



BRILL

CONTRASTIVE PRAGMATICS 1 (2020) 82–107

CONTRASTIVE
PRAGMATICS
A Cross-Disciplinary Journal
brill.com/jocp

Experimental Contrastive Pragmatics Using Robots

Kerstin Fischer

Department of Design and Communication, University of Southern
Denmark, Alsion 2, 6400 Sønderborg, Denmark
kerstin@sdu.dk

Alicja Depka Prondzinska

Department of Design and Communication, University of Southern
Denmark, Alsion 2, 6400 Sønderborg, Denmark
depkadana@gmail.com

Abstract

In this paper, we explore how robots can be used to study pragmatic strategies across a number of languages. Robots can assume many of the roles played by human interaction partners in a range of situations. They can be programmed to produce specific behaviours, each time repeating a behaviour in an identical way for as often as necessary. Thus, robots can be useful tools for investigating human behaviour in certain situations and even in cross-cultural contexts. We explore this use of robots in two case studies – one which investigates the delivery of bad news in Danish, German and English, and one which examines the giving of feedback in Danish, German and Polish. In both studies, systematic intercultural differences become apparent in the pragmatic strategies that are adopted. On the basis of the results, we discuss the advantages, potential pitfalls and possible solutions of using robots in the study of contrastive pragmatics.

Keywords

intercultural communication – human-robot interaction – bad news – feedback –
Danish – English – German – Polish

© KERSTIN FISCHER AND ALICJA DEPKA PRONDZINSKA, 2020 | DOI:10.1163/26660393-BJA10004

This is an open access article distributed under the terms of the CC BY-NC-ND 4.0 license.

1 Introduction

The study of social practices can be problematic as these practices are contingent on many contextual factors – in addition to the general difficulties posed by cross-linguistic studies (cf., for instance, Blum-Kulka et al., 1989). Interactional effects, such as the sequential structure of an interaction, interpersonal effects, i.e. the relationship between interlocutors, and the behavioural effects of certain linguistic choices, for example, the extent to which the linguistic choices make an utterance more persuasive, are generally difficult to study comparably, let alone across a number of languages and cultures.

In this paper, we propose that, by using robots to experiment with different behaviours, the interactional, interpersonal and behavioural effects of linguistic choices (Fischer, 2016a) can be identified. Robots are embodied agents and thus function as real interactional partners and, unlike humans, can be completely controlled. Their appearance, gender, height, behaviour and voice can be changed at will. Furthermore, robots can reproduce identical behaviours with each interactional participant, regardless of how often this is required – unlike human confederates, who are unable to control their behaviour to the degree that identical information is conveyed to each participant (Brennan and Hanna, 2009). For instance, Lockridge and Brennan (2002) replicated a study on the effects of providing feedback on storytelling that had been undertaken using confederates (Brown and Dell, 1987), but instead they used naïve subjects as listeners. They discovered significant differences between the two studies and thus concluded that the confederates, having heard the same story many times before, implicitly communicated that they did not need require further information. This skewed the results of the study, which could therefore not be replicated when using subjects who had not previously heard the story. Consequently, confederates are unable to control their behaviour to the degree that all participants are faced with comparable communication partners.

At the same time, robots are somewhat restricted in what they can produce; for instance, many robots have only a limited range of facial expressions (for example, few robots have eyebrows which can be raised, cf. Phillips et al., 2018). However, a robot would not be judged as negatively as a human being for not using facial expressions when communicating emotions. This allows for more controlled, and hence comparable, interactions than is possible in experiments with human confederates.

Moreover, confederates who are fluent in several languages can be difficult to find, whereas robots can be programmed to speak in almost any language,

using text-to-speech systems; many free speech synthesisers provide a wide range of different languages in similar voices.

Finally, because they are embodied and potentially autonomous agents, robots can assume many of the roles played by human interaction partners, thus allowing the controlled investigation of sequences of interaction and the identification of interpersonal and behavioural effects in a diverse set of situations. For example, Fischer et al. (2020) used two identical robots with two different speech profiles, based on the speech profiles of Steve Jobs and Mark Zuckerberg respectively, to make suggestions for places to visit during a trip to Paris. Participants were able to choose where to visit, and thus whose advice to follow. The results showed that participants more frequently followed the advice provided by the robot whose voice was based on the speech characteristics of Steve Jobs. Thus, robots not only permit the controlled investigation of how people perceive certain speech characteristics, but also allow the effect that these speech characteristics have on people's behaviour to be measured.

It is clearly evident that the use of robots in the study of contrastive pragmatics is not without its problems. In the following sections, we present two case studies. Firstly, we investigate the effect of employing empathy signals when delivering bad news and, secondly, we study the practices that are used for providing feedback in a number of languages. As a result of these case studies, we will be in a better position to assess the opportunities that are provided by the use of robots in this type of research. We will also consider any potential pitfalls and problems that may arise and will detail how these can be addressed.

2 Study 1: Delivering Bad News

The practice of delivering bad news has been rather extensively studied from a conversation analytic perspective by Maynard (e.g. 1991, 2017; Maynard and Frankel, 2006). Maynard (1997), for instance, identifies four phases of news delivery, namely announcement, response, elaboration and assessment. In the announcement phase, the speaker provides the listener with an indication of how s/he should respond, depending on whether the news is 'good' or 'bad'; this allows the listener to predict the form of response that is expected by the speaker. While Maynard investigates many different situations, including medical scenarios, his findings are based on US contexts, and thus it is unclear to what extent the delivery of bad news differs across countries.

From our own intercultural experience of working in the Danish-German border region, we discovered situations where native Danish speakers violated our own expectations (the first author having been raised in Germany) regarding the provision of bad news. For instance, during a visit at the university's IT-service desk, we encountered messages such as "we can't help you", but from a German perspective, we would have expected a phrase such as "I'm sorry, but ...". Informal conversations with our Danish colleagues regarding this observation suggested a possible link with assuming (a lack of) responsibility for the problem, and we therefore decided to investigate this observation further.

To confirm whether there is a cross-cultural difference, we first created a brief discourse completion task; this method, although widely used in contrastive pragmatics (e.g. Labben, 2016), has also been criticised for not producing similar data to authentic data collection (e.g. Cyluk, 2013). However, in this case, we only used this method to initially investigate potential differences, and to elicit formulations which could be used in subsequent experiments.

