Towards alignment strategies in human-agent interactions based on measures of lexical repetitions

Guillaume Dubuisson Duplessis, Caroline Langlet, Chloé Clavel, Frédéric Landragin

Towards alignment strategies in human-agent interactions based on measures of lexical repetitions

AUTHORS' DRAFT

Guillaume Dubuisson Duplessis Caroline Langlet
Chloé Clavel Frédéric Landragin

Abstract

Alignment of communicative behaviour is an important feature of Human-Human interaction that directly affects the collaboration and the social connection of conversational partners. With the aim of improving the communicative abilities of a virtual agent, and in particular its strategies related to (lexical) verbal alignment, this article focuses on the alignment of linguistic productions of dialogue participants in task-oriented dialogues. We propose a new framework to quantify both the lexical alignment and the self-repetition behaviours of dialogue participants from dyadic dialogue transcripts. The framework involves easily computable measures based on repetition of lexical patterns automatically extracted via a sequential pattern mining approach. These measures allow the characterisation of the nature of these processes by addressing various informative aspects such as their variety, complexity, and strength. This framework is implemented in the freely available and open-source software dialign. Using these measures, we present a contrastive study between Human-Human and Human-Agent dialogues on various corpora that reveals major differences in the lexical alignment and self-repetition behaviours. Lastly, we address the challenge of integrating lexical alignment capabilities in artificial agents. To this end, we describe guidelines and we discuss the integration of the proposed framework in a real-time dialogue system.

1 Introduction

Human-agent (H-A) interaction requires the definition of interaction strategies that define how the agent has to respond to the user's verbal and non-verbal behaviours. Modulating the agent's response based on the user's behaviours helps the agent to be more effective in the task assigned to it, and to build a richer relationship, thereby improving its ability to spark sympathy [11] and feelings of proximity [2] in the user. In this context, the concept of engagement plays a crucial role in H-A interaction: the establishment of engaging agents is essential, both for H-A one-shot interactions and for those dedicated to the construction of long-term relations between the user and the agent. Many applications of H-A interaction such as training [21], health [3] [32], or museums [38] [12] show the importance of the engagement paradigm.

The strategies for maintaining the user's engagement can focus either on verbal or non-verbal behaviours. For example, many studies show the importance of feedback and backchanneling for fostering the user's engagement by providing the agent with listening behaviour [39]. Another example of more verbal-focused interaction strategies can be found in the politeness strategies that provide the agent with a social intelligence [64], making it perceived as more engaged in the interaction [30]. For instance, [18] provides a model centered on verbal content allowing the agent to align with the level of politeness and formality of the user. Alignment processes have been widely explored in linguistic studies dedicated to corpus observation [60]. In recent years, we have seen an increase in interest in the implementation of alignment processes in human-machine interaction. As alignment can be a clue to user engagement [30, 14], measuring such a phenomenon in a H-A interaction context can be essential for building interaction strategies focusing on engagement. Grounding a dialogue system on alignment strategies is important because it can help to foster the user's engagement. For example, verbal alignment strategies using the production of other-repetition help to produce appropriate feedback [58].

This article presents easily computable measures of lexical alignment designed to be used in H-A interaction platforms. Verbal alignment is here considered through its lexical materialization, both at the intra and inter speaker levels. In a previous study [23], we provided a first set of measures for inter-speaker alignment. The present paper is an extension of the proposed model, providing: *i)* improved measures of other-repetitions that better integrate the complexity of the shared lexicons built; *ii)* the quantification of a new communicative behaviour—self-repetitions; *iii)* the extension of the study of the lexical alignment process to a new H-A corpus—the HAI Alice Corpus [62].

Section 2 is dedicated to related work presenting existing studies of verbal alignment in both human-human (H-H) and H-A dialogues. Section 3 presents our lexical alignment model and the underlying algorithms for measuring other and self-repetitions. Section 4 provides an analysis of our outputs and thus an in-depth study of lexical alignment processes in H-A interactions. Finally, Section 5 presents the first step in the integration of repetition measures in a dialogue system focusing on the natural language generation module.

2 Related Work

2.1 Convergence of Behaviours in Dialogue and Lexical Alignment

Convergence of behaviours is an important feature of H-H interaction that occurs both at a low level (e.g., body postures, accent and speech rate, prosody, word choice, repetitions) and at a high level (e.g., mental, emotional, cognitive) [29]. In particular, it has been argued that dialogue participants (DPs) automatically align their communicative behaviour at different linguistic levels including the lexical, syntactic, and semantic ones [50]. Indeed, interactive alignment is a subconscious phenomenon that naturally occurs in H-H dialogue [50]. Studies have shown that speakers reuse lexical as well as syntactic structures from previous utterances [51, 65]. It has been observed that DPs are more likely to continue using the same words with the same addressee than with a new addressee [7]. Put differently, there is much less variation *within* than *between* dialogues. Notably, alignment theory predicts the existence of patterns of repetition via a priming mechanism stating that "encountering an utterance that activates a particular representation makes it more likely that the person will subsequently produce an utterance that uses that representation" [50]. One consequence of successful alignment at several levels between DPs is a certain repetitiveness in dialogue and the development of a lexicon of fixed expressions established during dialogue [50]. In other words, DPs tend to automatically establish and use fixed expressions that become *dialogue routines* via a process called *routinization*. By fostering the development of a common language between collaborators, verbal alignment facilitates successful task-oriented dialogues [45, 26, 28, 55]. Thus, a key ability in dialogue is to be able to align to show a convergent, engaged behaviour.

The alignment process encompasses many aspects of communicative behaviour leading to different related notions. This article focuses on lexical alignment and more specifically on the repeated linguistic structures between DPs involving lexical and syntactic levels. On the other hand, other notions differ in the way they integrate the temporal and dynamic aspects. For example, the notion of *mimicry* is defined as the direct imitation of what the user produces [1] whereas *synchrony* is defined as the reciprocal and dynamic adaptation of temporal structures of behaviours between interactants [19].

There is a debate in the psycholinguistics community about whether convergence of communicative behaviours derives from *partner adaptation* or from *recency adaptation* [6, 8, 5, 56]. Partner adaptation refers to an adaptation of behaviour based on a model of the partner. It is referred to as audience design or entrainment [7]. Verbal alignment, on the other hand, refers to recency adaptation, i.e., an on-the-fly adaptation in the course of dialogue due to the activation of linguistic structures via a priming mechanism [50]. This distinction matters from the perspective of designing an agent able to adapt to a human since it directly impacts the choice of whether to rely on the sole dialogue history for the adaptation behaviour, or whether to build a model of

the human partner in order to adapt. However, both types of adaptation lead to the convergence of behaviours observable in lexical and syntactic cues from dialogue transcriptions. As such, this distinction is of lesser importance when measuring the convergence of behaviours from dialogue transcriptions. Indeed, this kind of measure provides an indicator of the convergence of behaviours that can contribute to a large variety of adaptation models (whether relying on partner adaptation or on recency adaptation only).

Lastly, recent work argues that the convergence of behaviours observed in task-oriented dialogues may not be the norm in ordinary conversations, where divergence rather than convergence occurs [35, 51]. Thus, patterns of repetition may be specific to task-oriented dialogues and may not generalise to ordinary conversation in H-H interactions.

2.2 Lexical Alignment Measures

Measuring lexical alignment between DPs is a challenging task that has been addressed by qualitative and quantitative corpus studies. In this context, computational approaches have been developed to quantify lexical alignment. These approaches are mainly dedicated to corpus studies of H-H conversations and task-oriented dialogues. As such, they do not envision an online usage in an artificial agent interacting with a human. First, some measures focus on the lexical level by considering the sharing and reuse of single words. For instance, [13] quantified lexical alignment based on vocabulary overlap (VO) between DPs[1], while [45] quantified verbal alignment based on high-frequency words, i.e., tracking the reuse of the most common words in the corpus. This approach was further extended to multiparty task-oriented H-H dialogues in [26]. Next, several approaches aimed at studying structural priming, i.e., the reuse of lexical items as well as syntactic structures. [51] focused on regression models to study priming effects within a small time window in a given H-H dialogue. This approach was further extended to take into account prosodic convergence in tutorial dialogues [65]. Meanwhile, [35] computed similarity at the syntactic and lexical levels on windows of a fixed number of turns to question the convergence behaviour in H-H conversations. While previous approaches considered adaptation within a dialogue, [56] used a frequency-based approach [15] to measure adaptation *between* dialogues.

Recently, methods of nonlinear data analysis have been investigated to study the dynamic coupling of DPs during dialogue [27]. Several studies have employed cross-recurrence quantification analysis (CRQA) to quantify interactive alignment and interpersonal synergy at the lexical and prosodic levels [28] as well as emotional contagion [63]. CRQA enables the analysis of the shared dynamics of two time series. It can be viewed as a non-linear equivalent of cross-correlation: it quantifies the strength, but also the form and complexity of the shared dynamics of two systems [27].

[1] VO is the ratio of shared tokens between locutor S_1 and locutor S_2. It is given by: $\frac{\#(\text{Tokens}_{S_1} \cap \text{Tokens}_{S_2})}{\#(\text{Tokens}_{S_1} \cup \text{Tokens}_{S_2})}$. The higher it is, the more vocabulary is shared between S_1 & S_2.

2.3 Implementations of Lexical Alignment Capabilities in Dialogue Systems

Many existing H-A interaction systems implement lexical alignment capabilities. The main goal is to enable a two way alignment (rather than just the human aligning to the system). All the systems focus on adapting the next system utterance given the preceding dialogue history. Most of the existing systems target specific verbal alignment phenomena and are (partially) rule-based. They rely on explicit modelling of alignment based on carefully crafted rules. Some virtual agents aim at high-level alignment on politeness [18] or appreciation [14]. [18] presents a virtual guide which is able to align on the level of politeness based on the theoretical work of Brown and Levinson [9].

[14] presents a virtual agent interacting with a human partner in a museum setting. The agent is able to align (or not) at the level of appreciations (e.g., "I like Picasso too!"). It is based on carefully crafted rules and templates to align (or not) both at the level of appreciations and at the level of the surface form of the expression of the appreciation. [42] presents a spoken dialogue system for an information-providing task in the public transport domain in English. The authors specifically focus on lexical entrainment, i.e., the tendency to use the same terms when DPs refer repeatedly to the same objects [6].

Some systems introduce verbal alignment capabilities in existing natural language generation modules taking as input a specification of the meaning of the utterance (e.g., a logical form) and outputting several alternative surface forms. For instance, the SPUD prime system [10] is a micro-planner capable of both syntactic and lexical alignment designed to investigate psycholinguistic hypotheses in a controlled way.

The PERSONAGE-PRIMED system [36] extends the PERSONAGE language generator [43] with a dialogue context representation including many features of prior human utterances on which the system can align. It includes the user's referring expression, tense and modality, various specific lexical selections (verb, noun, cue words) and syntactic template selection.

Some recent preliminary work aims at developing fully trainable natural language systems able to align with the user [25]. Notably, these efforts aim at modelling alignment implicitly by learning the generation process from data. However, this approach relies on data that are arguably hard to obtain.

