, we are here talking about constraining the output. There are at least the following two possibilities for handling the situation: 1) Let the generation module of the MT system directly generate properly constrained output, or 2) post-process the output to adhere to the CL.

The first approach likely would take advantage of the recent advances in the field of natural language generation (NLG) (ACL2002, 2002). NLG traditionally uses a pipelined architecture (Reiter and Dale, 2000): First identify the communicative goal, then plan what to say, and decide how to say it. For MT, the communicative goal arises from the source language input to the MT system, obviously. In NLG, deciding how to say it often involves a so-called sentence-planner, and this would be the natural place for the MT system to constrain the output. Walker et al. (2002) specifically address the issue of sentence-planning for SDSs. Letting the MT system's generation module directly take care of constraining the output might be most appropriate for interlingual MT systems (Hutchins and Somers, 1992) such as the KANT system described in (Mitamura et al., 1991; Nyberg and Mitamura, 1992), or other generation-heavy systems such as the one proposed by (Habash, 2002; Habash and Dorr, 2002). The transfer-driven MT system described by (Yamada et al., 2000) involves some amount of controlled generation that takes place during transfer to handle the issue of politeness when translating spoken dialog from English to Japanese. The LSI-Trans engine described in (Montgomery and Li, 2002) is moving away from the earlier interlingual approach described in (Stalls et al., 1994) to a transfer-based approach, as part of streamlining the system to handle spontaneous dialogs. The issue of prompt management is not addressed in these papers, though.

The other approach, viz. making the MT output conform to the CL rules in a post-processing step, may use techniques such as the ones described in (Bernth, 1998) or (McCord and Bernth, 1998). The techniques in both cases involve exploring a parse tree. In one case, the parse tree is directly explored to identify undesirable constructions, and the text directly reformulated and substituted in the text; in the other case, a number of tree transformations are applied to the transfer tree structure as part of restructuring transfer in a transfer-based MT system. For controlling the output, the transformations could be applied in a post-processing step or as part of restructuring.

A completely different post-processing approach involves considering the task a type of automatic post-editing (Chander, 1998; Knight and Chander, 1994; Allen and Hogan, 2000; Allen, In press). In particular, Allen and Hogan (2000) make the connection between post-editing and controlled language, and describe an approach using tri-text files. The idea is to statistically train an automatic post-editor based on a set of source text, MT output, and post-edited texts. Viewing controlled generation as a kind of post-editing task is particularly appealing because it shows the parallelism with the common use of CL for *pre*-editing.

Brown (1998) describes an approach for post-editing SDS output that gives the user the option of making selections among alternative translations and updating the knowledge base dynamically, and then *back-translating* the resulting target utterance into the source language for verification purposes. Whereas this has the potential for giving a measure of confidence in the translation, it does not really address the matter of dialog management.

As can be seen, there are many different possibilities for implementing the control of the target text; the objective of this paper is to point out the need for controlling it, regardless of implementation, and in addition to suggest some concrete rules. The proposal is to control the MT output to provide a suitable prompt. Could the control be applied to the source text instead? Probably in some cases. However, depending on the differences between the source and target language, the problems may not show up until the translation. An example of this is shown in (6):

(6) English target: *Can I take the bus going to London?*
a. Spanish source: *Puedo tomar el bus que va a Londres?*
(Lit.: *Can I take the bus that goes to London?*)
b. Spanish source: *Puedo tomar el bus cuando voy a Londres?*
(Lit.: *Can I take the bus when I go to London?*)

The attachment of the present participle in the English target text is ambiguous and could in fact originate in two different (and unambiguous) Spanish source sentences.

Since the prompt appears in the target language, it seems more fruitful to make sure that the target utterance is suitable. Traditionally, CL has considered controlling vocabulary and syntax of *source* texts in order to reduce ambiguity. There is no reason that ambiguity should not also be considered for the target text.

4 Some CL rules for Speech-to-Speech MT

In this section we give some suggestions for what rules for speech-to-speech MT systems could look like, with reference to the examples given in Section 2. Of course, some of these rules will be language-specific; here we just show what they could look like for English target.

Example 1 gives rise to Rule 1.

Rule 1 *Do not translate disjunctive yes/no questions into disjunctive yes/no questions. Give the disjunctive content in the question as a hypothetical statement, followed by a yes/no question relating to only one of the disjuncts. If the answer is "no", then try the other disjunct.*

Hence our sentence from Example 1, "Do you have an e-ticket or a boarding pass?", would be handled as follows: "You could have an e-ticket or a boarding pass. Do you have an e-ticket?" If the reply is not affirmative, the system will ask "Do you have a boarding pass?"

