Automatic Measures to Characterise Verbal Alignment in Human-Agent Interaction

Guillaume Dubuisson Duplessis
Sorbonne Universités,
UPMC Univ Paris 06,
CNRS, ISIR,
75005 Paris, France
gdubuisson@isir.upmc.fr

Chloé Clavel
LTCI, Télécom ParisTech,
Université Paris-Saclay
75013 Paris, France
clavel@enst.fr

Frédéric Landragin
Lattice Laboratory, CNRS, ENS,
Université de Paris 3,
Université Sorbonne Paris Cité,
PSL Research University
Paris/Montrouge, France
frederic.landragin@ens.fr

Abstract

This work aims at characterising verbal alignment processes for improving virtual agent communicative capabilities. We propose computationally inexpensive measures of verbal alignment based on expression repetition in dyadic textual dialogues. Using these measures, we present a contrastive study between Human-Human and Human-Agent dialogues on a negotiation task. We exhibit quantitative differences in the strength and orientation of verbal alignment showing the ability of our approach to characterise important aspects of verbal alignment.

1 Introduction

Convergence of behaviour is an important feature of Human-Human (H-H) interaction that occurs both at low-level (e.g., body postures, accent and speech rate, word choice, repetitions) and at high-level (e.g., mental, emotional, cognitive) (Gallois et al., 2005). In particular, dialogue participants (DPs) automatically align their communicative behaviour at different linguistic levels including the lexical, syntactic and semantic ones (Pickering and Garrod, 2004). A key ability in dialogue is to be able to align (or not) to show a convergent, engaged behaviour or at the opposite a divergent one. Such convergent behaviour may facilitate successful task-oriented dialogues (Nenkova et al., 2008; Friedberg et al., 2012). Our goal 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 believability, to increase interaction naturalness and to maintain user's engagement (Yu et al., 2016). In this paper, we aim at providing measures characterising verbal alignment processes based on repetitions between DPs. We propose a framework based on repetition at the lexical level which deals with textual dialogues (e.g., transcripts), along with automatic and generic measures indicating verbal alignment between interlocutors. We offer a study that contrasts H-H and Human-Agent (H-A) dialogues on a negotiation task and show how our proposed measures can be used to quantify verbal alignment. We confirm quantitatively some predictions from previous literature regarding the strength and orientation of verbal alignment in Human-Machine Interaction (Branigan et al., 2010).

Section 2 presents and discusses the related work. Section 3 describes the proposed model and outlines its main features. Next, Section 4 presents the corpus-based experimentation protocol and states the main investigated hypotheses. Then, Section 5 presents the quantitative analysis and discusses the main results. Finally, Section 6 concludes this paper.

2 Related Work

When people are engaged in a dialogue there is evidence that their behaviours tend to converge (Gallois et al., 2005) and automatically align at several levels (Pickering and Garrod, 2004). This includes non-linguistic levels such as facial expressions and body postures as well as linguistic levels such as lexical, syntactic and semantic ones. In particular, 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" (Pickering and Garrod, 2004). Thus, DPs tend to reuse lexical as well as syntactic structure (Reitter et al., 2006; Ward and Litman, 2007). One consequence

Proceedings of the SIGDIAL 2017 Conference, pages 71–81,
Saarbrücken, Germany, 15-17 August 2017.

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 (Pickering and Garrod, 2004). DPs tend to automatically establish and use fixed expressions that become dialogue routines via a process called "routinization". Recent work argues that these patterns of repetition may be specific to task-oriented dialogues and do not generalise to ordinary conversation in H-H interactions (Healey et al., 2014). Here, we are specifically interested in verbal alignment in H-H and H-A task-oriented interactions. We use the term alignment to say that DPs converge at the lexical level by using the same words and expressions (e.g., by employing the expression "that's not gonna work for me" to reject a proposition).

Studies point out evidence that lexical items and syntactic structures used by a system are subsequently adopted by users (Brennan, 1996; Stoyanchev and Stent, 2009; Parent and Eskenazi, 2010; Branigan et al., 2010). (Branigan et al., 2010) argue that linguistic alignment should occur in Human-Machine interaction. 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. In this work, we bring quantitative evidence supporting the fact that human align more with a virtual agent than with another human based on a study contrasting H-H and H-A interactions at the level of repetition of expressions. While previous studies have mainly focused on H-H dialogues, we offer in this work an analysis of verbal alignment in H-A dialogues based on a corpus.

