Creating Humanlike Chatbots: What Chatbot Developers Could Learn from Webcare Employees in Adopting a Conversational Human Voice

Christine Liebrecht[1(✉)] and Charlotte van Hooijdonk[2]

[1] Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands
C.C.Liebrecht@tilburguniversity.edu
[2] VU Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
C.M.J.Van.Hooijdonk@vu.nl

Abstract. Currently, conversations with chatbots are perceived as unnatural and impersonal. One way to enhance the feeling of humanlike responses is by implementing an engaging communication style (i.e., Conversational Human Voice (CHV); Kelleher 2009) which positively affects people's perceptions of the organization. This communication style contributes to the effectiveness of online communication between organizations and customers (i.e., webcare), and is of high relevance to chatbot design and development. This project aimed to investigate how insights on the use of CHV in organizations' messages and the perceptions of CHV can be implemented in customer service automation. A corpus study was conducted to investigate which linguistic elements are used in organizations' messages. Subsequently, an experiment was conducted to assess to what extent linguistic elements contribute to the perception of CHV. Based on these two studies, we investigated whether the amount of CHV can be identified automatically. These findings could be used to design humanlike chatbots that use a natural and personal communication style like their human conversation partner.

Keywords: Conversational Human Voice · Linguistic elements · Tool development · Chatbots

1 Introduction

Customer service plays an important role in organizations' ability to generate revenue. In recent years customer service has transformed from mediated communication (e.g., contact by phone) to computer-mediated communication (e.g., contact via social media channels; i.e. 'webcare'; Van Noort and Willemsen 2012) to human-AI interaction (e.g., contact using of chatbots). This transformation also occurs in the Netherlands: in 2016 4.7% of the organizations used chatbots to supplement their customer services. This number has tripled in the last two years (Van Os et al. 2016; 2018), because chatbots provide 24/7 customer service and save time and money by reducing the number of service employees (Gnewuch et al. 2017).

A. Følstad et al. (Eds.): CONVERSATIONS 2019, LNCS 11970, pp. 51–64, 2020.
https://doi.org/10.1007/978-3-030-39540-7_4

However, chatbot technology does not live up to its full potential yet. Much effort is put on the accuracy and performance of conversational AI, such as language recognition (Coniam 2008; Shawar and Atwell 2005), recall of previously mentioned topics (Jain et al. 2018), and the introduction of new topics or follow-up questions (Schuetzler et al. 2018; Silvervarg and Jönsson 2013), but currently, people perceive their conversations with chatbots as unnatural and impersonal (Drift, SurveyMonkey Audience, Salesforce, Myclever 2018).

One way to enhance the feeling of natural and personal chatbot responses, is by implementing a Conversational Human Voice (CHV, Kelleher 2009; Kelleher and Miller 2006). This communication style reflects human communication attributes, such as personally addressing the stakeholder, using informal speech, and being open to dialogue. Webcare research shows that CHV in organizational messages positively affects people's perceptions of the organization (e.g., Kerkhof et al. 2011; Park and Lee 2013). However, we have insufficient knowledge regarding the adoption of CHV in chatbots.

In a project funded by a NWO KIEM grant for creative industries we investigated how insights on the use of CHV in webcare messages and the perceptions of CHV can be implemented in customer service automation. We developed an online monitoring tool that enables webcare employees to respond with an appropriate communication style to customers' messages. This monitoring system may be useful as a basis for developing humanlike chatbots.

2 Theoretical Background

2.1 Chatbots as Social Actors

According to the Computers Are Social Actors (CASA) paradigm people tend to respond socially to computers, similarly to other humans, even when aware they are interacting with a computer (Nass et al. 1994). This implies that people automatically apply social rules, expectations, and scripts known from interpersonal communication in their interaction with computers (Nass and Moon 2000; Reeves and Nass 1996). These social reactions to computers in general (Nass and Moon 2000) and to chatbots in particular (von der Pütten et al. 2010) increase when more social cues are provided, such as communication style (Verhagen et al. 2014). For example, a customer service chatbot using informal speech increased the perception of the chatbot as being humanlike (Araujo 2018). A communication style that could be applied to chatbots is the Conversational Human Voice (Kelleher 2009; Kelleher and Miller 2006).

