# Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in humancomputer dialogue

Cowan, Benjamin R.; Branigan, Holly P.; Obregón, Mateo; Bugis, Enas; Beale, Russell

*DOI:*
10.1016/j.ijhcs.2015.05.008



*Document Version*
Peer reviewed version

*Citation for published version (Harvard):*
Cowan, BR, Branigan, HP, Obregón, M, Bugis, E & Beale, R 2015, 'Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in humancomputer dialogue', *International Journal of Human-Computer Studies*, vol. 83, pp. 27-42. https://doi.org/10.1016/j.ijhcs.2015.05.008


www.elsevier.com/locate/ijhcs

# Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in Human-computer dialogue

Benjamin R Cowan, Holly P. Branigan, Mateo Obregon, Enas Bugis, Russell Beale

PII: S1071-5819(15)00102-0
DOI: http://dx.doi.org/10.1016/j.ijhcs.2015.05.008
Reference: YIJHC1964

To appear in: *Int. J. Human-Computer Studies*

Received date: 26 June 2014
Revised date: 1 May 2015
Accepted date: 28 May 2015

Cite this article as: Benjamin R Cowan, Holly P. Branigan, Mateo Obregon, Enas Bugis, Russell Beale, Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in Human-computer dialogue, *Int. J. Human-Computer Studies,* http://dx.doi.org/10.1016/j.ijhcs.2015.05.008

# Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in human-computer dialogue

**Benjamin R Cowan[1,3]*, Holly P. Branigan[2], Mateo Obregón[2], Enas Bugis[1], Russell Beale[1]**

***Corresponding Author**

**[1] HCI Centre, School of Computer Science, University of Birmingham, Edgbaston Campus, Birmingham, B15 2TT**

**[2] Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ**

**[3] School of Information and Library Studies, University College Dublin, Belfield, Dublin 4, Ireland**

**Email: benjamin.cowan@ucd.ie**

**Tel: +353 (0)17167660**

**Abstract**

The growth of speech interfaces and speech interaction with computer partners has made it increasingly important to understand the factors that determine users' language choices in human-computer dialogue. We report two controlled experiments that used a picture-naming-matching task to investigate whether users in human-computer speech-based interactions tend to use the same grammatical structures as their conversational partners, and whether such *syntactic alignment* can impact strong default grammatical preferences. We additionally investigate whether beliefs about system capabilities that are based on partner identity (i.e. human or computer) and speech interface design cues (here, voice anthropomorphism) affect the magnitude of syntactic alignment in such interactions. We demonstrate syntactic alignment for both dative structures (e.g., *give the waitress the apple* vs. *give the apple to the waitress*), where there is no strong default preference for one or other structure (Experiment 1), and noun phrase structures (e.g., *a purple circle* vs. *a circle that is purple*), where there is a strong default preference for one structure (Experiment 2). The tendency to align syntactically was unaffected by partner identity (human vs. computer) or voice anthropomorphism. These findings have both practical and theoretical implications for HCI by demonstrating the potential for spoken dialogue system behaviour to influence users' syntactic choices in interaction. As well as verifying natural corpora findings, this work also highlights that priming and cognitive mechanisms that are unmediated by beliefs about partner identity could be important in understanding why people align syntactically in human-computer dialogue.

**Keywords:** Human-computer dialogue, syntactic alignment, speech interaction, user behaviour, psycholinguistics, interlocutor modelling

# 1. INTRODUCTION

Recent innovations in consumer electronics have led to a rapid increase in the frequency of spoken dialogue interactions between people and artificial systems, where users use natural speech to command devices or to query systems, and devices and systems in turn use natural speech to respond. Speech and human-computer dialogue interactions are now common in mainstream technology products; 87% of IPhone 4S users have reported using Siri at least once a month (Barrett & Jiang, 2012) and many other services such as Google Now, in-car systems and Smart TVs are using speech as an interaction modality. The future growth of human-robot interaction as well as the use of embodied conversational agents highlights that natural dialogue interactions between computers and humans are likely to become more prominent. With this in mind, recent calls have been made for HCI and speech-based researchers to combine efforts to understand what governs our interactions with speech technology to design more effective speech interface interactions (Aylett, Kristensson, Whittaker, & Vazquez-Alvarez, 2014).

Although a vast body of knowledge has been accumulated about the factors that govern spoken dialogue behaviours between two people (*human-human dialogues; HHD*), comparatively little is known about spoken dialogues between people and artificial systems (*human-computer dialogues; HCD*). In this paper, we focus on one particular factor that has been hypothesised to strongly influence speakers' behaviour in HHD: linguistic *alignment*, or the tendency for conversational partners to converge on common language choices.

We investigate whether users show alignment of grammatical structure (*syntactic alignment*) in speech-based HCDs under controlled experimental conditions using a game in which a participant and their partner alternately describe and match pictures. The game includes images that can be described by two different grammatical structures. In the game, the participant's partner (a 'confederate') uses specific grammatical structures (*primes)* to describe their images and we observe whether participants then tend to use the same structure rather than the alternative structure in their subsequent descriptions. The games therefore gives participants an opportunity to make choices around the structure they use to describe the images whilst allowing experimenters to observe the effect of the prime on their syntactic choice. Although previous more naturalistic research has suggested that users do align syntactically in speech-based HCD (Stoyanchev & Stent, 2009), such work does not control potential confounds that may affect the magnitude of alignment (e.g. effects associated with repetition of particular words, recency effects, natural frequency of structures and speech recognition errors). Using a controlled experimental paradigm such as the picture-description-matching game allows a precise focus on the causal impact of particular variables while controlling such confounds in the game materials. Indeed such studies are important in validating more naturalistic corpora work (Gilquin & Gries, 2009). The current study also expands previous laboratory-based alignment research on text (Branigan, Pickering, Pearson, McLean, & Nass, 2003) to speech-based dialogue interactions, reflecting the increased prominence of speech as an interaction modality in popular devices. Previous research has also highlighted higher sharing of syntax with the computer in speech-based than text-based interactions (Le Bigot et al., 2007). Text-based studies of alignment may therefore underestimate the

magnitude of alignment in spoken HCD, making it important to explore syntactic alignment in a speech-based HCD context.

Importantly, we examine alignment for two types of syntactic alternation that differ in their default structural preferences. Experiment 1 focuses on Double Object (DO; e.g. "*The cowboy offering the robber the banana*) and Prepositional Object (PO; e.g. "*The cowboy offering the banana to the robber"*) structures. These are relatively evenly balanced in terms of default preferences in a non-biasing context [1] when people are describing dative events (i.e., events involving transfer of possession; roughly 60% PO, 40% DO, based on natural language corpora: Gries & Stefanowitsch, 2004; see also Pickering, Branigan, & McLean, 2002 for similar evidence from experimental studies). Experiment 2 focuses on noun phrase structures, specifically Adjective-Noun (AN; e.g., *the red circle*) and Noun-Relative Clause (RC; e.g., *the circle that's red*) structures. People have been shown to have a strong default preference for using AN structures in a non-biasing context when they are describing relevant items such as coloured and patterned shapes (around 95%, based on available evidence from experimental studies; Branigan, McLean, Messenger, & Jones, in preparation). Studying two syntactic alternations allows us to verify the generalizability of our findings and test whether mechanisms of syntactic alignment in HCI are sufficiently influential to impact strong intrinsic structural preferences. In addition it also allows us to explore whether syntactic alignment might be mediated more strongly by beliefs about the conversational partner, or *interlocutor,* for structure choices in which one alternative is strongly favoured (as in

[1] In this case, a non-biasing context refers to a context where people have not just been systematically exposed to one or other structure produced by their conversational partner (i.e. a prime).

noun phrase structures). Under these circumstances, the choice between a strongly favoured and a strongly disfavoured alternative might be particularly salient, and might therefore be more amenable to strategic decisions based on beliefs about interlocutors' likely understanding or preferences.

The work also adds insight onto the role that partner type (i.e. computer or human) and design choices (such as voice anthropomorphism) have on levels of syntactic alignment. Current findings in HCD suggest that many of our language behaviours are mediated by our perceptions of computers as effective communication partners (Amalberti, Carbonell, & Falzon, 1993; Bell & Gustafson, 1999; Brennan, 1998; Le Bigot et al., 2007). In particular, research on alignment of lexical choice in HCD (Bergmann, Branigan, & Kopp, in press; Branigan, Pickering, Pearson, McLean, & Brown, 2011) suggests that users adapt their lexical choices to accommodate their partner's perceived limitations as an interlocutor, with greater adaptation to partners perceived as less able. Work on anthropomorphic robotic agents suggests that we see such agents as more intelligent and capable than non-anthropomorphic agents (Kiesler, Powers, Fussell, & Torrey, 2008; King & Ohya, 1996). This raises the possibility that anthropomorphic cues in HCD scenarios may lead users to adapt less in these contexts than when interacting with a computer partner with less anthropomorphic cues.

Validating the occurrence of syntactic alignment in speech-based HCD under controlled experimental conditions, and demonstrating that characteristics of the partner affect syntactic alignment, would provide evidence that computer partner utterances as well as design can act as a means of inducing users to use predictable structures that the system can process successfully. Importantly the work also has implications for the

understanding of what guides our linguistic choices in HCD. Demonstrating that syntactic alignment is impacted by the anthropomorphism of the partner (and indeed by whether the partner is a computer or human) would show that syntactic alignment, like other language behaviours in HCD, is adaptive and influenced by our perceived limitations of the system as a dialogue partner. In contrast, if we found that syntactic alignment occurs in HCD but is unaffected by partner type, this would tentatively support the notion that cognitive architectures involved in language comprehension and production and the priming of language representations may play a role in syntax choice in HCD (Pickering & Garrod, 2004).

In both experiments, native English speakers played a picture-naming and -matching game with either a human, a computer with an anthropomorphic voice, or a computer with a robotic voice. The computer voices used were shown to yield significant differences in user perceptions of partner ability, with the anthropomorphic voice leading people to see a computer partner as being more advanced, flexible and competent than those hearing the robotic voice (see section 3). The participants and their partners took turns describing pictures of dative events (Experiment 1) or colored patterned shapes (Experiment 2), and choosing pictures in response to their partner's descriptions (in all experiments the partner was either a human confederate or a computer controlled remotely by a member of the experiment team). Two types of structural alternation were tested across the experiments, each with two alternatives that were primed using the partner's descriptions (Experiment 1: Prepositional Object (*PO*) or Double Object (*DO*); Experiment 2: Adjective-Noun (*AN*) or Noun-Relative clause (*RC*)) with the experimenter noting the structure that the participants used when producing their own

immediately subsequent description. Syntactic alignment is said to occur when participants used the same structure as the prime they were previously exposed to. We found that syntactic alignment occurred in both experiments, yet this effect was not significantly impacted by the partner conditions. This supports the notion that automatic priming of linguistic representations may play a significant role in alignment of syntax choice in HCD.

## 2. BACKGROUND

### 2.1 Alignment in Human-Human Dialogue

A large body of evidence from HHD has shown that conversational partners influence each other's behaviour. In particular, conversational partners show a robust tendency to converge on, or *align*, their non-linguistic and linguistic behaviour, such as posture and gestures (Chartrand & Bargh, 1999; Van Baaren, Janssen, Chartrand, & Dijksterhuis, 2009), as well as semantic, lexical, and syntactic choices (Branigan, Pickering, & Cleland, 2000; Brennan & Clark, 1996; Clark & Brennan, 1991; Garrod & Anderson, 1987; Pickering & Branigan, 1998). Alignment of language has been hypothesised to play a causal role in successful communication: by aligning their linguistic representations in production and comprehension, interlocutors also come to develop aligned situation models, or shared semantic representations of the topic under discussion, and hence mutual understanding (Garrod & Pickering, 2009; Pickering & Garrod, 2004). By corollary, communication is likely to be less successful if speakers do not align their language use (Reitter & Moore, 2007). Behavioural alignment has also been argued to act as a social glue, heightening social bonds and increasing liking

between interlocutors (Chartrand & Bargh, 1999; Giles, Coupland, & Coupland, 1991; Van Baaren et al., 2009).