Several studies aim at providing virtual agents with the ability to verbally align with the user in order to improve credibility, naturalness, and also to foster user engagement (Clavel et al., 2016). It involves high-level alignment such as politeness (De Jong et al., 2008) or aligning on appreciations (Campano et al., 2015). Work on convergence in the spoken dialogue system community has mainly focused on lexical entrainment, i.e. the tendency to use the same terms when DPs refer repeatedly to the same objects (Brennan and Clark, 1996). Several entrainment models have been proposed to let the system entrains to user utterances (e.g., (Brockmann et al., 2005; Buschmeier et al., 2010; Hu et al., 2014; Lopes et al., 2015)). These models are completely or partially rule-based and focus on specific aspects of entrainment. Recent work aims at introducing entrainment in a fully trainable natural language system by exploiting the preceding user utterance (Dušek and Jurcıcek, 2016).

Several metrics have been employed to automatically measure linguistic alignment in written corpora. At the word or token levels, (Nenkova et al., 2008) quantify verbal alignment based on high-frequency words while (Campano et al., 2014) quantify verbal alignment based on vocabulary overlap between DPs. (Healey et al., 2014) compute similarity at the syntax and lexical levels on windows of a fixed number of turns. (Fusaroli and Tyln, 2016) employ (cross-)recurrence quantification analysis to quantify interactive alignment and interpersonal synergy at the lexical, prosodic and speech/pause levels. (Reitter et al., 2006; Ward and Litman, 2007) focus on regression models to study priming effects within a small window of time in single dialogues. (Stenchikova and Stent, 2007) use a frequency-based approach (Church, 2000) to measure adaptation *between* dialogues. In this paper, we propose global and speaker-specific measures based on the automatic construction of the expression lexicon built by the DPs. An originality of our approach is to consider lexical patterns predicted by the routinization process of the interactive alignment theory. These measures rely on efficient algorithms making an online usage in a dialogue system realistic. They indicate both verbal alignment at the level of repetitions and the orientation of verbal alignment between DPs in single dialogues.

3 Model: Expression-based Measures of Verbal Alignment

To address the problem of detecting (possibly overlapping) repetitions between DPs, we propose a framework defining key features of repeated expressions, along with an efficient computational mean of building an expression lexicon.

In this work, we define an *expression* as a surface text pattern at the utterance level that has been produced by both speakers in a dialogue. In other words, it is a contiguous sequence of tokens that appears in at least two utterances produced by two different speakers. An expression may be a single

token (e.g., "you", "I"). However, an expression should contain at least one non-punctuation token. Thus, sequences like "?", "!", "," are not expressions. An *instance of an expression* can either be free or constrained in a given utterance[1]. A *free* instance is an instance of an expression that appears in an utterance without being a subexpression of a larger expression. A *constrained* instance is an expression that appears in a turn as a subexpression of a larger expression. The *initiator* of the expression is the interlocutor that first produced an instance of the expression either in a free or constrained form. Lastly, an expression is established as soon as the two following 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. Eventually, the *expression lexicon* of a dialogue is the set of established expressions that appear in this dialogue. Importantly, the expression lexicon contains all expressions that appear in a dialogue at least once in a free form. Expressions that are always constrained (i.e. which instances are always a subpart of a larger expression) are discarded.

Table 1 presents an excerpt of dialogue extracted from the corpus used in this work. In this example, "that's not gonna work for me" is an expression initiated by A in turn 1 and established in turn 4. This expression is free in this excerpt, and it belongs to the expression lexicon. Similarly, "work for" is an expression initiated by A in turn 1 and established in turn 2. It appears in a constrained form in the expression "that's not gonna work for me" in turns 1 and 4, and in a free form in turn 2. It belongs to the expression lexicon. The expression "that's not gonna" occurs in a constrained form in turns 1 and 4, and never occurs in a free form. This expression is never established (contrary to its parent expression "that's not gonna work for me") and thus is not included in the expression lexicon.

The automatic extraction of expressions from a dialogue is an instance of sequential pattern mining (Mooney and Roddick, 2013) applied to textual dialogues. In this work, we follow a similar approach than (Dubuisson Duplessis et al., 2017)

[1]This terminology is borrowed and adapted from the textual data analysis field and the notion of "repeated segment" (Lebart et al., 1997)

Loc.	Utterance
A_1	well, that's an interesting idea. but no, that's not gonna work for me.
B_2	what will *work for* you?
A_3	*what* do *you* think about me getting two chairs and one plate and *you* getting one chair, one plate, and the clock?
B_4	*that's not gonna work for me*
A_5	well which of these items would be your first choice?
B_6	*well* i don't want *the clock*
A_7	oh really?