2.2 Operationalization of Conversational Human Voice

In order to enable chatbot designers to develop conversational agents that adopt CHV, it is important to understand which linguistic elements contribute to this communication style. Van Noort et al. (2014) distinguished three strategies to create CHV in messages, that were operationalized into several conversational linguistic elements by van Hooijdonk and Liebrecht (2018). The first strategy is Message Personalization: the degree to which a specific individual (organization and stakeholder) can be addressed in a message (cf. Walther 2011), such as greeting the stakeholder (*Hi Peter!*) and using personal pronouns (*you, your*) (van Hooijdonk and Liebrecht 2018). The second strategy is Informal Speech: casual, everyday language that differs from formal, corporate language (cf. Kelleher and Miller 2006), such as the adoption of non-verbal cues (*veeeery, :-)*) and interjections (*haha*) (van Hooijdonk and Liebrecht 2018). The third strategy is Invitational Rhetoric: to what extent the organization's communication style stimulates stakeholders to engage in conversations and creates mutual understanding between the parties (cf. Foss and Griffin 1995), such as acknowledging (*thanks for the message*) and showing sympathy/empathy (*I can imagine this is disappointing*) (van Hooijdonk and Liebrecht 2018).

It has been shown that the adoption of CHV by chatbots is beneficial for organizations. Liebrecht and van der Weegen (to appear) found that customer service chatbots using multiple conversational linguistic elements from all three strategies enhanced brand attitude and perceived warmth of the chatbot. These relations were mediated by the perceived social presence: people's perceptions of actually communicating with another human being (Short et al. 1976). Thus, the adoption of CHV in chatbots can diminish customers' feelings of unnatural and impersonal service contact.

2.3 Aim of This Paper

To facilitate the development of humanlike chatbots, several design issues should be addressed. In this paper, we focus on two aspects from webcare research that could inform the development of conversational agents that adopt a humanlike conversation style. First, following the principles of Communication Accommodation Theory (CAT; Giles et al. 1991), a chatbot's communication style should match the communication style of the customer (Jakic et al. 2017). This requires that the chatbot can automatically identify conversational linguistic elements in the customer's messages. In order to train a conversational agent to recognize these elements, we first needed to establish whether human coders can identify them reliably. Furthermore, the identification also results in a list of conversational linguistic elements that can be used to train the conversational agent on the recognition of CHV. We therefore conducted a corpus analysis to investigate which conversational linguistic elements webcare employees of the Netherlands Red Cross use in their messages to various stakeholders (e.g., benefactors, collectors, emergency workers, etc.) on public and private social media channels (i.e., Study 1). This study is a replication of van Hooijdonk and Liebrecht's (2018) study, who conducted a corpus analysis on conversational linguistic elements in webcare messages of Dutch municipalities on Twitter.

Second, the contribution of specific conversational linguistic elements to the perception of CHV also needs to be investigated. Although the presence of conversational linguistic elements seems to contribute to perceived CHV (van Noort et al. 2014), the weighted contribution of each linguistic element is unknown. Several experimental studies investigated the relation between linguistic elements in webcare messages and perceived CHV (e.g., Park and Lee 2013; Barcelos et al. 2018), but there are considerable differences in the type and number of linguistic elements used. For example, Park and Lee (2013) found that the perceived CHV increased by only one personalization element (i.e., signature, such as ^CL), whereas Barcelos et al. (2018) concluded that a combination of personalization elements and informal speech increased the perceived CHV. These results are also relevant for the design of chatbots' communication style. Liebrecht and van der Weegen (to appear) included multiple conversational linguistic elements from all three strategies, but it is unclear which elements contribute to what extent to the perception of CHV, and consequently to people's perceptions of the chatbot and the organization. To examine how conversational linguistic elements are related to the perceived CHV, an experiment was conducted in which webcare employees evaluated the perceived CHV of messages (i.e., Study 2). Finally, the findings of both Study 1 and Study 2 were used to investigate whether the amount of CHV in messages can be identified automatically (i.e., Study 3).

3 Study 1: Identification of Conversational Linguistic Elements

3.1 Method

The OBI4wan monitoring tool[1] was used to collect a random sample of webcare dialogues from March 2017 until October 2017 between the Netherlands Red Cross and their stakeholders. The sample included both public as well as private channels. The public conversations were collected from Instagram (35), Facebook (75), and Twitter (81). The private conversations were collected from WhatsApp (80), Facebook Messenger (72), and Twitter DM (80). The total corpus contained 423 dialogues (895 stakeholders' messages and 689 webcare messages).

We only collected Dutch webcare conversations and anonymized them by deleting names, addresses, and phone numbers. Thereafter, the linguistic elements were manually coded by five coders and (partly) double coded by one of the authors of this paper. We used a slightly adjusted version of the identification instrument of van Hooijdonk and Liebrecht (2018): Informal Speech categories Shortenings and Abbreviations were merged and one Message Personalization category (i.e., Addressing the webcare employee) and one Invitational Rhetoric category (i.e., Well-wishing) were added.

