Facing Formality: Investigating the Influence of Formality of Restaurant Recommendation Chatbot Responses on User Experience

Amy Oey
a.oey@uu.students.nl
7082770

Kathleen de Boer
k.k.deboer@students.uu.nl
6439608

Sunny Hsieh
s.hsieh@students.uu.nl
6534856

Andrea Roijen
a.roijen@students.uu.nl
5726042

Figure 1

ABSTRACT

The presence of Chatbots and dialogue systems into the daily life has grown a lot since the past few years. Previous research has demonstrated the importance for users of these systems to have the right communication style, in particular the formality of the chatbot system's responses. The aim of this study was to determine whether the formality of responses also has an influence on the user experience of a task-based dialogue system that implements a restaurant recommendation system. To investigate this, 58 people participated in survey to measure six different aspects of user experience (e.g., naturalness, quality). In addition, objective measurements (number of turns and total time of the conversation) were analyzed as well. All participants conversed once with the system in a formal and once in an informal setting. By performing a paired t-test, the results of this experiment show that there is no significant difference in any of the user experience aspects between the informal and formal speech setting of the chatbot. Based on these results, the formality of responses does not seem to be a relevant factor to the overall user experience when using a restaurant recommendation dialogue-system.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

ACM Reference Format:
Amy Oey, Kathleen de Boer, Sunny Hsieh, and Andrea Roijen. 2021. Facing Formality: Investigating the Influence of Formality

of Restaurant Recommendation Chatbot Responses on User Experience. In *Proceedings of* . ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The popularity of chatbots has skyrocketed in the last couple of years. Chatbots start to appear everywhere in our daily lives; in customer services of webshops, IT and HR departments within small and big organisations, on mobile devices and much more [1, 4, 15]. Most of these are task-based dialogue agents. This type of conversational agent has the goal to help users with a specific task, for example to book a flight, find a recommendation, or provide information that the user is looking for. Depending on the simplicity of the task and environment, these systems can be based on relatively simple rule-based systems, while others have complex artificial intelligence systems running in the background. The implementation of these systems reduce operating costs by taking over these relatively simple kinds of tasks from employees that get overloaded with repeating tasks [16]. The implementation of these systems are therefore relevant, though at the same time not straight forward. Research that looks into the user experience of these systems provides useful insights into the challenges that these systems are still facing.

Recommendation Dialogue Systems

An example of an application of a goal-based dialogue system that still faces challenges, is one that implements a recommendation system. Non-dialogue recommendation systems are ubiquitous nowadays. For instance, a complex recommendation system such as a YouTube video recommendation system uses various learning algorithms to give an optimal recommendation based on user data and activity [2]. Google has a restaurant recommendation system integrated in Google Assistant and the google search engine. This recommendation system can be used for various tasks and it works relatively well by using a large amount of data and underlying algorithms. At the same time, there is a technological gap between this type of recommendation systems and dialogue agents that needs to be tackled when designing a recommendation dialogue system. This mostly lies in the collaboration between user and recommendation system when a recommendation must fulfill the user's expressed needs [13]. The collaboration relies on the communication through processing and generation of natural human language. Having the right communication style is something that is an intuitive task for humans, whereas this type of language use in dialogue systems is still a big challenge in the field of Artificial Intelligence. At the same time, communication by use of natural language is something that these systems cannot function without and are completely based on. Not only is it the task of these systems to extract and provide the correct information using natural language, but they should also do this in a way such that customers experience their communication with the chatbot as natural and easy, like how it would have been with a human employee. Preceding research has demonstrated this, by showing that customers experience their conversations with chatbots as unnatural and impersonal [3, 11]. It was also shown that 43% of customers said they prefer dealing with an actual person (which was the number one potential barrier to using chatbots) and 24% would stop using a chatbot if it wasn't able to interact in a friendly manner [3]. A chatbot's informal communication style induced a higher perceived social presence which in turn positively influences quality of the interaction and brand attitude [11]. This preceding work points out the relevance of dialogue systems having an appropriate communication style, such as the right amount of (in)formality. For this reason, the current research focuses on the influence of formality of responses from a restaurant recommendation dialogue-system on user experience. The next subsection first gives a brief overview on the topic of formality of language.

