Investigating the impact of interlocutor voice on syntactic alignment in human-computer dialogue

Benjamin R. Cowan
HCI Centre
School of Computer Science
University of Birmingham
Edgbaston
B15 2TT
b.r.cowan@cs.bham.ac.uk

Holly P. Branigan
Department of Psychology
7 George Square
University of Edinburgh
Edinburgh
EH8 9JZ
holly.branigan@ed.ac.uk

Russell Beale
HCI Centre
School of Computer Science
University of Birmingham
Edgbaston
B15 2TT
r.beale@cs.bham.ac.uk

Language is at the core of most social activity. Psycholinguistic research has shown that our conversational partners influence our linguistic choices be it syntactic or lexical, a concept termed *alignment*. As our interaction with computer interlocutors become more frequent recent efforts have been made to understand how and what impacts alignment with computers, showing that our perceptions of computer systems impact on alignment with computer interlocutors. This work looks to identify the impact of how spoken dialogue system design characteristics, specifically system voice type, impact user linguistic behaviour in terms of syntactic alignment in human-computer dialogue. Additionally we wished to identify whether syntactic alignment levels can be used as a behavioural indicator of interaction satisfaction. The research used a wizard of oz experiment design paired with a confederate-scripting paradigm commonly used in psycholinguistics research. We found that there was no significant effect of voice type on syntactic alignment, although there was a significant effect of voice type on interaction satisfaction. Participants rated their experiences with a basic computer voice significantly lower in satisfaction compared to human based and advanced voice computer conditions. The results are discussed in terms of the conceptual nature of syntactic alignment and the impact of item stimuli on alignment levels. Future plans for research are also discussed.

Keywords: Syntactic Alignment, Human-computer Dialogue, Satisfaction, Controlled Experiment, Psycholinguistics.

1. INTRODUCTION

As social beings we interact frequently. Language is usually at the core of this socialisation. Through language we convey thoughts, feelings and ideas to our conversational partners. Our conversational partners also influence how we express ourselves linguistically. Over a decade of research has highlighted that our conversational partners' speech has an effect on our language choices in that we converge on the syntax we use (syntactic alignment) (Branigan et al. 2000; Pickering & Branigan 1998) or the word choices we make to describe concepts and objects (lexical alignment) (Brennan 1996; Brennan & Clark 1996). This convergence or co-ordination in dialogue is termed *alignment*.

Much of the alignment literature focuses on our interactions with other human interlocutors (human-human dialogue). Yet as dialogue interactions with computers become more frequent, an understanding of how computer interlocutors influence our speech in human-computer dialogue (HCD) becomes an interesting proposition for psycholinguistic and human-computer interaction researchers. The understanding of user linguistic behaviour is of high practical value to spoken dialogue interface designers with such knowledge being able to inform the design and build of such systems. Additionally, as explained further in this section and explored in this work, alignment in dialogue could also act as a potential behavioural indicator of natural and satisfying interactions with such interfaces.

This potential is partly the reason for the recent incorporation of alignment research in human-computer dialogue scenarios (Branigan et al. 2011; Branigan et al. 2003; Branigan et al. 2010; Pearson et al. 2006), although much of this research is still used to primarily understand the mechanisms by which we align in human-human dialogue. As we

hold dialogue with computers on a regular basis using spoken dialogue systems it is surprising very little is understood about the impact of interlocutor design in this scenario has on our linguistic alignment. Recent research has identified that lexical alignment is higher in interactions with a computer that people feel is basic compared to that which is seen as advanced (Branigan et al. 2011; Pearson et al. 2006). This heightened alignment is thought to be due to designing utterances for the audience to increase likelihood of successful communication (termed mediated alignment in the literature). The research presented in this paper opens interesting questions about how design characteristics of spoken dialogue systems that may infer or lead us to assume capabilities, such as the type of voice used, impacts on users' alignment. Indeed if too much alignment occurs it may signal that users have a negative view of the system and its capabilities and thus feel they need to tailor their utterances to the audience (i.e. the system) they are communicating to. In this instance alignment could have potential as a metric highlighting difficult, unnatural and unsatisfying interactions. For instance interactions which are more natural, similar to human-human dialogue and potentially more satisfying could lead to similar levels of human-human alignment whereas those which are unnatural and less satisfying could see a heightened alignment above natural levels. The user will likely conform to the linguistic concepts present in the computer interlocutor's speech so as to guarantee communication success in poor interactions. The research in this paper aims to identify the potential for using alignment in this fashion and to further investigate how design decisions may impact alignment in HCD.

