ISCA Archive
http://www.isca-speech.org/archive

International Workshop on
Spoken Language Translation 2011
San Francisco, CA, USA
December 8-9, 2011

Speaker Alignment in Synthesised, Machine Translated Communication

Anne H. Schneider, Saturnino Luz

School of Computer Science and Statistics
Trinity College Dublin, Ireland
schneia@scss.tcd.ie, luzs@scss.tcd.ie

Abstract

The effect of mistranslations on the verbal behaviour of users of speech-to-speech translation is investigated through a question answering experiment in which users were presented with machine translated questions through synthesized speech. Results show that people are likely to align their verbal behaviour to the output of a system that combines machine translation, speech recognition and speech synthesis in an interactive dialogue context, even when the system produces erroneous output. The alignment phenomenon has been previously considered by dialogue system designers from the perspective of the benefits it might bring to the interaction (e.g. by making the user more likely to employ terms contained in the system's vocabulary). In contrast, our results reveal that in speech-to-speech translation systems alignment can in fact be detrimental to the interaction (e.g. by priming the user to align with non-existing lexical items produced by mistranslation). The implications of these findings are discussed with respect to the design of such systems.

1. Introduction

Understanding user expectations and behaviour when communicating with machines has long been regarded as an important element in the design of spoken language dialogue systems, being seen by many as having an impact comparable to that of good speech recognition and synthesis technology on the success of such systems [1]. Human factors research into the way people communicate in everyday dialogues has therefore drawn on linguistics and cognitive science theories (e.g. Grices' maxims, speech act theory) and empirical findings (e.g. memory constraints, lexical and syntactic alignment) to create system design guidelines. Similarly, applications such as speech-to-speech (S2S) translation, which employ speech technology to mediate human communication, stand to gain from the results of human factors research.

In (same-language) human communication, dialogue partners often align their linguistic behaviour to one another in terms of lexical and grammatical choices. Successful communication is more likely to occur when they become well aligned [2]. This phenomenon[1] has been extensively studied by linguists and psychologists. Findings of these studies have been taken up by language technology researchers and designers of spoken language dialogue systems, who exploited the fact that people can be shaped by the system's output in their lexical choices and phrase structures [3, 4, 5]. Techniques developed along these lines have been used for language modelling and the design of repair strategies. Given this background, one might expect that other forms of computer mediated dialogues such as communication in speech-to-speech translation could also benefit from the phenomenon of alignment. Contrary to that expectation, our study suggests that even though the phenomenon of alignment can be exploited to good effect in traditional spoken language dialogue systems, the same phenomenon can be detrimental to the performance of S2S translation systems. In other words, in certain situations, alignment is something designers of S2S translation systems will need to guard against.

The study presented below was motivated by an initial exploratory Wizard-of-Oz experiment designed to test the effects of machine translated output on the dialogue behaviour of users. In that study, we observed that participants regularly repeated the erroneous machine translation output when answering the system's questions. In order to investigate this incident of alignment to wrong and even non-existing words in greater depth, we performed a targeted study in which thirty participants were confronted with ten machine translated and speech synthesised questions. Five of those questions contained errors (e.g. non-existing words) in the main descriptors of the options from which the participants had to choose, and the other five contained only correctly worded options. The participants of the study were German speakers, and the lexicon included items such as word compounds, which are a likely source of mistranslation (in this case, from English).

We were interested in finding out whether people would align their answers to the wording presented in the questions or if they would correct the given alternatives or engage in some sort of attempted repair. Somewhat surprisingly, the results indicated that when presented with synthesised machine translation output people are indeed likely to adopt the mis-

[1] Different terms (e.g., entrainment, priming, lexical convergence) are used to refer to this phenomenon in the literature. For the remainder of this paper we will, however, only use the term 'alignment'.

takes they hear as part of their own speech and repeat these mistakes in their answers. This behaviour is likely to have a negative impact on the processing of the user response, by inducing speech recognition errors and their cascading effects on the machine translation module.

As S2S translation applications become more widespread, alignment issues need to be revisited so that their implications for the design of such interactive systems can be better understood. In particular, designers need a greater understanding of how to preserve the positive effects of alignment currently exploited in dialogue systems while avoiding its pitfalls.