[1] The OBI4wan monitoring tool enables organizations to monitor and manage stakeholders' messages on multiple public and social media channels (e.g., Twitter, Instagram, Facebook, Facebook Messenger, and WhatsApp).

3.2 Results

Table 1 shows the identification instrument and the intercoder reliability scores per subcategory. In accordance with Van Hooijdonk and Liebrecht's (2018) findings, the identification instrument turned out to be reliable. The codings of all Message Personalization subcategories resulted in perfect reliability scores. Regarding Informal Speech, the intercoder reliability of interjections was perfect. The reliability of non-verbal cues, and shortenings and abbreviations was substantial. The intercoder reliability scores of the Invitational Rhetoric subcategories varied from perfect to fair. Whereas apologizing, acknowledging, and well-wishing resulted in perfect reliability scores, joking, sympathy/empathy, and stimulating dialogues had poor scores. This was possibly due to its limited presence in the double coded sample.

Table 1. Identification instrument of linguistic elements, the Krippendorff's alpha scores per subcategory, their absolute and relative frequency in the corpus ($N_{webcaremessages} = 689$)

Linguistic element	Krippendorff's alpha	Frequency	Example
Message Personalization			
Greeting	.98	239 (34.7%)	Hi Peter!
Addressing stakeholder	.92	448 (65.0%)	you, your, Anna
Addressing webcare*	.92	352 (51.1%)	I, we, my, us
Signature	.92	570 (82.7%)	^WP
Informal Speech			
Shortenings/abbreviations*	.70	53 (7.7%)	pls, ok, LOL, DM
Non-verbal cues	.88	53 (7.7%)	??, veeery, :-)
Interjections	1.00	27 (3.9%)	haha, oh
Invitational Rhetoric			
Acknowledging	.96	190 (27.6%)	thanks for the message
Apologizing	1.00	20 (2.9%)	I am sorry
Sympathy/empathy	.59	179 (26.0%)	I can imagine this is disappointing
Stimulating dialogues	.32	38 (5.5%)	Let us know what you think
Joking	.66	9 (1.3%)	#joke, just kidding
Well-wishing*	.89	113 (16.4%)	Have a nice day!

Note. The asterisks represent categories that are adjusted to the van Hooijdonk and Liebrecht (2018) identification instrument.

Table 1 also shows the presence of linguistic elements in webcare conversations of the Netherlands Red Cross. Message Personalization was frequently used. Especially signatures of employees were frequently employed and webcare employees often address stakeholders personally. Informal Speech, on the other hand, was less frequent in webcare messages. If webcare employees used informal speech, they mostly employed non-verbal cues or shortenings and abbreviations. Regarding Invitational Rhetoric, acknowledging, showing sympathy/empathy and well-wishing were often present.

The corpus enabled us to compare the usage of linguistic elements in webcare responses across public and private social media channels. To do this, we aggregated the identified linguistic elements per webcare tweet into an average score per webcare conversation (see Table 2). The analyses showed significant differences between the social media channels for all Message Personalization categories: personal greetings of the stakeholder ($F(5,417) = 42.82$, $p < .001$, $\eta_p^2 = .34$.), addressing stakeholder ($F(5,417) = 17.98$, $p < .001$, $\eta_p^2 = .18$), addressing webcare employee ($F(5,417) = 25.24$, $p < .001$, $\eta_p^2 = .23$), and signatures ($F(5,417) = 64.02$, $p < .001$, $\eta_p^2 = .43$). The first three categories appeared more often in private social media channels than in public social media channels. Regarding signatures, pairwise Bonferroni comparisons showed that these appeared least on Instagram compared to the other channels (Twitter: $p < .001$; Facebook: $p < .001$; WhatsApp: $p < .001$; Twitter DM: $p < .001$, Facebook Messenger: $p < .001$).

Table 2. Mean presence of linguistic elements in webcare conversations per social media channel (standard deviations between brackets).