We created three settings in which the participant assumes the role of a person delivering bad news. The tasks differ according to the participant's level of involvement in the institution that is responsible for the problem: in one task, the participant is an employee of the institution, in another task, s/he is only loosely connected to the institution, and in the third task, there is no direct connection between participant and institution:

1. You are working for DSB/DB/Amtrack and a passenger comes up to ask you about a train that should arrive any minute. You know that the train will be two hours late (because of a broken engine) and that there is no alternative train. You say (in Danish/German/English):
2. A customer comes into a shop in which you help out occasionally. She asks for almond chocolate. While the shop normally carries almond chocolate, you happen to know that the supplier has failed to deliver it three times in a row. You answer (in Danish/German/English):
3. You are waiting at a bus stop when a bus comes in that you don't want to take. While the bus is beginning to move, a person comes running after the bus, which however does not stop for her. Desperate she turns to you and asks when the next bus. You happen to know that this was the last bus. You say (in Danish/German/English):

We distributed this informal discourse completion task to employees of the University of Southern Denmark, varying the order of the above three situations. We received 47 responses, examples of which are given below:

scenario 1	<p>Toget er desværre forsinket på grund af en teknisk fejl.</p> <p>Det er forsinket 2 timer, du er nødt til at vente. Der er ingen alternative afgange.</p> <p>Jeg har desværre lige hørt at toget er forsinket to timer på grund af en defekt motor. Der er ingen alternative tog, så jeg er bange for at du er nødt til at vente.</p>	<p>Because of a technical problem, the train is unfortunately delayed.</p> <p>It is 2 hours late, you have to wait. There are no alternative trains.</p> <p>Unfortunately, I have just heard that the train is delayed because of engine failure. There is no alternative train, so I'm afraid that you have to wait.</p>
scenario 2	<p>Vi har normalt mandelchokolade, men vores leverandør har i et par uger desværre ikke kunne overholde aftalen.</p> <p>Den har vi ikke, vi har problemer med leverandøren.</p> <p>Jeg er ked af det, men vi har gennem længere tid ikke kunnet få den chokolade fra vores faste leverandør. Vi håber, der snart kommer noget.</p>	<p>We normally have almond chocolate, but our supplier has unfortunately been unable to fulfil the contract.</p> <p>We don't have it, we have problems with the supplier.</p> <p>I'm sorry, but we have been unable to get this chocolate from our regular supplier for a long time. We hope that it will come in soon.</p>
scenario 3	<p>Det der, var den sidste bus i dag.</p> <p>Der er desværre ikke flere busser i dag.</p> <p>Jeg tror desværre det var den sidste.</p> <p>Desværre. Det var den sidste bus i den retning.</p> <p>Jeg er bange for at det var den sidste!</p>	<p>This one, it was the last bus for today.</p> <p>There are unfortunately no further buses today.</p> <p>I believe, unfortunately, that this was the last one.</p> <p>Unfortunately, it was the last bus going in that direction.</p> <p>I'm afraid that this was the last one!</p>

The responses we received contain many examples of the use of *desværre* (unfortunately) and *beklageligvis* (unfortunately), and occasionally Danish speakers included empathy signals or other expressions of emotion (like *jeg er ked af det*; *jeg er bange for*). However, many responses were just statements of facts. A relationship between the speaker's degree of responsibility and the number and type of empathy signal does not become apparent since in the third task, a number of empathy signals were found, even though the speaker was only a bystander and therefore had no responsibility for the running of the buses.

We distributed the same discourse completion task to German friends and colleagues who were able to read the English instructions (to ensure that the tasks were comparable). While some statements of facts were found,

German respondents produced many empathy signals and other mitigation devices, such as *leider* (unfortunately), *bedauerlicherweise* (unfortunately), *ich befürchte* (I'm afraid), *tut mir leid* (I'm sorry), *ich muss Ihnen leider mitteilen* (unfortunately I have to tell you) or alternative offers (can I offer you something else), for instance:

scenario 1	Tut mir leid, aber die Lokomotive ist kaputtgegangen. Wir müssen den Schaden erst beheben. Sie können erst in zwei Stunden weiterfahren. Leider muss ich Ihnen mitteilen, dass Ihr Zug wegen eines Maschinenschadens zwei Stunden Verspätung haben wird. Bedauerlicherweise gibt es keine(n) alternative(n) Zug(verbindung). (Eine Entschuldigung und Hinweis auf die Berechtigung zur anteiligen Erstattung des Fahrpreises und ein Getränkutschein wäre noch eine schicke Hinzufügung ...)	I'm sorry, but the train has broken down. We have to repair the damage first. You can only continue your trip in two hours. Unfortunately I have to tell you, that your train will be delayed for two hours because of engine failure. Unfortunately, there is/are no alternative connection(s). (an apology and a comment on the right to a partial refund of the ticket fee and a voucher for a drink would be a nice addition ...)
scenario 2	Bedauerlicherweise sind wir erneut bei der Belieferung nicht bedacht worden. Darf ich Ihnen eine andere Sorte anbieten?	Unfortunately we have not received any supply. Can I offer you another type?
scenario 3	Tut mir leid, aber ich befürchte, dass das war für heute der letzte Bus in diese Richtung war.	I'm sorry, but I'm afraid that was the last bus going in that direction.

These responses suggest that Danes and Germans tend to handle these situations differently, although it is clear that no categorical differences can be identified, since in both languages factual statements and statements portraying emotional expressions and other mitigation devices are reported. However, factual statements are more common in the Danish data, while the use of emotional expressions is more common in the German data.

This discourse completion task provides: (1) initial findings on how the different strategies of providing bad news are distributed in the two languages and (2) linguistic vocabulary which inspires the construction of stimuli for use in a subsequent experiment to test the hypothesis that, cross-culturally, differences exist when delivering bad news, such that Danes use fewer empathy signals.

To test this hypothesis, we developed an experiment in which bad news was delivered to the participants in two ways, that is, with and without empathy signals. To create a situation where participants felt that they were receiving bad news, we created an expectation which was subsequently left unfulfilled, i.e. the participants were left disappointed. We addressed this by presenting the participants with four images and descriptions of different types of chocolate, and then asking each participant to select which type of chocolate they would want for a year-long supply. A robot then delivered the bad news that the chosen type of chocolate was unavailable, after which the participants were asked to rate the robot in a questionnaire, where the focus was on the likeability of the robot and not on the delivery of bad news. Thus, we decided against evoking metalinguistic cognition, but instead focused on the interpersonal effects realised by the message. In addition, we collected demographic information, namely the age, gender, nationality and native language of each participant.

A small humanoid robot, the JD Humanoid called EZ-bot, was used to deliver the messages. This robot is easy to operate remotely, and a video was created in which the robot moves its head and arms. This video was then combined with sound files for the two different message versions (with and without empathy), with a free text-to-speech tool, iSpeech, for synthesising the speech. The text-to-speech tool iSpeech was chosen because it produces appropriate intonation contours in all three languages being considered. As only a female voice was available in Danish, female voices were chosen for all three languages.¹

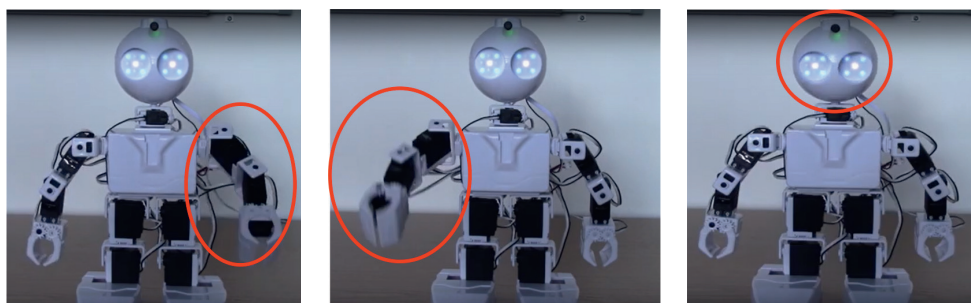


FIGURE 1 Some gestures produced by the EZ-bot

- 1 Unfortunately, the subjunctive/hypothetical forms that were used (*müssten* in German, *would* in English, *må* in Danish) can also perform politeness functions, and there is therefore some ambiguity between the intended hypothetical reading (people do not get the chance to make another choice) and a politeness reading. Since these forms are used in both conditions, they do not influence the results; however, it is possible that any differences between these two conditions would be even more marked in the absence of the mitigating politeness signal.