This example illustrates an important difference between "traditional" NLG and controlled generation. "Traditional" NLG employs the technique of aggregation (Horacek, 2002; Lemon et al., 2002), whereas CL typically employs the technique of syntactic cues (Kohl, 1999). Aggregation reduces the text by combining phrases (or sentences) that share information into a single construction. Syntactic cues expand text by introducing elided items such as articles and heads in coordination. The two processes can be viewed as inverse. Aggregation is applied to make the text more natural; syntactic cues are added to make the text less ambiguous.

Example 2 causes us to suggest Rule 2:

Rule 2 *Do not translate yes/no questions with negations literally. Remove the negation.*

The negation may serve to set up a presupposition in only one of the two speakers in the dialog or differing presuppositions in both speakers. This presupposition may be removed without greatly affecting the propositional content of the question. If we apply this rule to Example 2, we simply get “Will I need a reservation?” instead of “Won’t I need a reservation?”

Examples 3 and 4 suggest Rule 3:

Rule 3 *Remove the wh-raising verb when it only serves as a politeness indicator.*

This verb serves as a politeness indicator and can be removed without greatly affecting the meaning. Thus, the question “Do you know if there is an Italian restaurant nearby?” could be rephrased as “Is there an Italian restaurant nearby?” And the question “Do you know when the next train to London leaves?” simply becomes “When does the next rain to London leave?”

Example 5 gives rise to Rule 4:

Rule 4 *Rewrite any polite requests involving modal verbs into an imperative of a suitably polite form.*

For English, the politeness might be indicated by “please”; hence “Can you lift that suitcase for me?” becomes “Please lift that suitcase for me.”

5 Conclusion

Earlier attempts at controlling MT output, e.g. the Translation Confidence Index described in (Bernth, 1999; Bernth and McCord, 2000), have focused on filtering out bad translations so that the user, typically a professional translator, would not be bothered with post-editing output that would be more trouble to correct than the effort involved in translating the segment from scratch. The current paper brings a new dimension to controlling MT output by arguing that the output of speech-to-speech MT systems needs to be controlled so as to supply well-crafted prompts useful for a spoken dialog system. Other uses for controlled generation include controlling the vocabulary, sentence length, or syntax to fit the reading skills of the intended audience.

Acknowledgements

The basic idea for this paper arose during a conversation with Kurt Godden, and the author wishes to thank him for his contribution as well as for valuable feedback on a draft of the paper. The author also wishes to thank Hideo Watanabe for help with the Japanese example, and Nelson Correa for help with the Spanish example, as well as the anonymous reviewers for helpful comments and suggestions.

References

ACL2002. 2002. *Proceedings of the Second International Natural Language Generation Conference (INLG-02)*. Association for Computational Linguistics.

Jeffrey Allen and Christopher Hogan. 2000. Toward the development of a postediting module for raw machine translation output: A controlled language perspective. In *Proceedings of the Third International Workshop on Controlled Language Applications, (CLAW-2000)*, pages 62–71, Seattle, WA.

Jeffrey Allen. In press. Post-editing. In Harold Somers, editor, *Computers and Translation: A Handbook for Translators*. Benjamins, Amsterdam.

Arendse Bernth and Claudia Gdaniec. 2001. MTranslatability. *Machine Translation*, 16:175–218.

Arendse Bernth and Michael C. McCord. 2000. The effect of source analysis on translation confidence. In John S. White, editor, *Envisioning Machine Translation in the Information Future, 4th Conference of the Association for Machine Translation in the Americas*, number 1934 in Springer Lecture Notes in Artificial Intelligence, pages 89–99, Cuernavaca, Mexico, October. AMTA.

Arendse Bernth. 1998. EasyEnglish: Addressing structural ambiguity. In David Farwell, Laurie Gerber, and Eduard Hovy, editors, *Machine Translation and the Information Soup, Third Conference of the*

Association for Machine Translation in the Americas, number 1529 in Lecture Notes in Artificial Intelligence, pages 164–173, Langhorne, PA, USA. Association for Machine Translation in the Americas, Springer.

Arendse Bernth. 1999. Controlling input and output of MT for greater user acceptance. In *Translating and the Computer 21*, ASLIB Conference Proceedings, page no page numbering, London, England. ASLIB.

Ralph Brown. 1998. Improving embedded machine translation with user interaction. Workshop on Embedded MT Systems, AMTA 98. Available on the web at http://www-2.cs.cmu.edu/ ralf/papers/amta98.ps.gz.

Ishwar Chander. 1998. *Automated Postediting of Documents*. Ph.D. thesis, University of Southern California.