The paper is structured as follows. In the following section, we briefly survey the related work on the phenomenon of alignment and its exploitation in human-computer interaction, with emphasis on dialogue systems. We then describe our preliminary Wizard-of-Oz study and the initial observations it prompted. The targeted controlled question-answering experiment is presented in the following section, along with detailed reporting of its results and a general discussion. The paper concludes with the implications of our findings for the development of S2S translation systems, and directions for future work.

2. Related work

The phenomenon of linguistic alignment has been researched in the fields of linguistics (especially in the area of corpus linguistics) and cognitive science. Alignment in human-human communication can range from the word level to sentence level. Brennan and Clark use the term *lexical entrainment* to refer to the fact that when two people repeatedly discuss the same object, their lexical choices tend to converge. In doing so people achieve conceptual pacts, or shared conceptualizations, which they mark by using the same terms [6]. Evidence of alignment on a sentence level can also be found. In such cases the term *priming* is usually used, which refers to a process that influences linguistic decision-making, where a linguistic choice (prime) of a speaker influences the recipient to make the same decision, i.e. re-use the structure, at a later choice-point [7].

In human-computer conversation, data have also been collected which confirm alignment. Brennan reports on a Wizard-of-Oz experiment using a database query task in which *lexical convergence* with computers is discovered [4]. Results of an experiment by Branigan et al. show that alignment occurs whether participants believe they are interacting with another human participant or with a computer [8]. These studies traditionally involve grammatically correct and error free language. However, Bortfeld and Brennan [9] observed a degree of *conceptual entrainment* in human-human dialogues, where native English speakers produce non-idiomatic referring expressions in order to ratify a mutually achieved perspective with non native speakers. Although this phenomenon is somehow related to the participant behaviour we observed in our study, we are not aware of any other studies where alignment is investigated in the context of computer-mediated spoken language translation.

As mentioned above, the research on alignment, especially the observation that people can be shaped in their lexical choices so as to use vocabulary and phrase structures aligned to certain system outputs, has influenced work in the area of spoken language dialogue systems. In the early nineties, a study tested whether people can be shaped to use vocabulary and phrase structure of a program's output [3]. The study found that users of natural language programs will model the program's output, that there is no difference in modelling or shaping effects between spoken and typed inputs, and that it is easier for people to model both the length and the vocabulary of a terse computer output than of a conversational computer output. Attacking the vocabulary problem, i.e., the potential for enormous variability in dialogue, Brennan summarizes the results of a series of experiments, where visually separated pairs of people had to line up identical sets of picture cards in the same order, and two Wizard-of-Oz experiments using a database query task. The results of these studies are discussed with respect to their implications to modelling and constraining lexical variability in spontaneous human-computer dialogue [4]. The vocabulary problem is also targeted by Gustafson et al., who report on Wizard-of-Oz experiments, which show that people mostly adapt their lexical choices to system questions, and investigate the possibility to add an adaptive language model to the speech recognizers in a spoken dialogue systems [5].

It would be desirable to assess the implications of those findings with respect to other applications that also focus on computer mediated communication. The research in speech-to-speech translation, which aims at producing real-time translation for people who do not share a common language, for example, has expanded considerably in recent years. Many research projects in this area have been conducted, including VERBMOBIL, ATR'S MATRIX project and NESPOLE! (see [10] for an overview). This development was triggered by an increasing demand for translingual communication in a globalised world and on advances in component technologies such as machine translation and speech recognition. A similar process was provoked by the emergence of the European Union with projects like EUROMATRIX[2], EUROMATRIXPLUS[3] and META-NET[4].

Broadly speaking, S2S translation systems usually comprise three modules through which data are processed [11]. The input of a human speaker is recognised by an automatic speech recognizer (ASR) and transcribed into natural text. This text is translated into the communication partner's language using machine translation (MT), and finally synthesised using a text-to-speech system (TTS). Traditionally, evaluation of S2S systems has been done either end-to-end like in the TC-STAR project [12], or focussed on the intrin-

[2] http://www.euromatrix.net/
[3] http://www.euromatrixplus.net/
[4] http://www.meta-net.eu/

sic performance of the individual modules using automatic metrics like the BLEU score [13] or word error rate (WER) for example. In addition to the efforts to evaluate the combination of ASR and MT, namely spoken language translations, in evaluation campaigns, there have been investigations on the influence of machine translation on the intelligibility of TTS output in S2S translation systems [14] and on how the MT output can be altered to improve the performance of the TTS [15]. However, the fact that communication using a S2S system is a two way process that involves humans is often overlooked. Our study indicates that it is also necessary to take into account the influence of the S2S output on the ASR component.