Formality of language

A formal style language is characterized by detachment, accuracy, rigidity and heaviness; an informal style is more flexible, direct, implicit, and involved, but less informative [9]. Heylighen and Dewaele (1999) propose that formality becomes larger when the distance in space, time or background between the interlocutors increases. They also state that formality is preferred when the speaker is male, introverted or academically educated [9]. The remark on distance can also relate to the fact that an informal communication style in human customer service messages was perceived appropriate for familiar brands but inappropriate for unfamiliar ones [7]. Formality differs among topics, area and culture. In economics, the middle east and science, formal language is preferred over informal language. Nevertheless, in the topics of entertainment and fun, the preference tends to be more towards informal language [12]. Different languages prefer different levels of formality in different situations. For example, in Russian, Japanese and German people address different persons with different levels of formality. This means that the level of formality from the chatbot is reflective of the type of relationship between the user and the chatbot. This is important for the overall experience of the user.

Research Question and Hypothesis

It is hard to say for developers what level of formality is suitable for a restaurant recommendation dialogue system given the previous information. To create a high-quality system that provides an optimal user experience, the question remains what type of formality the chatbot needs to use

in its utterances. Therefore, this research investigates the importance of formality of language in a restaurant recommendation dialogue-system for user experience with the following research question: *"Does the formality of the restaurant recommendation chatbot's responses influence the user experience of the chatbot?"*
The study of Van der Weegen's (2019) [17] addressed the effects of a conversational human voice (CHV) usage in chatbot conversations on brand attitude and support the assumption that CHV and an informal communication may be equally stimulating dialogues. Also, in the study from Sander (2020) [14] it was found that a chatbot's informal communication style positively influences quality of the interaction and brand attitude if participants perceived high levels of social presence [11]. Therefore, the hypothesis is that the informal version of the chatbot system will be preferred by the user. This is based on the assumption that the social distance between the user and chatbot will be smaller by using an informal setting of the system, in such a way that the user feels more comfortable.

In order to find an answer to this research question, user surveys and objective measurements are used to obtain insights in user experience during a formal and an informal conversation with our restaurant recommendation chatbot called *KASA*. In the next section, a detailed explanation of the approach is presented. In the third section, results that followed from this are presented. And finally, in section 4 the implications and limitations of the current study are discussed, in order to converge to the conclusion of this research paper.

2 METHOD

Participants

58 participants (28 female, 30 male) took part in this experiment. The participants ranged in age from 17 to 69 years ($M = 27, SD = 10.3$). The distribution is presented in Figure 2. All participants were native speakers of Dutch and have acquired English as their second language. To prevent the effects of an insufficient language proficiency to process the formality of English as their second language, the aim was to control as much as possible for a sufficient proficiency in English. This was done by setting a minimum of level B1, according to the *Common European Framework of Reference for Languages* in 2001 (this is comparable to the level of English after having achieved a HAVO certificate in the Netherlands). In addition to this, participants were asked to rate their own English proficiency on a scale of 1 to 5 (where 1 is poor and 5 is very good). The self-proclaimed proficiencies in English ranged from 2 to 5 ($M = 4.2, SD = 0.8$). It is important to note that none of the participants had any previous experience with the chatbot or insights into the system that is used in the chatbot, to prevent that this has any influence on the experiment.