Although it is uncontroversial that alignment of language is widespread and robust in HHD, there is less agreement concerning its underlying mechanisms. One account suggests that alignment is largely automatic and unconscious (Pickering & Garrod, 2004, 2006). Under this account, alignment in HHD occurs because linguistic representations are activated whenever conversational partners produce or comprehend utterances, and residual activation or implicit learning of these representations leads to an increased chance of their subsequent use (Branigan et al., 2000; Chang, Dell, & Bock, 2006). That is, alignment arises from automatic priming of linguistic representations that occurs in non-interactive as well as interactive contexts (e.g., Chang et al., 2006; Meyer & Schvaneveldt, 1971). This account makes no reference to non-linguistic factors (such as speakers beliefs' about their interlocutors), and characterizes alignment as an automatic consequence of the cognitive architecture of language processing.

An alternative account proposes that alignment on particular linguistic choices in HHD may be mediated by speakers' beliefs about their interlocutors (e.g. Branigan, Pickering, Pearson, McLean, & Brown, 2011; Brennan & Clark, 1996). In this account, alignment is seen as a manifestation of audience design (Bell, 1984). One facet of audience design is that speakers plan their utterances with reference to their beliefs about what the listener will understand. Thus they choose between linguistic alternatives according to their model of their interlocutors' knowledge and abilities. This *interlocutor model* may be based on assumptions about the communities to which they believe their interlocutor belongs (e.g., resident of Edinburgh, non-native speaker of English, wine

aficionado) and the knowledge that these communities are assumed likely to have (Clark, 1996; Fussell & Krauss, 1992), as well as calculations about what the interlocutor is likely to understand given their previous observed language use (Branigan et al., 2011). For instance, if a non-native speaker has previously used an unconventional name for an object (e.g., *chair that goes backwards and forwards* instead of *rocking chair*), her interlocutor may align on the same name to enhance the likelihood of mutual understanding, on the basis that the non-native speaker probably does not understand the conventional name (otherwise she would have used it) but clearly does understand *chair that goes backwards and forwards* (as evidenced by the fact that she has just used that term; Bortfeld & Brennan, 1997).

Of course, these two explanations for alignment effects are not mutually exclusive. Alignment may well have both unmediated and mediated components that manifest themselves to differing extents in different contexts for different aspects of language (Branigan, Pickering, Pearson, & McLean, 2010). For example, in communicative contexts in which mutual understanding between interlocutors is paramount (e.g., safety-critical situations), beliefs about what the interlocutor is likely to understand correctly may play a particularly strong role in alignment.

Recently, studies have used HCD interactions to explore these theoretical positions, comparing levels of alignment with human and computer partners (further details of this work is included in section 2.2). This is based on the fact that computers are more likely to be judged as less communicatively able compared to human partners (Branigan et al., 2003). Our research extends this work on how syntactic alignment is impacted by perceptions of partner abilities by using spoken HCD interactions.

Furthermore, rather than solely observing overall differences between computers and human partners, we also explore how design cues within HCD interactions may affect these alignment levels and as such how these theoretical accounts operate within an HCD context. Evidence that syntactic alignment is affected by partner type (e.g. human vs. computer) and by computer partner design would support a more mediated account to syntactic alignment in HHD and HCD, whereas no effect of partner would lend support to a more automatic mechanism being influential in syntactic alignment in HCD interactions.

## 2.2 Alignment in Human-Computer Dialogue

Recent research has shown that alignment occurs for some aspects of language in HCDs (see Branigan et al., 2010 for a review). For example, speakers show alignment of voices with computer interlocutors on prosodic and acoustic speech features (Bell, Gustafson, & Heldner, 2003; Levitan et al., 2012; Oviatt, Darves, & Coulston, 2004; Suzuki & Katagiri, 2007). Work has also shown that users align at a lexical level with computer partners (Bergmann et al., in press; Branigan et al., 2011; Brennan, 1996; Stoyanchev & Stent, 2009). In addition, both more naturalistic (Stoyanchev & Stent, 2009) and laboratory research investigating text-based dialogues (Branigan et al., 2003) has shown that speakers align syntactically in HCD. Research by Stoyachev & Stent (2009) using the Let's Go! dialogue system found that more action verbs were present in user responses when the system used such verbs in their system prompts. The trend to reuse a partner's syntax was also noted in recent research observing children playing a dialogue game with robot partners (Nalin et al., 2012). Controlled experimental studies of

syntactic alignment in text-based HCD using a similar picture naming and matching game as in the current study (Branigan et al., 2003) also showed that people tended to align syntactically with computers. As noted above, much of the previous literature has been focused on more naturalistic interactions where confounding factors that might affect the magnitude of alignment are difficult to control. Previous controlled experimental work on syntactic alignment in HCD has also concentrated on text-based interactions, where research has shown lower levels of syntactic structure sharing compared to speech-based interactions. Our work validates and extends previous naturalistic work by using controlled experimentation in a speech-based interaction context to control for confounds through the experimental materials (e.g. ensuring no boost to alignment associated with lexical repetition [see Branigan, Pickering and Cleland 2000], controlling the turns between prime and target, balancing the exposure to prime structures across the game) and set up (e.g. using a wizard of oz procedure to ensure no impact of speech recognition errors). In addition the use of speech-based interaction increases the relevance of the work to the growing use of speech as a popular interface modality.



## 2.3 Partner effects on alignment in Human-Computer Dialogue

Evidence from alignment of lexical choices in HCDs highlights the possibility for effects of partner identity on alignment under at least some circumstances. Branigan et al. (2011) had participants take part in a picture-naming and -matching task with a partner that they believed to be a human or a computer (in fact, it was always a pre-scripted computer). With both kinds of partner, participants tended to name objects using the

same name that their partner had previously used (e.g., calling an object a *seat* vs. a *bench*), in both text-based and spoken interaction. The tendency to align lexical choice was robust and persistent, occurring even when the name was normally strongly disfavoured (produced spontaneously less than 20% of the time in a non-biasing context), and when their partner had named the object eight turns earlier. Crucially, however, alignment was stronger when participants believed that they were interacting with a computer than a human. This pattern was replicated in German by Bergmann et al. (in press), and contrasts with previous suggestions that lexical alignment occurs at similar levels in HHD and HCD (Brennan, 1996; although note that unlike Branigan et al.'s (2011) and Bergmann et al.'s (in press) experiments, Brennan's study did not statistically compare HHD and HCD directly). Moreover, participants' tendency to align with a computer partner was affected by superficial aspects in the interaction (i.e., aspects unrelated to the system's actual behaviour): participants who began the task by viewing a start-up screen with a 1987 copyright along with a fictitious review from a computer magazine stating its limited abilities ('Basic' computer) showed a stronger tendency to align than participants who viewed a start-up screen with a current year copyright and review stating the system's sophisticated technology ('Advanced' computer).

Branigan et al. (2011) suggested that participants' tendency to align on lexical choice was influenced by their beliefs about what their interlocutor would be likely to understand based on perceived identity and non-functional aspects of the interaction. Overall, participants took an interlocutor's prior use of a name as evidence that the interlocutor would understand that name correctly; participants therefore chose to use that name, to facilitate successful understanding. But they did so to a greater extent when

interacting with a computer because computers are generally believed to be less communicatively able than humans, and therefore more prone to misunderstanding (Branigan et al., 2003). Moreover, the stronger alignment with a 'Basic' than an 'Advanced' computer suggests that participants' beliefs about interlocutor ability (and in turn participants' linguistic behaviour) were affected by superficial cues, specifically a system suggesting age and limited functions versus a system suggesting modernity and extensive functions. These beliefs were apparently established at the outset of the interaction and not updated on the basis of the interlocutors' actual behaviour during the interaction (in all conditions, the interlocutor always displayed successful understanding of the participants' lexical choice by correctly choosing the object named by the participant). These results therefore suggest not only that people may display different linguistic behaviour when interacting with computers than with humans, but also that system design could engender these behavioural differences by affecting users' expectations about the computer's abilities as an interlocutor in HCD.

The finding that people aligned to different extents depending on whether they believed their interlocutor to be a computer or a human is consistent with previous research suggesting significant differences in linguistic behaviour in HCDs versus HHDs. Amalberti, Carbonell, & Falzon (1993) showed that when people took part in a telephone conversation concerning air-fares and timetables, their linguistic behaviour differed depending on whether they believed their partner to be a human or a computer. Thus when interacting with a computer, people tended to use fewer fillers and coherence markers, provided less information, used more words, and tended to solve problems on their own rather than using linguistic means to clarify ambiguities and increase

understanding. Kennedy, Wilkes, Elder, & Murray (1988) reported similar findings, whereby participants tended to use 'simpler' utterances (reduced use of pronominal anaphors, more basic lexical choices, and shorter utterances) in HCD than in HHD. Similarly, people tend to use simple syntactic structures when interacting linguistically with animated computer-based agents (Bell & Gustafson, 1999). People's preconceptions of the system's capability have also been shown to be integral to how users form their speech when interacting with speech dictation software, for example with respect to hyperarticulation as well as phonological and lexical adjustments (Meddeb & Frenz-Belkin, 2010).

Such findings suggest users' models of the computer's competencies as a dialogue actor may guide their linguistic behaviour in HCD. In fact, users' perceptions of their interlocutors' abilities have consistently been proposed as a significant determinant of linguistic adaptation in HCD (Amalberti et al., 1993; Brennan, 1998; Le Bigot et al., 2007). That is, linguistic behaviour in HCDs is assumed to be fundamentally guided by beliefs about the characteristics of the computer (and failures in establishing accurate beliefs have been identified as a particular cause of communicative breakdown in HCDs; Brennan, 1998), although it is not clear how such beliefs may be established in the first place. Branigan et al. (2011)'s experiments suggest that system design could play some role by influencing people's perceptions of system abilities, with superficial (non-functional) features giving rise to different interlocutor models, and as such could impact levels of alignment in dialogue.

In sum, much of the research on language use in HCD has highlighted a significant impact of partner modelling on our language choices. Many of the studies

have observed how our language use varies when interacting with humans and computers, be it the use of simpler syntactic structures, anaphora or lexical choice. Yet little attention has been given to the role partner modelling may play specifically in alignment of syntactic structure, and how these may be affected by design decisions. As in previous experimental research on lexical alignment, we wish to see whether syntactic alignment is sensitive to partner type and superficial cues, consistent with other HCD research. Importantly, rather than focusing on the superficial cues of system age and reviews as in previous lexical alignment research, we examine the role that design of the interlocutor may have on perceptions and levels of syntactic alignment. Understanding this has significant practical value for those wishing to leverage syntactic alignment effects in guiding spoken dialogue system user inputs.

## 2.4 Voice anthropomorphism and partner modelling

As highlighted, the work of Branigan et al. (2011) supports a potential role for non-functional, superficial cues to impact levels of alignment in HCD, through impacting our model of the partner's dialogue competence. Such models may be impacted by the design of the interlocutor. Research on anthropomorphism in robot and computer agents has shown that people rate anthropomorphised robotic agents as more lifelike (Kiesler et al., 2008), and that anthropomorphic agents are rated as more intelligent and capable (King & Ohya, 1996). Users have also been shown to treat computer partners using anthropomorphic prompts more similarly to a human social partner, using more second person pronouns compared to computer partners using other, less anthropomorphic, prompts (Brennan & Ohaeri, 1994). Indeed seminal research in HCI highlights our

tendency to behave similarly towards computer partners as we do towards human partners in aspects such as politeness (Fogg & Nass, 1997; Nass, Steuer, & Tauber, 1994) and using voice as a social actor identity cue (Nass et al., 1994). This work led us to focus on the potential of voice anthropomorphism in affecting user assumptions of partner ability in spoken HCD interactions. Using an anthropomorphic voice for a computer dialogue partner may make the computer appear more akin to a human conversational partner not only in the form of the output, but also in ascribed communicative competences. In other words, having a human-like voice may lead users to believe that the computer has more advanced communicative capabilities.

Although previous work suggests that anthropomorphism is likely to affect user attributions, it is not specific about the aspects of the agent that impact such attributions. The work presented here therefore tests the role of the voice specifically (see section 3: Manipulation Check below). This not only disambiguates the impact voice may have from other anthropomorphised attributes of an agent but also addresses a design decision that is highly relevant to speech interfaces more generally. To preview our results, we show that people rate an *Anthropomorphic* computer voice as more advanced, flexible and competent than a less anthropomorphic *Robotic* computer voice. This supports previous research mentioned above and demonstrates that voice anthropomorphism specifically affects users' perception of interlocutor competence and thus may have an impact on levels of alignment in HCD.