3. Background: Preliminary Wizard-of-Oz Experiment

An initial Wizard-of-Oz (WOZ) experiment was conducted, which mainly aimed at comparing extrinsic and intrinsic MT evaluation methods in an interactive context [16]. In this study, participants conversed with a (simulated) dialogue system in order to get information about an Internet broadband product. The utterances of the system were synthesised German machine translations of appropriate English answers. One of the questions asked by the system required the participant to decide between one of two options. The original utterance was *'Are you looking for mobile Internet or a landline connection?'*. Option A, *mobile Internet*, and Option B, *landline connection*, where mistakenly translated to *bewegliches Internet* and *Überlandleitungsverbindung* respectively. The first mistake stems from the ambiguity of the word mobile, which in this context should have been translated to *mobil*, The machine translation system, however, interpreted mobile in the sense of *agile* or *flexible*. The second mistake stems from the rule based nature of the machine translation software we used. The system binds the translations for *landline* (*Überlandleitung*) and *connection* (*Verbindung*) into a compound that does not exist in the German language. The correct translation would have been *Festnetzanschluss*.

These mistakes are obvious and would have been easily noticed by fluent (native) speakers of the German language. Therefore, we were surprised to find that all participants in the experiment (all native speakers of German) repeated the wrong words in their answers. In a post-experiment interview, we inquired the participants about this behaviour. Most of them stated that they knew that the words were incorrect and they were aware of the intended correct meaning due to the context, but they concluded the system would understand these words since it came up with them in the first place. They even explained that they assumed that the system would only understand these mistakes.

4. Targeted question-answering study

In order to investigate this issue further, we conducted a study in a more constrained setting, with a simple targeted question-answering set-up. Participants were asked to answer questions similar to the one in the preliminary WOZ experiment in which we made our initial observation. The questions were set up so as to enable us to assess the extent to which the alignment behaviour observed in the WOZ study would persist.

4.1. Methods

Ten dual choice questions were prompted to the participants, who were asked to choose one out of two options (e.g., *Would you rather travel to Rome or to Athens?*). The questions were translated from English into German using the online machine translation service of SYSTRAN[5]. Our aim was to investigate how people would react to obviously wrong translations, in particular, whether they would correct the mistakes or adopt them in their answers. Therefore, we had to force the participants to choose between incorrect options. In order to do so, we adjusted the questions as follows: in five of the sentences we adjusted the two options, so that they would both contain mistakes, and in the other five questions the two options were corrected so that they contained no errors. This setting ensured that in five cases participants had to select an erroneous option. The five sentences containing error-free options served as fillers and helped to mask the true intent of the study.

We kept the machine translations and only changed single words (the options). We did not alter the rest of the sentence. Therefore, mistakes of the translation system were preserved. In the ten cases (=5 sentences $\times$ 2 options), where we adjusted the options we emulated mistakes that were seen before, since predicting errors of the machine translation system is not a trivial endeavour. The options that were changed were simple nouns. We wanted our introduced mistakes to resemble the mistakes observed in the preliminary study, especially the one where the system produced a word that cannot be found in a German dictionary. We had to make sure that the mistakes would be identifiable when uttered by the participants for our analysis. However, another aim was to keep the wrong words comprehensible. Therefore, we only introduced additional letters or syllables into the words or changed an umlaut. We also took care that there were similar mistakes for both options within one question.

Following the above described alterations, the resulting ten questions were synthesised using the MUSE text-to-speech system [17] and recorded. An overview of the German questions and their original English version can be seen in Table 1.