Table 1: Excerpt of dialogue extracted from the H-A corpus (described in Section 4.1). Expressions are coloured. Established expressions are in italic.

by employing a generalised suffix tree in order to solve the multiple common subsequence problem (MCSP) (Gusfield, 1997) to extract frequent surface text patterns between utterances, and then filtering patterns used by both DPs. Notably, the MCSP is solved in linear time with respect to the number of tokens in a dialogue (Gusfield, 1997).

3.1 Properties of Expressions

An expression has a *frequency* which corresponds to the number of utterances in which the expression appears. For example, the expression "work for" has a frequency of 3 because it appears in utterance 1, 2 and 4. Next, the *size* of an expression is its number of tokens (e.g., expression "the clock" has size 2). Then, the *span* of an expression is the number of utterances between the first production and the last production of this expression in the dialogue (including the first and last utterances). The minimum span is 2, meaning the expression has been established in two adjacent utterances. For instance, the expression "the clock" has a span of 4 because it appears first in utterance 3 and last in utterance 6. We derive the *density* of an expression which is given by the ratio between its frequency and its span. For instance, the density of the expression "well" is 0.5. Eventually, the *priming* of an expression is the number of repetitions of the expression by the initiator before being used by the other interlocutor (either in a free or constrained form). For example, the expression "well" has a priming of 2 because it is repeated by speaker A in utterance 1 and 5 before being established in utterance 6.

3.2 Measures

Globally, we derive the following measures from the model:

Expression lexicon size (ELS) the number of items in the expression lexicon, i.e. the number of established expressions in the dialogue

Expression variety (EV) the expression lexicon size normalised by the total number of tokens in the dialogue. It is given by: $\text{EV} = \frac{\text{ELS}}{\text{\# Tokens}}$. This ratio indicates the variety of the expression lexicon relatively to the length of the dialogue. The higher it is, the more there are different expressions established between DPs.

Expression repetition (ER) the ratio of produced tokens belonging to an instance of an established expression, i.e. the ratio of tokens belonging to a repetition of an expression. It is given by: $\text{ER} = \frac{\text{\# Tokens in an established expr.}}{\text{\# Tokens}}$, $\text{ER} \in [0, 1]$. The higher the ER is, the more DPs dedicate tokens to the repetition of established expressions.

We also derive the following measures for each speaker S:

Initiated expressions (IE_S) number of expressions initiated by S (and further established) normalised by the expression lexicon size. It is given by: $\text{IE}_S = \frac{\text{\# Expr. initiated by S}}{\text{ELS}}$, $\forall$ S, $\text{IE}_S \in [0, 1]$. Note that in a dyadic dialogue involving speaker S_1 and S_2, $\text{IE}_{\text{S}_1} + \text{IE}_{\text{S}_1} = 1$.

Expression repetition (ER_S) ratio of produced tokens belonging to an instance of an established expression, i.e. ratio of tokens belonging to a repetition of an expression. It is given by: $\text{ER}_S = \frac{\text{\# Tokens from S in an established expr.}}{\text{\# Tokens from S}}$, $\forall$ S, $\text{ER}_S \in [0, 1]$

Eventually, we also consider a measure independent of the model: the *Token Overlap* (TO) which is the ratio of shared tokens between locutor S_1 and locutor S_2 in a dialogue. It is given by: $\text{TO} = \frac{\#(\text{Tokens}_{\text{S}_1} \cap \text{Tokens}_{\text{S}_2})}{\#(\text{Tokens}_{\text{S}_1} \cup \text{Tokens}_{\text{S}_2})}$. The higher is TO, the more vocabulary is shared between S_1 and S_2.

4 Experimentation

Our methodology aims at comparing quantitatively both H-H and H-A task-oriented corpora at the level of the repetition of expressions.

4.1 Negotiation Corpora

The corpus of this study focuses on a negotiation task between two DPs and is detailed in (Gratch et al., 2016). It focuses on a common abstraction of negotiation known as the multi-issue bargaining task (Kelley and Schenitzki, 1972). 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. Negotiation can take two structures in this scenario. The integrative structure represents a negotiation that can turn out to be a win-win for both players (if they realise through conversation that this is a cooperative negotiation). On the other hand, the distributive negotiation represents a competitive (zero-sum) negotiation where players share the same interests in objects. However, players do not know in advance and often assume a distributive negotiation (i.e. their opponent wants the same thing as them) rather than an integrative negotiation. This corpus can be broken down into two parts: a H-H corpus and a H-A corpus. In both parts, people were given similar instructions, i.e. humans are told that they must negotiate with another player how to divide the contents of a storage locker filled with three classes of valuable items (such as records, lamps or painting).

In the H-H corpus, pairs of people performed one negotiation which was either distributive or integrative in structure. Independently, they were given information in the instructions that suggested the negotiation was integrative or distributive. Note that this condition does not affect the results presented below.