Linguistic element	Public channels			Private channels			Total (n = 423)
	Instagram (n = 35)	Facebook (n = 75)	Twitter (n = 81)	WhatsApp (n = 80)	Facebook Mess. (n = 72)	Twitter DM (n = 80)	
Greeting	.03 (.17)	.19 (.38)	.01 (.07)	.65 (.39)	.59 (.43)	.44 (.44)	.34 (.44)
Addressing stakeholder	.44 (.50)	.62 (.47)	.43 (.46)	.86 (.27)	.89 (.26)	.74 (.36)	.68 (.43)
Addressing webcare	.11 (.32)	.34 (.47)	.26 (.41)	.63 (.40)	.78 (.36)	.67 (.42)	.50 (.46)
Signature	.21 (.41)	.97 (.16)	.94 (.22)	.92 (.21)	.93 (.21)	.93 (.23)	.88 (.31)
Shortenings/abbreviations	.03 (.17)	.00 (.00)	.13 (.32)	.12 (.29)	.03 (.17)	.13 (.28)	.08 (.24)
Non-verbal cues	.34 (.48)	.03 (.16)	.10 (.28)	.06 (.19)	.06 (.22)	.07 (.22)	.09 (.26)
Interjections	.03 (.17)	.03 (.16)	.08 (.24)	.03 (.15)	.02 (.13)	.06 (.21)	.04 (.18)
Acknowledging	.31 (.46)	.45 (.49)	.28 (.43)	.25 (.36)	.34 (.42)	.32 (.43)	.32 (.43)
Apologizing	.00 (.00)	.00 (.00)	.03 (.17)	.02 (.08)	.08 (.23)	.03 (.15)	.03 (.14)
Sympathy/empathy	.41(.49)	.26 (.44)	.57 (.46)	.23 (.35)	.33 (.42)	.15 (.31)	.32 (.43)
Stimulating dialogues	.00 (.00)	.03 (.16)	.02 (.10)	.10 (.25)	.08 (.23)	.03 (.11)	.05 (.23)
Joking	.06 (.24)	.03 (.16)	.03 (.14)	.00 (.00)	.01 (.06)	.01 (.11)	.02 (.12)
Well-wishing	.06 (.24)	.05 (.20)	.27 (.43)	.15 (.30)	.24 (.37)	.18 (.33)	.17 (.34)

For Informal Speech significant differences between the social media channels were found for shortenings and abbreviations ($F(5,417) = 4.59$, $p < .001$, $\eta_p^2 = .05$), and non-verbal cues ($F(5,417) = 8.98$, $p < .001$, $\eta_p^2 = .10$). The former appeared less often on Facebook compared to Twitter ($p = .02$), Twitter DM ($p = .007$), and WhatsApp ($p = .02$). Non-verbal cues were more frequent on Instagram (compared to Twitter ($p < .001$), Facebook ($p < .001$), Twitter DM ($p < .001$), Facebook Messenger ($p < .001$), and WhatsApp ($p < .001$)). No differences were found between the social media channels in the mean number of interjections ($F(5,417) = 1.04$, $p = .39$). Regarding Invitational Rhetoric, the social media channels differed for all subcategories, with the exception of Acknowledging ($F(5,417) = 2.07$, $p = .07$, $\eta_p^2 = .02$), and

Joking ($F(5,417) = 2.28$, $p = .27$). However, the results did not show a consistent pattern between the public versus private social media channels.

In sum, webcare employees frequently adopted linguistic elements of Message Personalization in their messages. Invitational Rhetoric is also used regularly, but Informal Speech hardly appeared in webcare messages. Furthermore, public and private social media channels differ in the presence of linguistic elements.

4 Study 2: Contribution of Conversational Linguistic Elements to the Perceived CHV

4.1 Method

To examine to what extent each linguistic element contributes to the perception of CHV, an experiment was conducted. The experimental materials were developed on the basis of the webcare conversations of Study 1. The materials consisted of conversations between a stakeholder asking questions to a fictitious charity organization to which the organization responded. For these webcare responses, a basic response was formulated that contained an average amount of perceived CHV (which was determined in a pretest). An example of a conversation is shown in Fig. 1. Subsequently, the basic webcare response was adjusted by adding one of the linguistic element subcategories. For example, to include a non-verbal cue in the response, a smiley was added to the basic response. Nine CHV subcategories were included in the experiment (i.e., three subcategories per main category). For Message Personalization, greeting, addressing stakeholder, and signature were chosen. From the main category Informal Speech, shortenings and abbreviations, non-verbal cues, and interjections were included. Finally, showing sympathy/empathy, stimulating dialogue, and well-wishing were chosen from the Invitational Rhetoric category. In short, nine webcare responses per basic response were created by adding one of these nine CHV subcategories.

Stakeholder's message	Webcare response
Robin: @charityorganization Where can I find information about your projects? Can't find it on your website.	Thanks for the notification. This part of the website is under construction until tonight, after which the Projects page is completely up to date. Sufficient information will be available soon!

Fig. 1. Example of a basic webcare response.

The experiment conformed to a 1 (Stakeholder's Question) × 10 (Linguistic Element incl. basic response) within subjects latin square design. To avoid repetition of the questions' topics, ten customer service topics and accompanying webcare responses were created (10 topics * 10 webcare responses): each participant assessed one experimental condition per customer service topic (10 webcare responses in total).