## 2.5 Partner-based effects for syntactic alignment?

Although beliefs about an interlocutor may be decisive in determining some

linguistic choices, it may not play a strong role in determining others. There is evidence that this may be the case for grammatical choices in at least some circumstances. Branigan, Pickering, Pearson, McLean, & Nass (2003) had participants play a similar text-based picture-naming and -matching task to that used by Branigan et al. (2011), except that participants described and matched pictures of dative events (e.g., showing, giving) rather than naming individual objects. When participants described events that involved the same action as the event that their partner had just described, they showed a stronger tendency to align syntactic structure with a 'computer' than a 'human' interlocutor, as Branigan et al. (2011) found for lexical choices. But when participants described events that involved a different action to their partner's description, they aligned syntactic structure with their partner to the same extent irrespective of whether they believed that their partner was a computer or a human. Thus they were as likely to use a *double object* (*DO*) structure (e.g., *The waitress is showing the doctor the cup*) after their partner used a DO structure (e.g., *The cowboy is handing the jug to the clown)* and a *prepositional object (PO)* structure (e.g., *The waitress is showing the cup to the doctor*) after their partner used another PO structure when they believed they were interacting with a human as when they believed they were interacting with a computer. This discrepancy is intriguing, and suggests that speakers' linguistic choices in HCDs may not always be guided by beliefs about their interlocutors: when structural alternatives are not salient (as they may have been in Branigan et al., 2003, when the action - and hence verb - were repeated), speakers may not necessarily accommodate their interlocutors' perceived capabilities. Taking these results together with existing evidence about the role of partner modelling in affecting lexical alignment and adaptation in HCD more generally

(highlighted in section 2.3), it is clear that the role of partner modelling on users' syntactic choices in HCD, and specifically their tendency to syntactically align, needs further investigation.

## 2.6 Research Aims & Hypotheses

As described in section 1, the research aims to investigate magnitudes of syntactic alignment in HCD compared to HHD through controlled experimentation as well as its potential to impact user's strong default syntactic preferences. It also aims to explore the role of interlocutor identity and design cues such as voice anthropomorphism on the extent of syntactic alignment, independent of potential confounding factors. Such findings are not only informative about how design decisions about the system as a conversational partner causally impact user attributions and behaviour in HCD, but also whether these decisions, or indeed partner type (i.e. computer or human) itself, affect users' alignment behaviour, concordant with findings highlighting user-system adaptation in HCD and the sensitivity of lexical alignment to partner characteristics in HCD.

To facilitate comparisons with previous research, the two experiments that we report here use the same type of controlled experimental paradigm used in previous related research in HHD (Cleland & Pickering, 2003), which has also been shown to be sensitive to the effects of users' beliefs on language behaviour (Branigan et al., 2003). In our studies, we asked participants to interact with another human, a computer with a highly anthropomorphic voice, or a computer with a monotone 'robotic' voice (each tested in the manipulation check presented in section 3) to play a picture-description and matching game. The participants and their partners took turns describing pictures of

dative events (Experiment 1) or colored patterned shapes (Experiment 2) for their partner, and choosing pictures in response to their partner's descriptions. We manipulated the grammatical structure of the partners' (scripted) descriptions (Experiment 1: Prepositional Object (*PO*) or Double Object (*DO*)- similar in terms of default preference; Experiment 2: Adjective-Noun (*AN*) or Noun-Relative clause (*RC*)- AN being heavily preferred), and examined whether participants used the same structure that they had just heard when producing their own immediately subsequent description. We further examined whether any such tendency was affected by the identity of the partner, with greater alignment predicted for partners that might be believed to be communicatively less able, on the basis of identity (human vs. computer) or design (anthropomorphic vs. robotic voice).

We hypothesise that there will be a significant alignment effect in both experiments. Based on findings of lexical alignment and other HCD research we also hypothesise that there will be a significant effect of partner type on the magnitude of alignment, specifically that larger magnitudes of alignment will be seen in the computer partner conditions compared to the human partner condition, and that alignment will be significantly higher in the robotic compared to the anthropomorphic computer partner conditions. We also hypothesise that the influence of partner type on alignment may vary across the structures tested in Experiment 1 and 2 due to the difference in salience between structural alternatives. That is, the effect of partner type on alignment may be higher when structural alternatives used by the partner vary strongly in their use in a null context, making it more salient to the user when their partner is alternating between common and uncommon structures, with people aligning more with computer partners to

ensure communication success in this context. However if no statistically significant effects of partner are found, this may support a more automatic, priming-based view of syntactic alignment in HCD.

## 3. MANIPULATION CHECK

Before conducting the main research, people's initial beliefs about the abilities of a computer as a conversation partner were measured in a study to verify that design considerations such as voice anthropomorphism do significantly affect user judgments, as suggested by previous research.

A sample of 63 participants (35 women, 28 men) with a mean age of 24.33 years (S.D.= 4.13 years) were recruited via campus-wide emails from the University of Birmingham and Edinburgh staff and student communities to take part in the research. All participants were adult native English speakers.

The study involved participants listening to audio clips of one of six possible computer voices (a male and female version was created for each of the three voice types) describing 8 objects, in a between-participants design (i.e., each participant experienced only one voice). Similar to the experiments in section 4 and 5, a between-participants design was used. This was so that the study reflected as much as possible the scenario of interacting with a single dialogue partner, a common scenario in natural HCD interactions, rather than comparing multiple partners in one interaction. In the *Robotic voice* condition, descriptions were produced in a 'robotic' and monotone voice that lacked natural intonation. The audio recordings of descriptions used for this voice were created using the Fred (for the male version) and Kathy (for the female version) voice

options on the text-to-speech program Vox Machina 1.1 for Mac. The *Anthropomorphic voice* condition used audio recordings of the descriptions given by a computer voice producing human-like speech. The audio used for this condition was created using the voice Nick (for the male voice condition) and Nina (for the female voice condition) from the University of Edinburgh's Centre for Speech Technology Research (CSTR) Festival text-to-speech system (http://www.cstr.ed.ac.uk/projects/festival/). An extreme anthropomorphic voice (the *Human voice* condition) was also used in the experiment. It was created using recordings of a male and female member of the experiment team describing the same 8 objects. In all conditions participants were told that the voices heard were computer-generated voices.

Participants were asked to complete a questionnaire about the voices they heard, consisting of 15 items. Participants were given the statement "If a computer system used this voice to speak to me, I'd think it was ……" and were asked to rate the computer system on perceptions of its advanced nature (Basic-Advanced), capability (Capable-Incapable), cost (Cheap-Expensive), quality (Good-Bad), flexibility (Inflexible-Flexible), power (Lacking in Power- Powerful), speed (Quick-Slow), stability (Stable-Unstable), professionalism (Amateurish- Professional), modernity (Up to date- Old Fashioned), efficiency (Efficient-Inefficient), trustworthiness (Untrustworthy-Trustworthy), competence (Incompetent-Competent), controllability (Controllable-Uncontrollable) and complexity (Simple-Complex). All items were measured using a 7-point semantic differential scale. Items were taken from previous metrics used in the HCI (Hassenzahl, 2001) and wider literature (Osgood, 1957) in addition to specific items, such as the item probing the basic vs. advanced nature of the computer system, that were added for the

current study to test if the voices mapped onto perceptions of computers as basic and advanced interlocutors (Branigan et al., 2011). The presentation sequence of questionnaire items was individually randomized for each participant.

Participants were recruited via email. Participants were invited to take part in an online study investigating opinions of computer voices in which they would listen to 8 audio clips of a computer voice and then complete a short questionnaire about the voice they just heard. A link to the online questionnaire was also included in the original recruitment email. After accessing the online questionnaire through the link, users listened to 8 audio clips of one of six types of computer voice. They were then asked to rate the voice they had just heard on the 15 questionnaire items. After completing the questionnaire they were presented with a debrief page to explain the motivations of the study and were thanked for taking part in the research.

Due to the violation of multivariate normality, a robust version of MANOVA using permutation testing (Anderson, 2001) was run using the vegan package (Oksanen et al., 2015) in R (R Core Team, 2014) to test the effects of voice type and voice gender across the 15 items measured. Recent research has advised that robust statistical approaches should be used above classical statistical approaches due to their increase in statistical power and accuracy, especially (although not exclusively) in cases where assumptions are violated (for a discussion of robust methods and their procedures see Erceg-Hurn & Mirosevich, 2008; Field, Miles, & Field, 2012; Keselman, Algina, Lix, Wilcox, & Deering, 2008). These were therefore used throughout the data analysis for

this manipulation check[2]. The permutational MANOVA showed that there was a statistically significant main effect of voice type [F (2,57)= 11.21, p=.001]; however, there was no main effect of voice gender [F (1,60)=0.17, p>.05] or interaction between voice gender and voice type [F (2,60)=0.93, p>.05]. To identify the effects of voice on the dimensions of the questionnaire highlighted by the MANOVA, a robust One-Way ANOVA using 20% trimmed means with Winsorized variance and bootstrapping was run on each of the questionnaire items using the WRS2 package (Mair, Schoenbrodt, & Wilcox, 2014). Robust ANOVAs are used due to violation of the assumptions of normality and homogeneity of variance in much of the data analysed. In this situation modern robust approaches to ANOVA have been highlighted to be significantly more powerful than classic ANOVA approaches (see Erceg-Hurn & Mirosevic, 2008). Trimmed means and Winsorized variance are used to control for the potential influence of outliers, and the combination of these techniques and bootstrapping have been shown to result in better control of Type I error when classical test assumptions are violated (Keselman et al., 2008). For brevity only those that showed a statistically significant difference are reported.



There were significant differences between participants' ratings of the computer voices on the Basic-Advanced [$F_t$ =10.44, p =. 001], Capable-Incapable [$F_t$ =9.49, p<.001], Cheap-Expensive [$F_t$ =13.64, p<.001], Good-Bad [$F_t$ =16.52, p<.001], Inflexible-Flexible [$F_t$ =10.95, p=.002], Lacking in Power-Powerful [$F_t$ =20.04, p<.001], Amateur-Professional [$F_t$ =11.55, p=.002], Up to date-Old Fashioned [$F_t$ =31.07, p<.001],

[2] Classic parametric tests (i.e. MANOVA and One Way ANOVA) were also conducted concurrently with the robust analyses used in this section with similar results being attained. Due to the desire to control Type I error as well as maximize statistical power in the context of assumptions of normality and homogeneity of variance being violated, the results of the robust tests are reported.

Untrustworthy-Trustworthy [$F_t$ =11.26, p=.002], Incompetent-Competent [$F_t$ =9.88, p<.001] and Simple-Complex [$F_t$ =5.63, p=.006] items.

Robust post-hoc tests showed that, compared to the *Robotic* voice condition, participants rated a computer using the *Anthropomorphic* computer voice as significantly more advanced (p =.008), expensive (p=.008), good (p=.002), flexible (p=.002), powerful (p<.001), professional (p=.003), up to date (p<.001) and competent (p=.002).