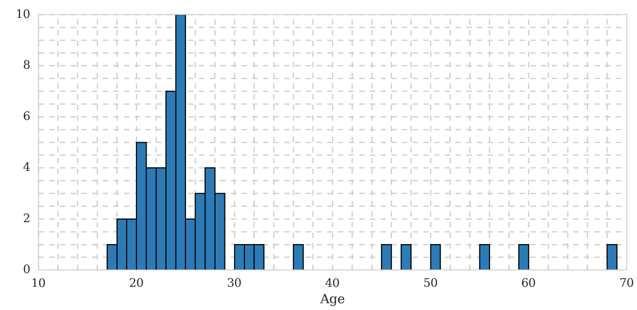

Figure 2: Age distribution

Materials

The dialogue system. The chatbot system *KASA* that is used for this experiment gives restaurant recommendations based specific preference types that are expressed by the user. The restaurants that it can recommend are based the data set containing 110 entries for restaurants in London that are specified by a number of properties (e.g., food type, area, price range, address and phone number). To do this, a transition system (dialog manager) was implemented, that guides the chatbot and user through different phases of the conversation. With a decision tree the system was trained to classify dialog acts of user utterances with an accuracy of 98% on the data set of Dialog State Tracking Challenge (DSTC 2) [8]. In order to process the content of the user utterances the chatbot uses a keyword matching system. The set of keywords is the set possible preferences that can match with any restaurant in the data set. These preference types were based on area, food type and price range. The state machine of the system checks whether the user provided enough preferences to recommend a restaurant. If this is not the case, it will continue to pose questions to the user. If, in the end, more than one restaurant matches the user preferences, the user will be asked if there are any additional requirements. The chatbot used a reasoning component to check whether any of these restaurant options matches the user's additional requirements. To generate responses, a fixed library of response options are used. This library consisted of a formal and an informal version of all responses. It is important to note that the formality setting only changed the formulations of the responses and not the content of the responses or the decisions that the chatbot made. The informal and formal formulations for the responses are based on the paper of Liebrecht, Sander, & Van Hooijdonk [**?**]. In the informal versions of the responses, punctuation marks and final letters of words are repeated in likeness to the informal sentences from Liebrecht, Sander, & Van Hooijdonk. Also, the use of emoticons is absent in the formal responses, whereas the informal responses use an emoticon sometimes, for example " :-) ". Other changes that have been applied are, for example, in the formal responses abbreviations are omitted as well as the use of 'slang' words. Examples of an informal and formal responses are given below:

> 'bye' formal mode: Thankyou, hopefully my service was useful. I look forward to our next meeting!
> 'bye' informal mode: Thanksss, hope it was useful. See yaa!!!
>
> 'nomatch' formal mode: Sorry, there's nothing matching your needs. Is there something that you would like to change about your order?
> 'nomatch' informal mode: Sorry, there's no restaurant matching your tastes :(. Wanna try something else???

Devices. Laptops were used to run the chatbot system, which enabled a clean chat window in the terminal in which participants could communicate their responses to the chatbot. To chat with the chatbot, the use of the laptop keyboard and a mouse(pad) sufficed.

Design

Each participant conversed once with the chatbot in the formal setting and once with the informal setting. This is a within-participant experiment with the dependent variable user experience being investigated by manipulating the independent variable formality. Half of the participant group began with the informal setting in the first task, and the formal setting in the second task. For the other half of the group it was the reverse, starting with the formal setting in the first task and the informal setting in the second task. To prevent the second conversation experience to go more easily as a result of the previous one, for both the first and the second conversation a different task was used (that is to say, a similar description with different preferences and requirements leading to another restaurant suggestion). The same set of tasks was used for all participants. Two additional tasks were created as a 'back up' if anything unexpected went wrong during the first or the second task.