Forty-seven webcare employees of different charity organizations in the Netherlands were recruited via their social media channels to participate in the study. The study consisted of two tasks. First, participants assessed the perceived CHV of every experimental condition on seven-point Likert scales. The perceived CHV was operationalized with 3 items: ‘The webcare response is personal/informal/distant’ (reversed item). The internal consistency of the items was high ($\alpha = .86$, $M = 4.44$, $p < .001$). Subsequently, the participants conducted a ranking task. Per main category, the basic response and the three manipulated responses (i.e., each response included one of the three subcategories) was shown. Participants ranked the tone of voice of the four responses from least to most human. Consequently, participants could write their own webcare response and underpin their choices regarding their ranking of the webcare responses.

4.2 Results

Table 3 shows the findings of both tasks. The scores of the ranking task were transformed into scores on a 4-point scale ranging from 1 (least human) to 4 (most human) To investigate whether the mean scores differ significantly, we conducted repeated measures ANOVAs with simple contrasts effects.

Table 3. Means of the perceived CHV per linguistic element (standard deviations between brackets). The assessment task used a 7-point scale, the ranking task used a 4-point scale.

CHV category	Assessment task	Simple contrast effects, $F(1,46)$	Ranking task	Simple contrast effects, $F(1,46)$
Basic response	3.80 (1.37)			
Personalization	4.78 (1.04)	$F = 29.77$, $p < .001$	Basic = 1.26 (.77)	
Greeting	5.30 (1.21)	$F = 43.24$, $p < .001$	3.60 (.83)	$F = 115.48$, $p < .001$
Addressing stakeholder	4.31 (1.44)	$F = 5.79$, $p = .02$	2.36 (.64)	$F = 46.87$, $p < .001$
Signature	4.72 (1.43)	$F = 17.08$, $p < .001$	2.79 (.72)	$F = 102.08$, $p < .001$
Informal Speech	4.31 (1.12)	$F = 8.58$, $p = .005$	Basic = 2.13 (.90)	
Shortenings/abbreviations	3.80 (1.44)	$F < 1$, $p = 1.00$	1.74 (.90)	$F = 4.22$, $p = .046$
Non-verbal cues	4.46 (1.50)	$F = 7.96$, $p = .007$	3.28 (.85)	$F = 28.55$, $p < .001$
Interjections	4.68 (1.40)	$F = 12.25$, $p = .001$	2.85 (1.16)	$F = 8.00$, $p = .007$
Invitational rhetoric	4.43 (1.03)	$F = 9.87$, $p = .003$	Basic = 1.09 (.28)	
Sympathy/empathy	4.34 (1.51)	$F = 3.72$, $p = .06$	2.68 (.84)	$F = 147.52$, $p < .001$
Stimulating dialogues	4.42 (1.57)	$F = 6.21$, $p = .02$	3.32 (.84)	$F = 280.81$, $p < .001$
Well-wishing	4.52 (1.38)	$F = 9.80$, $p = .003$	2.91 (.83)	$F = 187.34$, $p < .001$

Regarding the assessment task, the basic response had an average perceived CHV score ($M = 3.80$). The results indicated that each main category differed significantly from the basic response in perceived CHV. Table 3 illustrates that Message Personalization contributed the most to the perceived CHV, whereas Informal Speech contributed least. A closer inspection of the subcategories of linguistic elements showed differences between them. Shortenings and abbreviations did not enhance the perceived CHV. Pairwise comparisons showed this subcategory differed from personal greetings ($p < .001$), signatures ($p = .009$), interjections ($p = .03$), and well-wishing ($p = .03$). Also, greetings enhanced the perceived CHV more than addressing stakeholder ($p < .001$), and showing sympathy/empathy ($p = .03$).

Regarding the ranking task, subcategories within Message Personalization enhanced perceived CHV compared to the basic response. Pairwise comparisons indicated that greetings resulted in a higher perceived CHV than addressing stakeholder ($p < .001$), and signatures ($p < .001$). A similar pattern is found for Invitational Rhetoric. The three subcategories significantly enhanced the perception of CHV compared to the basic response. Pairwise comparisons showed that stimulating dialogues induced a higher perceived CHV than showing sympathy/empathy ($p = .03$). However, a different pattern was found for Informal Speech. Pairwise comparisons showed that shortenings and abbreviations scored significantly lower than non-verbal cues and interjections on the perceived CHV (non-verbal cues $p < .001$; interjections $p = .001$).

In sum, it can be concluded that linguistic elements differ in their contribution to perceived CHV of webcare messages. Greeting the stakeholder, non-verbal cues, and stimulating dialogues contributed most to the perception of CHV.