2.4 Discussion and Main Contributions

This article focuses on lexical alignment in H-H and H-A task-oriented interactions. A major originality of this work is to provide an analysis of lexical alignment in H-A task-oriented dialogues while previous studies by other researchers have mainly focused on H-H task-oriented dialogues or conversations. We use the term alignment to refer to the fact that DPs converge at the lexical level by using the same words and expressions (e.g., by employing the expression "the piece of cake" to refer to an item in the book *Alice's Adventures in Wonderland*, or the expression "that's not gonna work for me" to reject a proposition in

a negotiation). Lastly, this work is related to convergence of behaviours from a recency adaptation perspective, i.e., mainly depending on the dialogue history, rather than from a partner adaptation perspective.

In this paper, we propose a new framework to quantify both the interactive lexical alignment process and the self-repetition behaviours of DPs from dyadic dialogue transcripts. An originality of our approach is to consider lexical patterns predicted by the routinization process of the interactive alignment theory. These lexical patterns go beyond single words and take into account syntactic cues to the extent of the patterns. We describe a computationally efficient method based on sequential pattern mining of the automatic construction of the expression lexicons built by the DPs during dialogue. We define measures based on these lexicons to quantify the lexical alignment process and the self-repetition behaviours of DPs. These measures address different granularities. They are suited to corpus study via measures characterising the entire dialogue as well as measures characterising the behaviours of each DP. In addition, our framework includes measures that make it possible to investigate the contribution of individual utterances to alignment and self-repetition. Importantly, our framework is not limited to corpus studies and has been designed to be embedded in interactive dialogue systems. Namely, it provides utterance-level measures that can be used in an online context such as a dialogue system. These measures rely on easily computable algorithms, enabling online usage in a time-constrained dialogue system.

3 Sequence Pattern Mining for Measures of Lexical Repetitions

We propose a framework to quantify the interactive lexical alignment process and the self-repetition behaviour of DPs in dyadic textual dialogues. This framework focuses on lexical patterns occurring in dialogue utterances. It distinguishes two main types of such patterns. The first type is *shared* lexical patterns between DPs, i.e., patterns that are initiated (or primed) by a DP, subsequently adopted by the other DP and possibly reused during the dialogue by any DP. These patterns are directly related to the interactive verbal alignment process, a particular type of on-the-fly linguistic adaptation. They can be seen as shared dialogue routines at the lexical level. They are a way to verbally align and ultimately share a common language to improve understanding, collaboration and social connection to a conversational partner. The second type is lexical *self*-repetition. Contrary to the previous type which considers patterns that are shared between DPs, self-repetition considers each DP in isolation. Self-repetitions [34] are lexical patterns appearing at least twice in the dialogue utterances of a given DP, independently of the other DP's utterances. Self-repetitions are directly related to the self-consistency of the linguistic production of a given DP.

The general idea of the framework is depicted in Figure 1. The main con-

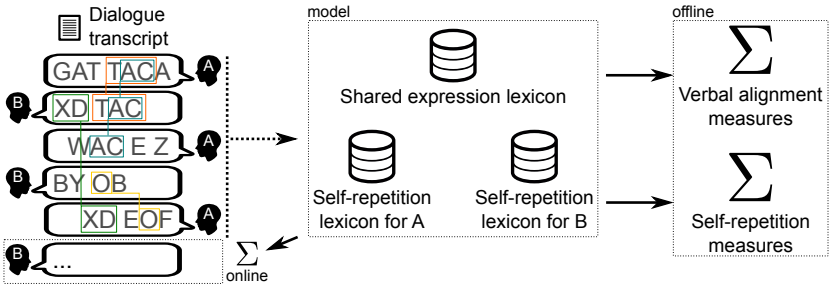


Figure 1: Proposed framework: automatic building of the shared expression lexicon and the self-repetition lexicons to derive offline and online measures of verbal alignment and self-repetition behaviour. Shared lexical patterns are shown on the dialogue transcript.

cept behind our model is the automatically built *lexicon* detailed in Section 3.1. Lexicons and the dialogue transcript are leveraged by deriving offline and online measures to quantify aspects of the verbal alignment process and the self-repetition behaviour of DPs. Offline measures are intended to be used for past dialogue interactions (e.g., corpus studies presented in Section 4) while online measures are intended for use in a dialogue system (see Section 5).

3.1 Automatic Building of the Lexicons using Sequential Pattern Mining

Our model considers three lexicons: a shared expression lexicon for shared lexical patterns, and two self-repetition lexicons (one for each DP). It relies on the following definitions that are illustrated in Table 1 which presents an excerpt of dialogue extracted from one of the corpora used in this work.

– A *shared expression* is a surface text pattern inside an utterance that has been produced by both speakers in a dialogue (or a portion of dialogue). In other words, it is a sequence of contiguous tokens that appears in at least two utterances produced by two different speakers. A shared expression may be a single token (e.g., "sorry" in Table 1). However, it should contain at least one non-punctuation token. Thus, sequences like "?", "!", "," are not expressions. An *instance of a shared expression* refers to the lexical patterns that constitute the shared expression but that appear in the dialogue transcript (i.e., in a dialogue utterance). An instance can be either free or constrained in a given utterance[2]. A *free* instance is an instance of a shared expression that is not placed under the syntactic dependency of another syntactic segment. Conversely, a *constrained* instance is an instance of a shared expression that appears in an utterance within a dependency relationship with another syntactic segment. For example, the expression *a piece of* in the segment *I would rather*

[2]This terminology is borrowed and adapted from the textual data analysis field and the notion of "repeated segment" [41].

Table 1: Excerpt of dialogue C7 extracted from the H-A Alice corpus (described in Section 4.1). The agent is a Woz system. Instances of shared expressions are coloured. Established expressions are in italics.

Loc.	Utterance
H_1	okay well but the books like there are many eh facts in the book which are against the law of physics like the piece of cake which is which makes eh Alice bigger or smaller i don't remember but which
A_2	indeed
H_3	yeah eh does the piece of cake makes Alice bigger or smaller ?
A_4	*the piece of cake* grows *Alice*
H_5	*grows Alice* okay and what makes her smaller?
A_6	sorry could you repeat that ?
H_7	what makes eh *Alice* smaller ?
A_8	throughout *the book there are many* making *her smaller*
H_9	i'm *sorry* i didn't understand
A_{10}	throughout *the book there are many* passages where *Alice* shrinks

a piece of cake is a constrained instance because placed under the dependency of the noun *cake*. However, in the same sentence, the segment*rather a piece of cake* would be considered as a free form.

The *initiator* of a shared expression is the interlocutor that first produced an instance of the expression either in a free or constrained form. Lastly, a shared expression is established as soon as the following two criteria are met: (i) the expression has been produced by both interlocutors (either in a free or constrained form), and (ii) the expression has been produced at least once in a free form. The first turn in which these criteria are all met is the *establishment turn* of the expression. Finally, the *shared expression lexicon* of a dialogue is the set of established shared expressions that appear in this dialogue. Importantly, the shared expression lexicon contains all the shared expressions that appear in a dialogue at least once in a free form. Expressions that are always constrained (i.e., instances that are always a subpart of a larger instance) are discarded. In Table 1, "the piece of cake" is a shared expression initiated by H in Turn 1 and established in Turn 4. This shared expression is free in this excerpt, and it belongs to the shared expression lexicon. Similarly, "Alice" is a shared expression initiated by H in Turn 1 and established in Turn 4. It appears in a free form in Turns 1, 3, 7 and 10; and in a constrained form in the other expression "grows Alice" in Turns 4 and 5. It belongs to the shared expression lexicon. Finally, the lexical pattern "the piece of" occurs in a constrained form in Turns 1, 3 and 4, and never occurs in a free form. This lexical pattern is never established (contrary to its parent expression "the piece of cake") and thus is not included in the shared expression lexicon.

– A *self-repetition* is a surface text pattern at the utterance level that has

been produced at least twice by the same DP. For example, in Table 1, "throughout the book there are many" is a self-repetition of locutor A occurring in utterances 8 and 10. On occasions, a lexical pattern may be both a self-repetition and a shared expression. However, self-repetitions and shared expressions are conceptually different. The main difference is that self-repetitions regard the utterances from a given DP in isolation from the utterances of the other DP contrary to the essence of shared expressions. This difference is manifested by the existence of two self-repetition lexicons (one for each DP). The lexicons only retain the largest self-repetitions. That is, subparts of self-repetitions that always occur inside a larger self-repetition are discarded. For example, this is the case of the self-repetition "throughout the book" in Table 1 that always occurs inside the larger repetition "throughout the book there are many". While the latter is included in the lexicon, "throughout the book" is not.

The automatic extraction of shared expressions and self-repetitions from a dialogue is an instance of sequential pattern mining [44] applied to textual dialogues. In this work, we follow a similar approach to [22, 23] by employing a generalised suffix tree in order to solve the multiple common subsequences problem [33] to extract frequent surface text patterns between utterances. Notably, this problem is solved in linear time with respect to the number of tokens in a dialogue [33]. Shared expressions are extracted from recurrent lexical patterns occurring across the utterances of both DPs. Frequent text patterns are filtered out to retain patterns that are shared between DPs and that appear at least once in a free form. Self-repetitions are extracted from frequent text patterns occurring across the utterances of a given DP (utterances from the other DP are not considered). The algorithms and software used in this work are currently freely available at the following URL: `https://github.com/GuillaumeDD/dialign`.

3.2 Easily Computable Measures of Lexical Alignment and Self-Repetition Behaviour at the Lexical Level

As shown in Table 2, the proposed measures aim at characterising and quantifying several aspects of the interactive verbal alignment process and the self-repetition behaviours by considering (repeated) lexical patterns. It leverages offline and online measures. Offline dialogue-level measures are computed for an entire dyadic dialogue. Offline measures related to the shared expression lexicon can be broken down into speaker-independent ones and speaker-centered ones (measure_S). Online utterance-level measures (measure_U) operate on an ongoing dialogue and rely on an incrementally built lexicon. They provide an utterance with context-aware scores that take into account the preceding dialogue history. Measures can also be related to the verbal alignment process, or to the self-repetition behaviour of individual DPs (SEV_S, SER_S, SENTR_S, SL_S and SLMAX_S are the speaker-specific equivalents for the self-repetition lexicons of, resp., the EV, ER, ENTR, L and LMAX measures). We also provide a speaker-centered alignment measure, which considers shared expressions initiated by Speaker S, IE_S. The proposed measures aim to inform the following complementary aspects: variety, complexity, stability, strength, and orientation of the

Table 2: Measures of Repetitions : a synthetic presentation (explained in Section 3.2)

	Offline measures		Online Measures
Speaker-independent	Strength	ER	DER_U
	Variety	EV	
	Complexity	$ENTR$	
	Extension/Stability	L $LMAX$	
Speaker-centered	Strength	ER_S SER_S	$DSER_U$
	Variety	SEV_S	
	Complexity	$SENTR_S$	
	Extension/Stability	SL_S $SLMAX_S$	
	Alignment Orientation	IE_S	

studied processes. While some measures were previously introduced in [23], the ENTR, L, LMAX, SEV_S, SER_S, SENTR_S, SL_S and SLMAX_S measures are new contributions of this article.

3.2.1 Measures of the Variety of the Lexicons

The *variety* of the constituted lexicons is a measure of the variety of shared expressions or self-expressions emerging during a dialogue relative to its length. That is, it is directly related to the number of unique expressions in a lexicon. However, variety is not impacted by the length of the expressions in the lexicons. It is indicated by two off-line dialogue-level measures. The higher these measures are, the more varied the lexicons are.