Procedure

At the beginning of the experiment, each participant was provided a small introduction to the chatbot and an explanation of how they are expected to communicate with it. The document with this introduction and task descriptions should be read first. This could also be altered during the conversation during the tasks. Each task contained a description of preferences and an additional requirement that the participant should be supposedly looking for in a restaurant. This was done in order to prevent the system from failing to find a suggestion, which would undesirably influence the user experience. The participants were not informed about the difference in formality of the chatbot or anything else that explains the system behind the chatbot. In the introduction, participants were asked to talk to the chatbot as they would have normally done in a situation that they would be using a chatbot to find a restaurant based on their preferences. They were also asked not to use punctuation marks. This is because the system is not able to process words that are attached to punctuation marks, unfortunately. When everything was clear, the participants started with the task 1. During the task, the participants should not be directed by any additional instructions, unless they wanted to end the task due to getting stuck in the conversation or any other error. If the instructions are followed correctly, and no unexpected errors occurs, the system provided a restaurant suggestion. At that point, the participant was asked to confirm the suggestion as soon as the system gives one, which also marked the end of each task. After each task, the participants were asked to fill out the survey in which they provided information about their demographics, as well as an evaluation of their conversation with the chatbot. For task 2, the exact same procedure was followed, ending with the same survey on their experience with the chatbot. The total experiment took around ten minutes per participant.

To give an example of one task description, the description of task 1 was as follows:

> You want to search for an <u>expensive Indian</u> restaurant in the <u>east</u> of the city. Try to see if KASA can find any restaurant that meets these preferences. If you get a restaurant recommendation you can confirm that this is the one you are looking for. This is the end of your conversation and you can type: *bye*.
> If KASA asks for any extra requirements, let KASA know that you would like to bring your <u>children</u>.

The underlined preferences match the keywords of the system with which it will always find a restaurant recommendation. This is done to make the user aware of the importance of these preferences and to promote the use of these specific words.

Measurements

Objective measurements. For each task, the number of turns of the participants was tracked, as well as the total time of their conversations with the chatbot. This was in order to obtain objective measurements of the experiment.

Subjective measurements. In the evaluation survey at the end of each task, questions about different aspects of user experience were posed. Participants were asked to fill in their demographics (anonymized names, age, and sex) and a rating of their own English proficiency, as mentioned earlier. For the evaluation, participants were asked to evaluate their experience with the Chatbot on a 5-Likert scale, where (1) is strongly disagree and (5) is strongly agree. The survey questions on user experience were as given below (based on [6], [5], [10]):

- How **natural** was your conversation with KASA?
- How **clear** were the answers and questions from KASA?
- What was the **quality** of the conversation with KASA?
- How **human-like** were the answers and questions of KASA?
- How would you rate your **experience** using KASA for a **restaurant recommendation**?
- How likely are you to **recommend** KASA?

At the end, one last open question was posed to prevent any potential problems during the conversations to influence the results:

- Did you experience any difficulties during your conversation? If so, what happened?

3 RESULTS

For both formal and informal users rated Naturalness (of the conversation), Clearness (of the answers), Quality (of the chatbot), Human-Likeness (of the chatbot), Experience (overall experience), Recommendation (whether someone would recommend this chatbot). Interaction (amount of dialog turns) and Time (total time spend on the conversation) were kept up. On each of the variables individually, the mean (μ) and standard deviation (σ) were calculated and a paired t-test was performed. For none of the variables, a significant effect of formality of the conversation was found. Below, as well as in Table 1, the mean, standard deviation and correlations are presented of all variables, together with the t-values and p-values that are also presented in Table 2. A visualisation of the presented results can be found in Figure 3.

Objective measurements. No difference was found between the number of interactions during the formal settings of the chatbot (μ = 7.6 and σ = 2.4) and during the informal settings (μ = 8.1, σ = 2.9), with t = -1.042373 and p = 0.299444. The number of interactions represent the number of responses that were given by the user. Similarly, no difference could be found for the duration of the conversation, measured in seconds, between the formal settings (μ = 268, σ = 133) and the informal settings (μ = 355 and σ = 565), with t = -1.134916 and p = 0.258791.