Compared to the *Robotic* voice condition, participants also rated a computer using the *Human* voice condition as more advanced (p<.001), capable (p<.001), expensive (p<.001), good (p<.001), flexible (p<.001), powerful (p<.001), professional (p<.001), up to date (p<.001), trustworthy (p<.001), competent (p<.001) and complex (p=.01).

| Dimension | Voice | N | Mean | S.D. |
|---|---|---|---|---|
| Basic-Advanced | Anthropomorphic | 21 | 3.23 | 0.85 |
| | Robotic | 21 | 2.08 | 0.81 |
| | Human | 21 | 4.08 | 0.92 |
| Capable-Incapable | Anthropomorphic | 21 | 3.46 | 0.51 |
| | Robotic | 21 | 4.00 | 0.84 |
| | Human | 21 | 2.62 | 0.51 |
| Cheap-Expensive | Anthropomorphic | 21 | 3.31 | 0.50 |
| | Robotic | 21 | 2.08 | 0.81 |
| | Human | 21 | 4.31 | 0.87 |
| Good-Bad | Anthropomorphic | 21 | 3.54 | 0.51 |
| | Robotic | 21 | 4.92 | 1.26 |
| | Human | 21 | 2.46 | 0.51 |

| | | | | |
|---|---|---|---|---|
| Inflexible-Flexible | Anthropomorphic | 21 | 3.46 | 1.17 |
| | Robotic | 21 | 2.15 | 0.77 |
| | Human | 21 | 4.08 | 0.87 |
| Lacking in Power-Powerful | Anthropomorphic | 21 | 3.62 | 0.51 |
| | Robotic | 21 | 2.39 | 0.51 |
| | Human | 21 | 4.39 | 0.83 |
| Quick-Slow | Anthropomorphic | 21 | 3.77 | 0.48 |
| | Robotic | 21 | 4.46 | 1.59 |
| | Human | 21 | 3.54 | 0.51 |
| Stable-Unstable | Anthropomorphic | 21 | 3.23 | 1.24 |
| | Robotic | 21 | 3.54 | 1.17 |
| | Human | 21 | 2.85 | 0.83 |
| Amateurish-Professional | Anthropomorphic | 21 | 3.92 | 0.87 |
| | Robotic | 21 | 2.31 | 0.81 |
| | Human | 21 | 4.54 | 1.21 |
| Up To Date-Old Fashioned | Anthropomorphic | 21 | 4.15 | 1.30 |
| | Robotic | 21 | 6.08 | 0.81 |
| | Human | 21 | 2.69 | 0.87 |
| Efficient-Inefficient | Anthropomorphic | 21 | 3.31 | 1.12 |
| | Robotic | 21 | 3.85 | 1.27 |
| | Human | 21 | 3.23 | 0.85 |
| Untrustworthy-Trustworthy | Anthropomorphic | 21 | 4.31 | 0.50 |
| | Robotic | 21 | 3.62 | 0.51 |
| | Human | 21 | 5.31 | 0.87 |

| | | | | |
|---|---|---|---|---|
| Incompetent-Competent | Anthropomorphic | 21 | 4.77 | 0.79 |
| | Robotic | 21 | 3.69 | 0.87 |
| | Human | 21 | 5.31 | 0.50 |
| Controllable-Uncontrollable | Anthropomorphic | 21 | 3.77 | 0.79 |
| | Robotic | 21 | 2.92 | 0.87 |
| | Human | 21 | 3.15 | 0.89 |
| Simple-Complex | Anthropomorphic | 21 | 2.46 | 0.51 |
| | Robotic | 21 | 1.92 | 0.87 |
| | Human | 21 | 3.39 | 0.87 |

*Table 1: Trimmed means and Winsorized standard deviations for each item by condition*

A computer using the *Human* voice condition was rated as significantly more capable (p=.025), expensive (p=.008), good (p=.002), up to date (p=.01), trustworthy (p=.007) and complex (p=.03) compared to the *Anthropomorphic* voice condition. All other comparisons were not statistically significant (p >.05). The trimmed means and Winsorized standard deviations of the sample are displayed in Table 1.

The findings of the manipulation check provide evidence that participants judged computers that use the voices in the experiment differently on dimensions that are likely to impact their views of the computers' abilities as effective interlocutors. Importantly the *Anthropomorphic* computer voice was rated as more advanced, flexible and competent than the *Robotic* voice, with the *Anthropomorphic* and *Human* voices not differing statistically on these dimensions. The *Anthropomorphic* condition also led participants to rate a computer using this voice as more expensive, good, powerful, professional and up-

to-date than participants experiencing the *Robotic* voice condition. The most extreme anthropomorphic computer voice (i.e. the *Human* voice) led people to rate a computer as more capable, competent and flexible when compared to a more robotic-sounding voice. Unsurprisingly the *Human* voice also led to users to believe it to be more trustworthy, capable and up-to-date and expensive when compared to the speech synthesized *Anthropomorphic* voice.

The manipulation check extends previous research suggesting that people see anthropomorphic agents as more capable and intelligent (Kiesler et al., 2008; King & Ohya, 1996) by showing that people can make judgments about anthropomorphism of agents on the basis of voice, and that these judgments in turn affect judgments of attributes such as ability. It also demonstrates that the voices used in the following experiments map onto advanced and basic judgments that are suggested to influence lexical alignment (Branigan et al., 2011). The following experiment (Experiment 1) used the *Anthropomorphic* and *Robotic* voice conditions tested in this manipulation check to observe the potential impact of user's beliefs on syntactic alignment, because they were generated using speech synthesis and thus most comparable to voices likely to be used in current speech interface design.



## 4. EXPERIMENT 1

Experiment 1 set out to establish whether alignment of syntactic structure occurs in speech-based HCI, and whether this is influenced by judgments of partner competence based on superficial (i.e. non-functional) aspects such as voice type. The study used the dative (PO/DO) alternation, which has been extensively studied in previous speech-based

HHD and text-based HCD alignment research (Branigan et al., 2003). If people behave in speech–based HCD in the same way as in speech-based HHD, then we would expect to find a significant syntactic alignment effect, so that participants would be more likely to use a PO structure if their partner had just used another PO structure than if their partner had just used a DO structure. Moreover, if this alignment effect were affected by beliefs about the communicative ability of the dialogue partner, then we would expect significantly stronger alignment with computer partners than with human partners (as in Branigan et al., 2011); if such beliefs were in turn affected by interlocutor design considerations such as voice anthropomorphism, then we would further expect differences between the anthropomorphic and robotic voice conditions when compared to the human condition, with greater alignment predicted with robotic voices than anthropomorphic voices.

## 4.1 Method

### 4.1.1 Participants

A sample of 42 participants (23 women, 19 men) with a mean age of 23.34 years (S.D. = 4.19 years) took part in the research. All participants were recruited from the University of Birmingham community with both staff and students from a wide range of disciplines taking part in the research. All participants were adult native English speakers. They were given £5 as an honorarium for taking part in the research.

### 4.1.2 Communication Game

Participants completed a communication game with a partner. Conversational partners took turns to describe images to their partner (*describing turn*) and to choose from a pair of displayed pictures an image that matched their partner's description (*matching turn*). The dyad comprised of a naïve participant and a confederate (human or pre-scripted computer). The confederate used pre-specified grammatical structures (primes) when describing their pictures. The participants were not made aware that their partner was a confederate until after the end of the session. On a matching turn, participants listened to their partner's (i.e. the confederate's) utterance (the prime) and clicked on the picture that matched that description from the images in front of them. On a describing turn, participants described the image displayed in front of them (the target).

### 4.1.3 Communication Game Items

24 experimental items were included in the game, each comprising a description of a picture (a *prime sentence,* uttered by the confederate), a *match picture* (a picture that matched the confederate's prime sentence, seen in the participant's matching turn), a *distractor picture* (displayed with the match picture during the participant's matching turn), and a *target picture* (displayed on the participant's describing turn for the participant to describe - see Figure 1 for an example image). The 24 prime sentences occurred in two conditions (PO: e.g. *The chef handing the jug to the waitress* vs. DO: e.g. *The chef handing the waitress the jug*). The match and target pictures each depicted a dative event involving an agent, patient and beneficiary. Below each picture was a present-tense verb in capital letters (which participants were instructed to use in their

target descriptions). There were four prime sentences (and 4 related match pictures) and four target pictures for each of the six verbs (*give, hand, offer, sell, show, throw*). The event depicted in the target picture always involved different entities and a different action from the event depicted in the prime picture. Distractor pictures involved a mixture of dative and monotransitive events, and were selected randomly on each trial from a pool of 48 filler pictures also used for filler trials (see below; 30 monotransitive events involving 18 monotransitive verbs, each used between two and four times: *pull, kick, hit, hold, lift chase, kiss, punch, eat, scold, shoot, drop, push, catch, tickle, touch, polish, follow*; 18 dative events involving the six experimental verbs).


*Figure 1* - Example experiment item picture. Such a picture can be described either as *"the cowboy offering the robber the banana"* (Double Object-DO) or *"the cowboy offering the banana to the robber"* (Prepositional Object-PO).

48 filler items were also included in the game. These items were used to mask the focus of the game being on the experimental items. As with the experimental items, they comprised of a description of a picture (description of a monotransitive event uttered by the confederate), a match picture (seen by the participant in a matching turn that was a match to the confederate's description), a distractor picture (involving a monotransitive or dative event, displayed with the match picture to the participant in a matching turn), and a target picture to be described by the participant in their describing turn to the partner (i.e. the confederate). Crucially the target picture was of a monotransitive event, rather than a ditransitive event as in the experimental items (see Figure 2). Pictures for the target and match pictures in these items were taken from the pool of 48 filler pictures described above.


*Figure 2*- Example monotransitive filler item picture "*The waitress kicking the robber*".

We constructed two lists, each containing one version of each experimental item: 12 of each prime condition (PO or DO prime), as well as all the filler items. Experiment items in list 1 that had DO as their prime had PO as their prime in list 2 and vice versa. The prime condition was within subjects so as to observe participants' likelihood of using a particular structure in conditions where their partner in the same dialogue primed both that structure and an alternative grammatically acceptable structure equally. This helps rule out explanations for any priming effect, such as participants imprinting on one specific structure, which would exist if structures were primed between subjects. The list received by participants was balanced as much as possible within each condition. The order of experimental items and filler items was fixed for all participants with the distractor pictures being randomly assigned for each item. At least two filler items separated experimental items in each list. A flowchart of the turns for the confederate and the participant in the game are included in Figure 3.

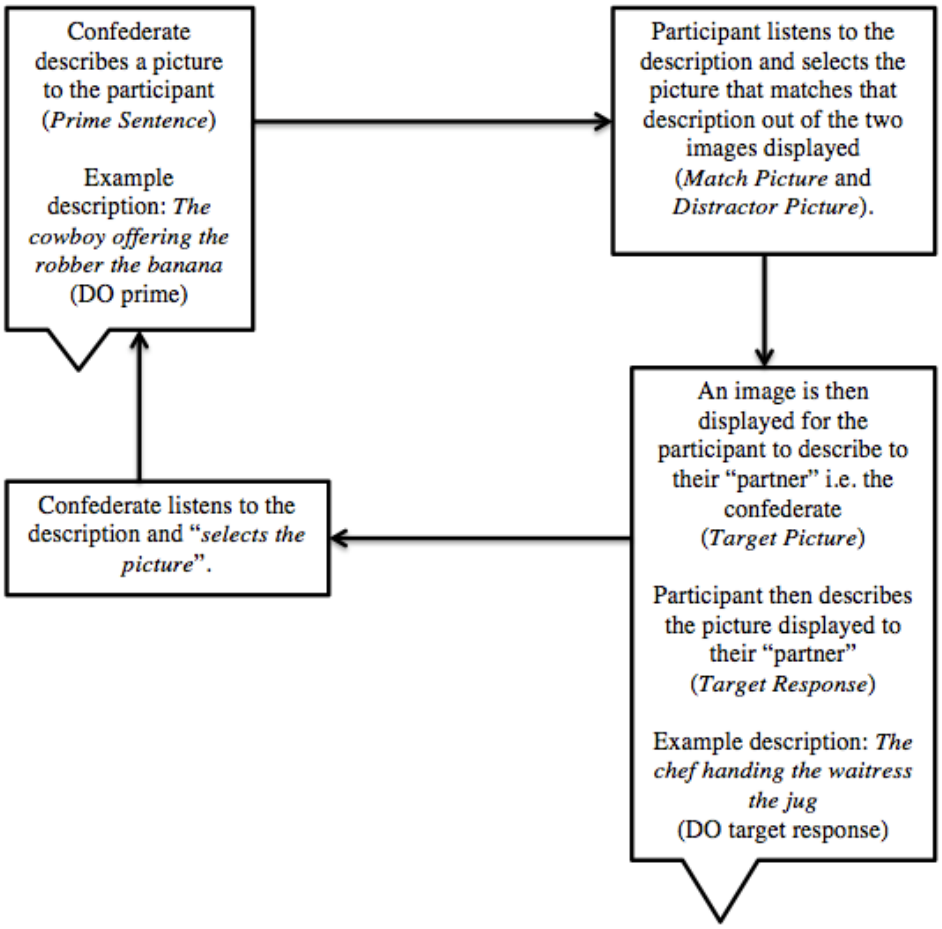


*Figure 3*- Flowchart of game interaction for experiment items. Filler items are identical apart from monotransitive descriptions are used in the prime sentence position and target pictures in the participant turn being of monotransitive events.