The formulations elicited from the discourse completion task were used to create experimental stimuli. As the discourse completion task focuses on lexical means, the videos featuring EZ-bot differ in respect to the lexical expressions of empathy. We chose a professional speech register to ensure that additional social signals were not introduced into the experiment, and no prosodic adjustments were made so that the stimuli were as similar as possible across the three languages. In both conditions (with and without empathy), the robot's speech was created using the same text-to-speech system:

- Condition 1: Due to a technical problem, the chocolate you have chosen is not available. You would have to choose another one.
- Condition 2: Oh, I'm really sorry, due to a technical problem, the chocolate you have chosen is unfortunately not available. I'm afraid you would have to choose another one, sorry!

As we were not actually offering chocolate and the choice was presented as being hypothetical from the start, there is no form of 'resolution' after the video was shown. The participants were only required to complete a questionnaire about the robot.

The stimuli created for the German participants are:

- Condition 1: Aufgrund eines technischen Fehlers ist die von Ihnen ausgewählte Schokolade nicht verfügbar. Das bedeutet, dass Sie eine andere Sorte wählen müssten.
- Condition 2: Oh, es tut mir total leid, aber aufgrund eines technischen Fehlers ist die von Ihnen ausgewählte Schokolade unglücklicherweise nicht verfügbar. Ich fürchte, dass Sie leider eine andere Sorte wählen müssten.

The stimuli created for the Danish participants are:

- Condition 1: På grund af en teknisk fejl kan chokoladen ikke vælges. Du må vælge en anden.
- Condition 2: Av, jeg er meget ked af det, men på grund af en teknisk fejl kan chokoladen desværre ikke vælges. Jeg er bange for, at du må vælge en anden. Jeg beklager!

After the participants had watched one of the two videos² featuring the delivery of bad news, they were asked to rate the robot according to the following features: friendliness, politeness, warmth, degree of engagement, formality

² Videos of the robot delivering bad news in the two conditions can be found at:
 German: youtu.be/M-rEyTKs9Yo and youtu.be/452aC9B1iBE;
 Danish: youtu.be/QzC5IaTbTo and youtu.be/ohTD2JfjS4;
 English: www.youtube.com/watch?v=8nEbvM.

and professionalism on a 7-point Likert scale. These features represent the interpersonal functions that would generally be affected by the use of empathy signals; on the one hand, this involves the evaluation of the speaker as being friendly, polite, warm and engaging and, on the other hand, the speaker's degree of formality and his/her professional role.

We piloted the survey several times to remove as many inconsistencies and errors as possible. The questionnaire was then sent to *Gesprächsforschungsliste*, a mailing list for German conversation analysts, to various social networks and to students at Cornell university. As our recruitment process involved a wide range of participants, the questionnaire was also sent to native speakers of languages other than the three under consideration in this study, who were asked to complete the English version of the questionnaire. The two groups of English speakers, native and non-native, were analysed separately.

In total, 1195 people accessed the questionnaire, of which 372 completed the survey, and it is only these latter responses that are included in the analysis. With regard to the gender of the participants, 195 are female, 113 are male and 64 chose not to provide gender information; these groups are evenly distributed across the two conditions and the native languages. In all, 40 participants are native Danish speakers, 55 are native English speakers and 196 are native German speakers. A further 81 participants reported that they had other native languages and are treated as a single group in this study.

We first performed an Analysis of Variance (ANOVA) to identify statistically significant differences for the two conditions, the different genders and between the different languages. Post hoc comparisons were then conducted to identify specific differences between the languages.

The results show that there are significant differences between the two conditions across the four language groups, namely the empathetic delivery of bad news was rated higher with regard to friendliness ($F(1, 371) = 21.986, p < .001$) (see Figure 5); warmth ($F(1, 371) = 7.579, p = .0062$); politeness ($F(1, 371) = 12.34, p = .00049$); and engagement ($F(1, 371) = 6.458, p = .0112$). In contrast, we found no evidence that the robot's use of emotional expressions influenced the ratings given to formality and professionalism.

Furthermore, significant differences are witnessed between the different language groups for four of the six features. In particular, the four language groups evaluated the robot differently with regard to warmth ($F(3,369) = 11.86; p < .001$), engagement ($F(3,369) = 3.074; p = .02784$), formality ($F(3,369) = 3.836; p = .009988$) and professionalism ($F(3,369) = 2.823; p = .0387$). Differences in friendliness and politeness are not statistically significant.

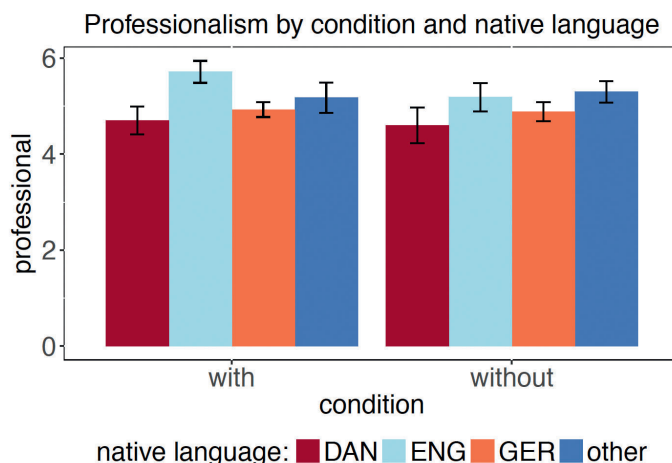


FIGURE 2 Professionalism ratings given by each native language group for the two conditions (with and without empathy)

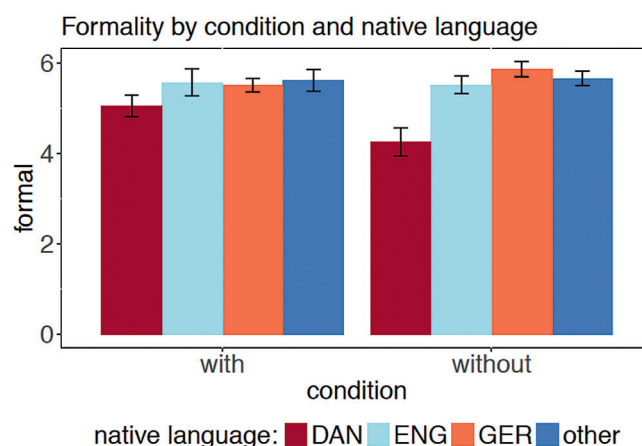


FIGURE 3 Formality ratings given by each native language group for the two conditions (with and without empathy)

As Figure 2 illustrates, the use of empathy signals does not influence how native German or Danish speakers perceive the robot's professionalism, but in the case of native English speakers, these signals increase the perception of professionalism.