Recent work on linguistic alignment in HCD has shown not only that users tend to align in computer interaction but that this effect can be more prominent in human-computer interactions than in human-human interactions. A recent study researched the impact of verb repetition on users descriptions in both human-human and human-computer dyads (Branigan et al. 2003). Study participants played a picture matching game with a computerised conversational partner (effectively a computer confederate if we are to reflect the structure of the confederate scripting paradigm used in human-human alignment research (Branigan et al. 2000)). The conversation partners (the computer and the participant) took turns in describing images using text for the other to match with their pictures on screen. In both the experiment conditions the confederate was played by a computer, although the participants in one condition were led to believe that they were interacting with a computer and those in the other believed they were interacting with a human partner. The research found that beliefs about the interlocutor influenced the amount of alignment present in the participants' sentences. Participants syntactically aligned (used the same verb phrase structure) more in human-computer interaction than human-human interaction when verbs used by participants' were the same to those previously used by the confederate, but not when the verb was not repeated. Such research suggests heightened syntactic alignment in human-computer scenarios when compared to human-human dyad interactions.

Research on lexical alignment in HCI contexts has also highlighted expectations towards system capabilities influence users' lexical alignment (Branigan et al. 2011; Pearson et al. 2006). Users were made to believe that they were interacting with either a "basic" or "advanced" computer when playing a picture matching game with a computer interlocutor. In the basic condition participants saw a start up screen stating the computer interlocutor was a basic version of the software, with a 1987 copyright, paired with a review highlighting its limitations. The advanced condition saw users exposed to a start up screen proclaiming the software to be the "Advanced Version: Professional Edition", with a current year copyright and a review stating its sophistication and wide range of features. The study found users aligned more towards lexical representation used by the computer in both conditions, but that this alignment was significantly larger when interacting with the "basic" computer (Branigan et al. 2011; Pearson et al. 2006). Both pieces of research highlight the impact of computer interlocutors on linguistic alignment and how speakers' beliefs of the system impact this concept. Rather than being more of an unconscious priming effect (i.e. unmediated) linguistic alignment in HCD is seen as more mediated in nature (Branigan et al. 2011). In other words, the characteristics of the interlocutor are likely to influence the linguistic behaviours the user feels is necessary to enhance communication effectiveness within the dyad, analogous to the psycholinguistic concept of audience design in dialogue (Bell 1984). Therefore the way the system is viewed by users is likely to impact on how they align in dialogue.

The findings from the research above, and indeed this current study, are of practical value to HCI, spoken dialogue interface researchers and designers. They give causal insight into how the design of these interactions impact people's linguistic behaviour in interaction, knowledge relevant to gaining a deeper understanding of these interactions. The findings above also highlight the potential for alignment to be used as an interaction quality barometer. One would assume that when alignment is higher than human levels, there might be problems in the perception and indeed the user

experience of the interface since users have to be more considered and mediated with their speech structures as they see the system capabilities as more basic.

To date, much of the research on alignment in HCI has been based on textual input and lexical alignment rather than dialogue interaction and syntactic alignment. This work aims to add knowledge by researching syntactic alignment in spoken dialogue interactions and how interface design choices that may impact user perceptions of the system (such as the naturalness of the type of voice used by the interlocutor) impacts on alignment in HCD. The paper also aims to identify whether alignment can be indicative of a users satisfaction with the interface. We hypothesise that participants' alignment will be significantly affected by the type of interlocutor experienced. We also hypothesise that participants will also vary significantly in their interaction satisfaction, with the trend being the inverse of the alignment findings.

2. METHOD

2.1 Research Design

The research used a wizard of oz experiment methodology. A wizard of oz methodology involves the simulation of a system interaction rather than interacting with a final system. Such methods tend to be used to simulate future technologies or functionality of technologies not currently developed to high enough standards to simulate the desired interaction. At present existing systems are not sufficiently flexible or intelligent enough to offer the full subtleties required to observe syntactic alignment for the aims above. For instance the use of an automated system (rather than one controlled by an experimenter) leaves the experiment liable to potential errors when the computer takes their turn. To control this a wizard of oz method was used.

2.2 Participants

58 participants (30 females and 28 males) took part in the experiment. All were members of the University of Birmingham community.