### 4.1.4 Interlocutor Conditions

The study included 3 interlocutor conditions (the levels of the independent variable *Interlocutor*) in a between-participants design. A between-participants design was used to ensure that the experiment session reflected as much as possible an interaction with a single dialogue partner, thus lending our findings increased ecological validity in simulating the context of more natural HCD scenarios. A within-participants

design would lead to comparisons to the other partner conditions and thus would impact the ecological validity of any potential partner-based effects to real world HCD contexts (in which people do not sequentially carry out the same task with human and computer partners). In addition we wished to ensure that user behaviours were not impacted by boredom or practice effects, a significant issue in using within-participant research in this context. In the *Human* condition, participants completed the task with a co-present human partner. This condition was included effectively as a control condition against which we could compare levels of alignment in the computer interlocutor conditions. In the *Robotic* computer condition, participants completed the task with a computer that projected a robotic voice. As in the manipulation check, the audio recordings of descriptions used for this voice were created using the Fred voice option on text-to-speech interface Vox Machina 1.1 for Mac. In the *Anthropomorphic* computer condition, participants completed the task with a computer using the anthropomorphic voice (Nick from the University of Edinburgh's Centre for Speech Technology Research (CSTR) Festival text-to-speech system). During the sessions the computer confederates were simulated using a wizard of oz procedure whereby a member of the experiment team controlled remotely the utterances that they used. A connection to the computer in the experiment session was established using Windows Remote Assistance. From this, the experimenter was able to control the computer in the experiment room remotely. The experimenter listened into the session using Skype on the participant's laptop and played audio clips of the relevant descriptions needed for participants to match their pictures on the lab-based machine, thus simulating a computer interlocutor being present in the room, similar to the human-human condition.

### 4.1.5 Procedure

Native English speaking participants were recruited via email from across the University of Birmingham staff and student community and were randomly assigned to one of the three conditions. Upon arrival, they were welcomed by the experimenter, given information about the task being conducted in the study and asked to give consent to take part in the research. The experimenter also checked whether the participant had taken part in any similar studies previously and if so they were informed that they could not take part in the research. The experimenter then informed the participant that they were leaving to get their partner ready and would return soon. The experimenter then returned and took the participant to the experiment lab where they were asked to take a seat on one side of a table. Upon initially entering the lab they could see their partner (and therefore identify whether it was a human or computer partner). During the experiment itself, the table was divided by a screen so that the participant could not see their interlocutor during the dialogue (and thus could not use non-verbal signals for communication). They were then asked to complete a demographic questionnaire (gathering data about their age, gender, whether they were a native English speaker and whether they suffered from any medical condition that would affect their ability to view computer screens safely). The criteria for participation were made clear in the recruitment email. The questions in the demographic questionnaire were used as a final check of these criteria. If participants stated that they were not native English speakers or suffered from a medical complaint, they were informed that they could not take part in the research.

Upon completion of the demographic questionnaire, participants were given information verbally and in written form by the experimenter about the game they were

about to play with a partner. The experimenter instructed the participant (and confederate in the human condition) that the game involved each of them taking turns in being the matcher and the describer of pictures. They were told the aim of the game was to describe the picture in front of them (on describing turns), and to select the correct item described to them by their partner (on matching turns), as quickly and as accurately as possible. They were explicitly informed in these instructions whether they were playing with another human participant (in the Human condition) or a computer (in the Robotic and Anthropomorphic conditions). To familiarize participants with the game, they completed a practice trial of four items.

The confederate always took the role of the describer (i.e. took a describing turn) first and always understood participant's descriptions and matched the pictures. The experimenter noted the syntactic structure (PO, DO or Other) of the participant's target responses when describing the target item in the experiment item pair. A description was scored as a "PO" if the theme of the action immediately followed the verb and was followed by the preposition "to" and the beneficiary. A description was scored as a "DO" if the beneficiary immediately followed the verb and was followed by the theme. Responses not scored as either POs or DOs were scored as "Other". This data acts as the categorical dependent variable *Target Response* in the analysis below. The sessions were audio recorded so that on the rare occasion that the experimenter did not note down the target responses they could be recovered. The experimenter then thanked and debriefed the participants as to the motivations of the study.

## 4.2 Results

Of the 1008 target responses, 665 (65.97%) were coded as PO and 327 (32.44%) were coded as DO. There were 16 target responses (1.59%) coded as Other. These 16 data points were removed from the *Target Response* variable. The LME analysis used to analyze the data (see below) is robust to the inclusion of NA data in the dependent variable.

Table 2 shows the proportion of PO target responses as well as the number of PO target responses by condition. This is also shown graphically in Figure 4. The *alignment effect* is calculated as the difference between the proportion of PO target responses in the PO and DO prime conditions[3].

| Condition | N | PO Primes | DO Primes | Alignment Effect |
|---|---|---|---|---|
| Human-Human | 14 | .72 (120) | .61 (101) | .12 |
| Human-Robotic | 14 | .80 (131) | .65 (106) | .15 |
| Human-Anthropomorphic | 14 | .73 (120) | .52 (87) | .20 |
| Total | 42 | .75 (371) | .59 (294) | .16 |

*Table 2*- Proportion and number of Prepositional Object (PO) target responses by condition

[3] These represent the proportion of PO responses in the total number of PO and DO responses within each condition and as such the proportion of DO responses in each condition can be identified by subtracting the proportions displayed from 1.

Mixed effects logistic regression analysis was run on the data using the *lme4* package (Version 1.1-7) (Bates, Maechler, Bolker, & Walker, 2014) in R (R Core Team, 2014) (Version 3.1.2). The analysis models the impact of fixed effects (e.g. prime, interlocutor and interactions) on the log odds of a specific outcome (e.g. a PO target response) occurring. It also facilitates the inclusion of random effects in the model that can consider participant and item variation (by-participant and by-item random intercepts) as well as the varying impacts of the fixed effects within these units (by-participant and by-item random slopes) (see Barr, Levy, Scheepers & Tily, 2013 for a detailed discussion). This allows us to more fully model potential individual item and participant effects within the analysis as well as negating the need for separate item and participant analyses previously used in psycholinguistic research (Barr, Levy, Scheepers, & Tily, 2013; Clark, 1973).

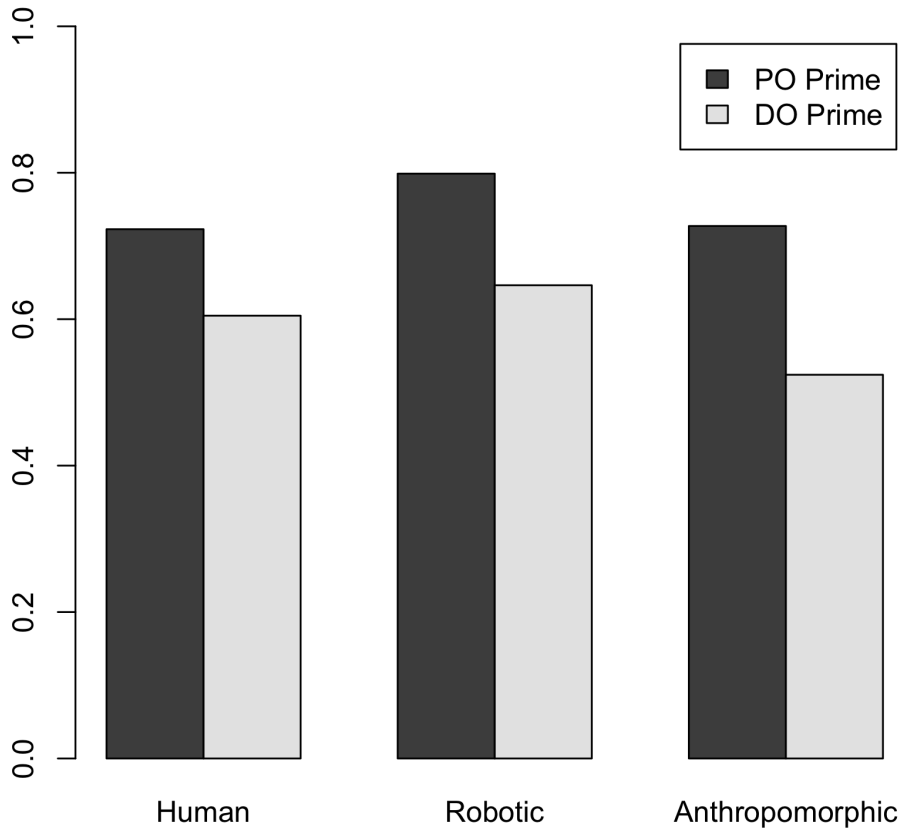


*Figure 4* - Proportion of Prepositional Object (PO) target responses for Prepositional Object (PO-Black) and Double Object (DO-Grey) primes by Interlocutor condition.

The outcome variable *Target Response* was releveled (using the relevel () function) to ensure that the model output refers to the likelihood of PO production. The Interlocutor and Primes variables were also releveled to ensure that the Human and DO prime conditions acted as the base categories for comparison. The model and related *lme4* syntax are shown in Table 3. Due to issues with model convergence identified with using maximal models in mixed effects logistic regression analyses (Barr, Levy, Scheepers, &

Tily, 2013), the higher order within-item random slope for the *Prime:Interlocutor* interaction was removed to facilitate convergence. The final model includes within-participant random slopes for *Prime* and within-item random slopes for *Prime* and *Interlocutor*.

Model: *Target_Response ~ Prime + Interlocutor + Prime:Interlocutor + (1+Prime | Participant) + (1+ Prime | Item) + (1+ Interlocutor| Item)*

| Fixed Effects | Estimates | SE | Wald Z | P value |
|---|---|---|---|---|
| Intercept | 0.67 | 0.48 | 1.38 | .167 |
| Prime (PO) | 1.39 | 0.46 | 3.05 | .002 |
| Interlocutor (Anthropomorphic) | -0.51 | 0.66 | -0.77 | .442 |
| Interlocutor (Robotic) | 0.24 | 0.65 | 0.37 | .715 |
| Prime (PO): Interlocutor (Anthropomorphic) | 0.54 | 0.56 | 0.97 | .332 |
| Prime (PO):Interlocutor (Robotic) | 0.30 | 0.55 | 0.54 | .591 |

| Random Effects | SD |
|---|---|
| *Participant* | |
| Intercept | 1.54 |
| Prime (PO) | 0.80 |
| *Item* | |
| Intercept | 0.22 |
| Prime (PO) | 0.92 |
| *Item* | |

| | |
|---|---|
| Intercept | 0.71 |
| Anthropomorphic | 0.73 |
| Robotic | 0.19 |

*Table 3* - Summary of fixed and random effects for Experiment 1 LME model

The model shows that there was a statistically significant increase in the likelihood of PO target descriptions being used in the PO prime condition compared to the DO prime condition (z= 3.05, p=.002) [4]. There were no significant interactions between the Prime and Interlocutor levels (PO-Robotic: z=0.54, p >.05; PO-Anthropomorphic: z=0.97, p >.05). Thus there was no statistically significant difference between the alignment effect in the human and computer-based conditions, nor any effect of voice type on alignment levels when compared to the human condition. A summary of the fixed and random effects of the model is shown in Table 3.

## 4.3 Discussion

Experiment 1 found evidence of syntactic alignment in both human-human and human-computer speech-based dialogues. Participants showed a reliable tendency to more likely produce PO descriptions after hearing a PO description than after hearing a DO description. This tendency occurred to the same extent irrespective of whether participants interacted with a human or a computer interlocutor, and irrespective of whether the computer interlocutor's voice was anthropomorphic or robot-like.

[4] To check that alignment and partner effects did not vary across the experiment an analysis including a fixed effect of *Time* (first vs. second half of communication game) was also conducted. Time did not significantly affect alignment or the effect of partner on alignment.