In the H-A corpus, the human participant engaged in two negotiations with two different virtual agents (a male called Brad and a female called Ellie). The first negotiation was a cooperative/integrative negotiation while the second was a competitive/distributive negotiation. The order of interaction with the agents (Brad-Ellie or Ellie-Brad) was randomly chosen. The interaction was framed. Half of the human participants was told they were interacting with an autonomous agent while the other half was told they were interacting with a human wizard (though the agent was always controlled by a wizard). The Woz system controlling virtual agents has been designed to be

Table 2: Figures about the H-H corpus and the H-A corpus. U = Unique, T/Utt.=Tokens per Utterance, med. = median

	H-H	H-A
Dialogue	84	154
Utterance (U)	10319 (7840)	17125 (6109)
... avg (std)	122.8 (84.1)	111.2 (57.5)
Token (U)	79396 (2516)	90479 (1335)
T/Utt.		
avg/med. (std)	7.7/6.0 (7.4)	5.3/4.0 (5.7)
avg (std)	7.7 (7.4)	5.3 (5.7)
min/max	1/66	1/154

as natural as possible (DeVault et al., 2015). 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 possible utterances (more than 11,000) along with a specific interface enabling the human operator to rapidly select among those (DeVault et al., 2015). For both virtual human agents, wizards were rather free but followed some guidelines. First, the goal in both negotiations is for the agent to win. Next, in the distributive condition, wizards were requested to be soft, polite and vague trying hard to get the human participant to make the first offer and avoiding revealing what they wanted (unless the human directly asks). In the integrative condition, wizards could share preferences and were not requested to be vague. However, they were requested to try getting the human share first and make the first offer. Table 1 presents an excerpt from a competitive negotiation from the H-A corpus.

Figures about both corpora can be found in Table 2. Globally, dialogues in both corpora contains more than 100 utterances. It shows that H-A dialogues are a bit shorter than H-H dialogues but still comparable. Besides, utterances are shorter in terms of tokens in the H-A dialogues than in the H-H dialogues.

4.2 Randomised Corpora

To investigate hypotheses stated in Section 4.3, we constituted two randomised corpora HH_R and HA_R respectively for the randomised version of the H-H corpus and the H-A corpus. This randomisation process is similar to the ones adopted by various work investigating verbal alignment (e.g., (Ward and Litman, 2007), (Healey et al., 2014), (Fusaroli and Tyln, 2016)). To constitute the HH_R corpus, the following process is performed for each dialogue of the initial corpus: each interlocutor's real turns in sequence are interleaved with turns randomly chosen from the H-H corpus. A similar process is followed for the HA_R corpus with the exception that each human turn is substituted by a random human turns from the H-A corpus when keeping the sequence of wizard turns; while each wizard turn is substituted by a random wizard turns from the H-A corpus when keeping the sequence of human turns. In all, two dialogues are generated by these processes for each original H-H/A dialogue (one for each locutor). These surrogate corpora lack the coherence of dialogues in the H-H and H-A corpora. Indeed, utterances are no longer in their original relationship with their response utterances. We thus expect to find reduced verbal alignment at the level of expressions in these corpora.

4.3 Hypotheses

4.3.1 "Above Chance" Hypotheses

Our first hypothesis is that DPs should verbally align at the level of expressions in both the H-H corpus and the H-A corpus more than would be expected by chance. This hypothesis can be expressed in the following way:

routinization DPs should constitute a richer expression lexicon than they would by chance (this should be indicated by the EV measure)

repetition DPs should repeat expressions more often than chance (this should be indicated by the ER and the TO measures)

4.3.2 H-H VS H-A Hypotheses

Following Branigan et al's hypothesis (Branigan et al., 2010), we should expect more verbal alignment at the level of expressions in the H-A corpus than in the H-H corpus. Besides, we should expect more verbal alignment from the human participant than from the agent. Indeed, the human participant both has the ability to verbally align (contrary to the agent) and may be influenced by beliefs about the communicative limitations of the agent. This hypothesis can be expressed in the following way:

routinization DPs should constitute a richer expression lexicon in the H-A corpus than in the H-H corpus (this should be indicated by the EV measure)

repetition DPs should dedicate more tokens to the repetition of established expressions in the H-A corpus than in the H-H corpus (this should be indicated by the ER and the TO measures)

orientation the human participant should repeat more expressions initiated by the agent than the other way around (this should be indicated by the IE_S and the ER_S measures)

4.3.3 H-A-specific Hypotheses

In this study, we also consider conditions that affects only the H-A corpus. First, interactions with the virtual agent were randomly "framed" meaning that, prior interactions, the human participant was either told that the agent was controlled by a human operator (72 dialogues) or that it was autonomous (82 dialogues). This condition affects the mediated component of verbal alignment i.e. the beliefs of the human participant about the communicative capabilities of the agent (e.g., in terms of understanding). This leads us to the following hypothesis:

framing framing should impact verbal alignment in the routinization, repetition and orientation aspects.