The experiment was set up as a question-answering session. We prepared twelve presentation slides. The first and last slide contained an introduction and a debriefing respec-

[5]http://www.systranet.com/

Sentences with correct options	
1	Würden Sie mögen eher nach Rom oder nach Athen reisen? *(Would you rather like to travel to Rome or to Athens?)*
2	Finden Sie das Schach oder Fussball aufregender? *(Do you find chess or football more exciting?)*
3	Würden Sie eher Albert Einstein oder Marie Curie treffen vorziehen? *(Would you rather want to meet Albert Einstein or Marie Curie?)*
4	Mögen Sie eher zum Rockmusik oder Popmusik hören? *(Do you prefer to listen to Rock or Pop music?)*
5	Mögen Sie eher Früchte oder Gemüse? *(Do you prefer fruits or vegetables?)*
Sentences with incorrect options	
6	Finden Sie **olivengrün** oder **goldengelb** schöneres Farbe? *(Do you like olive green or golden yellow more?)*
7	Essen Sie eher mögen **Blauenbeerpfannkuchen** oder **Schweinerhaxe**? *(Would you rather eat blueberry pancakes or knuckle of pork?)*
8	Lesen Sie eher **Liebe-Geschichten** oder **Verbrechen-Bücher**? *(Do you prefer to read love stories or detective stories?)*
9	Konnten dir Sie eher vorstellen, um **Frösche-Schenkel** zu essen, oder zu essen brodelnde **Fledermause**? *(Could you rather imagine to eat frog legs or boiled bats?)*
10	Tun Sie mögen **Montagen** oder **Donnerstagen** besser? *(Do you prefer Mondays or Thursdays?)*

Table 1: The ten sentences used in the experiment. The erroneous options in the last five sentences were highlighted. The English version is shown in brackets after each question.

tively. One of the ten questions was on each of the remaining ten slides. The order in which the questions were presented was randomized, so as to control for order effects. The participants were instructed that they would get one of ten machine translated and synthesised questions at a time, which they had to answer aloud before they could proceed to the next question. The transition to the next question automatically triggered the playback of the synthesised question. All answers were audio recorded and later transcribed.

Based on the results of the initial experiment, we hypothesised that the participants would repeat the erroneous options rather than attempt to make a correction.

4.2. Results

The experiment was conducted with thirty native speakers of the German language, all aged between 25 and 42. Thirteen of the participants were female, seventeen were male. We do not expect gender or age to have an impact on the results of this study.

In total, each participant had to choose ten out of twenty options. Five times, they were forced to choose between options described by sentences containing incorrect words as the main descriptor. In the other five cases, they had to choose from correct options. Some of the options were picked only by few participants (e.g., *Fledermausen*, the incorrect translation of bat) whereas others were quite popular (e.g., *Blauenbeerpfannenkuchen*, the incorrect version of blueberry pancake). Figure 1 shows an overview how often the options in the five sentences with incorrect options were chosen.

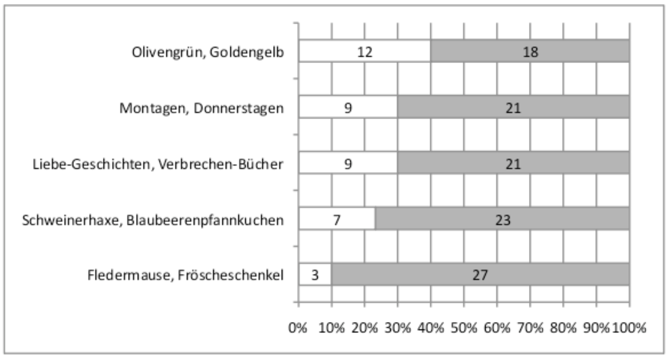

Figure 1: Distribution of incorrect options as selected by the participants.

In what follows, we will concentrate on the responses to sentences containing incorrect options. In about half the cases, participants corrected the mistakes in the wrong options. In 70 out of the 150 (= five sentences $\times$ thirty participants) cases, in which participants were facing a wrong translation, they aligned their answer to the word used in the wrong option. Only six participants corrected all five wrong options, three of the participants, on the other hand, aligned to the options completely.

Participant 13 intended to correct *Verbrechen-Bücher* but failed by uttering the word *Verbrecher-Bücher*, which also does not exist. The correct translation would have been *Kriminalroman* or *Krimis*. Participant 6 (who answered in whole phrases) not only repeated the mistakes in four out of five cases but also copied wrong phrases from the questions with the right options. In a question rendered by the system in the wrong word order, for example, he repeated the option and the verb phrase. The option was lexically correct but the participant aligned to the wrong word order (e.g. question: *Would you Marie Curie or Albert Einstein rather meet?* participant's answer: *Marie Curie rather meet*.). Participant 30 repeated the mistakes first and later corrected them (e.g., '*Fröscheschenkel ... Froschschenkel*'). A similar observation was made when participant 26, who corrected all options, started to utter *Verbrechen...*, then paused and corrected the answer to *Krimis*. We counted these occurrences as corrections.