2.3 Conditions

The independent variable *Partner* (3 conditions- between subjects) varied in terms of the humanness of voice of the partner. As a control condition participants interacted with a *human* partner. This was as to be able to compare levels of human-human alignment with those in the human-computer conditions. In the *basic* computer voice condition participants interacted with a computer producing an artificial sounding computer generated voice. This voice was distinct in its lack

of naturalness in intonation in comparison to human based speech. The speech for this condition was created using Vox Machina for Mac. The third condition was the *advanced* computer voice condition. In this condition participants interacted with a computer that produced human-like speech. This was simulated using voice recordings of members of the experiment team to ensure that the voice was as human like as possible. Each session was run so that the experimenter was not the same person used as the voice of the advanced computer.

2.4 Materials and Measures

2.4.1 Game and Items

Participants were asked to take part in a communication game similar to the referential communication task used in alignment research with human-human dyads (Branigan et al. 2000). The game involves partners taking turns in the role of the describer and the matcher. When being the describer the partner would describe a picture displayed on a laptop in front of them to the other partner in the game. The other partner would at this point be playing the role of the matcher where their task is to find the picture being described to them in the two pictures presented before them. To denote which turn it was an icon of either a speech bubble (to indicate turn as the describer) or a character listening (to indicate the matcher turn) were used. These icons were explained at the start of the experiment. Example screenshots of the game when participants were both the matcher (upper) and describer (lower) are shown in Figure 1.
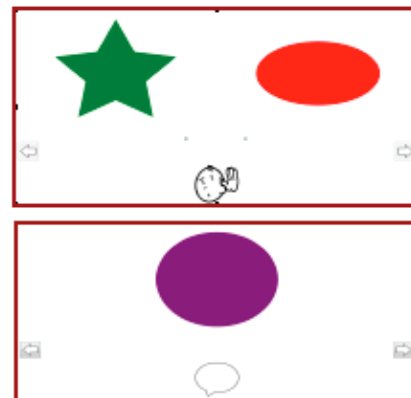
Figure 1: Example screenshots of communication game when participant plays the role of the matcher (upper) and when they play the part of the describer (lower).

However, throughout all of the experiment sessions one of the partners was a confederate (either computer based or a human confederate who was a member of the experiment team) and would

simulate playing the game as another participant. As such only the participant received the game on the laptop as presented in Figure 1. Instead of the game images the confederate had a Powerpoint slideshow displaying slides with scripted utterances. When being the describer the confederate would use relative clause structured descriptions (e.g. "circle that's red"- henceforth referred to as RC utterances) or adjective noun structured descriptions (e.g. "red circle"- henceforth referred to as AN utterances) of shapes. If alignment occurred it would be expected that the naïve participant would use the same grammatical structure in their next describer turn as the one they had just heard from the confederate when they (the participant) were playing the role of the matcher.

The game involved the participant describing a total of 192 pictures (termed items) over 4 sets (48 items per set). Each set had 18 experiment items. The experiment items were pictures of shapes (Triangle, Square, Heart, Oval, Diamond and Star) varying in colour (Orange, Red, Blue, Purple, Green and Yellow) similar to those used in (Cleland & Pickering 2003). These were the items that would be monitored for syntactic alignment with the primed syntactic structures (RC and AN). Each experiment item (the target item) was paired with a description (the prime item), given by the confederate, when the participant was playing the role of the matcher. These prime items directly preceded the participants' turn as describer of a target item. 30 filler items were also used per set. Filler items were pictures made up of various multiples of the shapes in the experiment items (2,3,4 or 5 shapes), patterns (stripy, wavy, dotted, chequered, zigzag or pitted), colours (orange, red, blue, purple, green and yellow) and combinations of the colours and patterns mentioned. These were used to control for potential carry over priming effects and to hide the focus of the experiment being on the experiment item descriptions. The filler items were included between prime-target item pairings.

Each possible combination of colour and shape in the experiment items were described by the participant twice throughout the experiment, once after an AN prime and once after an RC prime. This was to ensure any priming effect found was not due to any natural differences in how to describe each item. Each combination of colour and shape was also presented twice in the matching side to provide the correct match from the description of the partner.

To ensure that any priming effect was not due to lexical similarity between primes and target utterances, the shapes and colours of each prime-target pair were never the same. There was always at least one filler describer-matcher pair between each prime-target item pair.

When participants played the matcher, they were asked to click the radio button of the item that matched the item described to them. When matching the item when the confederate had described an experiment item (i.e. the participant had heard a prime), the two shapes displayed (see Figure 1 upper) varied systematically in terms of their similarity in colour and shape. 25% of the shape pairs had the same colour as the other shape displayed, 25% had the same shape as the other shape displayed and 50% varied in colour and shape. This was to ensure that there was no consistency between the display of the shapes in the matching turn so as to make sure that participants could not use this cue to identify the most felicitous way to describe the experiment items to their partner when they were taking their turn as the describer (e.g. "blue one" rather than "blue oval"). The filler items were randomly paired with each other and also displayed in sets of two in the matching turn.