– **EV** (Expression Variety), a speaker-independent measure which measures the variety of the shared expression lexicon by taking into account the length of the dialogue. It corresponds to the shared expression lexicon size normalized by the length of the dialogue (which is its total number of tokens in the dialogue). The higher EV is, the higher the proportion of different shared expressions established between DPs.

– $\textbf{SEV}_S$ (Self-Expression Variety for locutor S) is the speaker-specific equivalent for the self-repetition lexicons of the EV measure, and indicates the variety of self-repetitions by S. It equals the self-repetition lexicon size normalized by the total number of tokens produced by S in the dialogue.

3.2.2 Measures of the Complexity of the Lexical Patterns

The *complexity* of the (shared) lexical patterns that are employed by DPs during dialogue is here directly linked to the size of the expression instances. Complexity indicates the variety of the types of lexical patterns (a similar idea can be

found in the CRQA approach [28]) and is here featured by the following Shannon entropy measures [54]. High entropy indicates the presence of a wide range of lexical patterns relative to their lengths in number of tokens (e.g., ranging from a single word to a full sentence). On the contrary, low entropy indicates the predominance of one type of lexical pattern.

– **ENTR** (Shannon entropy of the lengths in tokens of the shared expression instances) is a measure of the *complexity* of the interactive verbal alignment process. Low ENTR indicates a poor variety of sizes while high ENTR indicates a greater variety of sizes.

– $\mathbf{SENTR}_S$ (Shannon entropy of the length in tokens of the self-repetitions by S) is the equivalent for the self-repetition process. It indicates the complexity of the self-repetitions by S.

3.2.3 Measures of *Extension* and *Stability* of the Lexical Patterns

The *extension* and *stability* of the (shared) lexical patterns (a similar idea can be found in the CRQA approach [28]) are also related to the size of the lexical patterns. The extension indicates the size of the lexical patterns. The longer it is, the more extended the lexical pattern is. Extension is directly linked to the stability of the processes since the more extended the patterns are, the more stable the processes are. The following off-line measures can be viewed as measures of the *stability* of the lexical alignment process:

– **L** (average length in tokens of the shared expression instances). It indicates the extension of shared expression instances, that is if the instances used are rather short (e.g., a single token) or longer (e.g., from several tokens to full sentences). High L indicates long repeated lexical patterns (and thus a more stable lexical alignment) while low L indicates short repeated lexical patterns (and thus a more flexible lexical alignment).

– **LMAX** (Maximum length in tokens of the shared expression instances) is the longest length in tokens of the shared lexical patterns established between DPs. It serves as an indicator of stability. The higher LMAX is, the more stable the interactive lexical alignment process is.

– $\mathbf{SL}_S$ (Average length in tokens of the self-repetitions by S) indicates the stability of the self-repetition process of locutor S.

– $\mathbf{SLMAX}_S$ (Maximum length in tokens of the self-repetitions by S) can be interpreted as an indicator of the stability of the self-repetition behaviour.

3.2.4 Measures of the *Strength* of Repetition of the Lexical Patterns

The *strength* of repetition of the (shared) lexical patterns, i.e., how much the patterns are reused is featured by both off-line and on-line measures of lexical alignment and self-repetition behaviour.

– **ER** (Expression Repetition) is the proportion of tokens which DPs dedicate to the repetition of shared expressions ($\mathrm{ER} \in [0, 1]$). It measures to what extent the lexicon is used by the DPs. The higher ER is, the more tokens DPs dedicate to the repetition of shared expressions. It can be considered an indicator of the

strength of repetition for the lexical alignment process at the global dialogue level.

– $\mathbf{ER}_S$ (Expression Repetition for locutor S) is the proportion of tokens produced by S belonging to a repetition of a shared expression ($\forall$ S, $\mathrm{ER}_S \in [0,1]$). It is the speaker-specific version of ER. The higher ER_S is, the more S dedicates tokens to repeating shared expressions. It can be viewed as an indicator of the strength of repetition for the lexical alignment process at the speaker level.

– $\mathbf{DER}_U$ (Dynamic Expression Repetition at utterance level) is the proportion of tokens dedicated to the repetition of shared expressions in the considered utterance u. It measures the shared expression lexicon usage for a given utterance u and a given preceding dialogue history. It measures the proportion of tokens of u that repeats shared expressions previously established in the dialogue history ($\forall$ u, $\mathrm{DER}_u \in [0,1]$). The higher it is, the more u reuses shared expressions. Note that $\mathrm{DER}_u = 1$ does not necessarily imply a verbatim repetition of a previous utterance since an utterance can be composed of multiple shared expressions leading to this score. It can be viewed as an indicator of the strength of repetition for the lexical alignment process at the utterance level (on-line measure).

– $\mathbf{SER}_S$ (Self-Expression Repetition for locutor S) is the proportion of tokens which locutor S dedicates to self-repetition. It can be interpreted as an indicator of the strength of repetition for the self-repetition behaviour at the speaker level.

– $\mathbf{DSER}_U$ (Dynamic Self-Expression Repetition) is the proportion of tokens dedicated to self-repetition in the considered utterance u. It can be interpreted as an indicator of the strength of repetition for the self-repetition behaviour at the utterance level. It is the equivalent of DER_u but for the self-repetitions.

3.2.5 Measures of the *Orientation* of the Interactive Lexical Alignment Process

The *orientation* of the interactive lexical alignment process can be derived from speaker-dependent measures. Orientation is mainly concerned with the (a)symmetry of this process. Symmetry refers to similar behaviours between DPs with regards to the studied interactive process. It can be observed by the absence of measurable differences in the speaker-dependent measures. On the contrary, asymmetry refers to dissimilar behaviours between DPs. It is evidenced by the presence of measurable differences in the speaker-dependent measures. The previously described off-line speaker-dependent measure $\mathbf{ER}_S$ and on-line measure $\mathbf{DER}_U$ can be leveraged to study the similarity or the difference of the process between DPs. We propose the following additional measure of orientation:

– $\mathbf{IE}_S$ (Initiated Expression for locutor S) is the ratio of shared expressions initiated by locutor S that indicates the orientation of the interactive verbal alignment process ($\forall$ S, $\mathrm{IE}_S \in [0,1]$). It is a speaker-dependent measure. Note that in a dyadic dialogue involving speaker S_1 and S_2, $\mathrm{IE}_{\mathrm{S}_1} + \mathrm{IE}_{\mathrm{S}_2} = 1$. The

higher IE_S is, the more S initiates shared expressions. A symmetric behaviour regarding this measure means that $\text{IE}_{\text{S}_1} \approx \text{IE}_{\text{S}_2}$ whereas asymmetry means that $\text{IE}_{\text{S}_1} \neq \text{IE}_{\text{S}_2}$.

3.3 Relationship with the (C)RQA Approach.

Our framework entertains a close connection to the application of the (C)RQA method to the study of interaction as illustrated by similar measures (such as L, LMAX – see Section 3.2.3 – and ENTR – see Section 3.2.2). The main difference concerns the underlying representation of the data under study. While our approach relies on sequential pattern mining and considers lexical patterns as sequences of tokens, the (C)RQA approach views the whole dialogue as two time series (one for each DP). We argue that viewing lexical patterns as sequences of tokens rather than as a subpart of a time series is more easily interpretable and allows for clearer definitions of shared lexical patterns. In addition, our framework does not rely on intricate parameters that are difficult to tune (such as the embedding dimension, the delay, and the radius) while allowing us to compute measures similar to those of the (C)RQA approach (namely, L, LMAX, and ENTR) and even additional measures (e.g., EV, ER, IE_S). Then, our framework considers both offline measures (e.g., for corpus studies) and online measures that can be integrated into a real-time dialogue system, whereas the (C)RQA method is intended for offline use. Lastly, we would argue that our framework and the (C)RQA method for studying Human-Agent interaction are complementary. The method should be chosen on the basis of what is the most adequate point of view when seeing the studied objects either as a time series (e.g., when studying the acoustic signal) or as a sequence of items (e.g., when studying verbal content such as lexical patterns).

4 Results: Quantitative and Contrastive Corpora Study

4.1 Materials: Interaction Corpora

While more and more transcribed dialogue corpora are available involving H-H interactions (e.g., see [53] for an extensive study), H-A interaction corpora are still rare.

Our study is based on one H-H negotiation corpus and two H-A corpora (one on a negotiation task and comparable to the H-H corpus, and one on an information-providing task). The H-A corpora used in this work can be considered as *idealised* H-A interactions because the agent is operated by a human operator. Such corpora represent more fluent and efficient conversations with fewer communication problems (which can be caused by a noisy automatic speech recognition system for example) than real H-A interactions. The goal of this study was thus to compare H-H and *idealised* H-A interactions and in particular what is different when the user believes he is interacting with a

virtual agent. In such *idealised* H-A interactions, the human participant is very likely to adapt his communicative behaviour with regard to his beliefs about the communicative capabilities of the agent [4]. These three interaction corpora make it possible to compare H-A interactions on two different tasks:

– *Human-Human and Human-Agent Negotiation Corpora* are based on a negotiation task between two DPs and involve H-H interactions as well as H-A interactions. These corpora are detailed in [20, 31]. They focus on a common abstraction of negotiation known as the multi-issue bargaining task [37]. Here, it requires two interlocutors to find an agreement over the amount of a product each player wishes to buy. Each player receives some payoff for each possible agreement, usually unknown to the other party. The payoff is assumed to be additive.

In the H-A corpus, the human participant is engaged in two negotiations with two different virtual agents (a male called Brad and a female called Ellie) operated by Wizards of Oz (Woz) [17].

The Woz system controlling virtual agents has been designed to be as natural as possible [20]. It involves low-level functions carried out automatically (such as the selection of gestures and expressions related to speech) and high-level decisions about verbal and non-verbal behaviour carried out by two wizards. Notably, it includes a large number of preset possible utterances (more than 11,000) along with a specific interface enabling the human operator to rapidly select among them [20]. The Woz system is thus constrained in its verbal output and cannot intentionally produce verbal alignment.

The Woz system is controlled by two human operators making high-level decisions about the agent's verbal and nonverbal behavior [31].

– The *Human-Agent Information-Providing Corpus: HAI Alice Corpus* is the publicly available HAI Alice corpus detailed in [62]. It focuses on an information-providing task between a human user and an embodied conversational agent operated by a Woz. This corpus was collected as part of the ARIA VALUSPA project which targets multimodal virtual agents for information retrieval that can deal with unexpected situations [61].

The conversational domain of this corpus is the novel *Alice's Adventures in Wonderland* written by Lewis Carroll. The virtual agent, a middle-aged male standing in a library, answers questions from the human participant and also suggests topics (e.g., talking about "the white rabbit"). Table 1 presents an excerpt from this corpus. The interaction scenario involves the virtual agent, operated by a Woz, and a human participant, unaware of the Woz setup and supposedly believing that they are interacting with an autonomous agent. The human participant was instructed to ask questions about the book (e.g., events, characters, the author Lewis Carroll, adaptations into movies). The interaction duration is fixed to 7 minutes.

The wizard can see and hear the participant through a video-conference connection and has immediate and almost full control over the agent's verbal and nonverbal behaviour. The virtual agent is rendered in real-time and sent over the connection to the human participant. The wizard is trained to speak like a robot to mimic generated speech. Additionally, his voice is passed through audio

Table 3: Descriptive statistics about the interaction corpora. U = Unique, T/Utt.=Tokens per Utterance, med. = median.