5 Study 3: Automatic CHV Identification

In order to explore whether it is feasible to implement the insights on the usage and perceptions of linguistic elements to customer service automation (e.g., chatbots) we examined to what extent the amount of CHV can be identified automatically. We therefore developed a beta-version of a tool together with OBI4wan[2]. In this section, we report the development of the tool and the first qualitative results.

5.1 Method

The findings of Study 1 and Study 2 informed the development of the automatic indication of the amount of perceived CHV in webcare responses. The codings of the linguistic categories in Study 1 allowed us to compile a list of Illocutionary Force Indicating Devices (IFIDs; Houtkoop and Koole 2000) that indicate the potential presence of a subcategory. For example, 'you', 'your', 'yours', are words that were often used to address the stakeholder. This list contained all linguistics elements found in Study 1, supplemented with synonyms from (online) sources. Also, standardized lists containing first names, abbreviations, and emoticons were used. The tool was trained on the basis of these lists to identify the linguistic elements.

To calculate the amount of perceived CHV in a message, we created a ranking and a formula based on the average scores in Study 2. For example, within the main

[2] The beta-version of the tool can be tested on request by the authors and OBI4wan.

category Message Personalization, greeting the stakeholder contributed most to the perceived CHV. Therefore, the presence of this linguistic element in a webcare message contributed more to the perceived CHV than presence of other Message Personalization categories, such as a signature. To investigate whether the tool was able to indicate the amount of perceived CHV in webcare messages, the webcare messages of Study 1 were used as input.

5.2 Results

In Table 4 three webcare messages are shown which the tool qualified as having a high amount of perceived CHV. The first example contains all subcategories of Message Personalization, and stimulating dialogue (the subcategory within Invitational Rhetoric that contributes most to the perception of CHV). The second example contains several linguistic elements of Invitational Rhetoric, and two linguistic elements of Message Personalization. In the third example, multiple linguistic elements of all three main categories are present. Within their categories, the smiley and stimulating dialogues contributed the most to the perception of CHV.

Table 4. Webcare messages the tool qualified as having a high (examples 1–3) versus low (examples 4–6) amount of CHV.

Webcare message	CHV elements
1. Hi Dave, how can we help you? Greetings, Niels. [WhatsApp]	Message Personalization: greeting, addressing stakeholder, addressing webcare, signature. Invitational Rhetoric: stimulating dialogue
2. Apologies, it is not our intention to irritate you. Thank you for the support you have already given. Have a nice #spring day [Twitter]	Message Personalization: addressing stakeholder, addressing webcare. Invitational Rhetoric: apology, sympathy, acknowledgement, well-wishing
3. No problem! We are happy to help as far as we can in this case:) Have a nice weekend. Greetings, Ilse [Twitter DM]	Message Personalization: addressing webcare, signature. Informal Speech: non-verbal cues. Invitational Rhetoric: stimulating dialogue, well-wishing
4. The information can be found here. [hyperlink] Greetings, Ilse [Facebook Messenger]	Message Personalization: signature
5. Thanks for your support! [Instagram]	Message Personalization: addressing stakeholder. Invitational Rhetoric: acknowledging
6. That is true Carmen:) ^Caroline [Facebook]	Message Personalization: addressing stakeholder, signature. Informal Speech: non-verbal cue

Table 4 also shows three webcare messages which the tool qualified as having a low amount of perceived CHV. Despite the presence of several linguistic elements, these webcare messages will be perceived as less personal and engaging, because their relative contribution to the perception of CHV is low. This is illustrated in example 4 in which the webcare message only contains a signature. In example 5 an acknowledgement is expressed and the stakeholder is addressed personally. However, only addressing the stakeholder was taken into account in the calculation of the perceived CHV. The final example contains three subcategories, but only the non-verbal cues had a relatively high contribution to the perceived CHV.

Although these first qualitative results of the beta-version are promising for CHV recognition, not all qualifications of the tool correspond with our own observations. First, the identification of the linguistic elements can be improved. Although extensive lists are used to inform the tool, some linguistic features were not identified, or identified incorrectly. For example, first names that did not occur in our lists were not identified. Second, the current beta-version is programmed to identify all linguistic element categories, but only the categories that are measured in Study 2 are included in the calculation of the CHV score. As a result, messages that do contain several CHV subcategories could still be qualified as having a low amount of CHV. Finally, the amount of CHV is only calculated for one webcare message. However, a webcare conversation can consist of multiple webcare messages, and the position of these messages within the conversation influences the linguistic elements used.