Regarding formality, speakers of German, English and other languages rate the robot equally, with means of approximately 5.5 on a 7-point Likert scale. In contrast, Danes differ significantly from Germans ($p = .009$), speakers of other languages ($p = .020$) and native English speakers ($p = .04969$) (see Figure 3) in both conditions. We can therefore conclude that an expression of empathy

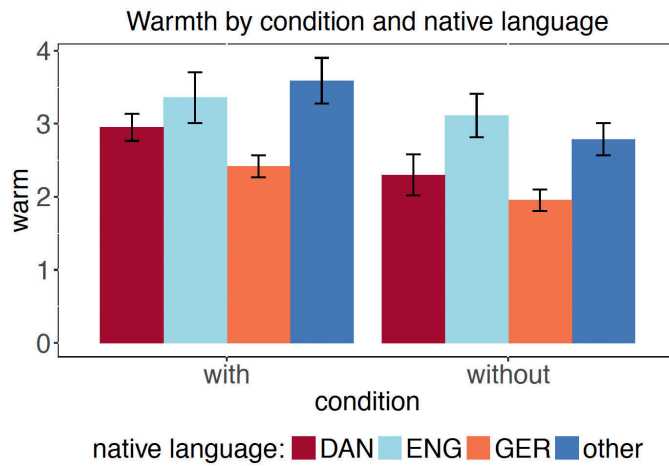


FIGURE 4 Warmth ratings given by each native language group for the two conditions (with and without empathy)

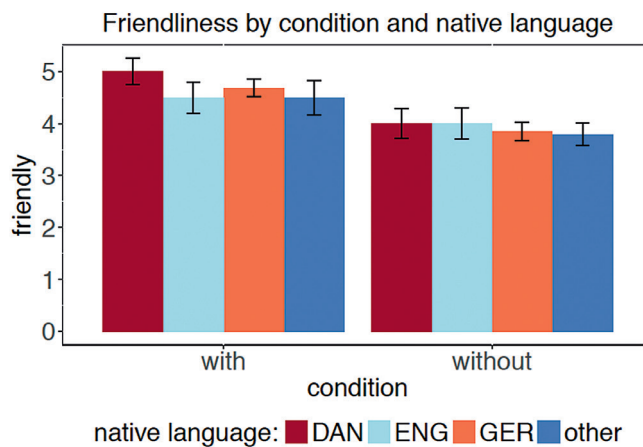


FIGURE 5 Friendliness ratings given by each native language group for the two conditions (with and without empathy)

serves as a marker of formality for the Danish speakers, whereas for speakers of the other languages, an expression of empathy does not influence how they view the formality of the situation, i.e. as formal or informal.

As Figure 4 illustrates, the use of empathy signals has a significant effect on the degree of warmth that the participants attribute to the robot; in the empathetic condition, participants rate the robot more highly with respect to warmth. However, the ratings of the four language groups demonstrate considerable differences; in particular, Germans and Danes attribute less warmth to the robot in both conditions than native and non-native English speakers.

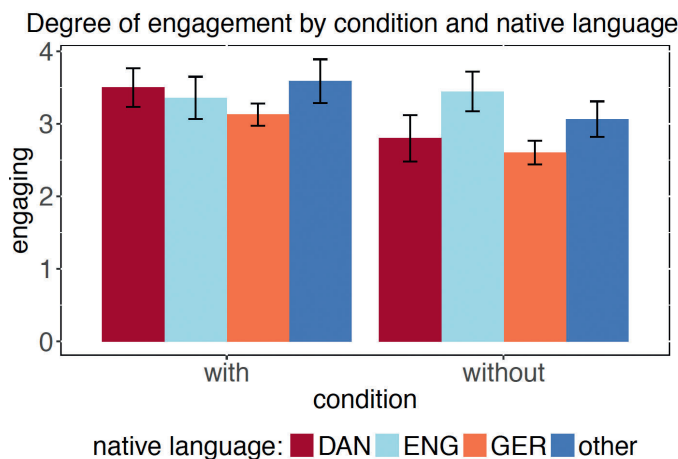


FIGURE 6 Engagement ratings given by each native language group for the two conditions (with and without empathy)

Figure 5 shows that the robot is generally rated as being friendly. All speakers, particularly the Danes, rate the robot as being less friendly when it does not use empathy signals. As described above, the differences between the two conditions are highly significant.

Furthermore, the ANOVA results confirm that there are significant differences in the ratings given by the participants for engagement, in terms of both the use/non-use of empathy signals and across the different languages. However, as illustrated in Figure 6, the engagement ratings given by the native English speakers remain the same in the two conditions, and are thus unaffected by the use of empathy signals. In contrast, the ratings given by the other native language groups, particularly the Danish speakers, are influenced to a greater degree by empathy.

As illustrated in Figure 7, the politeness ratings given by each native language group are influenced by the use of empathy signals, none more so than the Danish ratings.

In summary, this study has revealed that judgments made by the participants about the (robotic) speaker are due to the presence or absence of lexical expressions of empathy, and significant intercultural differences were observed. Restricting the analysis to only lexical signs, as we have done in this study, may have neglected other ways in which empathy can be expressed and which have a role to play in, for example, Danish interactions. However, in our real-life encounters, no specific prosodic realisations were identified. While a systematic investigation of the delivery of bad news in Danish without employing empathy is still necessary, the results show that not using lexical

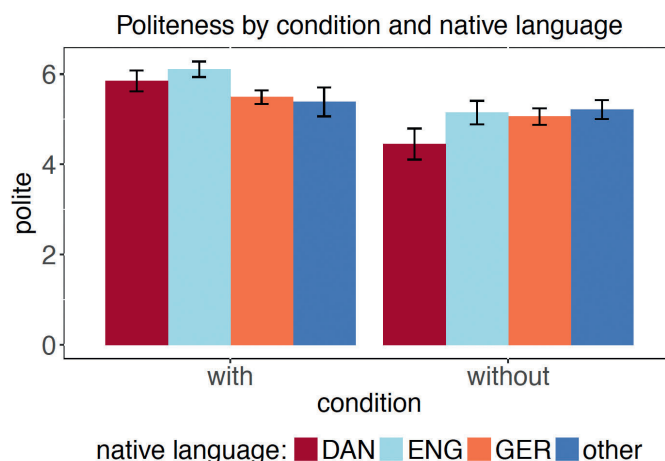


FIGURE 7 Politeness ratings given by each native language group for the two conditions (with and without empathy)

signs of empathy significantly lowers how, in some cultures, the speaker is rated in terms of friendliness, warmth, professionalism and engagement. Furthermore, the results may have been influenced somewhat by the translations of the stimuli. For instance, in German, an utterance-final expression of empathy like the final 'sorry' in English or the utterance-final 'desværre' in Danish is not common; we demonstrated this by including an instance of 'leider' (unfortunately) in the last robotic utterance but, of course, we cannot be certain that the effects are exactly the same. Similarly, the translations of questionnaire items can evoke different associations in different languages. For example, 'warmherzig' might create different associations for German speakers than 'warm' creates for native English speakers, and therefore the differences regarding the perception of the bad news delivered by the robot are not intercultural differences, but differences in how the questionnaire items have been understood. The results can thus be influenced by translational choices.

There may also be limitations due to the prosodic delivery of the respective utterances, which can differ across languages. Since most text-to-speech systems are trained by using read speech, they generally use intonation contours which are common in read speech but are inappropriate in conversational speech. We chose to use iSpeech in this study because, in general, it produces a relatively natural rendering of the robot's utterances, but the quality may differ between the different languages; for example, less text-to-speech training material is available for Danish than for English.