More specifically, "human" framing should lead to a more "human-like verbal alignment" while "agent" framing should lead to a "HMI-like verbal alignment" (Branigan et al., 2010).

Moreover, the human participants interacted with two versions of the virtual agent. One was Ellie, a female agent, while the other was Brad, a male agent. Interaction order was random (Brad-Ellie or Ellie-Brad). This condition leads us to the following hypothesis:

gender gender matching (Male-Male or Female-Female) or unmatching (Male-Female, Female-Male) should not impact verbal alignment

Lastly, interactions involved two types of negotiations (integrative and distributive). We study the impact of the negotiation type on the verbal alignment at the level of expressions.

5 Quantitative Analysis and Results

5.1 Comparisons to the Surrogate Corpora

We compare the H-H and H-A corpora of real interactions to the surrogate HH_R and HA_R corpora to ensure that established expressions in the dialogues are actually due to the coherent sequence of utterances and are not incidental.

We investigated whether DPs in the H-H corpus verbally align at the level of expressions more than would be expected by chance by comparing it to the HH_R corpus (following hypotheses stated in Section 4.3.1). First, the expression variety is significantly higher for the H-H corpus (mean=0.118, std=0.023) than for the HH_R corpus (mean=0.110, std=0.015). Statistical difference is checked by a Wilcoxon rank sum test ($U = 8951$, $p = 0.00051 < 0.001$, $r = 0.22$)[2]. This indicates that H-H interactions lead to a richer expression lexicon. However, the expression repetition is not significantly different ($p = 0.3446$) between the H-H corpus (mean=0.436, std=0.107) and the HH_R corpus (mean=0.420, std=0.108). This means that the amount of tokens dedicated to the repetition of expressions is similar between the H-H corpus and the HH_R corpus. An explanation of this may be that the dialogues happen in a closed domain on a specific task (negotiations of a set of objects) and thus in a constrained vocabulary. This inevitably leads random dialogues to include repetitions though in a lesser variety. This is confirmed by the token overlap that is significantly higher for the H-H corpus (mean=0.316, std=0.073) than for the HH_R corpus (mean=0.276, std=0.058) ($U = 9468.5$, $p = 9.781 \times 10^{-6} < 0.001$, $r = 0.28$). DPs share a richer vocabulary than what would happen by chance.

We performed a similar analysis by comparing the H-A corpus and the HA_R corpus. It turns out that both the expression lexicon variety and the expression repetition are significantly higher in the H-A corpus than in the HA_R corpus. Indeed, the expression variety is significantly higher ($U = 30126$, $p = 2.155 \times 10^{-6} < 0.001$, $r = 0.22$) for the H-A corpus (mean=0.134, std=0.022) than for the HA_R corpus (mean=0.124, std=0.020). Besides, the expression repetition is significantly higher ($U = 28124$, $p = 0.0011 < 0.01$, $r = 0.15$) for the H-A corpus (mean=0.416, std=0.086) than for the HA_R corpus (mean=0.386, std=0.088). This is comforted by the fact that the token overlap is significantly higher ($U = 30164$, $p = 1.875 \times 10^{-6} < 0.001$, $r = 0.22$) for the H-A corpus (mean=0.322, std=0.06) than for the HA_R corpus (mean=0.293, std=0.06).

All in all, it turns out that both H-H and H-A di-

[2]For each test, we report the test statistics (U/W), the p-value (p) and the effect size (r).

alogues constitute a richer expression lexicon than they would by chance (routinization hypothesis). As for the repetition hypothesis, DPs clearly repeat expressions more often than chance in the H-A corpus. However, repetition in the H-H corpus is comparable to what would happen by chance in closed domain task-oriented dialogues. All things considered, our indicators show that both corpora tends to verbally align at the level of shared expressions more than they would by chance.

5.2 Differences between H-H/A Interactions

We compare verbal alignment at the expression level between the H-H corpus and the H-A corpus globally, per speaker and at the lexicon level.