	Nego. H-H	Nego. H-A	HAI Alice (H-A)
Dialogue	84	154	16
Utterance (U)	10319 (7840)	17125 (6109)	1023 (864)
... avg (std)	122.8 (84.1)	111.2 (57.5)	63.9 (9.5)
Token (U)	79396 (2516)	90479 (1335)	10166 (829)
T/Utt.			
avg/med. (std)	7.7/6.0 (7.4)	5.3/4.0 (5.7)	9.9/8.0 (6.9)
min/max	1/66	1/154	1/60

filters to make it sound robotic. The facial animations of the agent are controlled using the wizard's facial expressions (see [62] for further technical details). The Woz system from the HAI Alice corpus can be said to be loosely constrained. While the utterances of the agent are still prepared beforehand based on possible questions that the agent might get during the interaction, the human operator has the right to paraphrase the utterances and is not limited by having to select from a pool of constrained utterances. The present work used the complete version of the HAI Alice corpus (16 dialogues) and included one additional dialogue compared to the public release. This dialogue is not part of the public release due to permission restrictions. In addition, the interruption phases of the dialogue[3] with the experimenter were not considered in this study and were discarded. The Woz system is controlled by a single human operator [62].

Descriptive figures about both corpora can be found in Table 3. Globally, dialogues in both negotiation corpora contain more than 100 utterances and H-A negotiation dialogues are a little shorter than H-H negotiation dialogues but still comparable. Utterances are shorter in terms of tokens in the H-A negotiation dialogues than in the H-H negotiation dialogues. In the HAI Alice corpus, each 7-minute dialogue contains on average 63.9 utterances (without the brief interruption phase). Compared to the negotiation corpora, the dialogues are shorter when looking at the number of utterances (more than 100 utterances on average for the negotiation dialogues, around 60 for the information-providing dialogues). However, utterances contain more tokens (on average, less than 8 tokens for the negotiations; on average, more than 8 tokens for the information-providing dialogues).

4.2 Hypotheses

Following our previous work [23], we hypothesise that the interactive lexical alignment process differs between H-H interactions and H-A interactions. Indeed, several studies provide evidence that lexical items and syntactic structures

[3]In order to study interruption phenomena, 4 minutes after the start, an external interruption was triggered by the experimenter to offer a drink to the human. Then, the interaction proceeded as before the interruption.

used by a spoken dialogue system are subsequently adopted by users [6, 57, 48]. However, one straightforward reason for a difference in lexical alignment is that both human participants are able to verbally align in a standard H-H dialogue whereas the agent usually does not (yet) have the ability to adopt the expressions primed by the human user. Nevertheless, this reason may not be the only one to justify differences as pointed out by other researchers [4]. In particular, they outline the fact that the strength of alignment may be dependent on the human's belief about the communicative capability of the machine. As such, alignment might be stronger from a human participant who believes that it might improve communication and understanding.

First, we investigate common points and differences between H-H and H-A interactions regarding the interactive alignment process at a coarse-grained level using the previously described off-line speaker-independent measures. Second, we take a finer-grained perspective by analysing individual DPs and their utterances using off-line and on-line speaker-dependent measures. We investigate whether the behaviours of the measures are similar or different with respect to the interactive alignment process and the self-repetition behaviours. In particular, we analyse the interactions regarding the variety, complexity, stability, and strength, investigating four hypotheses:

- **H1, H-H interactions *vs* H-A interactions:** we expect the lexical alignment process to be more important in H-A interactions than in H-H interactions. The main reason is the belief of the human partner about the limited communicative capabilities of the agent. As such, we expect a more stable and a stronger lexical alignment process in H-A interactions than in H-H interactions. To investigate this hypothesis, we extend the comparison of the nego. H-H and nego. H-A corpora initiated in previous work [23].

- **H2, variations in terms of lexical alignment among broad categories of H-A interactions:** we expect that the scenarii behind the H-A interactions will impact the lexical alignment process. By applying the same measures to, on the one hand, information-providing dialogues (HAI Alice corpus) and to, on the other hand, negotiation dialogues (Nego. H-A corpus), we will be able to characterise whether the verbal alignment process is similar between these two different scenarii. We can see at least two factors that may enter into consideration: (i) the difference in the task underlying the dialogues (information-providing *vs* negotiation), and (ii) the constraints of the Woz (resp., loosely constrained *vs* strictly constrained).

- **H3, differences between dialogue participants:**

  - **H3-a, shared vocabulary:** We expect that the measures will show a difference between participants in the same H-A interaction regarding the shared expressions they use and initiate. In particular, we expect the initiator to be more likely to be the agent. The main

Table 4: Average values and the corresponding standard deviation of the dialogue-level and speaker-independent measures on the interaction corpora. Voc.=Vocabulary Overlap.

	HAI Alice	Nego. H-A	Nego. H-H
EV	$.096 \pm .014$	$.134 \pm .022$	$.118 \pm .023$
ER	$.336 \pm .048$	$.416 \pm .086$	$.436 \pm .107$
ENTR	$1.065 \pm .258$	$1.039 \pm .201$	$.893 \pm .237$
L	$1.359 \pm .127$	$1.339 \pm .101$	$1.268 \pm .108$
LMAX	$4.563 \pm .892$	4.811 ± 1.230	4.179 ± 1.372
Voc.	$.274 \pm .042$	$.322 \pm .060$	$.316 \pm .073$

reason is that the human participant is fully able to verbally align while the virtual agent is constrained in its linguistic productions.

– **H3-b, self-repetitions:** We expect that the measures will show a difference between participants in the same H-A interaction regarding their self-repetitions. We envision three main factors that may impact self-consistency: (i) the constraints of the Woz in the production of its utterances, (ii) the natural adaptation of the human participant that may increase self-repetition and self-consistency in order to improve communication with a supposedly limited virtual agent, and (iii) the constraints of the task underlying the dialogue.

4.3 Results: Interactive Lexical Alignment Process

In order to investigate hypothesis **H1**, we contrast H-H negotiations (Nego. H-H) with H-A negotiations (Nego. H-A) given the EV, ER, ENTR, L, LMAX and voc. overlap measures (see Tables 4 and 5). This extends previous work [23] which did not consider the complexity and stability of the verbal alignment process but was limited to strength, variety and orientation. For these indicators, statistical significance was tested by a Wilcoxon rank sum test with continuity correction (reported in Tables 5 and 9). The most striking difference lies in the indicators related to the nature of the shared expression instances, namely, ENTR, L and LMAX, which can be respectively related to the complexity and stability of the lexical alignment process. Indeed, ENTR, L and LMAX are consistently lower for the H-H negotiations compared to the H-A negotiations. All in all, this shows that H-A negotiations exhibit a more complex interactive lexical alignment process than H-H negotiations in that they involve a wider range of shared expression instances in terms of their length. H-A negotiations exhibit a more stable interactive lexical alignment process than H-H negotiations in that they involve longer instances. In other words, H-H negotiations involve shorter shared expression instances with less variation in sizes.

To investigate **H2** about the variability of the lexical alignment process in H-A interactions, we contrasted the two H-A corpora. On the one hand we studied information-providing dialogues about *Alice's Adventures in Wonder-*

Table 5: Contrastive comparisons between the interaction corpora based on the dialogue-level and speaker-independent measures. Statistical difference was tested by a Wilcoxon rank sum test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, "=" means that the difference is not statistically significant.

Comparison	EV	ER	ENTR	L	LMAX	Voc.
Nego. H-H / Nego. H-A	***(<)	=	***(<)	***(<)	***(<)	=
Nego. H-A / HAI Alice	***(>)	***(>)	=	=	=	***(>)

land (HAI Alice corpus). On the other hand, we studied negotiation of a definite set of objects (Nego. H-A corpus). Overall, H-A negotiations differ from H-A information-providing dialogues in terms of EV, ER and voc. overlap, which show clear and significant differences. These measures are higher in the negotiation corpus than in the information-providing corpus, indicating that negotiation leads to more lexical alignment at the level of shared expressions. In negotiation, DPs constitute a richer expression lexicon and dedicate more tokens to the repetition of shared expressions. This is corroborated by the voc. overlap indicator. On the contrary, H-A corpora appear homogeneous with regards to complexity (ENTR) and stability (L, LMAX). Measurable differences in the variety and the strength of the lexical alignment process may indicate that this process is dependent on the type of underlying dialogue task. These underlying tasks obviously impact the vocabulary and thus the patterns of repetition of shared expressions found in utterances. In the Nego. H-A corpus, for instance, we mainly find expressions of preferences and propositions (e.g., "I like the lamps and the records") or negotiation-specific expressions (e.g., "it's a pleasure doing business with you"). In the information-providing dialogues, we mainly observe shared expressions related to the task (e.g., "would like to know") and the domain (e.g., referring expressions such as "Alice", "the Cheshire cat", "the white rabbit"). However, a deeper analysis would be required to confirm to what extent the lexical alignment process is affected by the underlying task (e.g., with corpora on more varied tasks and domains). Notably, other factors may also impact this process such as the constraints of the Woz (one being strictly constrained in the H-A Nego. corpus and the other one being loosely constrained in the HAI Alice corpus) or the interaction style of the human operator controlling the Woz system.

We investigated **H3-a** by taking a closer look at the difference between DPs in the same dialogue at two different levels: globally over the entire dialogue, and at the level of individual utterances. Statistical significance was tested by a Wilcoxon signed rank test with continuity correction (reported in Tables 6, 7 and 10).

– *Dialogue-level Perspective.* We investigated lexical alignment at the level of shared expressions by having a closer look at each speaker in a dialogue in terms of initiated expressions (IE_S), expression repetition (ER_S), the number of tokens produced ($Tokens_S$) and the relative shared vocabulary (SV_S). Table 6 reports the observed statistical differences based on the dialogue-level and

Table 6: (A)symmetry between speakers on the interaction corpora based on the dialogue-level and speaker-dependent measures. Statistical difference was tested by a Wilcoxon signed rank test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, "=" means that the difference is not statistically significant. A is the software agent/Woz, H is the human. SV=Shared Vocabulary.

Corpus	IE_S	ER_S	$Tokens_S$	SV_S	SEV_S	SER_S	$SENTR_S$	SL_S	$SLMAX_S$
HAI Alice	*** A > H	=	=	=	*** A < H	=	** A > H	*** A > H	** A > H
Nego. H-A	*** A > H	*** A < H	=	=	*** A < H	*** A < H	* A < H	=	*** A ≠ H
Nego. H-H	=	=	=	=	=	=	=	=	=

Note: Relative SV for S_1 is computed as follows: $SV_{S_1} = \frac{\#(Tokens_{S_1} \cap Tokens_{S_2})}{\#(Tokens_{S_1})}$

speaker-dependent measures. First, it should be noted that DPs play a symmetrical role with regards to the number of tokens produced and the relative shared vocabulary for all corpora. That is, DPs produce a similar number of tokens in a dialogue; and they share a similar amount of their vocabulary. For H-A interactions, this means that the Woz and the human tend to produce the same number of tokens, and that the Woz does not share more of its vocabulary than the human participants, and conversely. This observation goes further for the H-H corpus since both speakers play a symmetrical role at the level of shared expressions. First, they initiate a similar amount of expressions (approx. 50%). Next, they dedicate the same amount of tokens to the repetition of shared expressions.