Wrong word	alignment	cor.	false cor.
Verbrechen-Bücher	17	3	1
Olivengrün	11	1	0
Goldengelb	10	8	0
Frösche-Schenkel	9	18	0
Blauenbeerpfannkuchen	7	15	1
Donnerstagen	6	15	0
Montagen	4	5	0
Liebe-Geschichten	3	6	0
Fledermause	2	1	0
Schweinerhaxe	1	6	0
Total	70	78	2

Table 2: Overview of the ten erroneous options sorted by the number of participants who aligned with them (first column). The second column shows how often the participants corrected (cor.) the option, and the last column shows the number of false corrections (i.e., failed attempts by the participant to correct a wrong word).

Certain words were singled out for alignment more often than others. The option *olive green (olivengrün)*, for example, was chosen twelve times out of thirty and only corrected once. An overview of how often participants aligned to the options or corrected them can be found in Table 2. Due to the varying frequency of choice, we calculated the alignment rate relative to the number of times that every option was chosen. For an overview how often an option was chosen and how often participants aligned to the chosen option, as well as the relative alignment to the number of times that the option was chosen, please refer to Figure 2.

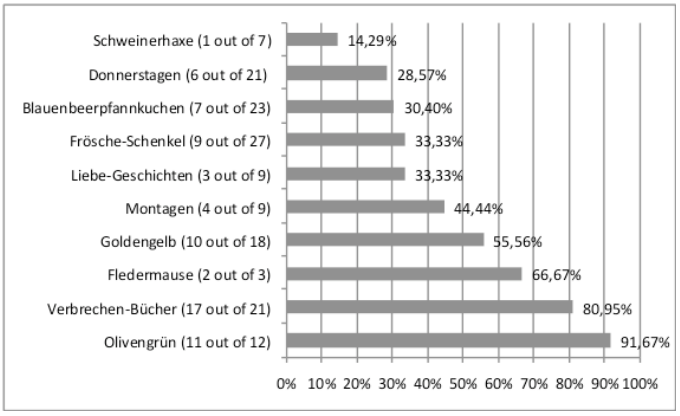

Figure 2: Alignment values relative to the number of times that the participants chose the option.

On the other hand, it is also interesting to notice the cases in which participants corrected the mistakes most often. For example *knuckle of pork (Schweinerhaxe)*, which was chosen seven times, was corrected in six cases. Table 3 shows an overview of the options that were corrected in more than 50% of the cases in which the respective answers were picked.

Wrong option	picked	correction (rel. value)
Schweinerhaxe	7	85.71%
Donnerstagen	21	71.43%
Blauenbeerpfannkuchen	23	69.57%
Frösche-Schenkel	27	66.67%
Liebe-Geschichten	5	66.67%
Montagen	9	55.56%

Table 3: Overview of options that were corrected in more than half of the cases in which they were picked. The second column shows how often the participants picked the option. The third column shows the relative value of correction.

5. Discussion

The results of the targeted question-answering experiment support our initial hypothesis that people would align their behaviour and repeat incorrect words (and occasionally incorrect syntax) when presented with machine translated and synthesised text.

In 47% of the cases, where participants had to choose between two options described by incorrect words they simply used the wrong terms instead of attempting a repair. One participant went as far as copying the wrong word order rendered by the MT system in the questions. Only 20% of the participants attempted repairs in all cases, with one failing to correct an option. A failed attempt to correct a word (false correction) could also be observed in another case.

The results also show that some words are more likely to be corrected than others. We speculate that this is due to the specific type of error introduced in the option. The translation of *blueberry pancake*, for example, was corrected in most of the cases. This might be due to the introduction of an extra syllable at the end of the translation of the word *blue* (*blau*). This syllable increases the effort needed to utter the word and disrupts its fluency. In the other cases, however, the fluency of the option was preserved. Although it is tempting to attribute this observation to a general phonetic constraint, we note that since the experiment employed only a small number of questions to be answered in an otherwise de-contextualised situation a generalisation along these lines would be premature. A different experimental setting and further data would be needed in order to better investigate this hypothesis.