Moreover, some of the observed effects may be due to the fact that people were interacting with a robot; the ways in which people from different cultural

backgrounds respond to a robot might not correspond with how they would interact with another person.

Finally, despite our best efforts to create a longing for large amounts of delicious chocolate, some people might not be disappointed by the bad news being delivered by the robot, whereas in real-life situations, when more is at stake than a questionnaire study, people may respond differently to the two versions of feedback.

The results obtained in this study need to be interpreted in light of these issues, which could have an effect on the validity of our conclusions. However, from our specific experiment, we can conclude that when a robot presents certain utterances which employ empathy i.e. when the robot demonstrates that it understands that the situation may have negative consequences for the communication partner and that it feels for him/her, the robot is generally rated as being more friendly, warm, polite and engaging than a robot that presents the same information without these signals. At the same time, native Danish speakers do not only rate the robot that uses empathy signals as being more friendly, polite, warm and engaging, but also as significantly more formal. Thus, empathy signals carry an additional pragmatic function in Danish, compared with the other languages under consideration.

3 Study II: Providing Feedback

In our second study, we address strategies for giving feedback. While providing feedback involves many different phenomena, our focus here is on providing feedback in response to a certain performance. The procedure we have adopted for this study is similar to that of Study I, in that we first elicited qualitative data, which we then tested in a questionnaire study. However, in contrast to the first study, the data was elicited experimentally, and the manipulations were not concerned with only one selected feature, like empathy signals, but rather holistic feedback-giving strategies. In the following, we detail how the data was elicited to form the hypotheses, and how it was analysed and then exploited to generate the stimuli. Finally, we present the questionnaire study and its results.

In this study, we focus on the countries of Germany, Denmark and Poland, i.e. three European countries in close spatial proximity. Previous research on the intercultural differences between neighbouring countries is somewhat inconclusive. For instance, the three countries being investigated in this study are categorised by Meyer (2014) as being rather similar with regards to feedback. In her book *The Culture Map*, Meyer (2014) argues for a multi-dimensional understanding of intercultural differences and distinguishes,

for example, between how directly something is communicated, drawing on Hall's (1976) distinction between high- versus low-context cultures. According to Meyer (2014), directly communicating cultures can be expected to use upgraders (such as 'this is absolutely, totally, strongly unprofessional') whereas indirectly communicating cultures prefer to use downgraders (e.g. 'this is not quite there yet') to mitigate feedback. According to her, Denmark, Germany and Poland are all directly communicating cultures.

Similarly, according to Hofstede Insights (2019), all three countries have an individualistic orientation, which, according to Stone-Romero and Stone (2002), greatly influences how feedback is received across the different cultures. Stone-Romero and Stone (2002) propose a cross-cultural model for the responses given to feedback, and this model describes how the cultural background of the person receiving the feedback influences the values to which this person subscribes, his/her self-esteem and cultural behavioural scripts, all of which influence how a person responds to feedback. In particular, the authors argue that the appraisal of feedback depends on whether negative feedback constitutes a threat to one's own self-esteem, or whether it is understood to be a threat to the welfare of the group; in turn, the interpretation of the feedback determines the resulting behaviour, such as rejection of the feedback, i.e. dysfunctional behaviour, or increased effort. Which values are invoked, what responses are triggered and which behaviours are activated depends on, according to the authors, whether the feedback receiving person has an individualistic or an allocentric cultural background. Consequently, since all three countries under consideration have individualistic orientations, their responses to negative feedback should be similar.

On the other hand, based on the dimensions developed in Hofstede (1983), the website Hofstede Insights (2019) identifies considerable differences between the three countries with regard to power distance, masculinity, uncertainty avoidance, long-term orientation and indulgence. The extent to which feedback is provided and received differently in the three countries, and the extent to which the various styles of feedback-giving are perceived differently, is thus unknown. Therefore, we conducted a study in which we first elicited feedback and then collected the responses to this feedback by speakers of the same and different languages.

First, in order to understand how feedback is given in the three languages, we designed a study in which participants were asked to provide feedback on two ideas for start-up companies. German, Danish and Polish participants were approached by confederates, who presented them first with an idea for a start-up company which we suspected would elicit positive feedback, namely a company that connects people together to share meals. The participants were then presented with an idea for a start-up company

which we expected to receive negative feedback, namely a retail cats clothing company. In both cases, our expectations were generally realised. We recruited three confederates per country, who each elicited feedback on these ideas from two of their friends or relatives. We used a number of different confederates to circumvent any potential fatigue effects (cf. Kuhlen and Brennan, 2013), and to minimise the impact of each confederate. All participants were recruited by word-of-mouth, so that the participants do not constitute a representative sample of the three cultures under consideration. The results of this study are therefore somewhat limited and need to be verified in the subsequent questionnaire study. However, it is evident that different strategies for providing feedback emerged during these interactions. The Danish participants came from Southern Jutland, the German participants from northern Germany and the Polish participants lived near Danzig.

We analysed the feedback strategies employed by the participants on the basis of previous work, and added features to the coding scheme as they became apparent during the comparison. In particular, in line with Meyer's (2014) suggestions, we looked for the presence of upgraders and downgraders as indicators of a more or less direct style of feedback. Furthermore, we used Brown and Levinson's (1987) politeness theory to investigate the extent to which criticism is bald on record and what devices are used to mitigate potentially face-threatening acts. In accordance with Lewis' (2018) model of feedback, we identified the response behaviours.

In general, we found that the German participants often used downgraders when delivering positive feedback and sometimes upgraders when delivering negative feedback, which would be, according to Meyer (2014), a typical feature of directly communicating countries. The Danish participants sometimes upgraded and sometimes downgraded positive feedback, but they always downgraded negative feedback. The Polish participants upgraded positive feedback and downgraded negative feedback. Whereas the Danish and Polish participants always presented the feedback as their own personal opinion, the German participants discussed the facts from several perspectives. The German negative feedback was more on the record, whereas the Danish participants tended to produce more off the record negative feedback. The Polish participants only produced off the record negative feedback. Accordingly, the level of relationship-building differed between these three groups; the Polish and Danish participants spent more time creating a friendly atmosphere, particularly when giving negative feedback, whereas the German participants focused solely on the topic at hand. For instance, based on the elicited data, the Polish positive feedback states a personal opinion and uses an upgrader, 'You have done great', whereas the German feedback

makes an objective statement and uses a downgrader, 'The task was done pretty well'.

Based on these findings, we created six feedback utterances, three positive and three negative, which were inspired by the patterns observed in the three languages. The feedback utterances were all translated into English so that the effects of each strategy on speakers of the original language and the other two languages could be measured. That is, in order to understand the effects that the different pragmatic feedback strategies had on speakers of the three languages, all interactions were carried out in English.

We then created two tasks which the participants had to complete in an online questionnaire. The first task asked them to choose three features (from a list of six) of a good leader. The second task was more difficult and involved matching different leadership styles with descriptions of these styles. Neither of these two tasks had obvious answers, but participants received positive feedback for the first task and negative feedback for the second task.