5.2.1 Global Interaction Analysis

It turns out that the expression variety is significantly lower for the H-H corpus (mean=0.118, std=0.023) than for the H-A corpus (mean=0.134, std=0.022). This is checked via a Wilcoxon rank sum test ($U = 4056.5$, $p = 2.035 \times 10^{-6} < 0.001$, $r = 0.31$). This indicates that DPs constitute a richer expression lexicon in the H-A corpus than in the H-H corpus. However, we noticed that there is no significant difference between the H-H corpus and the H-A corpus in terms of expression repetition and token overlap. Indeed, the expression repetition is not significantly different between the H-H corpus (mean=0.436, std=0.107) and the H-A corpus (mean=0.416, std=0.086) by a Wilcoxon rank sum test ($p = 0.1261$). Besides, the token overlap is not significantly different between the H-H corpus (mean=0.316, std=0.073) and the H-A corpus (mean=0.322, std=0.06) by a similar test ($p = 0.6618$).

H-A interactions lead to a richer expression lexicon than the H-H interactions (routinization hypothesis). This indicates more verbal alignment at the level of shared expressions in H-A dialogues. However, DPs do not dedicate more tokens to the repetition of established expressions in the H-A corpus than in the H-H corpus (repetition hyp.).

5.2.2 Speaker Perspective Analysis

We investigated verbal alignment at the level of expressions by having a closer look at each speaker in a dialogue in terms of initiated expressions (IE) and expression repetition (ER). In the H-H corpus, both speakers play a symmetrical role at the level of expressions. First, they initiate a similar amount of expressions. Indeed, IE_{S_1} and the IE_{S_2} are not significantly different (Wilcoxon signed rank test, $p = 0.5978$). Next, they dedicate the same amount of tokens to the repetition of expressions (see Figure 1). In fact, ER_{S_1} and the ER_{S_2} are not significantly different ($p = 0.9875$).

On the contrary, the H-A corpus shows an asymmetrical role at the level of expressions between the Woz and the human participant. First, the Woz initiates more expressions than the human participant. Indeed, IE_{Woz} (mean=0.596, std=0.116) is significantly higher than IE_{H} (mean=0.404, std=0.116) (Wilcoxon signed rank test, $W = 10161$, $p < 2.2 \times 10^{-16} < 0.001$, $r = 0.87$). Then, the human participant dedicates more tokens to the repetition of an established expression than the Woz (see Figure 1). As a matter of fact, ER_{Woz} (mean=0.347, std=0.104) is significantly lower than ER_{H} (mean=0.492, std=0.086) (Wilcoxon signed rank test, $W = 545$, $p < 2.2 \times 10^{-16} < 0.001$, $r = 0.87$). Notably, this asymmetry does not appear when considering the number of tokens produced by each speaker, i.e. the Woz and the human tend to produce the same amount of tokens. Indeed, there is not a significant difference in the proportion of tokens produced by the Woz (mean=0.483, std=0.134) and by the human participant (mean=0.517, std=0.134) (Wilcoxon signed rank test, $p = 0.08067$). Besides, a closer look at the shared vocabulary shows that there is not a significant difference in the proportion of vocabulary shared by the Woz (mean=0.4853, std=0.116) and by the human participant (mean=0.515, std=0.093)[3] (Wilcoxon signed rank test, $p = 0.08029$). That is, globally, the Woz does not share more of its vocabulary than the human participants, and conversely.

It turns out that verbal alignment at the level of shared expressions is symmetrical in the H-H corpus. On the contrary, it is asymmetrical in the H-A corpus (orientation hypothesis) where it indicates that the human participant verbally align more by (i) adopting more Woz-initiated expressions (than the Woz adopting Human-initiated expressions), and (ii) dedicating more tokens to the repetition of established expressions.

5.2.3 Expression Lexicon Analysis

Eventually, we took a closer look at the expression lexicon produced in the H-H corpus and the

[3] Relative shared vocabulary for S_1 is computed as follow: $\text{SV}_{\text{S}_1} = \frac{\#(\text{Tokens}_{\text{S}_1} \cap \text{Tokens}_{\text{S}_2})}{\#(\text{Tokens}_{\text{S}_1})}$

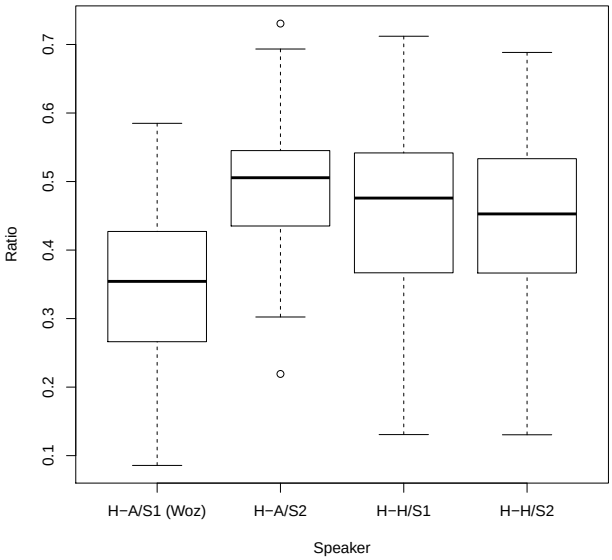

Figure 1: Comparison of the H-H/A corpora for ER_S. Difference is significant for H-A ($p < 0.001$), not for H-H (cf. Section 5.2.2).