On the contrary, all the H-A corpora studied exhibit an asymmetrical role at the level of shared expressions between the Woz and the human participant, thus confirming and extending the results previously presented in [23]. In both H-A corpora, the Woz initiates more shared expressions than the human participant. This is mainly explained by the limited communicative capabilities of the Woz who cannot straightforwardly adopt lexical patterns from the human (either due to technical limitations or data collection scenario constraints). Then, in the Nego. H-A corpus, the human participant dedicates more tokens to the repetition of established shared expressions than the Woz (while this is not the case for the HAI Alice corpus). It turns out that a closer look at the utterance-level perspective is needed to better understand the repetition of shared expression in the HAI Alice corpus.

The proposed speaker-dependent measures thus make it possible to quantify a difference in nature in lexical alignment in terms of (a)symmetry between H-H and H-A interactions, which does not appear when looking at classical measures such as the number of tokens or the vocabulary overlap.

– *Utterance-level Perspective.* We took a finer grained approach by looking at the lexical alignment behaviour at the individual utterance level (and taking into account the previous utterances in the dialogue history). We investigated differences between speakers in terms of lexical alignment at the level of

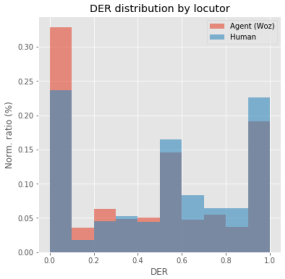
shared expressions by having a closer look at the dynamic utterance-level measure DER_u, a context-aware score for an utterance that takes into account the previous dialogue history. Table 7 reports the statistical differences observed in the corpora based on the utterance-level and speaker-dependent measures. The situation between speakers at the utterance level is similar and symmetrical in the H-H corpus regarding those measures. In other words, DPs' utterances reuse the shared expressions in a quantitatively similar fashion. This confirms the observation previously shown at the dialogue level with the ER_S measure.

The situation is similar in the Nego. H-A corpus in that the asymmetry observed at the dialogue level is also observed at the utterance level. Put differently, the human participant reuses the shared expressions more than the agent participant. Going into more details, Figure 2 shows the distribution of DER_u per speaker for the Nego. H-A corpus. First, it shows two extreme peaks. One around DER_u=0.0 and one around DER_u=1.0. They correspond to specific utterances with regard to shared expressions, resp. utterances that do not reuse shared expressions (*disconnected-from-other* utterances) and utterances formed entirely of shared expressions (*plainly-connected-to-other* utterances). A clear difference between speakers can be observed in the Nego. H-A corpus. The agent appears to produce proportionally more disconnected-from-other utterances while the humans produce proportionally more plainly-connected-to-other utterances. Notably, those two extreme peaks can be observed in the H-H Nego. corpus as well. However, no significant difference in proportional use is observed between the two human participants.

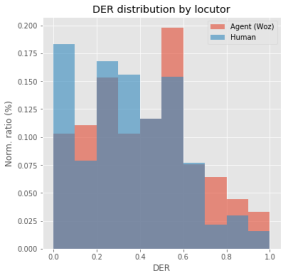
A closer look at the results for the HAI Alice corpus shows a paradoxical situation. While the behaviour at the dialogue level (ER_S) is symmetrical between DPs, it is asymmetrical at the utterance level where the agent appears to repeat more shared expressions. This paradox disappears once we analyse the results at the utterance level. The explanation lies in the fact that looking at the utterance level reduces the importance of the number of tokens dedicated to shared expression repetition. The distribution in terms of DER_u shows a peak of utterances with DER_u=0.0 for the human, which is stronger than the one observed for the agent (see Figure 2). While this type of utterances represents approx. 10% of the ones produced by the agent, it represents more than 17% of the utterances produced by the human. It turns out that there is a major difference in the size of utterances produced by the human participant regarding their DER_u scores. Globally, the size of these utterances is in average around 9.75 tokens (median=8, std=7.67, min=1, max=60). However, utterances with DER_u=0 are significantly shorter (mean=3.31, median=2, std=2.81, min=1, max=17) while utterances with $\text{DER}_u > 0$ are longer (mean=11.10, median=9, std=7.68, min=1, max=60). While ER_s offers a global point of view by considering tokens over the entire dialogue, DER_u takes a finer grained perspective at the level of utterances. In the latter perspective, tokens play a less important role in that short utterances (e.g., 1 or 2 tokens) are as important as longer ones (e.g., 60 tokens). To conclude, the utterance perspective uncovers a difference between speakers that is hidden at the dialogue level, namely: the human participant produces more utterances that do not reuse shared expressions (i.e.,

Table 7: (A)symmetry between speakers on the interaction corpora based on the utterance-level and speaker-dependent measures. Statistical difference was tested by a Wilcoxon rank sum test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, "=" means that the difference is not statistically significant. A is the software agent/Woz, H is the human.

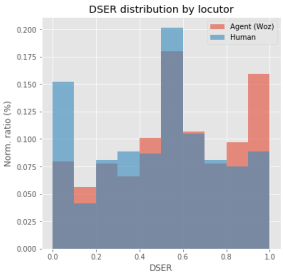
Corpus	DER_u	$DSER_u$
HAI Alice	*** $A > H$	*** $A > H$
Nego. H-A	*** $A < H$	*** $A < H$
Nego. H-H	=	=



DER_u per speaker for the Nego. H-A corpus.



DER_u per speaker for the HAI Alice corpus.



$DSER_u$ per speaker for the HAI Alice corpus.

Figure 2: Distribution of DER_u and $DSER_u$ per speaker in the H-A corpora.

DER_u=0.0) than the agent. These utterances are shorter than other human utterances.

To sum up, the utterance perspective reveals patterns of repetition of shared expressions that are different between the H-A corpora (illustrated by the distributions depicted in Figure 2), and uncovers specific patterns of (non-)repetition of shared expressions. We discussed two main types of utterances. *Disconnected-from-other* utterances do not reuse any shared expressions. They potentially cover a wide range of utterances such as short answers, utterances aiming at priming new expressions, and idiosyncratic utterances. They are mainly produced by the agent in the Nego. H-A corpus (> 31% of the utterances) while they are mainly produced by the human in the HAI Alice corpus (>17% of the utterances). On the other hand, *plainly-connected-to-other* utterances are composed entirely of shared expressions. They include verbatim repetition of full utterances but also utterances composed of many shared expressions. They represent more than 21% of utterances produced by the human participant and more than 18% of the ones produced by the agent in the H-A Nego. corpus. However, they are very scarce in the HAI Alice corpus (less than 4% of the utterances).

4.4 Results : Self-Repetition Behaviours (H3-b)

We investigated the self-repetition behaviour of each speaker in the interaction corpora (**H3-b**). Here, we are no longer interested in lexical patterns that are shared between DPs. Instead, we take a closer look at the self-consistency of speakers by only considering the sequence of their utterances in the whole interaction, without taking the sequence of the other speaker into account. Statistical significance was tested by a Wilcoxon signed rank test with continuity correction (reported in Tables 6, 7, 11, and 12).

– *Dialogue-level Perspective.* Table 6 reports the observed statistical differences based on the dialogue-level and speaker-dependent measures. First, DPs in the H-H corpus present a remarkable similarity regarding their self-repetition behaviour. They have a similar variety of self-expressions (SEV_S) and dedicate a similar amount of tokens to the repetition of self-expressions (SER_S). They also present a similar complexity ($SENTR_S$) and stability (SL_S, $LMAX_S$) of self-repetition behaviours. Self-repetitions of both DPs in dialogue exhibit a similar variety of sizes (in number of tokens) as well as similar average and maximum lengths.

The situation is more contrasted for H-A interactions and involves common points as well as notable differences. First, the human participant displays a greater variety of self-repetitions than the Woz (SEV_S) for all H-A corpora. In the Nego. H-A corpus, the human participant even dedicates more tokens to the repetition of self-expression (SER_S), i.e. the human participant repeats himself more than the agent. This may come as a surprise given that the Woz is constrained to a limited set of utterances while the human participant is not. Hence, we would expect more tokens to be dedicated to self-repetitions by the Woz than by the human participant. However, this should be seen in conjunction with the results presented in Section 4.3 for the interactive lexical alignment process, where it was observed that the human dedicated more tokens to the repetition of shared expressions (ER_S). A plausible explanation is that the human participant adopts more Woz-initiated expressions and repeats more shared expressions and self-expressions (which intersect). Thus, the human participant ends up self-repeating more than the Woz by verbally aligning and self-repeating more (probably to improve communication).

In the HAI Alice corpus, the self-repetition behaviour of the Woz is more complex (it involves a greater variety of lexical pattern sizes) as well as more stable (it involves larger lexical patterns) than the behaviour of the human participant. Notably, the human participant displays a greater variety of self-expressions (SEV_S) that however are less complex ($SENTR_S$) and less stable (SL_S and $SLMAX_S$). Nevertheless, these differences observed in the HAI Alice corpus are not observed for the Nego. H-A corpus. In fact, the tendency may be inverted with the human participant producing more complex self-repetition patterns ($SENTR_S$). But this tendency is not strongly marked. On the other hand, there is not a clear difference in stability between the DPs in the Nego. H-A corpus.

The most striking observation thus lies in the similarity of the self-repetition

behaviours between DPs in the H-H corpus. Our analysis did not yield any difference between DPs in the same dialogue in terms of variety, amount of tokens dedicated to self-repetitions, complexity and stability of the self-repetition behaviours. H-A corpora, on the contrary, exhibit self-repetition behaviours that are heterogeneous between the DPs (meaning that we have observed measurable differences with regards to speaker-dependent measures related to this behaviour). In both H-A corpora, the human participant exhibits a greater variety of self-repetition patterns. However, we did not observe a clear tendency regarding the strength of repetition of these patterns, the complexity and the stability of the behaviour. At the dialogue level, the Nego. H-A corpus shows a clear difference in repetition, with the human participant repeating more than the Woz. However, this observation does not hold for the HAI Alice corpus where the amount of repetition is similar between DPs. Finally, the Woz of the HAI Alice corpus involves more complex self-repetition patterns and a more stable process while there is no marked difference in the Nego. H-A corpus.

–*Utterance-level Perspective.* We took a finer grained approach by looking at the self-repetition behaviours at the individual utterance level. We investigated differences between speakers by scrutinizing the dynamic utterance-level measure DSER_u, a context-aware score of self-repetition for an utterance that takes the previous dialogue history into account. Table 7 reports the statistical differences observed in the corpora based on the utterance-level and speaker-dependent measures. The situation between speakers at the utterance level is similar in the H-H corpus regarding this measure. In other words, DPs' utterances reuse the self-expressions in a quantitatively similar fashion. This confirms the observation previously reported at the dialogue level with the SER_S measure. More specifically, the distribution of DSER_u shows a very clear peak of utterances with a value of 1.0 (including verbatim repetition of a previous utterance). 28.35% of the utterances produced have such a value.