As in Study 1, we synthesised both the feedback utterances and the instructions for the participants using a free text-to-speech system, this time fromtexttospeech.com, using a male voice with an American accent. These utterances were then connected to a video of the EZ-bot so that speech and robotic gestures were as compatible as possible.

The EZ-robot delivered the designed feedback to the participants in different combinations; the participants witnessed feedback from both their own and a different culture, which resulted in three conditions. In Condition 1, participants saw Danish-inspired (DK) positive and German-inspired (DE) negative feedback. In Condition 2, participants saw German-inspired positive and Polish-inspired (PL) negative feedback. Finally, in Condition 3, participants saw Polish-inspired positive and Danish-inspired negative feedback.

Since Stone-Romero and Stone (2002) hypothesise that individualistic cultures tend to respond with threat to self-esteem and defensively interpret negative feedback, we expected the speakers to attribute negative interpersonal characteristics to the robot. We therefore asked our participants to evaluate the robot regarding its perceived motivation, friendliness, politeness, dominance, empathy and engagement on a five-point Likert scale. 236 complete responses were gathered through social media; each version of the questionnaire was seen by between 19 and 31 participants in each of the three countries. Independent t-tests were applied in a pair-wise comparison of each of the features.

Interestingly, the feedback strategies did not always receive the best ratings from their own native speakers; for instance, Danish-inspired positive feedback is rated as significantly less motivating, empathetic and entertaining by

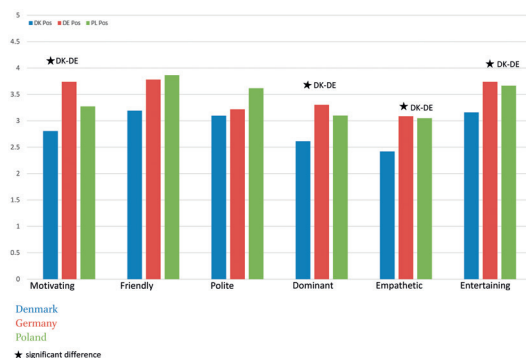


FIGURE 8
Responses given by Danish, German and Polish participants to Danish-inspired positive feedback

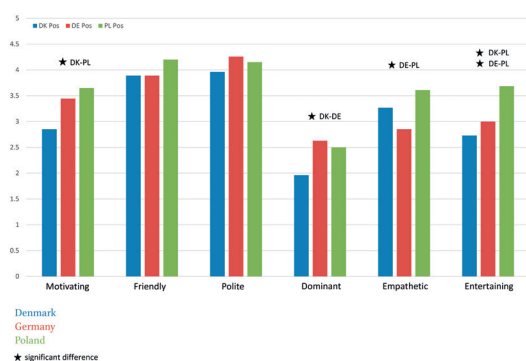


FIGURE 9
Responses given by Danish, German and Polish participants to Polish-inspired positive feedback

Danish participants than by German participants, but also less dominant. As Figure 8 illustrates, there are four significant differences: German participants rate the Danish-inspired positive feedback as significantly more motivating, empathetic and entertaining, but also significantly more dominant.

In contrast, Polish-inspired positive feedback receives the highest positive ratings from Polish speakers. In the case of motivating, empathetic and entertaining, these differences are significant. At the same time, all participants rate the Polish-inspired positive feedback as being rather low in dominance.

German-inspired positive feedback (see Figure 10) is judged to be significantly more dominant by the German participants than by the Danish participants, and Polish participants rate it to be significantly more motivating, friendly and polite than the German participants themselves. However, Danish participants rate German-inspired positive feedback as less motivating and less empathetic than the German participants. Thus, while the Polish participants rate the German feedback style positively, the Danish participants often rate it less favourably – with the exception of dominance, which receives the highest ratings from the German participants.

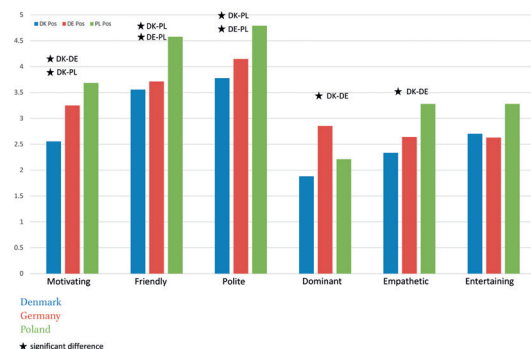


FIGURE 10
Responses given by Danish, German and Polish participants to German-inspired positive feedback

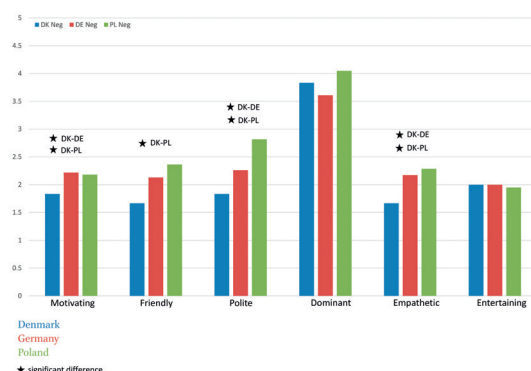


FIGURE 11
Responses given by Danish, German and Polish participants to German-inspired negative feedback

German-inspired negative feedback receives the least favourable response from all the participants (Figure 11), with the lowest ratings for motivation, friendliness, politeness and empathetic response being given by the Danish participants. In contrast, in the case of dominance, all three groups provide more or less equal ratings for the German-inspired negative feedback.

Many of the features of Danish-inspired negative feedback (Figure 12) were rated more or less equally by the language groups, with only the German participants rating the feedback as less motivating and entertaining. In addition, the mean ratings for empathy and entertainment are neutral in value. Similarly, Polish-inspired negative feedback (Figure 13) receives very similar, positive judgements from all the language groups. However, in terms of entertainment, it is rated as neutral, or worse, by the Danish participants.

The results show that the three different feedback profiles which were previously identified are received in significantly different ways by native speakers from the three countries under investigation. Negative feedback is evaluated more negatively than positive feedback in all three conditions. German-inspired negative feedback is generally not evaluated positively, particularly by Danish participants.

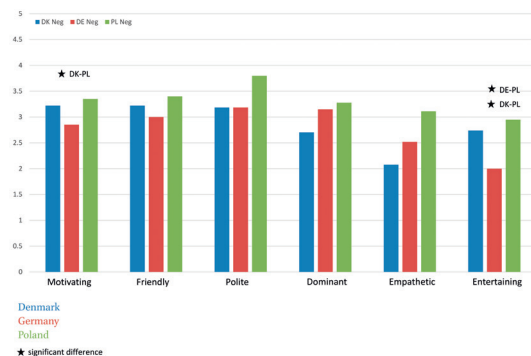


FIGURE 12
Responses given by Danish, German and Polish participants to Danish-inspired negative feedback

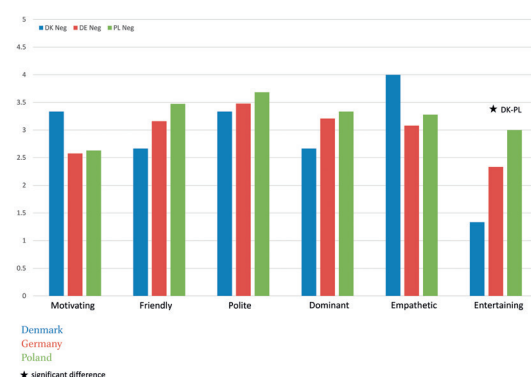


FIGURE 13
Responses given by Danish, German and Polish participants to Polish-inspired negative feedback

As in Study 1, the different text-to-speech systems might have created different artefacts in the three languages, and therefore the synthetic speech may have been acceptable to varying degrees in German, Danish and Polish.