Since user awareness of the occurrence of miscommunication is often crucial to the success of computer mediated communication, from a practical system's design perspective, a deeper investigation into error types and their likelihood to trigger alignment in a fully contextualised situation (e.g. a task-oriented dialogue mediated by MT) would also be desirable. A complementary study could also assess the effect of the output modality (text versus speech) on user alignment to incorrect words.

Although details on the nature and effects of linguistic alignment to translated content still need eliciting, general implications of the findings reported above to the design of S2S translation systems can be pointed out. It would seem necessary, for instance, to better integrate speech recognition and machine translation modules so that the latter is prevented from producing output that might lead to recognition errors further down in the interaction sequence. Recent work on MT-ASR integration might help address this issue, at least in part. Jiang et al., for instance, propose a method for tight coupling between ASR and MT to enhance speech translation performance [18]. The strength of their proposed strategy is that the MT system can recover from ASR errors before an ASR error is translated into an incorrect word. However, their approach carries information only one way, from ASR to MT. Our results indicate that there is also a need for a back channel from the machine translation to the ASR component.

Interaction design might also play an important role in avoiding problems due to alignment. Similarly to Brennan's recommendation for spoken dialogue systems, that the language system should present only terms in the output that the system can process as input [4], we suggest that in the first instance output of the MT component should be adjusted to the capabilities of the ASR component. Where this is not entirely possible, the system should implement robust mechanisms for meaning negotiation between the conservational partners. Given the bi-directional nature of the ASR and MT channels it might be interesting, for example, to keep a history of translations (and possible mistranslations) occurring in the dialogue so as to ensure consistency, thereby emulating at the system level the sort of alignment behaviour users tend to exhibit. However, this might sometimes be at the cost of lexical appropriateness in the use of terms by the system. Such mechanisms could then perhaps feed information back to the machine translation and speech components in order to mitigate further problems.

6. Conclusions

Machine translation is still far from perfect, and it is likely that in a S2S translation scenario mistakes in the MT output will occur. Our findings suggest that in such cases there is a good chance that people will adopt such mistakes as lexical items.

The goal of an ASR system is to map from an acoustic signal to a string of words. In order to perform this mapping, the acoustic signal is matched to an entry in a dictionary, which contains the word pronunciations, or the signal is split into phonemes (i.e., smallest segmental units of sound) and these are matched to graphemes (i.e., fundamental units in a written language). In situations where one communication partner repeats MT output that can not be found in the dictionary, the performance of the ASR module will degrade. In spoken language translation the ASR error rate should generally be lower than for other applications which involve ASR, such as information retrieval tasks for example, because the ASR output constitutes the input for the machine translation component [11]. However, even if recognizers were able to match the phonemes to the correct string of graphemes this would only pass the problem on to the machine translation module, which would produce an incorrect back translation. If the mistranslation of *landline connection (Überlandleitungsverbindung)*, for instance, was back translated by the same system the result would be *overhead power line connection*, which would obviously cause semantic problems to the conversational partners.

In [19], Brennan surveys aspects of interactive language use, among them the entrainment phenomena, and discusses implications for computational linguistics and human computer interaction. She suggests that future dialogue interfaces should include resources to enable users to negotiate meanings, model context and recognize which referring expressions are likely to index a particular conceptualization. The findings of the experiment we presented in this paper show that there is a flip-side to this argument, namely that potential problems due to alignment might arise when the machine assumes a mediating role. Their general implication is that countermeasures should be contemplated by designers of speech to speech translation systems.

The aim of our study was to confirm the occurrence of alignment effects in a MT mediated setting. In order to be able to do that we had to constrain the data employed in our experimental setting to a few clear-cut and comparable sentences. However, as we mentioned above the small linguistic sample and the de-contextualised nature of the task the participants were asked to perform prevents us from making broad generalisations. While this sort of trade-off between coverage and testability is quite common in studies and evaluations of applications of language technology, we acknowledge that considerably more investigation into human factors is needed in this area.

Therefore, we intend to further study issues concerning linguistic alignment in contexts involving computer mediation through machine translation, speech recognition and speech synthesis. Our more immediate plan is to investigate in which cases people are more likely to align to the mistakes. Results of this investigation could help to avoid such error types in the output. For the future, we also plan to examine if our conclusions also apply to other languages and cultural settings. The next logical step is to repeat the study with English speakers. Finally, we would like to explore the effects of presentation modality (speech, text, graphics) on the user's tendency to align to incorrect output.