6 Conclusion and Discussion

This project aimed to inform the development of humanlike chatbots that use an appropriate amount of CHV that matches the communication style of the conversation partner. We therefore obtained insights from the usage and perceptions of conversational linguistic elements by employees in webcare conversations, that can be adopted to customer service automation tools, such as chatbots. By learning from natural language use by humans, chatbot developers can design conversational agents that will be perceived more humanlike, which in turn might positively impact users' evaluations of the chatbot and the organization.

The first learning can be derived from our corpus study: Message Personalization should be adopted in chatbot conversations, because webcare employees frequently use these linguistic elements in their messages. Invitational Rhetoric was also used regularly, whereas Informal Speech was hardly employed. These findings support prior findings of van Hooijdonk and Liebrecht (2018). In addition, we showed that webcare employees employ linguistic elements differently in public and private channels. Private social media messages contained more personal greetings, addressing the stakeholder, and addressing the webcare employee, which is informative for the private nature of chatbot conversations.

Secondly, chatbot developers should be aware of the relative contribution of linguistic elements to the perception of CHV. Our experimental study showed that greeting the stakeholder induced the highest perception of CHV compared to the other subcatgories within Message Personalization. Within Informal Speech non-verbal cues contributed most to the perception of CHV. Finally, stimulating dialogues contributed most to the perception of CHV compared to the other subcategories within Invitational

Rhetoric. To our knowledge, this is the first study that systematically examined the relation between the use of single linguistic elements and the perception of CHV. If developers aim to create chatbots with a high amount of CHV, we advise to include personal greetings, non-verbal cues and sentences that stimulate dialogue.

Thirdly, it is possible to develop chatbots that use an appropriate communication style that matches the communication style of the human conversation partner. In Study 3, we showed that the amount of CHV in messages can be identified automatically. A first test showed that the tool was able to identify conversational linguistic elements and to calculate the amount of CHV in messages. Although more CHV categories must be added to the tool and some improvements are necessary, the findings are promising for customer service automation since it is shown that language accommodation positively impacts on people's perceptions (Jakic et al. 2017). However, the preferred organization's tone of voice should be taken into account as well. As distinguished in Giles et al.'s (1991) CAT, communication partners could also use a maintenance strategy, meaning that the interactant does not change the original communication style to the style of the conversation partner but sticks to the own, preferred communication style that matches the organization's image.

Finally, our findings can be used to research the usage and the effects of humanlike chatbots more systematically. On the one hand, our approach can be used to compare available chatbots on CHV or to monitor the same chatbot on CHV across time. On the other hand, people's perceptions of humanlike chatbots can be investigated. Feine et al. (2019) presented a taxonomy of cues that following CASA paradigm could impact on people's social reactions to chatbots. The conversational linguistic elements of our study can be seen as a concrete manifestation of verbal social cues, but little is known how these cues impact on users' perceptions and behavior. Given the differences of CHV elements to the contribution to the perceived CHV, it is important to investigate how human and personalized a chatbot should be. Designing chatbots that resemble humans may easily lead to users making wrong assumptions regarding the chatbot capabilities (e.g., Luger and Sellen 2016). We therefore need to evaluate which and how many CHV elements are considered appropriate and how they influence users' perceptions and use of chatbots.

References

Kelleher, T.: Conversational style, communicated commitment, and public relations outcomes in interactive online communication. J. Commun. **59**, 172–188 (2009)

Van Noort, G., Willemsen, L.: Online damage control: the effects of proactive versus reactive webcare interventions in consumer-generated and brand-generated platforms. J. Interact. Mark. **26**(3), 131–140 (2012)

Van Os, R., Hachmang, D., Akpinar, M., Derksen, M., Keuning, A.: Stand van Webcare 2018 (2018). https://www.upstream.nl/onderzoeken/stand-van-webcare-2018/

Van Os, R., Hachmang, D., Derksen, M., Keuning, A.: Stand van Webcare 2016 (2016). https://www.marketingfacts.nl/berichten/onderzoek-stand-van-webcare-nederland-2016

Gnewuch, U., Morana, S. Maedche, A.: Towards designing cooperative and social conversational agents for customer service. In: Proceedings of the International Conference on Information Systems (ICIS) (2017)

Coniam, D.: Evaluating the language resources of chatbots for their potential in English as a second language. ReCALL **20**(1), 98–116 (2008)

Shawar, B.A., Atwell, E.S.: Using corpora in machine-learning chatbot systems. Int. J. Corpus Linguist. **10**(4), 489–516 (2005)

Jain, M., Kumar, P., Kota, R., Patel, S.N.: Evaluating and informing the design of chatbots. In: Proceedings of the 2018 Designing Interactive Systems Conference, pp. 895–906 ACM (2018)