There were a total of 72 experiment items and 120 filler items described by each participant in the experiment.

2.4.2 Interaction Satisfaction Questionnaire

Perceived interaction satisfaction was measured using the Interaction Satisfaction Questionnaire (ISQ). The ISQ was created to measure users' satisfaction when interacting with their partner in the game. Although there are measures of spoken dialogue system satisfaction used in the literature (Davidson et al. 2004; Litman & Pan 2002; Möller et al. 2007) these do not use items that can easily be attributed to interactions with both humans and computers as they focus on usability of the system being tested. To the authors' knowledge, there are no questionnaire measures available to measure satisfaction with interlocutor interaction in general. We therefore decided to create a measure for this purpose. An initial list of 39 items were devised gaining inspiration from satisfaction measures from HCI research such as the QUIS (Chin et al. 1988), SUMI (Kirakowski & Corbett 1993), MINERVA (Dutton et al. 1993) and WUI (Cowan & Jack 2011) satisfaction measures as well as measures of Perceived Ease of Use (Roca et al. 2006; Venkatesh 2000). Items were then refined after discussions with the experiment team, which includes HCI design experts, HCI psychologists and psycholinguists. During the refinement process the items were circulated to the authors to identify items with poor face validity, redundant items and any concepts that had not been addressed in the existing items. From this 32 final items (17 negatively worded, 15 positively worded) were included in the measure administered to

participants. The items focus on the themes of *affect towards interaction* (e.g. "Interacting with my partner was fun"), *control in interaction* (e.g. "When interacting with my partner I didn't know what would happen next"), *ease of interaction* (e.g. "I found my partner easy to interact with"), *interaction quality* (e.g. "I felt the interaction needed improvement") and *voice quality* (e.g. "I disliked my partner's voice"). Items used a 5-point Likert scale from Strongly Disagree (1) to Strongly Agree (5). All negative items were reverse scored so the total score reflected the positive nature of the concept being measured. Questionnaires were randomised into 4 independent orders to control for potential confounds of item order.

2.5 Procedure

Participants were greeted by the experimenter and were asked to take a seat on one side of a table divided by a screen. On the other side sat the confederate. The participant (and the confederate in the human condition) completed a demographic questionnaire recording their age, gender, profession/subject of study, whether they were a native English speaker and whether they knowingly suffered from any medical condition that would influence their ability to safely view computer screens in the task. If the participant answered yes to either of the last questions they were informed that they were ineligible to take part and were thanked for showing an interest in the study.

Participants were then told they were going to be playing a game with a partner. They were then introduced to their partner. In the *Human* condition the participant and the confederate were introduced to each other as participants in the study, whereas in the computer conditions the participant was shown that there was a computer at the other side of the screen. It was explained that the aim of the game was to describe pictures on the screen to a partner and pick out the ones they describe to you as quickly and as accurately as possible.

A practice trial consisting of 4 items was completed so participants could get acquainted with the game. In this, as well as in each other set in the main game (of which there were 4), the confederate took the role of the describer first. During the game the experimenter noted participants' utterance structures during the session with sessions also being audio recorded to ensure utterance data could be retrieved if missed by the experimenter during the experiment.

During the experiment session the computer confederates were controlled remotely by a member of the experiment team using Skype and Windows Live remote assistant. Skype was used to listen in to the session and remote assistant was used to control the laptop in the experiment room so that the correct audio files could be played when the confederate needed to play the role of the describer.

After completing all 4 item sets participants were then asked to complete the ISQ, thinking about the interaction they had with their partner in the experiment. They were then thanked for participating and debriefed about the nature of the experiment.

3. RESULTS

### 3.1 The effect of partner on syntactic alignment

To highlight the pattern of utterances with each prime type the proportions of AN utterances in each condition are reported in Table 1. These proportions were calculated using the method described by (Cleland & Pickering 2003) whereby the AN utterances per condition were divided by the sum of the AN and RC utterances in the condition. The AN and RC proportions are therefore complimentary in that the proportion of RC responses can be gained by subtracting the AN proportions displayed from 1.0.