	Setting	Mean	SD	Correlation
Naturalness	*Formal*	3.88	0.83	0.40
	Informal	3.80	1.01	0.40
Clearness	*Formal*	4.36	0.80	0.40
	Informal	4.32	0.84	0.40
Quality	*Formal*	3.86	0.80	0.46
	Informal	3.84	0.91	0.46
Human-likeness	*Formal*	3.79	0.98	0.52
	Informal	3.69	0.95	0.52
Experience	*Formal*	3.91	0.84	0.56
	Informal	3.90	1.09	0.56
Recommendation	*Formal*	3.76	1.02	0.75
	Informal	3.71	1.10	0.75
Interaction	*Formal*	7.6	2.4	0.30
	Informal	8.1	2.9	0.30
Time	*Formal*	268	133	0.04
	Informal	355	565	0.04

Table 1: Mean, standard deviation and correlation

Subjective measurements. For the evaluation on Naturalness, no difference was found between the formal setting (μ = 3.88, σ = 0.83) and the informal setting (μ = 3.80, σ = 1.01), with t = 0.496485 and p = 0.620508. For the evaluation on Clearness, no difference was found between the formal setting (μ = 4.36, σ = 0.80) and the informal setting (μ = 4.32, σ = 0.84), with t=0.224247 and p=0.822967. For the evaluation on Quality, no difference was found between the formal setting (μ = 3.86, σ = 0.80) and the informal setting (μ = 3.84, σ = 0.91), with t = 0.107844 and p = 0.914309. For the evaluation of Human-Likeness, no difference was found between the formal setting (μ = 3.79, σ = 0.98) and the informal setting (μ = 3.69, σ = 0.95), with t = 0.572534 and p = 0.568087. For the evaluation of Experience, no difference was found between the formal setting (μ = 3.91, σ = 0.84) and the informal setting (μ = 3.90, σ = 1.09), with t = 0.094543 and p = 0.924844. For the evaluation of Recommendation, no difference was found between the formal setting (μ = 3.76 and σ = 1.02. In the informal setting (μ = 3.71 and σ = 1.10), with t = 0.347647 and p = 0.728746.

The critical value α was set to be α = 0.05, however, since there were 8 different t-tests we divide our α by 8, such that we remove the variable of any p-value from the t-tests occurs

	T-value	P-value
Naturalness	0.496485	0.620508
Clearness	0.224247	0.822967
Quality	0.107844	0.914309
Human-likeness	0.572534	0.568087
Experience	0.094543	0.924844
Recommendation	0.347647	0.728746
Interaction	-1.042373	0.299444
Time	-1.134916	0.258791

Table 2: Results of t-tests

to be smaller than α by chance (Bonferonni correction). In order to avoid a lot of spurious positives, the alpha value needs to be lowered to account for the number of comparisons being performed. This results in: $\alpha = 0.00625$. None of the t-test results has a p-value smaller than 0.00625. Moreover, none of the p-values is even smaller than 0.05. That is to say, still no significant differences is to be found between the two formality settings on the dependent variables.

4 DISCUSSION

Discussion of the results

The aim of the above experimentation was to find an answer to the research question on whether formality of the restaurant recommendation chatbot's responses influences the user experience. To do this, a survey on the user experience as well as objective measurements of the conversation with a restaurant recommendation system were conducted. The results that followed from the statistical analysis of the collected data showed that for none of the aspects to user experience a significant difference could be found due to the formality of the responses of the chatbot.
Results also showed that all features except time and interaction had a moderate strength of correlation. For Recommendation an even stronger strength of correlation was found. This implies that if the formal chatbot will more likely be recommended by a user, it is also more likely that the informal chatbot will be ed better. This does not mean there is necessarily a casual relationship, but it might be explained by the quality of the algorithm of the chatbot itself. So whether it is formal or not will not affect how well the chatbot is performing in suggesting restaurants, which is trivial since the algorithm of both cases are the same. The moderate to high correlation contributes a contrary factor to our hypothesis.