H-A corpus. Regarding the size in tokens of the expressions, there is no significant difference between the two corpora (Wilcoxon rank sum test, $p = 0.9897$). The majority of expressions contains less than 3 tokens. Around 70% of expressions are 1-token expressions, 20% are 2-token expressions, 5% are 3-token expressions, and the other 5% are 4-token and more expressions.

Considering the priming of an expression (i.e. the number of repetitions of the expression by the initiator before being used by the other interlocutor), most expressions have a priming of less than 3 repetitions in both corpora. However, there is a significant difference between the two corpora (Wilcoxon rank sum test, $U = 57185000$, $p < 2.2 \times 10^{-16} < 0.001$). The most striking one is about the proportion of 1-repetition priming expressions. 63% of expressions have a 1-repetition priming in the H-H corpus while it is higher in the H-A corpus at 72%. 20% of expressions have a 2-repetition priming in the H-H corpus while it is 17% in the H-A corpus. Lastly, 8% of the H-H expressions have a 3-repetition priming while it reaches 6% for the H-A corpus. The main reason of the difference at the priming level may be found in the functions that serve expression repetition in the corpora. This is supported by the study of the density of expressions (i.e. their ratio frequency/span) in both corpora. Expressions in the H-A corpus are denser (mean=0.174, std=0.238) than expressions in the H-H corpus (mean=0.146, std=0.206). This difference is significant (Wilcoxon rank sum test, $U = 45419000$, $p < 2.2 \times 10^{-16} < 0.001$). Expressions in the H-A corpus tend to occur more frequently between their first and last appearance in the dialogue than in the H-H corpus.

5.3 Other Conditions in Human-Agent Interactions

We studied the impact of the "human operator" framing against the "AI" framing on the verbal alignment at the level of expressions. It turns out there is no difference in the variety of the expression lexicon between the two framing modes. Indeed, the expression variety is not significantly different between "human operator" framing (mean=0.131, std=0.023) and the "AI" framing (mean=0.136, std=0.021) (Wilcoxon rank sum test, $p = 0.1338$). Study about repetition does not reveal any effect from the framing condition. As a matter of fact, the expression repetition is not significantly different between "human operator" framing (mean=0.423, std=0.087) and "AI" framing (mean=0.409, std=0.085) ($p = 0.2915$). Similarly, no effect is found at the token overlap. Besides, analyses on the expression initiation (EI) and the expression repetition at the speaker level (ER_S) yield the same results than the entire H-A corpus i.e. the verbal alignment is asymmetrical between the agent and the human. Contrary to our hypothesis, framing does not quantitatively impact verbal alignment at the level of expressions.

A similar analysis at the gender mismatch or match between the human participant and the agent (Brad or Ellie) does not reveal any difference at the expression variety, expression repetition (globally or by speaker), token overlap, and expression initiation. These analyses confirm our hypothesis that gender does not quantitatively impact verbal alignment at the level of expressions in our H-A corpus.

It turns out that some significant differences exist between the two types of negotiation (integrative and distributive) in the H-A corpus. First, distributive negotiation leads to longer dialogues in number of utterances (mean=144.3, std=58.757) than integrative negotiation (mean=82.5, std=41.09). Despite this difference in dialogue length, the expression variety is similar between the integrative negotiations (mean=0.133, std=0.022) and the distributive ones

(mean=0.133, std=0.020) (Wilcoxon signed rank test, $p = 0.9847$). However, a major difference can be observed at the expression repetition which is significantly higher for the distributive negotiations (mean=0.456, std=0.073) than for the integrative negotiations (mean=0.375, std=0.084) ($W = 142$, $p = 7.665 \times 10^{-10} < 0.001$, $r = 0.87$). All in all, this indicates that participants align more at the level of expressions in competitive negotiations than in cooperative ones. This may be due to the fact that they need to verbally align more on (counter-)propositions in competitive negotiations.