Table 1: Proportions of AN utterances per prime by condition

Condition	AN Prime	RC Prime	Alignment effect
Human	0.997	0.980	0.017
Basic	0.990	0.970	0.020
Advanced	0.984	0.947	0.037

In accordance with recent research highlighting the potential for spurious results using ANOVA with categorical data (Florian Jaeger 2008), a logit mixed effects model was used to statistically identify whether there was a significant effect of Partner on alignment. This is due to the categorical binomial dependent variable used to assess alignment (i.e. alignment=1, no alignment=0). Logit mixed effect models are similar to logistic regression (Florian Jaeger 2008) and are becoming more prominent in psycholinguistics research due the categorical nature of dependent variables and the difficulties with ANOVA recently highlighted. The analysis was run in R version 2.14.1 using the lme4 package.

The percentage of aligned and unaligned utterances for each condition are displayed in Table 2 and shown graphically in Figure 2.
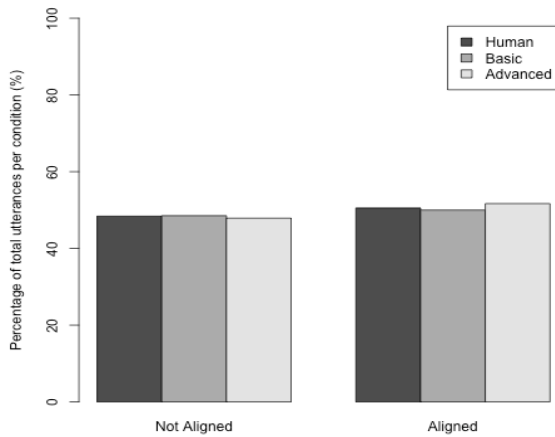
Figure 2: *Percentage of Not Aligned and Aligned utterances by condition*

As can bee seen in Table 2 only a small proportion of total utterances per condition were classed as *Other* across the experiment (Human: 1.0%; Basic: 1.5%, Advanced: 0.4%). The type of utterances classed as *Other* were those that did not adhere to the specific prime structures such as "Square, Blue" or "Green". Only those utterances categorised as *Aligned* or *Not Aligned* were included in the model.

Table 2: *Percentage of total utterances Aligned and Not Aligned by condition*

Condition	% Aligned	% Not Aligned	% Other
Human	50.6	48.4	1.0
Basic	50.0	48.5	1.5
Advanced	51.7	47.9	0.4

The logit mixed effects analysis showed that there was no significant impact of *Partner* on amount of alignment (Human-Basic Contrast: log likelihood= -500.93, Estimate= 0.046, SE= 0.350, Wald z= 0.131, *p= 0.896*; Human-Advanced Contrast: log likelihood= -501.45, Estimate= 0.331, SE= 0.343, Wald z= 0.963, *p= 0.335*; likelihood ratio test (χ^2 (2) = 1.049, *p=0.592, N*=4135).

3.2 The effect of partner on interaction satisfaction

3.2.1 ISQ Reliability:
6 items were excluded from the questionnaire ("When interacting with my partner I didn't know what would happen next", "I had to concentrate hard when interacting with my partner", "My partner didn't always do what I was expecting", "I felt under stress when interacting with my partner", "I got flustered when interacting with my partner", "I felt in

control when interacting with my partner") due to an increase in scale reliability with their removal. The remaining items in the questionnaire showed high reliability (Cronbach α= 0.92) above acceptable levels for psychometrics (Kline 2000). With the removal of the items the questionnaire scale ranged from 26 to 130.

3.2.2 Partner effects on Satisfaction
The scores on the questionnaire were summed and analysed using a One Way ANOVA to identify the difference between the interlocutor conditions on interaction satisfaction. The means and standard deviations are shown in Table 3. The means and standard error for each condition are also shown graphically in Figure 3.

Table 3: *Descriptive Statistics of the Interaction Satisfaction Questionnaire (ISQ) by condition*

Partner	N	Mean	S.D
Human	20	103.10	9.21
Basic	19	86.33	12.53
Advanced	19	97.73	11.08