Implications and Explanations

The implications of these results are that, for a restaurant recommendation dialogue system, an informal formulation

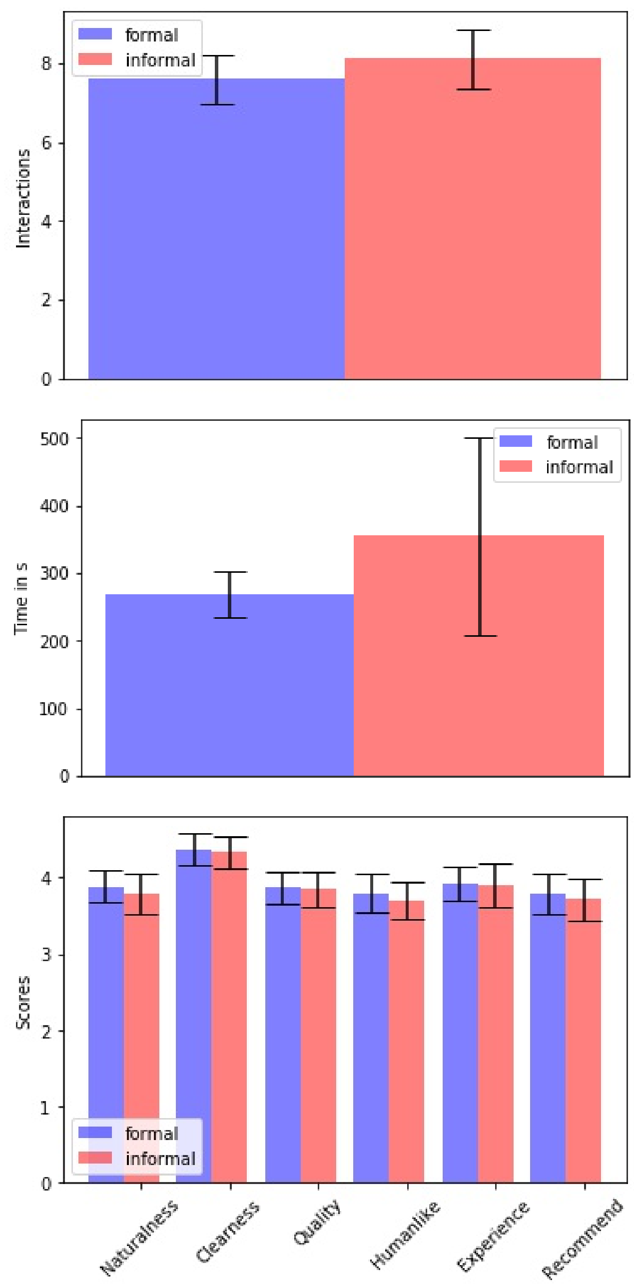

Figure 3: Bar plot of all dependent variables with error bars

of responses does not lead to the chatbot system being differently experienced by users than when the chatbot's responses are formally formulated. This is in contrast to what other research has pointed out on the human evaluation of chatbot systems. As explained in the first section of this paper, there have been plenty of results that point to the relevance of the communication style, and in particular, a chatbot such as in earlier mentioned research in section one

([9]) where an informal communication style induced the perception of social presence. These results led to the hypothesis that the same principles would hold for this type of dialogue system that recommends restaurants to people that indicate their preferences. However, based on the results of the current research, this hypothesis can be rejected.

Although previous research has found contrasting results for other but comparable systems, user preferences might not be the same when using different types of dialogue systems. As explained, formality is reflective of the distance and the relationship between the user and the system. If these distances and relationships to users differ for dialogue systems, then also the formality of the system might be experienced differently for these systems.

For starters, the type of goal of the system could be an explanation for not finding an effect of formality. Searching a restaurant could be considered a task that is not necessarily an informal or formal situation. The amount of distance or the specific relationship between the user and the chatbot might simply not be that clear in this case. Therefore, people might not be expecting or preferring one communication style over another when they are using a chatbot for restaurant recommendation. In other types of dialogue systems users may be more susceptible to the formality than for a restaurant recommendation chatbot. In the end of the day, a chatbot that is designed to provide personal medical information might be less expected to start its conversation with "hiyaaa what medicine are you looking for today????". More research is needed to gain more insight into the importance of formal and informal communication styles in this type of task-based dialogue systems.