Some of the effects witnessed may have been the result of translating the Danish, Polish and German strategies into English, and it is possible that not every participant was equally fluent in English. Furthermore, our recruitment strategy reached only a certain group of participants, namely young, internationally minded (English-speaking) internet users. Our results might have been quite different if other age groups and non-English speaking Danes, Germans and Poles had been recruited.

Moreover, it is unclear which of the features employed in the design of the robot utterances actually caused the different interpretations that were identified; since the feedback strategies identified in the first part of the study involved many different features, some of which are more global in nature and concern the design of the whole feedback utterance (such as humour in Danish-inspired feedback and directness in German-inspired feedback), whereas others, like the discourse markers identified, are more local in nature. Follow-up studies in

which only one feature is varied at a time could shed further light on the most relevant features. Previous work has suggested that the appropriateness of the criticism can differ between cultures, even if these cultures all support a direct style of communication (Meyer, 2014) – however, this feature alone does not explain why German-inspired positive feedback received lower ratings for empathy and entertainment, even from its own speakers.

Even though, in this study, we have been unable to ascertain what exactly makes feedback acceptable, it has revealed significant differences in the perception of different feedback strategies across a number of cultures; thus, the way in which feedback is delivered is important.

4 Experimental Contrastive Pragmatics Using Robots

The two studies detailed above have revealed a number of cross-cultural differences. These have the potential to lead to problems in intercultural communication, since pragmatic strategies that are common in one culture can lead to a less favourable interpretation by speakers from another cultural background. In these experiments, we used video recordings of a robot, which allowed all the participants to interact with the same embodied, although artificial, communication partner, with each participant being presented with identical stimuli in the same interactional context. From this perspective, we established the highest level of experimental control possible in an area of research which is generally hard to investigate in a controlled, yet ecologically valid manner.

Moreover, since robots are, as yet, unusual interactional partners and artefacts, i.e. they are manufactured and programmed by someone else, it is natural to ask to what extent they have certain characteristics; this is not necessarily the case with people, but even in our experiments of human-robot interactions, we found that, at times, Danes refused to evaluate another 'being' (i.e. the robot) after only a brief encounter. In this regard, it is perhaps easier for people to judge a robot than another human being.

However, in the two studies presented in this paper (but most likely in other experiments that involve robots as interaction partners (see, for instance, Fischer, 2016b; Fischer et al., 2020)), the response given by one of the interaction partners, namely the robot, is not contingent on the human participant's behaviour, which would be a normal part of human interaction (e.g. Sacks, Schegloff and Jefferson, 1974). The fact that the robot employs pre-synthesised, fixed utterances means that it produces behaviours irrespective of

the communication partner's actions; this can lead to differing interpretations of the robot's behaviour, since the participants might create different interactional contexts. For example, let us suppose that a participant in Study II has chosen not to reply to the first questionnaire task and has only given nonsense answers (in actual fact, we made it very difficult for the participants to do so by restricting the answer options). In this case, it is obvious that the robot's positive feedback is canned speech and is played irrespective of the participant's actual answer. Therefore, the feedback given by the robot performs a different pragmatic function than that performed when the participants assume that the feedback has been given in response to their actions. While it is unclear what the alternative might be, i.e. whether human confederates are able to cope more easily with these cases in a controlled way, this example shows that even identical stimuli in controlled experiments can have different interpretations, and, thus, this challenges the controllability of interactional experiments. In other words, since interactions are interactionally achieved (e.g. Schegloff, 1982), even the same robot utterance might not have identical pragmatic functions in these interactions, unless precautions are taken.

Furthermore, it is possible that people respond to robotic interaction partners in a different manner to human interaction partners. In human-robot interaction research, it has been found that, in general, people respond to social cues from robots and humans in similar ways – an observation that is known as 'media equation' (Reeves and Nass, 1996). According to the 'computers are social actors paradigm' (Nass and Moon, 2000), robots are generally responded to as social actors (cf. Andrist et al., 2014; Mutlu et al., 2012; Lohan et al., 2012, to name but a few). Nevertheless, there are considerable interpersonal differences in how people perceive robots (see Fischer, 2011; 2016b; 2020), which could have an impact on how people react to pragmatic strategies when they are presented by robots. Thus, robots, and computational interfaces in general, do not necessarily trigger social responses (e.g. Krämer, 2008; Lotze, 2016; 2018), and Fischer (2006) has suggested that human-computer interactions are themselves instances of intercultural communication, because people need to determine how to interact with a partner that is so different from themselves (and this applies to both robots and other computational interfaces). Thus, the results achieved in human-robot interaction experiments might not be transferable to interactions between humans. On the other hand, a large body of research suggests that social signals are indeed interpreted in similar ways, irrespective of whether they originate from humans or computers/robots (e.g. Hutchby, 2001; Fischer, submitted). Therefore, it is likely that the mechanisms behind human-human and human-robot interactions are similar;

however, whether the results obtained in this study can be generalised for human-human interactions would need to be proved by conducting experiments with human interaction partners – and this process is not straightforward, as we have previously argued.

In any case, what we can rely on, though, are relative differences. In particular, in our studies, the participants were faced with robots which had been programmed with different behaviours. The results that were obtained have revealed significantly different responses to these different robot behaviours. Thus, while people might well respond differently to robots than to other people, the results obtained all stem from comparable human-robot interactions. Even though there are many drawbacks to the way the utterances were delivered, given the problems with current speech technology, participants were confronted with equally restrictive and potentially problematic material.

Furthermore, some problems can arise from the subject area itself, namely the intercultural comparison. In Study I, we translated the questionnaire into the participants' native languages, and it is unclear whether all the terms used mean the same in each language being considered. In Study II, we translated the original feedback into English stimuli to compare the participants' responses to the feedback strategies. Thus, regardless of whether or not the speaker is a robot, contrastive pragmatics itself has many potential pitfalls, and employing robots can only help mitigate some of these pitfalls.

In summary, the aim of this paper has been to introduce human-robot interactions as a methodological paradigm for the experimental investigation of cultural differences. We believe that we have shown the benefits of this methodology and we have also discussed some potential problems with this approach. In addition, we have attempted to understand the functions of different pragmatic strategies across a number of different cultures. Despite the collection of quantitative data, our goal has been a qualitative assessment, namely to gain an understanding of the interpersonal effects of using a particular strategy and how these interpersonal effects differ in frequency across the different languages. The first study demonstrates that the functions of empathy signals differ across the cultures being considered; while all the language groups considered an expression of empathy to be polite, friendly, warm and engaging, some groups also associate these expressions with formality or professionalism. The second study revealed significant differences in the perception of feedback strategies across neighbouring cultures. Although these studies have some limitations, they have nevertheless revealed the pragmatic effects of linguistic choices, and have also shown that the use of robots is a promising methodology for experimental contrastive pragmatics research.