There was a significant main effect of *Partner* on the scores in the ISQ [F (2, 55) =11.75, p<0.001]. Bonferroni post hoc comparisons show that participants in the *Basic* computer voice condition (M=86.33) rated their satisfaction significantly lower than those in both the *Human* (M=103.10) (p<0.001) and *Advanced* computer voice (M=97.73) conditions (p=0.007). There was no significant difference between the *Advanced* and *Human* interlocutor conditions (p=0.40).
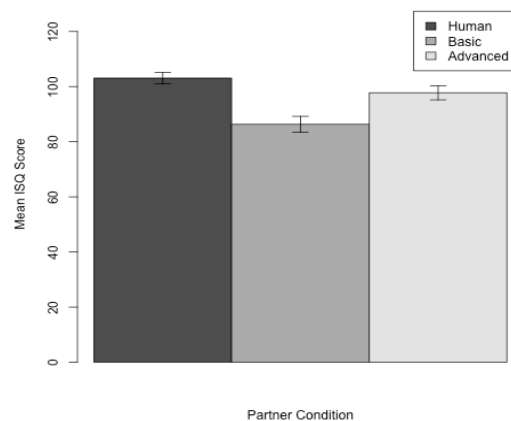
Figure 3: *Mean ISQ score per Partner condition*

As hypothesised, participants rated their experiences with the human and more advanced computer voice as more satisfying than that with the basic computer voice. If alignment could act as a behavioural metric for interaction satisfaction we would expect the alignment effect to be the direct opposite of what we have found in the satisfaction

measures. That is we would expect higher alignment in the basic condition when compared to the advanced and human interlocutor conditions. Yet as we saw in the analysis of alignment data, there was no significant effect of *Partner* on alignment.

4. DISCUSSION

The results of the research suggest that there was no significant effect of voice type on syntactic alignment in human-computer dialogue. Additionally, although there was a significant difference between our conditions in terms of interaction satisfaction, this difference was not reflected in the amount of alignment seen across the conditions. Our prediction that participants would syntactically align more in lower satisfaction interactions was therefore not supported. However our findings do suggest that users rated interactions with a more human like computer voice and a human as more satisfying than those with a basic voice.

It would seem therefore that design choices such as voice type have no significant effect on syntactic alignment in an HCD context and that syntactic alignment is not an effective metric of user satisfaction in this scenario. Especially due to recent evidence in lexical alignment and user perceptions, the lack of difference in alignment effect was surprising. Due to the large number of participants compared to other alignment work we can be confident of the power of our analysis to identify a statistically significant effect if such existed. An explanation for the lack of effect may lie in the difference between the likelihood of syntactic alignment being impacted by audience considerations. Although syntactic alignment is thought to have a mediated component in human-computer interactions (Branigan et al. 2003) research on syntactic alignment in human-human dialogue suggests that it is more of an automatic, unconscious priming process (Cleland & Pickering 2003). This could be an explanation for the lack of effect across the conditions. For instance in all conditions there is an equal assumption to align due to it being an unconscious process rather than consciously aligning to the conditions depending on voice type. As stated by (Branigan et al. 2010) unmediated alignment would suggest a lack of impact of interlocutor differences as it is less obvious what is felicitous in the interaction on a syntactic level compared to lexical alignment (Branigan et al. 2007). The nature of syntactic alignment with computers may be more unmediated than first anticipated. In addition research on syntactic alignment in human-computer interaction seemed to only find a heightened alignment effect in computer compared to human interlocutors when the verb was repeated in the prime-target pairs (Branigan et al. 2003). On reflection, our conditions are more analogous to the non-repetition of the verb in that we did not repeat the noun or adjectives across the prime-target items to ensure no linguistic boost to priming occurred. This could also explain the lack of difference in syntactic alignment in our findings.

Although this may explain a lack of difference between conditions in the experiment, our data seem to suggest no alignment effect in general. The proportions in Table 1 show that there is a major propensity to use the AN structure compared to RC structure no matter what the prime. This propensity may be due to the type of stimuli used. Although we used items and prime noun phrases similar to previous research (Cleland & Pickering 2003) that identified a marginal alignment effect in a similar human-human condition, the research also highlighted that such primes have a strong natural preference effect, that in which AN are highly preferred to RC phrases to describe such items. There is a large amount of variation in the size of alignment effect across studies likely because of task demands and structural preferences (Branigan et al. 2007). This high natural preference towards AN structures might have overruled any priming effect. The authors feel this is likely to be the main reason for a lack of alignment effect. Items focusing on prepositional (PO) and direct object (DO) verb phrase structures from other syntactic alignment research (Branigan et al. 2000; Branigan et al. 2007; Pickering & Branigan 1998) are planning to be used in further experiments to investigate whether this natural tendency did saturate any effect.