Related to this, formality may also be too dependent on language and culture to note any difference in the current research. As explained in the introduction of this paper, languages can differ in their preferences and levels of formal language. It would also be interesting to investigate whether there are differences between different cultures or native speakers of different language. Future studies can be followed by experimenting the same settings in a different language.

Another possible explanation for not finding any difference in user experience due to the formality might be that this system was designed for completing a relatively simple tasks for which only short conversations were needed. The simplicity of this task might make the communication style of the chatbot less important, because in both the formal and informal setting the task and responses of the chatbot are clear and successful. An interesting topic for future research could be to examine the user experience in the same settings of a dialogue system that has a more complex goal or a more complex message to communicate. For example, the goal could be to provide a recommendation that is less easily found by use of three simple preference types.

Furthermore, according to the results, the variable Quality had a p-value of $0.91 > 0.00625$. This means that the quality of the chatbot in both settings did not differ according to the users. Following the results, it is possible that both the formal and informal versions of the system were too similar for the user in all other perspectives. Both the formal and informal version of the chatbot worked on the same algorithm, so the efficiency and quality of finding a restaurant did not differ for these situations. That is to say, regardless of the communication style, both situations might have similarly led to successful collaborations between the system and the user, that the communication style in this collaboration was of less importance than it could have been.

Limitations

Language and Formality. Several limitations of this study should be noted. First of all, the most important limitation of this study is the fact that the experiment was conducted with native speakers of Dutch, while the system was in English. Although this was already taken into account by carefully selecting proficient speakers of English, a better alternative would have been that the system was in the native language of the participants. This might have had as a results that the manipulation of formality has gone too unnoticed by the users to have any influence on their experience. To adapt the entire system and translate everything into the native language of the participants was beyond the scope of the project, unfortunately. Another important limitation of this research was that the formal and the informal setting of the chatbot might have been not sufficiently manipulated to properly experience a difference in formality as a user. As explained, the chatbot contained a formal an informal version of each response. However, as mentioned in section 1, the formality of a conversation can differ in many perspectives. For example, a less informative but more flexible style is related to an informal style. To implement complete informal communication, the system should not only alter formulation of responses, but also in other ways. This was, however, beyond the scope of the current research.

Age and of participants. Another limitation of the current research was the limited distribution in age of the participants. Different ages could perceive and prefer different communication styles, such as formality. Most of the participants were between 20 and 30 years old. For this reason, the results of this research can not be generalized to other age groups, since it is not excluded that different age groups prefer different formality settings.

Overall quality of the chatbot. A better chatbot may help to lower the effect of the quality of the chatbot on the survey ratings of the users. Sometimes the chatbot made distracting errors that were unforeseen. A better chatbot would shift the focus from these specific mistakes to the formulation of the responses. Finally, to limit the duration of the experiment for all participants, they only performed one task per condition with the chatbot. However, this made results even more sensitive to the influence of the chatbot making a mistake one time. This is something that could be taken into account by following research.

Conclusion

All together, the results of this research imply that for a restaurant recommendation dialogue chatbot the formality of its responses do not make any difference in user satisfaction. This is not in line with what was hypothesized based on previous research. This difference could point at the difference in user preferences and expectations due to different types of dialogue systems. Future research could look further into this. Limitations of this research imply that the results cannot be generalized to other age or language groups. It was also the case that the limitation of having the system only available in the second language of the participants might have interfered with perceiving any difference in formality. These limitations ask for more extensive research in order to generalize these results for the influence of formality in recommendation chatbot systems.

REFERENCES

[1] Theo Araujo. 2018. Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior* 85 (2018), 183–189.