Schuetzler, R.M., Grimes, G.M., Giboney, J.S.: An investigation of conversational agent relevance, presence, and engagement. In: Proceedings of Americas' Conference on Information Systems (2018)

Silvervarg, A., Jönsson, A.: Iterative development and evaluation of a social conversational agent. In: 6th International Joint Conference on Natural Language Processing, pp. 1223–1229 (2013)

Drift, SurveyMonkey Audience, Salesforce, Myclever: The 2018 State of Chatbots Report. How chatbots are reshaping online experiences. https://www.drift.com/wp-content/uploads/2018/01/2018-state-of-chatbots-report.pdf. Accessed 01 Sept 2019

Kelleher, T., Miller, B.: Organizational blogs and the human voice: relational strategies and relational outcomes. J. Comput.-Mediat. Commun. **11**, 395–414 (2006)

Kerkhof, P., Beugels, D., Utz, S., Beukeboom, C.J.: Crisis PR in social media: an experimental study of the effects of organizational crisis responses on Facebook. Paper presented at the 61st Annual ICA Conference, Boston, USA (2011)

Park, H., Lee, H.: Show us you are real: the effect of human-versus-organizational presence on online relationship building through social networking sites. Cyberpsychol. Behav. Soc. Netw. **16**(4), 265–271 (2013)

Nass, C., Steuer, J., Tauber, E.R.: Computers are social actors. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 72–78. ACM (1994)

Nass, C., Moon, Y.: Machines and mindlessness: social responses to computers. J. Soc. Issues **56** (1), 81–103 (2000)

Reeves, B., Nass, C.I.: The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press, Cambridge (1996)

Von der Pütten, A.M., Krämer, N.C., Gratch, J., Kang, S.H.: It doesn't matter what you are! Explaining social effects of agents and avatars. Comput. Hum. Behav. **26**(6), 1641–1650 (2010)

Verhagen, T., van Nes, J., Feldberg, F., van Dolen, W.: Virtual customer service agents: Using social presence and personalization to shape online service encounters. J. Comput.-Mediat. Commun. **19**(3), 529–545 (2014)

Araujo, T.: Living up to the chatbot hype: the influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. Comput. Hum. Behav. **85**, 183–189 (2018)

van Noort, G., Willemsen, L.M., Kerkhof, P., Verhoeven, J.W.M.: Webcare as an integrative tool for customer care, reputation management, and online marketing: a literature review. In: Kitchen, P.J., Uzunoğlu, E. (eds.) Integrated Communications in the Postmodern Era, pp. 77–99. Palgrave Macmillan UK, London (2015). https://doi.org/10.1057/9781137388551_4

Van Hooijdonk, C., Liebrecht, C.: "Wat vervelend dat de fiets niet is opgeruimd! Heb je een zaaknummer voor mij? ^EK". Conversational human voice in webcare van Nederlandse gemeenten. Tijdschr. voor Taalbeheers. **40**(1), 45–82 (2018)

Walther, J.: Theories of computer-mediated communication and interpersonal relations. In: Knapp, M., Daly, J. (eds.) The Sage Handbook of Interpersonal Communication, 4th edn. Sage Publications (2011)

Foss, S., Griffin, C.: Beyond persuasion: a proposal for an invitational rhetoric. Commun. Monogr. **62**(1), 2–18 (1995)

Liebrecht, C., van der Weegen, E.: Menselijke chatbots: een zegen voor online klantcontact? Het effect van conversational human voice door chatbots op social presence en merkattitude. Tijdschrift voor Communicatiewetenschap (to appear)

Short, J., Williams, E., Christie, B.: The Social Psychology of Telecommunications. Wiley, London (1976)

Giles, H., Coupland, J., Coupland, N.: Accommodation theory: communication, context and consequences. In: Giles, H., Coupland, J., Coupland, N. (eds.) Contexts of Accommodation – Developments in Applied Sociolinguistics, pp. 1–68. Cambridge University Press, New York (1991)

Jakic, A., Wagner, M.O., Meyer, A.: The impact of language style accommodation during social media interactions on brand trust. J. Serv. Manag. **28**(3), 418–441 (2017)

Barcelos, R., Dantas, D., Sénécal, S.: Watch your tone: how a brand’s tone of voice on social media influences consumer responses. J. Interact. Mark. **41**, 60–80 (2018)

Houtkoop, H., Koole, T.: Taal in actie: hoe mensen communiceren met taal. Coutinho, Bussum (2000)

Feine, J., Gnewuch, U., Morana, S., Maedche, A.: A taxonomy of social cues for conversational agents. Int. J. Hum Comput Stud. **132**, 138–161 (2019)

Luger, E., Sellen, A.: Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM (2016)