Further to this the computer interlocutor was seen to interpret and successfully match the cards as easily when participants used RC as well as AN structures. The system in all interactions could interpret both structures of utterances and used both structures. The participant therefore may not have had any motivation to change their utterance structure from the one most prominently used, especially if we are to assume that alignment in HCD is more mediated than unmediated. Yet even if this was the case we would still expect to reflect the (albeit weak) syntactic alignment effect of previous research (Cleland & Pickering 2003) in the human-human interaction condition. It is hard to explain theoretically why an effect was not found in this research condition at least, although influences of differences in the set up of the experiment, such as number of items in the game and differences in the filler items used cannot be ruled out. Further work is being planned to explore the reasons for the lack of effect in terms of the impact of game length, item type and feedback on ability of the

computer understanding all structures on syntactic alignment.

Interestingly for HCI, it has also been proposed that alignment in HCD may have a social motivation. This is related to the idea that people use mimicry as a social glue (Chartrand & Bargh 1999) as well as consistent findings in HCI highlighting computers as social actors (Reeves & Nass 1998). From this we may assume that linguistic alignment could be used to strengthen social bond between the computer and participant. In this experiment we have taken a similar approach to others (Nass & Moon 2000) by extending an existing paradigm and methodology used in human-human interaction and included human-computer conditions to explore such phenomena in an HCI context. The inclusion of a "natural computer" condition and the lack of effect on alignment suggest a potential lack of impact of bond on syntactic alignment in this context. If bond and naturalness were going to impact on syntactic alignment we would expect to see a heightened alignment in the advanced computer condition (against our hypothesis set out in the introduction). Instead we found no effect across conditions, although it must be said this could again be due to the natural tendency to use AN structures over RC structures saturating any potential alignment effect. Again, further work using items with less of a natural linguistic bias are being planned at present to observe whether this did cover any effect.

Although this research found no syntactic alignment effect, studies of alignment in HCD have robustly demonstrated linguistic convergence, especially at the lexical level. Our work in this area now aims to not only look at reasons why an effect did not occur in this context but also how design considerations impact lexical alignment, a more clearly mediated phenomena (Branigan et al. 2011). We also aim to look at real world corpora of human-computer dialogue interactions and identify how alignment of syntactic and lexical structures occurs in these contexts. This would be an extension and validation of lab-based research in the area, which some have highlighted as restricted and task orientated (Howes et al. 2010). The authors feel that this restricted and task orientated interaction is more reflective of human-computer dialogue, yet analyzing real world corpora would give valuable insight into the extent of alignment in real world human-computer data.

Further work investigating the impact of the system's inability to understand specific structures may also be a fruitful avenue for research in this domain. Although syntactic alignment is not seen as critical or disadvantageous to communication success as lexical alignment in dialogue (Branigan et al. 2007) the question is raised about what if

such alignment became critical to success? This type of question is perhaps more of an issue in an HCD capacity rather than in human-human dialogue. Systems can have limitation on their interpretation and due to potential inflexibilities in the system design, errors may not be able to be resolved as easily as in a human-human based interaction, even if the interlocutor is not a fluent English speaker. One can imagine using other forms of communication such as gesture, intonation or breaking down utterances into more common phrase structures with more simple lexical items to resolve errors in human-human dialogue. Yet such flexibility needs to be built into a computer based dialogue system. If it has not been then the user may become more aware of their syntax in communication due to self-monitoring during dialogue (Horton & Keysar 1996) making syntax more critical and alignment more likely. A lack of success in interaction within an HCD scenario could lead to a significant shift in the user model of interaction away from that assumed from initial system perceptions, having implications for linguistic behaviour that move from being unconscious (i.e. unmediated) to consciously constructed. Such work is being planned at present by the authors.

The research presented was preliminary in nature and as such improvements and further investigation is needed to the design, the task and the metrics used in the experiment. For instance the ISQ, although showing high reliability and an ability to discriminate between voice conditions we would expect to impact user interaction satisfaction, needs further validating and factor analyzing. We therefore encourage other researchers to use the measure so that more detailed analysis can be conducted on the measure specifically. Additionally the number of items may have affected the alignment effect. With the description of 192 pictures participants may have got bored or lost focus towards the end of the trials. Although the order of the sets of items were counterbalanced across the experiment, further studies should aim to reduce the number of items used to identify whether this impacted on syntactic alignment levels.