[2] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, et al. 2010. The YouTube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*. 293–296.

[3] Drift, SurveyMonkey Audience, Salesforce, and Myclever. 2018. The State of Chatbots Report: How Chatbots Are Reshaping Online Experiences. https://www.drift.com/blog/Chatbots-report/

[4] Asbjørn Følstad and Marita Skjuve. 2019. Chatbots for customer service: user experience and motivation. In *Proceedings of the 1st international conference on conversational user interfaces*. 1–9.

[5] family=Zak given i=J. [n. d.]. UX Research Surveys: Best Questions to Ask for UX Research | Miro. https://miro.com/guides/ux-research/surveys-questions

[6] family=Team given i=M., given=Mjv. [n. d.]. 8 metrics for you to evaluate the success of your chatbot. https://www.mjvinnovation.com/blog/8-metrics-for-your-chatbot/

[7] Anaïs Gretry, Csilla Horváth, Nina Belei, and Allard CR van Riel. 2017. "Don't pretend to be my friend!" When an informal brand communication style backfires on social media. *Journal of Business Research* 74 (2017), 77–89.

[8] Matthew Henderson, Blaise Thomson, and Jason D Williams. 2014. The second dialog state tracking challenge. In *Proceedings of the 15th annual meeting of the special interest group on discourse and dialogue (SIGDIAL)*. 263–272.

[9] Francis Heylighen and Jean-Marc Dewaele. 1999. Formality of language: definition, measurement and behavioral determinants. *Interner Bericht, Center "Leo Apostel", Vrije Universiteit Brüssel* 4 (1999).

[10] Hotjar. [n. d.]. How to Create a UX Survey or Questionnaire | Get UX Feedback. https://www.hotjar.com/ux-design/surveys/

[11] Christine Liebrecht, Lena Sander, and Charlotte Van Hooijdonk. 2020. Too informal? How a chatbot's communication style affects brand attitude and quality of interaction. In *International Workshop on Chatbot Research and Design*. Springer, 16–31.

[12] Ellie Pavlick and Joel Tetreault. 2016. An empirical analysis of formality in online communication. *Transactions of the Association for Computational Linguistics* 4 (2016), 61–74.

[13] Dimitrios Rafailidis and Yannis Manolopoulos. 2018. The technological gap between virtual assistants and recommendation systems. *arXiv preprint arXiv:1901.00431* (2018).

[14] Lena Sander. 2020. Informal communication style of chatbots. 5 (2020), 123.

[15] Bayan Abu Shawar and Eric Atwell. 2007. Chatbots: are they really useful?. In *Ldv forum*, Vol. 22. 29–49.

[16] Abhishek Singh, Karthik Ramasubramanian, and Shrey Shivam. 2019. Introduction to Microsoft Bot, RASA, and Google Dialogflow. In *Building an Enterprise Chatbot*. Springer, 281–302.

[17] Evi van der Weegen. 2019. Humanlike chatbots: a blessing for online social customer service? *Tijdschrift voor Communicatiewetenschap* 47, 3-4 (2019), 23.

Appendices

A CONTRIBUTIONS

Contributions				
Task	Amy Oey	Kathleen de Boer	Andrea Roijen	Sunny Hsieh
Interim Report Part 2a	3h	0.5	5h	1h
Interim Report Part 2b	3h	-	-	-
Interim Report Part 2c	-	3h	-	-
Abstract	-	-	0.5h	1h
Introduction	9h	-	8h	1h
Method	5h	2h	8h	2h
Setup experiment	3h	3h	3h	2h
Research Question	1h	1h	1h	1h
Rewrite code for survey	7h	-	-	-
Rewrite text formality	-	3h	2h	3h
Data Collection	12h	12h	10h	10h
Data pre-processing	3h	3h	3h	2h
Statistical test	2h	3h	1h	4h
Results	1h	5h	3h	4h
Discussion	3h	2h	8h	5h