## 5. EXPERIMENT 2

Experiment 1 found no difference in the magnitude of people's syntactic alignment with computer versus human interlocutors, nor with computer interlocutors that had 'human-like' versus 'robot-like' voices. These results contrast with previous research on lexical alignment in human–computer dialogue, which showed stronger alignment with computer interlocutors than with human interlocutors, and with computers presented as more limited in ability than with computers presented as more advanced in ability (Branigan et al., 2011). This disparity might reflect a fundamental difference in the extent to which speakers draw on their interlocutor models when making lexical versus syntactic choices. However, an alternative explanation for the disparity between experiments may exist in differences in default preferences for the two alternatives between which speakers chose. Experiment 1 found equivalent alignment with computer and human interlocutors for syntactic choices that were relatively evenly balanced in terms of default preferences (roughly 60% PO, 40% DO; see Gries & Stefanowitsch, 2004; Pickering, Branigan, & McLean, 2002). In contrast, Branigan et al. (2011) found stronger alignment with less capable interlocutors, and with basic computers, for lexical choices that differed strongly in their default preferences (used spontaneously more than 80% vs. less than 20% in a non-biasing context).

In Experiment 2 we therefore examined syntactic alignment when one alternative structure was strongly favoured over the other. Specifically, rather than the Preposition (PO) and Double Object (DO) structures tested in Experiment 1, we compared syntactic alignment for *Adjective-Noun* (*AN*) and *Noun-Relative* (*RC*) clause structures (e.g., *the*

*red square* vs. *the square that's red*). Previous research on HHDs has shown that although syntactic alignment occurs for this structure pair, there is a very strong default preference for AN structures (around 95%) (Branigan et al., In Preparation), and the magnitude of alignment is correspondingly small (Cleland & Pickering, 2003). Experiment 2 therefore allowed us to test whether mechanisms of syntactic alignment in HCI are sufficiently influential to affect strong intrinsic structural preferences. In addition, under these circumstances the choice between a strongly favoured and a strongly disfavoured alternative might be particularly salient, and could therefore be more influenced by strategic decisions based on beliefs about interlocutors' likely understanding or preferences than the structures studied in Experiment 1. Thus we might expect to find more alignment with a computer than with a human interlocutor when one structure is normally strongly disfavoured, and – if the relevant beliefs are influenced by voice anthropomorphism – stronger alignment with a computer with a robotic voice than with a more human-like voice. However, if alignment occurred but identity of the interlocutor had no effect, this would further suggest that users' models of interlocutor abilities formed by superficial cues do not significantly impact speakers' syntactic choices in HCD under these conditions.



## 5.1 Method

### 5.1.1 Participants

A sample of 57 participants (30 women, 27 men) with a mean age of 21.30 years (SD= 3.76 years) from the University of Birmingham took part in the study. Participants were recruited from the staff and student community and came from a wide variety of

subject backgrounds. All were adult native English speakers. They were given a £7 honorarium for participation.

### 5.1.2 Communication Game Items

The communication game structure used in Experiment 1, where participants took turns to describe and match images, was again used in this study. In this experiment we prepared 72 experimental items, again with each experimental item comprising a prime description, a match picture, a distractor picture, and a target picture (see Figure 5 for example match picture used for prime descriptions). Rather than using the PO and DO structures in Experiment 2, the 72 prime descriptions occurred in two conditions (AN: *The red square* vs. RC: *The square that's red*). The materials used (i.e. the descriptions, match, distractor, and target pictures) varied from Experiment 1 in that they were modeled on those used in Cleland and Pickering (2003), and depicted a colored shape (shapes: *star, circle, square, heart, oval, diamond;* colors: *orange, red, blue, purple, green, yellow*). Each of the possible 36 combinations were used once as an RC prime and once as an AN prime for each participant. The target picture always involved a different color and shape from the prime picture. Distractor pictures differed from match pictures in color and shape (50%), shape (25%), or color (25%) to ensure that there was no consistency in the dimension(s) of difference that could lead to participants to assume that one or other description type would be more felicitous for their partner on their describing turn.

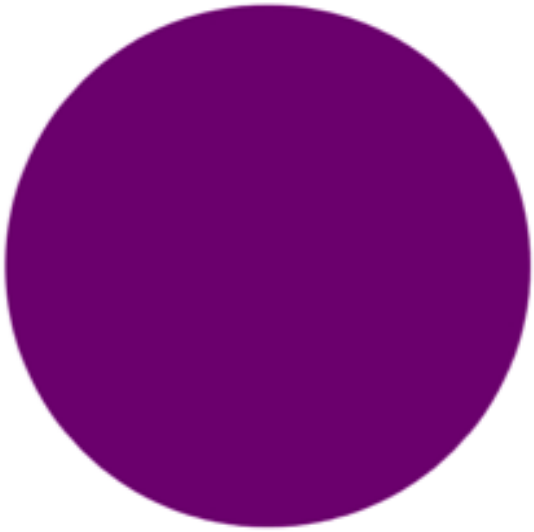

*Figure 5* - Example experiment target picture. The picture can be described using "*the purple circle*" (Adjective-Noun) or "*the circle that's purple*" (Relative Clause).

There were also 120 filler items with the same structure as the experimental items described above, thus containing a description by the confederate, a match picture, a distractor picture, and a target picture to be described (see Figure 6). Again, as in Experiment 1 these were used to mask the focus on the experiment being on the experiment items. To reflect similar dimension to the experiment items, the filler descriptions by the confederate (and the related match pictures), distractor and target pictures involved combinations of multiples of uncolored shapes (1, 2, 3, 4 or 5 shapes per picture), colors (*orange, red, blue, green, purple, yellow*), patterns (*stripy, wavy, dotted, chequered, zigzag, pitted*) and possible color-pattern combinations. This was so as to give consistency to the game. In total 192 items were experienced by each participant,

with the experimental items balanced for prime across the game (36 AN and 36 RC primes in total). The order of presentation was fixed for all participants with the constraint that at least one filler item separated each experimental item. Due to the dispreferred nature of RC structures, more items and participants were used in this experiment to increase the likelihood of RC structures being generated in the data, important to facilitating model convergence for the analysis.

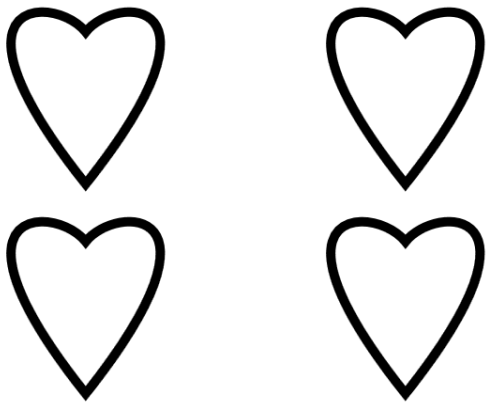

*Figure 6*- Example filler item target picture "*four hearts*".

### 5.1.3 Interlocutor Conditions

The study again included 3 interlocutor conditions: *Human*, *Robotic* and *Anthropomorphic*. As in Experiment 1, the *Robotic* condition used the voice in the pre-test for the *Robotic* condition. Audio recordings of the experimental team were used to simulate the computer interlocutor in the *Anthropomorphic* condition. This was so as to amplify the anthropomorphism of the interlocutor voice as well as maximize the difference in anthropomorphism of the computer interlocutor's voice compared to the *Robotic* condition. A computer using the type of *Anthropomorphic* voice used in this

experiment (i.e. the Human voice in the manipulation check mentioned in section 3) was rated as more advanced, flexible and competent compared to if a computer used the *Robotic* voice.

### 5.1.4 Procedure

The procedure was identical to Experiment 1. The experimenter noted the syntactic structure (AN, RC or Other) of the participant's target responses when describing the target picture in the experiment item. A description was scored as an "AN" if the adjective immediately preceded the noun (e.g. *the red circle* or *red circle*). A description was scored as an "RC" if it included a noun followed by a post-nominal phrase with the adjective (e.g. *the circle that's red*, *circle that is red, circle which is red*). Responses not scored as either ANs or RCs were scored as "Other".

## 5.2 Results

Of the total 4104 utterances, 3975 (96.86%) were AN and 88 (2.14%) were RC utterances. There were 41 (1.0%) target responses code as Other. These were removed from the *Target Response* variable. Table 4 shows the proportion of AN target responses in each *Interlocutor* and *Prime* condition. These are also shown graphically in Figure 7. The alignment effect was calculated as the difference between the proportion of AN target responses in the AN and RC prime conditions.

*Table 4*- Proportion and number of Adjective-Noun target responses by condition

| Condition | N | AN Primes | RC Primes | Alignment Effect |
|---|---|---|---|---|
| Human-Human | 19 | .997 (714) | .980 (695) | .017 |
| Human-Robotic | 19 | .990 (664) | .971 (657) | .019 |
| Human-Anthropomorphic | 19 | .985 (636) | .946 (609) | .039 |
| Total | 57 | .991 (2014) | .966 (1961) | .025 |

As in Experiment 1, mixed effects logistic regression was run using the *lme4* package, using the same model as Experiment 1. The outcome variable *Target Response* was releveled to ensure that the model assessed likelihood of producing an AN response[5]. The Prime and Interlocutor variables were also releveled as in Experiment 1.



[5] Mixed effect logistic regression analysis was also run to assess the effect of prime and interlocutor on the likelihood of producing RC target responses. The findings from this analysis showed the same pattern, i.e., only a significant effect of prime; note that this analysis required simplification of the random effects structure to facilitate convergence. To keep consistency with previous syntactic alignment research and to use the model with more detailed random effects, the model assessing the likelihood of AN target responses is presented.

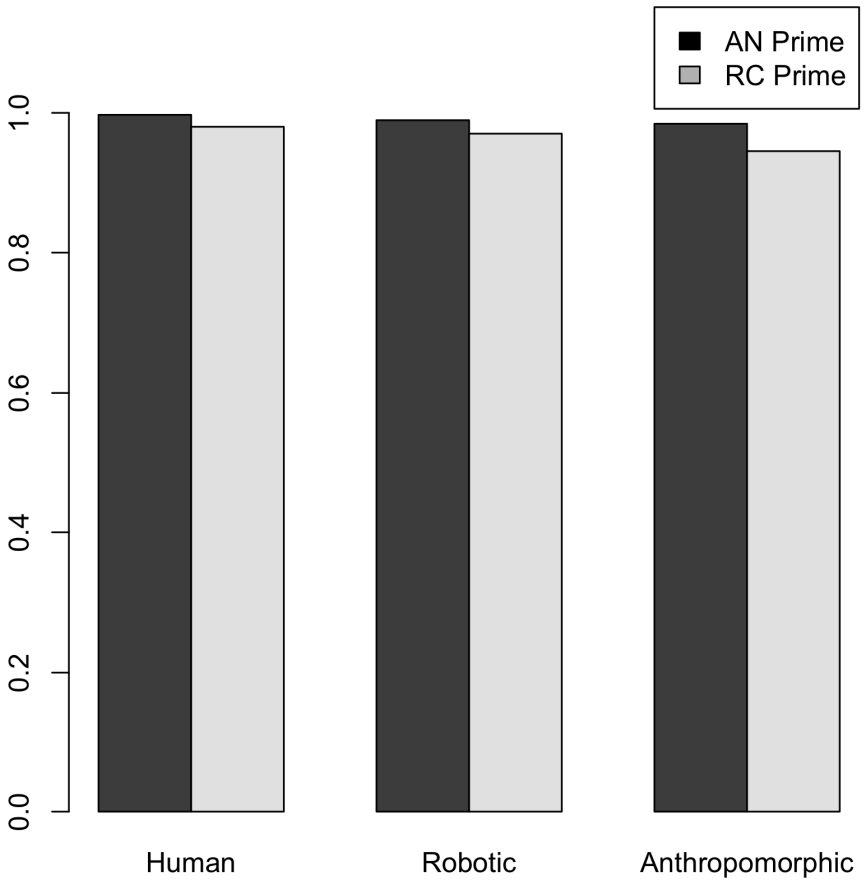


*Figure 7*- Proportion of Adjective Noun (AN) target responses for Adjective-Noun (AN-Black) and Relative Clause (RC-Grey) primes by Interlocutor condition.

The model showed that there was a significant effect of prime on likelihood of producing an AN target response, highlighting a higher likelihood of producing an AN response in the AN prime condition (z= 2.21, p=.027) than the RC condition [6]. Again there were no statistically significant interlocutor effects on alignment levels (AN-Robotic: z= -0.68, p=.50; AN-Anthropomorphic: z=-0.44, p=.66). The summary of fixed and random effects for the model is shown in Table 5.