7. Acknowledgements

This research is supported by the Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (www.cngl.ie) at Trinity College Dublin.

8. References

[1] L. Dybkjaer and N. O. Bernsen, "Usability Issues in Spoken Dialogue Systems," *Natural Language Engineering*, vol. 6, no. 3&4, pp. 243–271, 2001.

[2] M. J. Pickering and S. Garrod, "Toward a mechanistic psychology of dialogue," *The Behavioral and Brain Sciences*, vol. 27, no. 2, pp. 169–90, 2004.

[3] E. Zoltan-Ford, "How to get people to say and type what computers can understand," *International Journal of Man-Machine Studies*, vol. 34, no. 4, pp. 527–547, 1991.

[4] S. E. Brennan, "Lexical Entrainment in Spontaneous Dialog," in *Proceedings of 1996 International Symposium on Spoken Dialogue (ISSD'96)*, 1996, pp. 41–44.

[5] J. Gustafson, A. Larsson, R. Carlson, and K. Hellman, "How do System Questions Influence Lexical Choices in User Answers," in *Proceedings of Eurospeech'97*, 1997.

[6] S. E. Brennan and H. H. Clark, "Conceptual Pacts and Lexical Choice in Conversation." *Journal of Experimental Psychology. Learning, Memory, and Cognition*, vol. 22, no. 6, pp. 1482–93, 1996.

[7] D. Reitter, F. Keller, and J. D. Moore, "Computational Modelling of Structural Priming in Dialogue," in *Proceedings of the Human Language Technology Conference of the NAACL '06*, 2006, pp. 121–124.

[8] H. P. Branigan, M. J. Pickering, J. Pearson, J. F. Mclean, and C. I. Nass, "Syntactic alignment between computers and people : The role of belief about mental states," in *Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society*, 2003, pp. 186–191.

[9] H. Bortfeld and S. Brennan, "Use and Acquisition of Idiomatic Expressions in Referring by Native and Non-Native Speakers," *Discourse Processes*, vol. 23, no. 2, pp. 119–147, 1997.

[10] Y. Zhang, "Survey of Current Speech Translation Research," in *Multilingual Speech-to-Speech Translation Seminar*, Carnegie Mellon University, Pittsburgh, PA, 2003.

[11] A. Waibel and C. Fügen, "Spoken Language Translation," *IEEE Signal Processing Magazine*, no. May, pp. 70–79, 2008.

[12] O. Hamon, D. Mostefa, and K. Choukri, "End-to-End Evaluation of a Speech-to-Speech Translation System in TC-STAR," in *Proceedings of MT-Summit XI*, Copenhagen, 2007, pp. 223–230.

[13] K. Papineni, S. Roukos, T. Ward, and W.-j. Zhu, "BLEU: a Method for Automatic Evaluation of Machine Translation," in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL) 2002*, no. July, Philadelphia, 2002, pp. 311–318.

[14] L. M. Tomokiyo, K. Peterson, A. W. Black, and K. A. Lenzo, "Intelligibility of Machine Translation Output in Speech Synthesis," in *Proceedings of Interspeech'06*, 2006, pp. 2434–2437.

[15] A. Parlikar, A. W. Black, and S. Vogel, "Improving Speech Synthesis of Machine Translation Output," in *Proceedings of Interspeech'10*, 2010, pp. 194–197.

[16] A. H. Schneider, I. van der Sluis, and S. Luz, "Comparing Intrinsic and Extrinsic Evaluation of MT Output in a Dialogue System," in *Proceedings of the 7th International Workshop on Spoken Language Translation (IWSLT'10)*, Paris, France, 2010, pp. 329–336.

[17] P. Cahill and J. Carson-Berndsen, "MUSE: An Open Source Speech Technology Research Platform," in *Proceedings of the IEEE Workshop on Spoken Language Technology (SLT'10)*, 2010.

[18] J. Jiang, Z. Ahmed, J. Carson-Berndsen, P. Cahill, and A. Way, "Phonetic Representation-Based Speech Translation," in *Proceedings of Machine Translation Summit XIII*, Xiamen, China, 2011.

[19] S. E. Brennan, "Processes that Shape Conversation and their Implications for Computational Linguistics," in *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics (ACL '00)*, 2000.