The Nego. H-A corpus displays a certain heterogeneity at the utterance level, confirming that observed at the dialogue level. In other words, the human participant tends to repeat himself more than the agent participant. In detail, the distribution of DSER_u shows two very clear extreme peaks for the agent. The first one is around the 1.0 value and represents 27.71% of all agent utterances. The second one is around the 0.0 value and represents 27.51% of all agent utterances. Among agent utterances, there is therefore the same amount of *new* utterances (i.e., utterances which do not reuse previous self-expression) and *old* utterances (i.e., utterances which heavily reuse previous self-expressions). On the other hand, the distribution of DSER_u shows one very clear extreme peak at 1.0 for the human participant, accompanied by a less marked peak at 0.0. The peak around 1.0 represents 31.66% of all human utterances while the peak around 0.0 represents 20.33% of all human utterances.

An analysis of self-repetitions at the utterance level in the HAI Alice corpus shows a heterogeneity that does not appear at the dialogue level. That is, the agent tends to repeat itself more than the humans do. The distribution in terms of DSER_u shows a peak of utterances with DSER_u=0.0 for the human participant which is stronger than the one observed for the agent (see Figure 2).

While this type of utterance represents around 7.5% of the ones produced by the agent, it represents around 15% of utterances produced by the human (i.e., twice as much). It also shows a peak of utterances with DSER_u=1.0 (including verbatim repetition of a previous utterance) for the agent which is stronger than the one observed for the human. While this type of utterance represents around 8% of the ones produced by the human, it represents more than 15% of utterances produced by the agent. Once again, the fact that the asymmetry appears at the utterance-level and not at the dialogue-level can be explained by smaller utterances. In other words, these differences are not observed at the dialogue level but appear at the level of utterances. Regarding the human participant, the utterances with DSER_u=0.0 are smaller with on average 5.0 tokens (median=4, std=4.19, min=1, max=18) while utterances with $\text{DSER}_u > 0.0$ are on average 10.57 tokens (median=9, std=7.83, min=1, max=60). Regarding the agent, its utterances with $\text{DSER}_u = 1.0$ are on average 7.08 tokens (median=7, std=2.59, min=1, max=16) while its utterances with $\text{DSER}_u < 1.0$ are on average 10.66 tokens (median=9, std=6.24, min=1, max=50).

This analysis from the utterance-level perspective reveals clear self-repetition tendencies at two extremes. We have observed utterances which do not repeat a previous lexical pattern (i.e., utterances u where $\text{DSER}_u = 0.0$). They are produced by both the agent and the human in the Nego. H-A corpus, and by the human in the HAI Alice corpus (but much less so by the agent). We have observed utterances which largely repeat previous lexical patterns (i.e., utterances u where $\text{DSER}_u = 1.0$). They are produced by both the agent and the human in the Nego. H-A corpus, by the agent in the HAI Alice corpus (but not the human) and by the human participants in the Nego. H-H corpus.

All in all, two main factors may explain the differences between the two H-A corpora: the constraints of the Woz in the production of utterances, and the tasks and domains underlying the dialogue. The constraints of the Woz are a major source of self-repetitions. However, we have observed, somewhat surprisingly, that the most constrained Woz actually tends to self-repeat less than its human partner. The main explanation may be the effort of the human to verbally align and self-repeat, most likely to ensure self-consistency and thus improve communication and understanding in a negotiation with a virtual partner that is supposedly limited in its communicative capabilities. Another reason behind the differences in self-repetition in the H-A corpora lies in the tasks and domains. While the main reason for self-repetition is to improve understanding in the H-A negotiations, a significant source of self-repetition for the agent is answering and suggesting topics in the information-providing dialogues. Indeed, among the lexical patterns most frequently used by the agent, we find patterns such as "is there something you would like to know about ...", "is there anything you would like to know about ...", "is there something else you would like to know?", "sorry, I do not know that", and "sorry, I don't know the answer". In other words, H-A negotiations may push the human participant to align and self-repeat more to make sure that the agent understands his (counter-)propositions; while an information-providing agent may self-repeat more in its attempt to provide answers and suggest new topics to keep the interaction en-

gaging. Last but not least, part of the self-repetition behaviour may also be explained by the specific interaction style of the Woz operator. Exploring this possibility would require H-A interaction corpora with a larger variety of human Woz operators than the ones considered in this work.

5 Towards an Adaptive Natural Language Generation Process based on Measures of Repetition

This section addresses the challenge of integrating lexical alignment capabilities in agents interacting with humans. Our long-term objective is to provide a virtual agent with the ability to detect the alignment behaviour of its human interlocutor, as well as the ability to align with the user to enhance its credibility, to increase interaction naturalness, and to maintain the user's engagement [66]. The integration of lexical alignment capabilities in a virtual agent has many valuable advantages. First, it is a natural source of variation in dialogue that has the potential of allowing a virtual agent to produce richer linguistic content. Next, lexical alignment is a kind of on-the-fly adaptation that does not *a priori* require extensive user profiling. It is a way for DPs to share a common language and has thus the potential to positively impact a collaboration underlying the dialogue. In addition, the alignment process is seen as a sign of empathy and interpersonal skill [49, 55], and can therefore be considered a behaviour able to improve the perception of the agent's social competence. Finally, interactive alignment has the computational advantage of constraining the linguistic production of the DPs. One consequence of successful alignment at several levels is a certain repetitiveness in dialogue and the development of a lexicon of fixed expressions established during the dialogue [50]. This repetitiveness can be fruitfully exploited by a dialogue model of a software agent. We outline here the properties of the framework presented in this article with regard to this goal, and we develop the first steps towards an adaptive natural language generation process based on the previously described lexical alignment measures.

5.1 Designing an Adaptive Natural Language Generation Module based on Measures of Repetition

5.1.1 Lexical Alignment Capabilities for an Agent

Some important elements have to be kept in mind when designing lexical alignment capabilities for a virtual agent. First, the task and domain underlying the H-A interaction have a major impact on the lexical alignment capabilities. This has been outlined by our quantitative study which points out some notable differences between information-providing and negotiation H-A dialogues. A notable aspect is the task-oriented or non-task-oriented nature of the dialogue. As pointed out by researchers [35], this aspect may lead DPs in H-H

dialogue from converging (mainly in task-oriented dialogue) to diverging (e.g., ordinary conversations). As such, the virtual agent may need to display either a convergent behaviour or a divergent behaviour depending on the task.

Secondly, lexical alignment involves many levels from lexical, syntactic to higher levels such as politeness [18] or appreciation [14]. Consequently, many kinds of lexical alignment can be envisioned. In other words, integrating lexical alignment capabilities directly impacts various levels of the dialogue model of an agent. In particular, this directly affects the macro-planning (i.e., *what to say*) and the micro-planning (i.e., *how to say*) of the natural language generation process. Macro-planning is directly concerned by high-level lexical alignment at the decision level of the dialogue model. It affects the dialogue act selection. For instance, this is the case when the agent is designed to align on appreciations about paintings (e.g., "Me neither, I don't like Picasso") [14]. On the other hand, micro-planning is directly concerned by low-level lexical alignment at the surface form realization of a dialogue act. Here, the dialogue act has been selected and the goal is to generate a linguistic form that aligns (or not) at the lexical and syntactic levels with the user.

Lastly, a crucial aspect for a virtual agent is to ensure self-consistency of its linguistic production while at the same time verbally (un)aligning with its human partner. As we have seen in the previous corpus study, different configurations may emerge depending on the task and the communicative capabilities of the agent. An interesting perspective lies in the study of the connection between self-consistency and the lexical alignment process and its consequences on the natural language generation process.

5.1.2 Use of the Measures of Repetition

Our framework includes many useful features to contribute to the integration of engaging lexical alignment capabilities in the dialogue model of a virtual agent. Our model considers both the lexical alignment process and the self-repetition behaviour of the DPs at the level of (shared) lexical patterns. It relies on lexicons that keep track of lexical patterns and some valuable pattern features. These lexicons are automatically built for an entire dialogue or incrementally for a given dialogue history. Our model can usefully contribute to data-driven approaches to real-time lexical adaptation at the lexical level (and at the syntactic level to the extent of patterns). First, it allows the design of online context-aware indicators based on the lexicons and the dialogue transcript (e.g., DER_u and DSER_u). These scores can be leveraged to determine the lexical alignment and the self-repetition behaviours of both the system and the human partner at different granularities (last utterance, dialogue phases, entire dialogues). An interesting perspective lies in exploiting lexicons across many interactions between an agent and a given human partner. This could contribute to fostering the emergence of a common and personalised language between the agent and the user. In other words, lexicons created during a dialogue can be saved and reused for future interactions.

Our framework can be useful for the generation of surface forms of system

utterances that allows both lexical alignment and self-consistency. The generation of a surface form of an utterance can be viewed from the perspective of the lexicons in that the surface form of an utterance can be seen as exploiting a lexicon (or not). Exploiting a lexicon means that a (shared) lexical pattern is reused in the surface form. Depending on the lexicon that is exploited, it may indicate an attempt to verbally align, to be more self-consistent, or both. An utterance can also be seen as contributing to and expanding the lexicons (e.g., by adopting a lexical pattern primed by the human partner). The DER_u and DSER_u scores quantify the exploitation of the lexicon in individual utterances. A promising perspective lies in building finer-grained utterance scores that take into account the pattern features available in the lexicon. As such, we can envision scores quantifying the exploitation of the lexicon and which take into account the recency of use and the frequency of lexical patterns in order to model the decay of a shared linguistic structure.

All in all, the corpus study presented in Section 4 has pointed out some differences between H-H interaction and H-A interaction. Namely, the H-H interactions in our corpora are characterised by similar, symmetrical behaviours regarding the lexical alignment process and the self-repetition behaviours. On the other hand, H-A interactions are characterised by heterogeneous, asymmetrical behaviours. However, in order to validate these differences, this study should also be done on real H-A interactions – without a wizard of oz human operator. This would identify some potential bias produced by the operators' specific behaviours, for example in the choice among the scripted answers. Once this kind of bias is excluded, we believe that our framework can usefully contribute to more similar communicative behaviour in H-A interaction both at the lexical alignment and self-repetition behaviour levels.

5.2 A First Approach: Combining Alternative Utterance Generation and Measures of Repetition

In this section, we informally sketch out a first approach to integrate the lexical alignment measures into a Natural Language Generation (NLG) process. The presentation of this system is intended for illustrative purposes. Building a verbally aligning NLG module is challenging. The main reason is that it requires taking into account the context of the dialogue (i.e., the dialogue history). This dramatically increases the variety of surface forms a given system utterance can take. A standard data-driven approach to building such a process would require a dataset including the context (at least the previous user utterance) and multiple paraphrases of a system utterance with varying degrees of lexical alignment. Such corpora are very rare (see [24] for a dataset for the public transport information domain in English); limited to very specific tasks and domains, and to a very short preceding context; and arguably very hard to obtain because of the huge number of possible preceding dialogue histories and possible surface forms for the system responses.

To address this challenge, our proposition is a combination of an alternative utterance generation process with a selection module based on lexical alignment

measures. Whereas a standard NLG process relies on a dialogue act representation, the idea here is to modify an existing surface form of a system utterance assumed to be unaligned in order to obtain a set of alternatives with varying degrees of lexical alignment at the lexical level.

The sketched process is an instance of the overgenerate and rank paradigm [47]. The overgeneration is performed by the alternative utterance generation module. The ranking of alternative utterances is performed by using a context-aware score (such as DER_u and DSER_u). These scores take into account the dialogue history and rely on the automatically built lexicons proposed in our framework. The higher the score, the more the alternative utterance is verbally aligning at the lexical level. Notably, this process makes it possible to select converging alternative utterances as well as diverging alternatives. This is a key advantage when considering the dynamic lexical alignment strategy that may be required by an engaging virtual agent (discussed in the previous section).