[6] As in Experiment 1, we ran an analysis including *Time* (first versus second halves of the experiment) as a fixed effect to observe temporal effects of alignment, however the same model did not converge.

Model: *Target_Response ~ Prime + Interlocutor + Prime:Interlocutor + (1+Prime | Participant) + (1+ Prime | Item) + (1+ Interlocutor| Item)*

| Fixed Effects | Estimates | SE | Wald Z | P value |
|---|---|---|---|---|
| Intercept | 5.27 | 0.70 | 7.51 | <.001 |
| Prime (AN) | 2.29 | 1.03 | 2.21 | .027 |
| Interlocutor (Robotic) | -0.61 | 0.83 | -0.74 | .46 |
| Interlocutor (Anthropomorphic) | -1.20 | 0.80 | -1.49 | .14 |
| Prime (AN): Interlocutor (Robotic) | -0.68 | 1.00 | -0.68 | .50 |
| Prime (AN): Interlocutor (Anthropomorphic) | -0.45 | 1.02 | -0.44 | .66 |

| Random Effects | SD |
|---|---|
| *Participant* | |
| Intercept | 1.46 |
| Prime (AN) | 0.82 |
| *Item* | |
| Intercept | 0.23 |
| Prime (AN) | 0.23 |
| *Item* | |
| Intercept | 1.26 |
| Robotic | 0.43 |
| Anthropomorphic | 0.94 |

*Table 5* -Summary of fixed and random effects in Experiment 2 LME Model

## 5.3 Discussion

Experiment 2 again found evidence of reliable syntactic alignment in both human-human and human-computer speech-based dialogues. Participants were more likely to produce AN target descriptions after hearing their interlocutor produce an AN description than after hearing an RC description. This tendency occurred to the same extent irrespective of whether participants interacted with a human or a computer interlocutor, and irrespective of whether the computer interlocutor's voice was human-like (in fact, a recording of a human voice) or robot-like.

# 6. GENERAL DISCUSSION

Speech-based interfaces are becoming increasingly important in the interactions between people and artificial systems. Relatively little is known about the factors that determine people's language use in HCD, and speech-based HCD in particular, although previous research has suggested that users' models of the system's capabilities (interlocutor models) may play an important role in HCD generally. We examined through two controlled experiments whether people's grammatical choices in speech-based HCD are affected by their experience of the system's grammatical choices, so that they tend to use the same grammatical structures as the system has just used. The studies acted as experimental validation for previous naturalistic studies on syntactic alignment in HCD (Stoyanchev & Stent, 2009) by facilitating the control of potential confounds to syntactic alignment in such studies. It also allowed us to identify whether findings related to syntactic alignment in text-based HCD extend to speech-based interactions, a highly relevant context in current interaction modality developments. The testing of syntactic

constructs that vary in their default preferences across the experiments not only allowed us to generalize our findings, they also addressed the potential for alignment effects to impact strong intrinsic structural preferences in syntax use. We further investigated whether any such tendency might be influenced by users' beliefs about the system's capability, and specifically the possible role of system design, focusing on voice anthropomorphism. Participants interacted with a human or computer partner in a speech-based task that involved describing and selecting pictures showing dative events or colored patterned objects. The computer partners used voices that differed in their anthropomorphism and that were rated as differing in their characteristics along dimensions such as advanced nature, capability, modernity and efficiency.

The results of both experiments demonstrated that users' syntactic choices in speech-based dialogue were affected by their interlocutors' linguistic behaviour on a turn-by-turn basis. In Experiment 1, participants were more likely to produce PO descriptions of dative events immediately after hearing their interlocutor produce a PO description for an unrelated picture than after hearing a DO description; in Experiment 2, participants were more likely to produce AN descriptions of objects immediately after hearing their interlocutor produce an AN description for an unrelated object than after hearing an RC description. In both experiments, this tendency was unaffected by the perceived identity of the interlocutor: participants aligned to the same extent whether they were interacting with a human interlocutor or computer interlocutor; similarly, they aligned to the same extent with a computer interlocutor that had an anthropomorphic voice as with a computer interlocutor that had a robot-like voice.

These results add to the growing body of evidence that people tend to align

aspects of their language with their conversational partners not only in HHDs, but also in HCDs. Previous research has shown alignment of prosodic and acoustic features (Bell et al., 2003; Levitan et al., 2012; Oviatt et al., 2004; Suzuki & Katagiri, 2007) and lexical choice (Branigan et al., 2011; Brennan, 1996) in speech-based HCDs, and of lexical (Branigan et al., 2011; Brennan, 1996) and syntactic choice (Branigan et al., 2003) in text-based HCDs. Our results show that users also align syntactically in speech-based HCDs, and that this tendency occurs both for structural alternations in which the alternatives are relatively balanced in their default preferences (PO/DO structures), and for structural alternations in which one alternative is very strongly favoured (AN/RC structures). Previous research has shown that lexical alignment in human-computer interaction can affect very strong preferences (Branigan et al., 2011). The current study found much weaker alignment on disfavoured syntactic structures. Nevertheless, this increase was significant, and suggests that in HCDs as well as in HHDs, even very strong default preferences may be impacted by an interlocutor's linguistic behaviour.

This result has important implications for research on HCD, especially as the use of speech and natural dialogue as an interaction modality grows in popularity. One of the motivations for the current research was to investigate the potential for exploiting alignment to shape users' linguistic interactions with artificial systems (Bell et al., 2003; Stoyanchev & Stent, 2009). The fact that users syntactically align with their partners in speech-based HCD underlines the potential for the system's linguistic behaviours to implicitly guide the user into using specific syntactic structures in less constrained HCDs, leading to a predictable element of speech behaviours that can be modeled in speech recognition. Such modelling may lead to considerable reduction in recognition errors and

thus increase the likelihood of successful communication. Moreover, we have shown that syntactic alignment occurs in speech based interactions where the computer interlocutor and human partner are co-present, a relevant scenario to developments in speech-based dialogue interactions with devices as well as robotic and embodied conversational agents, supporting findings highlighting alignment in text-based interactions where partners are not co-present (Branigan et al., 2003). The work importantly lends experimental validation to more naturalistic studies that have found syntactic alignment in human-computer dialogue scenarios (Stoyanchev & Stent, 2009). From the alternative perspective of dialogue generation, our results support previous proposals (based on evidence from HHD) that engineering systems to produce output that aligned with their human interlocutors would yield more naturalistic dialogues (Brockmann, Isard, Oberlander, & White, 2005). Overall, our research suggests that syntactic alignment could be leveraged in automated interlocutor systems to improve recognition and comprehension of the users' behaviour, as well as to yield more naturalistic output by the system, and thus ultimately to improve communication success (Pickering & Garrod, 2004).



These experiments also contribute to understanding the mechanisms of language behaviours in HCDs. As natural speech grows as an interaction modality in HCI, we need to develop an understanding of the causal mechanisms that govern our linguistic behaviours within this modality, to give us a sound and generalizable basis for future systems development. Earlier less controlled and more naturalistic studies found differences in language use between HHDs and HCDs (Amalberti et al., 1993; Kennedy et al., 1988) that suggested that users' linguistic choices in HCD are affected by their

beliefs about the computer's abilities, and some researchers have accordingly suggested that interlocutor models strongly influence language use in HCD (Amalberti et al., 1993; Brennan, 1998). Evidence that people are more likely to repeat their interlocutor's word choices when they believe that their interlocutor is less capable (computer vs. human; 'basic' computer vs. advanced computer; Branigan et al., 2011) is consistent with this hypothesis. But our experiment-based research, similar in ethos to seminal HCI work by Nass, Steuer, & Tauber (1994) and Nass & Moon (2000), suggests that other factors may also influence language use in these contexts. Specifically, the finding that people tended to align with their interlocutors' syntactic choices, but to the same extent with computer as with human interlocutors (and irrespective of design cues that have been demonstrated to impact judgments of ability, i.e., computer voice), is consistent with current models of HHD that suggest part of the alignment effect may be due to automatic priming mechanisms that do not specifically involve interlocutor modelling in determining speakers' language behaviour (Garrod & Pickering, 2009; Pickering & Garrod, 2004). In this account, people tend to repeat their interlocutors' language choices partly due to the processing of those choices automatically facilitating their subsequent re-use. This account explains why speakers repeat syntactic choices in non-interactive contexts (Bock, 1986) as well as interactive contexts (Branigan et al., 2000; Cleland & Pickering, 2003). The fact that we found similar syntactic behaviour irrespective of interlocutor type gives tentative support to the importance of considering such an account in an HCD context. Our experiments indicate that a factor to be considered in our understanding of user language behaviour in HCD may be the relative accessibility of relevant structures, specifically facilitation of one alternative through prior exposure (i.e. priming); we would

similarly expect that other language-internal factors that have been shown to affect syntactic choice in HHDs (e.g., given vs. new information status; Clark & Haviland, 1977) could also affect syntactic choice in HCDs, although further research demonstrating this empirically is needed to support such a claim.

We stress that this does not mean that users' syntactic choices, and in particular their tendency to make the same choices as their interlocutor, is always automatic and impervious to beliefs. Our experiments examined speakers' choices between the PO/DO alternation and the AN/RC alternation. Although these structures differ in relative preferences (neither alternative is strongly favoured in the PO/DO alternation, whereas there is a strong preference for the AN in the AN/RC alternation), both alternations of the two structure types do not differ greatly in complexity (for example, all four structures are acquired relatively early in childhood; Brown, 1973; Campbell & Tomasello, 2001). Speakers might be influenced by their interlocutor models when they must choose between structural alternatives of markedly different complexity, for example active/passive structures. In such cases, the existence of structural alternatives, and the possible processing implications associated with each of these alternatives may be more salient to speakers (e.g., that passives may be more likely to be misunderstood because they involve atypical mappings of thematic roles to grammatical functions).

More importantly, features of the communicative context may determine the extent to which speakers consult their interlocutor models when making linguistic choices. For example, in our experiments there was no obvious penalty for misunderstanding, but in other contexts there may be an imperative requirement for guaranteed mutual understanding (such as in safety critical dialogues). This requirement

may lead the evaluation of interlocutor abilities to become highly salient, and thus give rise to effects of interlocutor modelling on syntactic choice, including the likelihood of syntactic alignment.

Equally, in our experiments the interlocutor always appeared to understand participants correctly (to ensure that variation in the interlocutor's comprehension behaviour did not confound our comparisons between different interlocutor identities and system design features). A limitation of taking this approach is that the computer partner is seen to understand both structures equally well, potentially leading there to be no motivation to the user to use their partner model to change their behaviour. However, if users' syntactic choices made reference to this partner model (in this case the belief that the partner could understand both structures equally well and either structure could therefore be used successfully without any danger of communication breakdown), we would expect users to either consistently imprint on the first structure that they encountered from the partner, or alternatively consistently use whichever structure they normally preferred to use in a non-biasing context. Contrary to this, we found a significant alignment effect in each study. This pattern of results is more consistent with an automatic priming account of syntactic alignment. Nevertheless, further research making the partner's limitations more salient through partner behaviour could lead to more definitive conclusions about the role that partner modelling plays in syntactic alignment, and syntactic choices in HCD more generally. For instance, if people experienced miscommunication with an interlocutor, such as comprehension errors, this might make the interlocutor's limitations more salient, so that speakers would show an increased tendency to take the partner's capabilities into account when formulating

subsequent utterances, and thus align more strongly to the communication partner's syntax. Communication breakdown might therefore trigger the use of interlocutor modelling to choose between linguistic alternatives (including syntactic choices), as Pickering and Garrod (2004) suggested.

In this research, response latencies of the confederate could not be measured effectively because the confederate and participant game systems were not linked. High turn-taking latencies are negatively correlated with levels of lexical as well as acoustic and prosodic alignment in the observation of entrainment in corpora (Levitan et al., 2012; Nenkova, Gravano, & Hirschberg, 2008). Although there is no evidence to suggest that such latencies affect syntactic alignment specifically (and there is evidence in the relevant psycholinguistic literature that priming effects may persist over many intervening utterances; Bock & Griffin (2000)), we note that varying latencies in the sessions may have impacted the levels of alignment in the experiment, and this remains an issue for further investigation. A further limitation is that, as is common in wizard of oz and confederate-based dialogue experiments in HCD, the confederates were not blind to the conditions being tested, potentially impacting their behaviours in the dialogue interactions. However we found no effect of partner in the studies, suggesting that such an effect is not likely to strongly impact the validity of our findings.