Nizar Habash and Bonnie Dorr. 2002. Handling translation divergences: Combining statistical and symbolic techniques in generation-heavy machine translation. In Stephen D. Richardson, editor, *Machine Translation: From Research to Real Users, 5th Conference of the Association for Machine Translation in the Americas*, number 2499 in Springer Lecture Notes in Artificial Intelligence, pages 84–93, Tiburon, CA, USA. AMTA.

Nizar Habash. 2002. Generation-heavy machine translation. In *Proceedings of the Second International Natural Language Generation Conference (INLG-02)*, pages 185–190. Association for Computational Linguistics.

Judith Hochberg, Nanda Kambhatla, and Salim Roukos. 2002. A flexible framework for developing mixed-initiative dialog systems. In *Proceedings of the Third SIGdial Workshop on Discourse and Dialogue*, pages 60–63, University of Philadelphia.

Helmut Horacek. 2002. Aggregation with strong regularities and alternatives. In *Proceedings of the Second International Natural Language Generation Conference (INLG-02)*, pages 105–112. Association for Computational Linguistics.

Willem-Olaf Huijsen. 1998. Controlled language – an introduction. In *Proceedings of The Second International Workshop On Controlled Language Applications, (CLAW-98)*, pages 1–15, Pittsburgh, PA, USA.

W. John Hutchins and Harold Somers. 1992. *An Introduction to Machine Translation*. Academic Press.

Kevin Knight and Ishwar Chander. 1994. Automated postediting of documents. In *Proceedings of the National Conference on Artificial Intelligence*, pages 779–784. AAAI.

John R. Kohl. 1999. Improving translatability and readability with syntactic cues. *TechnicalCOMMUNICATION*, pages 149–166, Second Quarter.

Oliver Lemon, Alexander Gruenstein, Alexis Battle, and Stanley Peters. 2002. Multi-tasking and collaborative activities in dialogue systems. In *Proceedings of the Third SIGdial Workshop on Discourse and Dialogue*, pages 113–124, University of Philadelphia.

Michael C. McCord and Arendse Bernth. 1998. The LMT transformational system. In *Proceedings of AMTA-98*, pages 344–355. Association for Machine Translation in the Americas.

Teruko Mitamura, Eric Nyberg, and Jaime Carbonell. 1991. An efficient interlingua translation system for multi-lingual document production. In *Proceedings of the Third Machine Translation Summit*.

Teruko Mitamura. 1999. Controlled language for multilingual machine translation. In *Proceedings of Machine Translation Summit VII*, pages 46–52, Singapore.

Christine A. Montgomery and Naicong Li. 2002. Approaches to spoken translation. In Stephen D. Richardson, editor, *Machine Translation: From Research to Real Users, 5th Conference of the Association for Machine Translation in the Americas*, number 2499 in Springer Lecture Notes in Artificial Intelligence, pages 248–252, Tiburon, CA, USA. Association for Machine Translation in the Americas.

Eric Nyberg and Teruko Mitamura. 1992. The KANT system: Fast, accurate, high-quality translation in practical domains. In *Proceedings of COLING-92*.

Ehud Reiter and Robert Dale. 2000. *Building Natural Language Generation Systems*. Studies in Natural Language Processing. Cambridge University Press.

Birte Schmitz. 1997. The translation objective in automatic dialogue interpreting. In Christa Hauenschild and Susanne Heizmann, editors, *Machine Translation and Translation Theory – Perspectives of Co-operation*, pages 193–210. Mouton de Gruyter, Berlin, New York.

Bonnie G. Stalls, Robert S. Belvin, Alfredo R. Arnaiz, Christine A. Montgomery, and Robert E. Stumberger. 1994. An adaptation of lexical conceptual structure to multilingual processing in an existing text understanding system. In *Technology Partnerships for Crossing the Language Barrier, Proceedings of the First Conference of the Association for Machine Translation in the Americas*, pages 106–113, Columbia, Maryland, USA. Association for Machine Translation in the Americas.

Marilyn A. Walker, Owen C. Rambow, and Monica Rogati. 2002. Training a sentence planner for spoken dialogue using boosting. Available on the Web at www.research.att.com/ walker/spot-csl-5.pdf.

Weiqun Xu, Bo Xu, Taiyi Huang, and Hairong Xin. 2002. Bridging the gap between dialogue management and dialogue models. In *Proceedings of the Third SIGdial Workshop on Discourse and Dialogue*, pages 201–210, University of Philadelphia.

Setsuo Yamada, Eiichiro Sunita, and Hideki Kashioka. 2000. Translation using information on dialogue participants. In *Proceedings of the 6th Applied Natural Language Processing Conference and 1st Meeting of the North American Chapter of the Association for Computational Linguistics*, pages 37–43, Seattle, WA, USA. Association for Computational Linguistics.