Here follows an example[4] of the output of the alternative utterance generation process associated with the use of lexical alignment measures to select convergent and divergent alternatives (U is the user, S is the system, changes in S2 and S3 have been emphasised):

U: Can you tell me something about the character, the white rabbit?

S1: The rabbit is being chased by Alice *[standard]*

S2: The *white* rabbit is being chased by Alice *[convergent]*

S3: The *bunny* is being chased by Alice *[divergent]*

This example is an instance of lexical alignment on referring expressions, often called entrainment [6]. S1 represents the standard system response without any adaptation. In other words, it is the raw output of the dialogue system. S1 is the starting point of the adaptive NLG process that we have sketched in this section. The alternative utterance generation is applied on S1 to generate a pool of alternative utterances with varying degrees of lexical alignment at the lexical level. S2 and S3 represent two possible outputs of this process. In S2, the adjective "white" has been added in front of "rabbit" to be closer to utterance U. On the contrary, the noun "rabbit" has been substituted by the synonym "bunny" in utterance S3. Then, the lexical alignment utterance scores that we propose can be leveraged to re-rank this pool of alternative utterances from the most convergent considering the dialogue history (here, utterance U) to the most divergent. The result of this process on this toy example would return S2 as the most convergent because it shares the lexical pattern "the white rabbit" with utterance U. On the other hand, S3 would be the most divergent because it only shares the token "the" with utterance U.

[4]Utterances U and S1 are taken from the HAI Alice corpus.

6 Conclusions

We have proposed a new framework to quantify both the lexical alignment interactive process and the self-repetition behaviours of DPs from dyadic dialogue transcripts. The lexical alignment process is a particular type of on-the-fly linguistic adaptation leading DPs to share a common language to improve understanding, collaboration and social connection with a conversational partner. On the other hand, the self-repetition behaviour is directly connected to the self-consistency of the linguistic productions of DPs.

The framework implements easily computable measures efficiently computed via a sequential pattern mining approach. These measures allow the characterisation of the nature of the studied processes by addressing various informative aspects such as their variety, complexity, stability, and strength; and by characterising the individual behaviours of DPs. We have applied these measures on a corpus study contrasting H-H dialogues and H-A dialogues. This has exposed major differences in the lexical alignment process and the self-repetition behaviours between H-H and H-A interactions as well as between H-A corpora. Among other things, we have shown through this framework that H-H dialogues involve a lexical alignment process that is more flexible than the one observed in H-A dialogues. H-H dialogues are characterised by similar DP behaviours with regard to lexical alignment and self-repetition while H-A dialogues are characterised by heterogeneous behaviours. Lastly, we have revealed measurable differences between H-A corpora in the nature of the studied processes and have discussed possible explanations for these differences. Finally, this framework includes measures that make it possible to investigate the contribution to alignment and self-repetition of a single utterance with regard to a given dialogue history. A notable consequence is that the proposed framework is not limited to corpus studies. Indeed, it has been designed to be embeddable in real-time interactive dialogue systems. To this end, we have outlined and discussed the properties of the framework with regard to the challenge of integrating lexical alignment capabilities in agents interacting with humans.

This work points towards many avenues for future research. A first direction is to deepen the corpus study presented here. This article has focused on alignment of communicative behaviour in task-oriented dialogues. A valuable perspective is to better understand the lexical alignment process in H-A interactions by comparing task-oriented and non-task-oriented H-A interactions at the level of convergence/divergence of communicative behaviour. Next, it is worth noticing that lexical alignment and self-consistency are not always compatible. As such, linguistic adaptation involves to some extent a trade-off between adopting a common language with a conversational partner and ensuring self-consistency. An interesting perspective aims at studying the connections between self-consistency and the interactive lexical alignment process as well as its consequences on the NLG process of a software agent.

As developed in Section 5, another direction is to aim at better lexical alignment capabilities in software agents. One of the future aims is to go deeper into the analysis of repetitions. We want to focus more on other-repetitions as

described by [58] and study how local this phenomenon is. Other-repetitions can be defined as specific shared vocabulary (one dialogue participant repeats a word or segment previously uttered by another dialogue participant) which can have several pragmatic functions (taking into account what has been just said, expressing an emotion, etc). The measures provided in the paper were a necessary first step for studying such a phenomenon, however we also need to consider the distance between the initiated segment and its repetition and analyse how this distance can impact the pragmatic effect of the repetition.

Another promising perspective lies in the enrichment of the automatically built lexicons with additional information about the lexical patterns. To this end, a wide range of additional information can be considered such as, for instance, information from sentiment analysis [16, 40], linguistic style [46] or functions behind repetitions [59, 52].

A Supplementary Materials

Table 8 presents the abbreviations and acronyms used in this article.

Table 9 presents the statistical values and p-values of the contrastive comparisons between the interaction corpora based on the dialogue-level and speaker-independent measures reported in Table 5.

Table 10 presents the statistical values and p-values of the (a)symmetry tests between speakers' verbal alignment behaviours on the interaction corpora based on the dialogue-level and speaker-dependent measures (reported in Table 6).

Table 11 reports the statistical values and p-values of the (a)symmetry tests between speakers' self-repetition behaviours on the interaction corpora based on the dialogue-level and speaker-dependent measures (reported in Table 6).

Table 12 presents the statistics and p-values for the (a)symmetry test between speakers on the interaction corpora based on the utterance-level and speaker-dependent measures (reported in Table 7).

acknowledgements

This work was supported by the European project H2020 ARIA-VALUSPA. We warmly thank Jonathan Gratch and David DeVault for sharing the negotiation corpora, and Catherine Pelachaud for valuable and enriching discussions. We also thank the reviewers of a previous version of this article for their relevant comments.

References

[1] Bertrand, R., Ferré, G., Guardiola, M., et al.: French face-to-face interaction: repetition as a multimodal resource. Coverbal Synchrony in Human-Machine Interaction p. 141 (2013)

Table 8: Table of Abbreviations and Acronyms

EV	Expression Variety
ER	Expression Repetition
ENTR	Entropy
L	Average length in tokens of the shared expression instances
LMAX	Maximum length in tokens of the shared expression instances
IE_S	Initiated Expression for locutor S
ER_S	Expression Repetition for locutor S
DER_U	Dynamic Expression Repetition at utterance level
$DSER_S$	Dynamic Self-Expression Repetition
SEV_S	Self-Expression Variety for locutor S
SER_S	Self-Expression Repetition for locutor S
$SENTR_S$	Shannon entropy of the length in tokens of the self-repetitions by locutor S
SL_S	Average length in tokens of the self-repetitions by S
$SLMAX_S$	Maximum length in tokens of the self-repetitions by S
VO	Vocabulary Overlaps
SV	Shared Vocabulary
H-H	Human-Human
H-A	Human-Agent
Nego. H-H	Corpus of Human-Human Negotiation Interactions
Nego. H-A	Corpus of Human-Agent Negotiation Interactions
DP	Dialogue Participant
CRQA	Cross-Recurrence Quantification Analysis
NLG	Natural Language Generation

[2] Bickmore, T., Schulman, D., Yin, L.: Maintaining engagement in long-term interventions with relational agents. Applied Artificial Intelligence **24**(6), 648–666 (2010)

[3] Bickmore, T.W., Puskar, K., Schlenk, E.A., Pfeifer, L.M., Sereika, S.M.: Maintaining reality: Relational agents for antipsychotic medication adherence. Interacting with Computers **22**(4), 276–288 (2010)

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

[5] Brennan, S., Schuhmann, K., Batres, K.: Entrainment on the move and in the lab: The walking around corpus. In: CogSci (2013)

[6] Brennan, S.E.: Lexical entrainment in spontaneous dialog. Proceedings of International Symposium on Spoken Dialogue (ISSD) **96**, 41–44 (1996)

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

Table 9: Statistical value and p-value report of the contrastive comparisons between the interaction corpora based on the dialogue-level and speaker-independent measures. Statistical difference was tested by a Wilcoxon rank sum test. W is the test statistic. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. NHA = Nego. H-A, NHH = Nego. H-H, AC = HAI Alice, Voc. Ov. = Vocabulary Overlap

Comparison		EV	ER	ENTR	L	LMAX	Voc. Ov.
NHH *vs.* NHA		***		***	***	***	
	W	4056.5	7245	4082	3940.5	4734	6245.5
	p-value	$2.0.10^{-6}$	.13	$2.6.10^{-6}$	$6.4.10^{-7}$	$4.2.10^{-4}$	.7
NHA *vs.* AC		***	***				***
	W	2261	1966	1120	1126	1358	1869
	p-value	$4.0.10^{-8}$	$9.1.10^{-5}$	.6	.6	.5	$6.8.10^{-4}$

Table 10: Statistical value and p-value report of the (a)symmetry between speakers' verbal alignment behaviours on the interaction corpora based on the dialogue-level and speaker-dependent measures. Statistical difference was tested by a Wilcoxon signed rank test. V is the test statistic. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. SV_S=relative Shared Vocabulary for speaker S.

Corpus	IE_S	ER_S	$Tokens_S$	SV_S
Nego. H-A	***	***		
V	10161	545	4931	4796
p-value	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	.08	.08
HAI Alice	***			
V	136	68	83	57
p-value	3.1×10^{-5}	1.0	.5	.6
Nego. H-H				
V	1730.5	1789	1890	1667
p-value	.6	1.0	.6	.6

Table 11: Statistical value and p-value report of the (a)symmetry between speakers' self-repetition behaviours on the interaction corpora based on the dialogue-level and speaker-dependent measures. Statistical difference was tested by a Wilcoxon signed rank test. V is the test statistic. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Corpus	SEV_S	SER_S	$SENTR_S$	SL_S	$SLMAX_S$
Nego. H-A	***	***	*		***
V	1564	1560	4645	6440	6614.5
p-value	2.0×10^{-15}	1.9×10^{-15}	.02	.4	1.1×10^{-6}
HAI Alice	***		**	***	**
V	5	99	124	136	115.5
p-value	3.1×10^{-4}	.1	2.1×10^{-3}	3.1×10^{-5}	1.7×10^{-3}
Nego. H-H					
V	1703	1789	1744	1724	1173.5
p-value	.7	1.0	.9	.8	1.0

Table 12: Statistical value and p-value report for (a)symmetry between speakers on the interaction corpora based on the utterance-level and speaker-dependent measures. Statistical difference was tested by a Wilcoxon rank sum test. W is the test statistic. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Corpus	DER_u	$DSER_u$
Nego. H-A	***	***
W	31590000	31509000
p-value	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$
HAI Alice	***	***
W	150520	149910
p-value	2.0×10^{-5}	5.1×10^{-5}
Nego. H-H		
W	13189000	13166000
p-value	.4	.3

[8] Brennan, S.E., Hanna, J.E.: Partner-specific adaptation in dialog. Topics in Cognitive Science **1**(2), 274–291 (2009)

[9] Brown, P., Levinson, S.C.: Politeness: Some universals in language usage, vol. 4. Cambridge university press (1987)