We suggest that future work should extend this research in terms of both the structures and the contexts investigated. First, it is important to examine a wider variety of more complex syntactic structures, which might be more amenable to influences of interlocutor modelling. Second, it is important to widen the context of study. The communication task used in this research used images that could easily be described

using the two structures under investigation in the respective experiments. This allowed controlled elicitation of the structures of interest, although of course it did not facilitate alignment itself (as both structures could potentially be used to describe the experimental items). However, it is important to also explore more naturalistic contexts and interactions where pictures are not the main stimuli and where stimuli are not deliberately designed to elicit the grammatical alternatives under investigation. Third, it is important to investigate communicative contexts where the salience of partner abilities is more marked than that tested here, as well as researching the impact of functional experiences of the system on alignment behaviour. If interlocutor models were not found to be impactful in these scenarios, it would further support the case to consider the role of low-level cognitive mechanisms in determining people's language behaviour in HCDs alongside interlocutor modelling suggested by previous research.

Finally, we suggest that our study has methodological implications for future HCI related dialogue research. In conjunction with previous work on alignment in HCD (Branigan et al., 2011, 2003), our experiments demonstrate that experimental psycholinguistic methodology can be harnessed to study the impact of interlocutor design in the development of dialogue systems, and moreover that it can be applied to the increasingly important context of natural spoken HCD in which the computer is present as an interlocutor. The laboratory–based approach adopted in the current study has benefits in allowing a carefully controlled study of the effects of manipulating factors such as voice anthropomorphism in ways that allow us to exclude potentially confounding factors (e.g., by manipulating beliefs about a partner's ability whilst keeping their actual behaviour constant; by controlling potentially important linguistic features

such as structural frequency). In addition, the use of similar methodology across both HCI and psycholinguistic fields allows for cross interpretation of findings that will likely accelerate the development of models, theoretical breakthrough and sharing of scientific knowledge across both the HCI and psycholinguistic domains. Controlled experiments of the kind reported here also offer a particularly powerful tool for validating findings from research focusing on more naturalistic contexts (Gilquin & Gries, 2009). Two important directions for future research are therefore to extend the current methodology to investigate other aspects of linguistic alignment in HCI contexts, and to examine whether the design-based findings replicate outside a laboratory context, for example by analyzing real-world corpora of HCD that use such design manipulations or by using experimental methods that place fewer restrictions on participants' language and interaction (Howes, Healey, & Purver, 2010) in such partner conditions.

In conclusion, we have shown through controlled experimentation that when people interact with computers using speech, they converge on their interlocutor's syntactic choices, supporting existing naturalistic research, and that the level of syntactic alignment is similar to when they interact with other people. However design aspects that have been shown to affect beliefs do not affect user syntactic choices, suggesting that levels of alignment of syntactic choices seem to be at least in part impacted by cognitive mechanisms rather than solely by interlocutor models.

## 7. ACKNOWLEDGEMENTS

The work was funded by the University of Birmingham Ramsay Fund and the British Academy. Holly P. Branigan was supported by a British Academy/Leverhulme

Senior Research fellowship. The authors would like to thank Charlie Pinder, Will Byrne and Manpreet Pangli for their data gathering work and Andrew Howes for his comments on previous drafts of the manuscript.

## 8. REFERENCES

Amalberti, R., Carbonell, N., & Falzon, P. (1993). User representation of computer systems in human-computer speech interaction. *International Journal of Man-Machine Studies*, *38*, 547–566.

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, *26*, 32–46.

Aylett, M. P., Kristensson, P. O., Whittaker, S., & Vazquez-Alvarez, Y. (2014). None of a CHInd: Relationship Counselling for HCI and Speech Technology. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (pp. 749–760). New York, NY, USA: ACM.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Barrett, J., & Jiang, Y. (2012). *Apple IPhone Siri Users* (Market Report). Dallas, Texas: Parks Associates.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. (Version 1.1-7).

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(02), 145–204.

Bell, L., & Gustafson, J. (1999). Interaction with an animated agent in a spoken dialogue system. In *Proceedings of the Sixth European Conference on Speech Communication and Technology* (pp. 1143–1146). Budapest, Hungary: ISCA.

Bell, L., Gustafson, J., & Heldner, M. (2003). Prosodic adaptation in human-computer interaction. In *Proceedings ICPhS 2003* (pp. 2453–2456). ISCA.

Bergmann, K., Branigan, H. P., & Kopp, S. (in press). Exploring the alignment space – lexical and gestural alignment to real and virtual humans. *Frontiers in Human-Media Interaction*.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387.

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: transient activation or implicit learning? *Journal of Experimental Psychology. General*, *129*(2), 177–192.

Bortfeld, H., & Brennan, S. E. (1997). Use and acquisition of idiomatic expressions in referring by native and non‐native speakers. *Discourse Processes*, *23*(2), 119–147.

Branigan, H. P., McLean, J. F., Messenger, K., & Jones, M. (In Preparation). The nature of children's lexico-syntactic representations.

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

Branigan, H. P., Pickering, M. J., Pearson, J. M., & McLean, J. F. (2010). Linguistic alignment between people and computers. *Journal of Pragmatics*, *42*(9), 2355–2368.

Branigan, H. P., Pickering, M. J., Pearson, J. M., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, *121*(1), 41–57.

Branigan, H. P., Pickering, M. J., Pearson, J. M., McLean, J. F., & Nass, C. (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* (pp. 186–191). Boston, MA: Erlbaum, Mahwah.

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

Brennan, S. E. (1998). The grounding problem in conversations with and through computers. *Social and Cognitive Psychological Approaches to Interpersonal Communication*, 201–225.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493.

Brennan, S. E., & Ohaeri, J. O. (1994). Effects of message style on users' attributions toward agents. In *Conference Companion on Human Factors in Computing Systems* (pp. 281–282). New York, NY, USA: ACM.

Brockmann, C., Isard, A., Oberlander, J., & White, M. (2005). Modelling alignment in affective dialogue. In *Proceedings of the UM-05 Workshop on Adapting the Interaction Style to Affective Factors*. Edinburgh.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272.

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

Clark, H. H. (1973). The language-as-fixed-effects fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335–359.

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Clark, H., & Haviland, S. (1977). Comprehension and the given-new contract. In *Discourse Production and Comprehension. Discourse Processes: Advances in Research and Theory* (Vol. 1, pp. 1–40). Norwood, NJ 07648: Ablex Publishing Corporation.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In *Perspectives on socially shared cognition* (pp. 127–149). American Psychological Association.

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*(2), 214–230.

Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: an easy way to maximize the accuracy and power of your research. *The American Psychologist*, *63*(7), 591–601.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Los Angeles; London: SAGE.

Fogg, B. J., & Nass, C. (1997). Silicon sycophants: the effects of computers that flatter. *International Journal of Human-Computer Studies*, *46*(5), 551–561.

Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, *62*(3), 378–391.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181–218.

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialogue. *Topics in Cognitive Science*, *1*, 292–304.

Giles, H., Coupland, J., & Coupland, N. (1991). *Contexts of accommodation: developments in applied sociolinguistics*. Cambridge University Press.

Gilquin, G., & Gries, S. (2009). Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory*, *5*(1), 1–26.

Gries, S. T., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective on `alternations'. *International Journal of Corpus Linguistics*, *9*(1), 97–129.

Hassenzahl, M. (2001). The effect of perceived hedonic quality on product appealingness. *International Journal of Human-Computer Interaction*, *13*(4), 481–499.

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*. Portland, Oregon.

Kennedy, A., Wilkes, A., Elder, L., & Murray, W. S. (1988). Dialogue with machines. *Cognition*, *30*(1), 37–72.

Keselman, H. J., Algina, J., Lix, L. M., Wilcox, R. R., & Deering, K. N. (2008). A generally robust approach for testing hypotheses and setting confidence intervals for effect sizes. *Psychological Methods*, *13*(2), 110–129.

Kiesler, S., Powers, A., Fussell, S. R., & Torrey, C. (2008). Anthropomorphic Interactions with a robot and robot–like agent. *Social Cognition*, *26*(2), 169–181.

King, W. J., & Ohya, J. (1996). The representation of agents: anthropomorphism, agency, and intelligence. In *Conference Companion on Human Factors in Computing Systems* (pp. 289–290). New York, NY, USA: ACM.

Le Bigot, L., Terrier, P., Amiel, V., Poulain, G., Jamet, E., & Rouet, J.-F. (2007). Effect of modality on collaboration with a dialogue system. *International Journal of Human-Computer Studies*, *65*(12), 983–991.

Levitan, R., Gravano, A., Willson, L., Benus, S., Hirschberg, J., & Nenkova, A. (2012). Acoustic-prosodic Entrainment and Social Behavior. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 11–19). Stroudsburg, PA, USA: Association for Computational Linguistics.

Mair, P., Schoenbrodt, F., & Wilcox, R. R. (2014). *WRS2: Wilcox robust estimation and testing.*

Meddeb, E. J., & Frenz-Belkin, P. (2010). What? I Didn't Say THAT!: Linguistic strategies when speaking to write. *Journal of Pragmatics*, *42*(9), 2415–2429.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234.

Nalin, M., Baroni, I., Kruijff-Korbayova, I., Canamero, L., Lewis, M., Beck, A., … Sanna, A. (2012). Children's adaptation in multi-session interaction with a humanoid robot. In *2012 IEEE RO-MAN* (pp. 351–357).

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, *56*(1), 81–103.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are Social Actors. In *Proceedings of CHI 94* (pp. 72–78). Boston, MA, USA: ACM.

Nenkova, A., Gravano, A., & Hirschberg, J. (2008). High Frequency Word Entrainment in Spoken Dialogue. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers* (pp. 169–172). Stroudsburg, PA, USA: Association for Computational Linguistics.

Oksanen, J., Guillaume Blanchet, F., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, B., … Wagner, H. (2015). vegan: Community Ecology Package (Version 2.2-1).

Osgood, C. E. (1957). *The Measurement of Meaning*. University of Illinois Press.

Oviatt, S., Bernard, J., & Levow, G. A. (1998). Linguistic adaptations during spoken and multimodal error resolution. *Language and Speech*, *41 ( Pt 3-4)*, 419–442.

Oviatt, S., Darves, C., & Coulston, R. (2004). Toward adaptive conversational interfaces: Modeling speech convergence with animated personas. *ACM Transactions on Computer-Human Interaction*, *11*(3), 300–328.

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

Pickering, M. J., Branigan, H. P., & McLean, J. F. (2002). Constituent Structure Is Formulated in One Stage. *Journal of Memory and Language*, *46*(3), 586–605.

Pickering, M. J., & Garrod, S. (2004). Towards a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169–225.

Pickering, M. J., & Garrod, S. (2006). Alignment as the Basis for Successful Communication. *Research on Language and Computation*, *4*(2-3), 203–228.

R Core Team. (2014). R: A language and environment for statistical computing (Version 3.1.2). Vienna, Austria: R Foundation for Statistical Computing.

Reitter, D., & Moore, J. D. (2007). Predicting success in dialogue. In *Proceedings of the 45th Annual Meetinng-Association for Computational Linguistics* (pp. 808–815). Prague, Czech Republic.

Stoyanchev, S., & Stent, A. (2009). Lexical and syntactic priming and their impact in deployed spoken dialog systems. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers* (pp. 189–192). Stroudsburg, PA, USA: Association for Computational Linguistics.

Suzuki, N., & Katagiri, Y. (2007). Prosodic alignment in human–computer interaction. *Connection Science*, *19*(2), 131–141.

Van Baaren, R., Janssen, L., Chartrand, T., & Dijksterhuis, A. (2009). Where is the love? the social aspects of mimicry. *Philosophical Transactions of the Royal Society B*, *364*, 2381–2389.

Highlights:

- Paper investigates syntactic alignment in spoken human-computer dialogue
- The role of partner modelling through partner type and voice is also explored
- Humans align similarly with human and computer partners, irrespective of voice
- Priming is an important mechanism to consider in explaining our HCD behaviours
- Syntactic alignment can affect strong default preferences and could be used to improve spoken dialogue technology