Acknowledgements

The work presented here was generously funded by the Danish Council for Independent Research under Grant Number 4180-00359B, which we gratefully acknowledge. We are furthermore grateful to Maria aus der Wieschen for her help with the data collection and analysis in Study I, and to the confederates and participants of Study II.

References

- Andrist, Sean, Xiang Zhi Tan, Michael Gleicher, and Bilge Mutlu. 2014. Conversational gaze aversion for humanlike robots. In: *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. ACM, 25–32.
- Blum-Kulka, Shoshana, Juliane House, and Gabriel Kasper (eds.). 1989. *Cross-cultural Pragmatics: Requests and Apologies*. Norwood, NJ.: Ablex.
- Brennan, Susan E., and Joy E. Hanna. 2009. Partner-specific adaptation in dialogue. *Topics in Cognitive Science* 1(2): 274–291.
- Brown, Paula, and Gary Dell. 1987. Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology* 19(4): 441–472.
- Brown, Penelope, and Stephen C. Levinson. 1987. *Politeness. Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Cyluk, Agnieszka. 2013. Discourse completion task: Its validity and reliability in research projects on speech acts. *ANGLICA-An International Journal of English Studies* 22(2): 101–112.
- Fischer, Kerstin. 2006. *What Computer Talk Is and Isn't. Human-Computer Conversation as Intercultural Communication*. Linguistics – Computational Linguistics 17. Saarbrücken, Germany: AQ-Verlag.
- Fischer, Kerstin. 2011. Interpersonal variation in understanding robots as social actors. In: *Proceedings of HRI'11*, March 6–9th, 2011. Lausanne, Switzerland, 53–60.
- Fischer, Kerstin. 2016a. Robots as confederates: How robots can and should support research in the humanities. In: Johanna Seibt, Marco Nørskov, and Søren Schack Andersen (eds.), *Proceedings of the Robophilosophy 2016 Conference*. Aarhus, Denmark: IOS Press, 60–66.
- Fischer, Kerstin. 2016b. *Designing Speech for a Recipient: The Roles of Partner Modeling, Alignment and Feedback in So-Called 'Simplified Registers'*. Amsterdam: John Benjamins.
- Fischer, Kerstin. 2020. Why Collaborative Robots Must Be Social (and even Emotional) Actors. *Techné: Research in Technology and Philosophy* 23(3): 270–289.

- Fischer, Kerstin. (submitted). Tracking Anthropomorphic Responses Dynamically in Human-Robot Interaction.
- Fischer, Kerstin, Oliver Niebuhr, Lars C. Jensen, and Leon Bodenhagen. 2020. Speech Melody Matters – How robots can profit from using charismatic speech. *ACM Transactions in Human-Robot Interaction* 9(1), Article 4: 1–21.
- Hall, Edward T. 1976. *Beyond Culture*. New York, London, Toronto, Sydney, Auckland: Anchor Books, Doubleday.
- Hofstede, Geert. 1983. The Cultural Relativity of Organizational Practices and Theories. *Journal of International Business Studies* 14(2): 75–89.
- Hofstede, Geert. 2019. Hofstede Insights. Retrieved June 2, 2019, from <https://www.hofstede-insights.com/country-comparison/denmark,germany,poland/>.
- Hutchby, Ian. 2001. *Conversation and Technology: From the Telephone to the Internet*. Cambridge: Polity.
- Krämer, Nicole C. 2008. Social effects of virtual assistants. A review of empirical results with regard to communication. *Proceedings of the 8th international conference on Intelligent Virtual Agents*, 507–508.
- Kuhlen, Anna K., and Susan E. Brennan. 2013. Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin & Review* 20(1): 54–72.
- Labben, Afef. 2016. Reconsidering the development of the discourse completion test in interlanguage pragmatics. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)* 26(1): 69–91.
- Lewis, Richard D. 2018. *When Cultures Collide* (4th ed.). London, Boston: Nicholas Brealey Publishing.
- Lockridge, Calion B., and Susan E. Brennan. 2002. Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review* 9(3): 550–557.
- Lohan, Katrin S., Katharina J. Rohlfing, Karola Pitsch, Joe Saunders, Hagen Lehmann, Christopher L. Nehaniv, Kerstin Fischer, and Britta Wrede. 2012. Tutor Spotter: Proposing a Feature Set and Evaluating It in a Robotic System. *International Journal of Social Robotics* 4(2): 131–46.
- Lotze, Netaya. 2016. *Chatbots – Eine linguistische Analyse*. Frankfurt am Main: Peter Lang.
- Lotze, Netaya. 2018. Zur sprachlichen Interaktion mit Chatbots – Eine linguistische Perspektive. In: Theo Hug, and Günther Pallaver (eds.), *Talk with the Bots – Gesprächsroboter und Social Bots im Diskurs*. Innsbruck: Innsbruck University Press, 29–50.
- Maynard, Douglas W. 1991. The perspective display series and the delivery and receipt of diagnostic news. In: Deirdre Boden, and Don H. Zimmerman (eds.), *Talk and Social Structure: Studies in Ethnomethodology and Conversation Analysis*. Cambridge, England: Polity Press, 164–192.

- Maynard, Douglas W. 1997. The news delivery sequence: Bad news and good news in conversational interaction. *Research on Language and Interaction* 30(2): 93–130.
- Maynard, Douglas W. 2017. Delivering bad news in emergency care medicine. *Acute Medicine & Surgery* 4(1): 3–11.
- Maynard, Douglas W., and Richard M. Frankel. 2006. On diagnostic rationality: bad news, good news, and the symptom residue. In: John Heritage, and Douglas W. Maynard (eds.), *Communication in Medical Care: Interaction between Primary Care Physicians and Patients*. Cambridge, UK: Cambridge University Press, 248–278.
- Meyer, Erin. 2014. *The Culture Map: Breaking Through the Invisible Boundaries of Global Business*. New York, N.J.: Public Affairs.
- Mutlu, Bilge, Takayuki Kanda, Jodi Forlizzi, Jessica Hodgins, and Hiroshi Ishiguro. 2012. Conversational Gaze Mechanisms for Humanlike Robots. *ACM Transactions on Interactive Intelligent Systems* 1(2): 1–33.
- Nass, Clifford, and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of Social Issues* 56(1): 81–103.
- Phillips, Elizabeth, Xuan Zhao, Daniel Ullman, and Bertram F. Malle. 2018. What is human-like?: Decomposing robots' human-like appearance using the Anthropomorphic roBOT (ABOT) database. In: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction. ACM, 105–113.
- Reeves, Byron and Clifford Nass. 1996. *The Media Equation*. Stanford: CSLI and Cambridge: Cambridge University Press.
- Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. A Simplest Systematics for the Organization of Turn-taking for Conversation. *Language* 50(4): 696–735.
- Schegloff, Emanuel A. 1982. Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In: Deborah Tannen (ed.), *Analysing Discourse: Text and Talk*. Washington: Georgetown University Press, 71–93.
- Stone-Romero, Eugene F., and Dianna L. Stone. 2002. Cross-cultural differences in responses to feedback: Implications for individual, group, and organizational effectiveness. *Research in Personnel and Human Resources Management, Vol. 21*. Bingley: Emerald Group Publishing Group, 275–331.