It seems therefore that the humanness of the interlocutor's voice had no significant statistical effect on syntactic alignment on noun phrase structures. Additionally, although participants rated the basic voice condition as significantly less satisfying than both the human and advanced computer voice, this was not reflective of the alignment effects suggesting in this instance alignment is not a good behavioural indication of satisfaction with the interlocutor. It must be noted though that due to natural preference for AN structures when describing the experimental

stimuli, it may be that the alignment effect was washed out by this preference. As such these findings cannot be generalised to other grammatical structures and must be interpreted specific to the utterance structure primed. Before it can be concluded that syntactic alignment is not affected by voice type and that alignment is an ineffective behavioural measure of satisfaction, we must first explore whether the findings are replicated with other syntactic structures with less natural preferences.

5. ACKNOWLEDGEMETS

5. REFERENCES

Bell, A. (1984) Language Style as Audience Design. *Language in Society*, 13, pp.145–204.

Branigan, H.P. et al. (2010) Linguistic alignment between people and computers. *Journal of Pragmatics*, 42, pp.2355–2368.

Branigan, H.P. et al. (2007) Syntactic alignment and participant role in dialogue. *Cognition*, 104, pp.163–197.

Branigan, H.P. et al. (2003) Syntactic alignment between computers and people: the role of belief about mental states. In *Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society* (CogSci '03), Boston, MA, pp. 186–191. Erlbaum, Mahwah, NJ.

Branigan, H.P. et al. (2011) The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, 121, pp.41–57.

Branigan, H.P., Pickering, M.J. & Cleland, A. (2000) Syntactic co-ordination in dialogue. *Cognition*, 75, pp.B 13–25.

Brennan, S.E. (1996) Lexical entrainment in spontaneous dialog. In *Proceedings of the International Symposium on Spoken Dialogue*. (ISSD-96). Philidelphia, USA, pp. 41–44. Acoustical Society of Japan.

Brennan, S.E. & Clark, H.H. (1996) Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, pp.1482–1493.

Chartrand, T. & Bargh, J.A. (1999) The chameleon effect: The perception-behaviour link and social interaction. *Journal of Personality and Social Psychology,* 76, pp.893–910.

Chin, J.P., Diehl, V.A. & Norman, K.L. (1988) Development of an instrument for measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (CHI '88) Washington, D.C., pp. 213–218. ACM Press, New York, NY.

Cleland, A. & Pickering, M.J. (2003) The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, 49, pp.214–230.

Cowan, B.R. & Jack, M.A. (2011) Exploring the wiki user experience: The effects of training spaces on novice user usability and anxiety towards wiki editing. *Interacting with Computers*, 23, pp.117–128.

Davidson, N., McInnes, F. & Jack, M.A. (2004) Usability of dialogue design strategies for automated surname capture. *Speech Communication*, 43, pp.55–70.

Dutton, R.T. et al. (1993) Identifying usability attributes of automated telephone services. In *Proceedings of the 3rd European Conference on Speech Communication and Technology* (EUROSPEECH '93), Berlin, pp. 1335–1338. ISCA.

Florian Jaeger, T. (2008) Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, pp.434–446.

Horton, W.S. & Keysar, B. (1996) When do speakers take into account common ground? *Cognition*, 59, pp.91–117.

Howes, C., Healey, P.G.T. & Purver, M. (2010) Tracking Lexical and Syntactic Alignment in Conversation. In *Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society* (CogSci 2010), Portland, Oregon.

Kirakowski, J. & Corbett, M. (1993) SUMI: the software usability measurement inventory. *British Journal of Educational Technology*, 24, pp.210–212.

Kline, P. (2000) *A Psychometrics Primer*. Free Association Books, London, UK,

Litman, D.J. & Pan, S. (2002) Designing and evaluating an adaptive spoken dialogue system. *User Modeling and User Adapted Interaction*, 12, pp.111–137.

Möller, S. et al. (2007) Evaluating spoken dialogue systems according to de-facto standards: A case study. *Computer Speech & Language*, 21, pp.26–53.

Nass, C. & Moon, Y. (2000) Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56, pp.81–103.

Pearson, J.M. et al. (2006) Adaptive language behavior in HCI. In *Proceedings of the SIGCHI conference on Human Factors in computing systems (CHI '06)*, Montreal, Quebec, Canada, p. 1177. ACM Press, New York, NY.

Pickering, M.J. & Branigan, H.P. (1998) The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39, pp.633–651.

Reeves, B. & Nass, C. (1998) *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places*. Cambridge University Press, Cambridge.

Roca, J.C., Chiu, C.-M. & Martinez, F.J. (2006) Understanding e-learning continuance intention: An extension of the Technology Acceptance Model. *International Journal of Human-Computer Studies*, 64, pp.683–696.

Venkatesh, V. (2000) Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the Technology Acceptance Model. *Information Systems Research*, 11, pp.342–365.