[10] Buschmeier, H., Bergmann, K., Kopp, S.: Modelling and evaluation of lexical and syntactic alignment with a priming-based microplanner. In: Empirical Methods in Natural Language Generation, pp. 85–104. Springer (2010)

[11] Callejas, Z., Griol, D., López-Cózar, R.: Predicting user mental states in spoken dialogue systems. EURASIP Journal on Advances in Signal Processing **2011**(1), 6 (2011)

[12] Campano, S., Clavel, C., Pelachaud, C.: "I like this painting too": when an ECA shares appreciations to engage users. In: Proceedings of the 14th international joint conference on Autonomous agents and multiagent systems (2015)

[13] Campano, S., Durand, J., Clavel, C.: Comparative analysis of verbal alignment in human-human and human-agent interactions. In: International Conference on Language Resources and Evaluation (LREC), pp. 4415–4422 (2014)

[14] Campano, S., Langlet, C., Glas, N., Clavel, C., Pelachaud, C.: An ECA expressing appreciations. In: International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 962–967. IEEE (2015)

[15] Church, K.W.: Empirical estimates of adaptation: the chance of two noriegas is closer to p/2 than p 2. In: Proceedings of the 18th conference on Computational Linguistics, vol. 1, pp. 180–186. Association for Computational Linguistics (2000)

[16] Clavel, C., Callejas, Z.: Sentiment analysis: from opinion mining to human-agent interaction. IEEE Transactions on affective computing **7**(1), 74–93 (2016)

[17] Dahlbäck, N., Jönsson, A., Ahrenberg, L.: Wizard of Oz studies - why and how. Knowledge-based systems **6**(4), 258–266 (1993)

[18] De Jong, M., Theune, M., Hofs, D.: Politeness and alignment in dialogues with a virtual guide. In: Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems (AAMAS), pp. 207–214. International Foundation for Autonomous Agents and Multiagent Systems (2008)

[19] Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., Cohen, D.: Interpersonal synchrony: A survey of evaluation methods across disciplines. Affective Computing, IEEE Transactions on **3**(3), 349–365 (2012)

[20] DeVault, D., Mell, J., Gratch, J.: Toward natural turn-taking in a virtual human negotiation agent. In: AAAI Spring Symposium on Turn-taking and Coordination in Human-Machine Interaction. AAAI Press, Stanford, CA (2015)

[21] D'Mello, S., Graesser, A.: AutoTutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. ACM Trans. Interact. Intell. Syst. **2**(4), 1?–39 (2013)

[22] Dubuisson Duplessis, G., Charras, F., Letard, V., Ligozat, A.L., Rosset, S.: Utterance Retrieval based on Recurrent Surface Text Patterns. In: 39th European Conference on Information Retrieval (ECIR), pp. 199–211. Aberdeen, United Kingdom (2017)

[23] Dubuisson Duplessis, G., Clavel, C., Landragin, F.: Automatic measures to characterise verbal alignment in human-agent interaction. In: Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL), pp. 71–81. Association for Computational Linguistics, Saarbrücken, Germany (2017)

[24] Dušek, O., Jurcıcek, F.: A context-aware natural language generation dataset for dialogue systems. In: Workshop on Collecting and Generating Resources for Chatbots and Conversational Agents-Development and Evaluation, pp. 6–9 (2016)

[25] Dušek, O., Jurcıcek, F.: A context-aware natural language generator for dialogue systems. In: 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), pp. 185–190 (2016)

[26] Friedberg, H., Litman, D., Paletz, S.B.: Lexical entrainment and success in student engineering groups. In: Spoken Language Technology Workshop (SLT), pp. 404–409. IEEE (2012)

[27] Fusaroli, R., Konvalinka, I., Wallot, S.: Analyzing social interactions: the promises and challenges of using cross recurrence quantification analysis. In: Translational recurrences, pp. 137–155. Springer (2014)

[28] Fusaroli, R., Tylén, K.: Investigating conversational dynamics: Interactive alignment, interpersonal synergy, and collective task performance. Cognitive Science **40**(1), 145–171 (2016)

[29] Gallois, C., Ogay, T., Giles Howard, H.: Communication accommodation theory: A look back and a look ahead. W. Gudykunst (red.): Theorizing about intercultural communication. Thousand Oaks, CA: Sage pp. 121–148 (2005)

[30] Glas, N., Pelachaud, C.: Politeness versus perceived engagement: an experimental study. In: Proceedings of the 11th International Workshop on Natural Language Processing and Cognitive Science (2014)

[31] Gratch, J., DeVault, D., Lucas, G.: The benefits of virtual humans for teaching negotiation. In: International Conference on Intelligent Virtual Agents (IVA), pp. 283–294. Springer (2016)

[32] Griol, D., Molina, J.M., Callejas, Z.: Modeling the user state for context-aware spoken interaction in ambient assisted living. Applied Intelligence pp. 1–23 (2014)

[33] Gusfield, D.: Algorithms on Strings, Trees and Sequences. Cambridge University Press, Cambridge, UK (1997)

[34] Healey, P., Purver, M.: Self-repetition in dialogue and monologue. In: AixDial: the 22nd SemDial Workshop on the Semantics and Pragmatics of Dialogue, Aix-en-Provence, France (2018)

[35] Healey, P.G., Purver, M., Howes, C.: Divergence in dialogue. PloS one **9**(6), e98,598 (2014)

[36] Hu, Z., Halberg, G., Jimenez, C.R., Walker, M.A.: Entrainment in pedestrian direction giving: How many kinds of entrainment? In: Situated Dialog in Speech-Based Human-Computer Interaction, pp. 151–164. Springer (2016)

[37] Kelley, H.H., Schenitzki, D.P.: Bargaining. Experimental Social Psychology. New York: Holt, Rinehart, and Winston pp. 298–337 (1972)

[38] Kopp, S., Gesellensetter, L., Krämer, N.C., Wachsmuth, I.: A conversational agent as museum guide–design and evaluation of a real-world application. In: International Workshop on Intelligent Virtual Agents, pp. 329–343. Springer (2005)

[39] Lambertz, K.: Back-channelling: The use of yeah and mm to portray engaged listenership. Griffith Working Papers in Pragmatics and Intercultural Communication **4**, 11–18 (2011)

[40] Langlet, C., Clavel, C.: Improving social relationships in face-to-face human-agent interactions: when the agent wants to know user's likes and dislikes. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), vol. 1, pp. 1064–1073 (2015)

[41] Lebart, L., Salem, A., Berry, L.: Exploring textual data, vol. 4. Springer Science & Business Media (1997)

[42] Lopes, J., Eskenazi, M., Trancoso, I.: From rule-based to data-driven lexical entrainment models in spoken dialog systems. Computer Speech & Language **31**(1), 87–112 (2015)

[43] Mairesse, F., Walker, M.A.: Controlling user perceptions of linguistic style: Trainable generation of personality traits. Computational Linguistics **37**(3), 455–488 (2011)

[44] Mooney, C.H., Roddick, J.F.: Sequential pattern mining – approaches and algorithms. ACM Computing Surveys **45**(2), 19:1–19:39 (2013)

[45] Nenkova, A., Gravano, A., Hirschberg, J.: High frequency word entrainment in spoken dialogue. In: Proceedings of the 46th annual meeting of the association for computational linguistics on human language technologies (ACL-HLT): Short papers, pp. 169–172. Association for Computational Linguistics (2008)

[46] Niederhoffer, K.G., Pennebaker, J.W.: Linguistic style matching in social interaction. Journal of Language and Social Psychology **21**(4), 337–360 (2002)

[47] Oh, A.H., Rudnicky, A.I.: Stochastic language generation for spoken dialogue systems. In: Proceedings of the 2000 ANLP/NAACL Workshop on Conversational systems, vol. 3, pp. 27–32. Association for Computational Linguistics (2000)

[48] Parent, G., Eskenazi, M.: Lexical entrainment of real users in the let's go spoken dialog system. In: INTERSPEECH, pp. 3018–3021 (2010)

[49] Pfeifer, J.H., Iacoboni, M., Mazziotta, J.C., Dapretto, M.: Mirroring others' emotions relates to empathy and interpersonal competence in children. NeuroImage **39**(4), 2076–2085 (2008)

[50] Pickering, M.J., Garrod, S.: Toward a mechanistic psychology of dialogue. Behavioral and brain sciences **27**(02), 169–190 (2004)

[51] Reitter, D., Keller, F., Moore, J.D.: Computational modelling of structural priming in dialogue. In: Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL (NAACL-HLT): Short Papers, pp. 121–124. Association for Computational Linguistics (2006)

[52] Schenkein, J.: A taxonomy for repeating action sequences in natural conversation. Language production **1**, 21–47 (1980)

[53] Serban, I.V., Lowe, R., Charlin, L., Pineau, J.: A survey of available corpora for building data-driven dialogue systems. arXiv preprint arXiv:1512.05742 (2015)

[54] Shanon, C.E.: A mathematical theory of communication. Bell System Technical Journal **27**, 379–423 and 623–656 (1948)

[55] Sinha, T., Cassell, J.: We click, we align, we learn: Impact of influence and convergence processes on student learning and rapport building. In: Proceedings of the 1st Workshop on Modeling INTERPERsonal SynchrONy And infLuence, pp. 13–20. ACM (2015)

[56] Stenchikova, S., Stent, A.: Measuring adaptation between dialogs. In: 8th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL) (2007)

[57] Stoyanchev, S., Stent, A.: Lexical and syntactic priming and their impact in deployed spoken dialog systems. In: Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL (NAACL-HLT): Short Papers, pp. 189–192. Association for Computational Linguistics (2009)

[58] Svennevig, J.: Other-repetition as display of hearing, understanding and emotional stance. Discourse studies **6**(4), 489–516 (2004)

[59] Tannen, D.: Talking voices: Repetition, dialogue, and imagery in conversational discourse, vol. 26. Cambridge University Press (2007)

[60] Truong, K.P., Heylen, D.: Measuring prosodic alignment in cooperative task-based conversations. Proceedings of Interspeech (2012)

[61] Valstar, M., Baur, T., Cafaro, A., Ghitulescu, A., Potard, B., Wagner, J., André, E., Durieu, L., Aylett, M., Dermouche, S., et al.: Ask Alice: an artificial retrieval of information agent. In: Proceedings of the 18th ACM International Conference on Multimodal Interaction, pp. 419–420. ACM (2016)

[62] van Waterschoot, J., Dubuisson Duplessis, G., Gatti, L., Bruijnes, M., Heylen, D.: HAI Alice - An Information-Providing Closed-Domain Dialog Corpus. In: International Conference on Language Resources and Evaluation (LREC 2018) (2018)

[63] Varni, G., Hupont, I., Clavel, C., Chetouani, M.: Computational study of primitive emotional contagion in dyadic interactions. IEEE Transactions on Affective Computing (2017)

[64] Wang, N., Johnson, W.L., Mayer, R.E., Rizzo, P., Shaw, E., Collins, H.: The politeness effect: Pedagogical agents and learning outcomes. International Journal of Human-Computer Studies **66**(2), 98–112 (2008)

[65] Ward, A., Litman, D.J.: Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora. In: Speech and Language Technology in Education (SLaTE2007), pp. 57–60 (2007)

[66] Yu, Z., Nicolich-Henkin, L., Black, A.W., Rudnicky, A.I.: A wizard-of-oz study on a non-task-oriented dialog systems that reacts to user engagement. In: 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), pp. 55–63 (2016)