5.4 Discussion

We have presented automatic and generic measures of verbal alignment based on an expression framework focusing on repetition between DPs at the level of surface of text utterances. This framework mainly takes into account lexical cues by building a lexicon of shared expressions emerging during dialogue, but also syntactic cues to the extent of expressions (other work on conversations report a strong correlation between lexical and syntactic cues regarding alignment (Healey et al., 2014)). The proposed measures make it possible to quantify the routinization process (via EV), the degree of repetition between DPs (via ER), and the orientation of the verbal alignment (via IE_S and ER_S) at the level of expressions. Besides, these measures are based on efficient algorithms (Gusfield, 1997) that make it realistic to envision an online usage in a dialogue system. They have made it possible to check quantitatively that verbal alignment was real in both H-H and H-A task-oriented interactions (i.e. it is not likely to happen randomly). Next, they have helped contrasting quantitatively H-H interactions from H-A interactions, showing that verbal alignment was symmetrical in H-H interactions while being asymmetrical in H-A (comforting previous hypotheses (Branigan et al., 2010)). Finally, we have observed that H-A verbal alignment was independent of the gender of the agent (male or female) and of the framing of the experiment (human operator VS AI). However, the proposed measures indicate more verbal alignment in competitive negotiations than in cooperative ones that may be due to the need to reach more agreements during competitive negotiations.

Nevertheless, this work is limited to automatically quantifying repetitions at the lexical level. Hence, it does not take into account other aspects of alignment such as linguistic style (Niederhoffer and Pennebaker, 2002) or higher level such as concepts (Brennan and Clark, 1996). However, the alignment theory proposes that alignment "percolates" between levels. As such, alignment at the level of repetition of expressions indicate alignment at other levels to some extent. Besides, this work does not consider the functions behind repetition such as conveying the reception of a message, appraising a proposal, introducing a disagreement, complaining (Tannen, 2007; Schenkein, 1980). A functional analysis could explain more in depth the differences between the H-H and the H-A corpora. Lastly, an interesting perspective would be to confirm these results on another corpora involving comparable H-H and H-A dialogues.

6 Conclusion and Future Work

This paper has presented a framework based on expression repetition at the surface text of dialogue utterances involving automatic and computationally inexpensive measures. These measures make it possible to quantitatively characterise the strength and orientation of verbal alignment between DPs in a task-oriented dialogue. A promising perspective of this work lies in the exploitation of these measures to adapt and align the verbal communicative behaviour of a virtual agent.

Acknowledgments

This work was supported by the European project H2020 ARIA-VALUSPA and the French ANR project IMPRESSIONS (ANR-15-CE23-0023). We warmly thank Jonathan Gratch and David DeVault for sharing the negotiation corpora, and Catherine Pelachaud for valuable and enriching discussions. We would like to thank the anonymous reviewers for their valuable comments and suggestions.

References

Holly P Branigan, Martin J Pickering, Jamie Pearson, and Janet F McLean. 2010. Linguistic alignment between people and computers. *Journal of Pragmatics* 42(9):2355–2368.

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

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

Carsten Brockmann, Amy Isard, Jon Oberlander, and Michael White. 2005. Modelling alignment for affective dialogue. In *Workshop on adapting the interaction style to affective factors at the 10th international conference on user modeling (UM-05)*.

Hendrik Buschmeier, Kirsten Bergmann, and Stefan Kopp. 2010. Modelling and evaluation of lexical and syntactic alignment with a priming-based microplanner. In *Empirical Methods in Natural Language Generation*, Springer, pages 85–104.

Sabrina Campano, Jessica Durand, and Chloé Clavel. 2014. Comparative analysis of verbal alignment in human-human and human-agent interactions. In *International Conference on Language Resources and Evaluation (LREC)*. pages 4415–4422.

Sabrina Campano, Caroline Langlet, Nadine Glas, Chloé Clavel, and Catherine Pelachaud. 2015. An eca expressing appreciations. In *International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, pages 962–967.

Kenneth W Church. 2000. 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*. Association for Computational Linguistics, volume 1, pages 180–186.

Chloé Clavel, Angelo Cafaro, Sabrina Campano, and Catherine Pelachaud. 2016. Fostering user engagement in face-to-face human-agent interactions: a survey. In *Toward Robotic Socially Believable Behaving Systems-Volume II*, Springer, pages 93–120.

Markus De Jong, Mariët Theune, and Dennis Hofs. 2008. Politeness and alignment in dialogues with a virtual guide. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems (AAMAS)*. International Foundation for Autonomous Agents and Multiagent Systems, pages 207–214.

David DeVault, Johnathan Mell, and Jonathan Gratch. 2015. 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*.

Guillaume Dubuisson Duplessis, Franck Charras, Vincent Letard, Anne-Laure Ligozat, and Sophie Rosset. 2017. Utterance Retrieval based on Recurrent Surface Text Patterns. In *39th European Conference on Information Retrieval (ECIR)*. Aberdeen, United Kingdom, pages 199–211.

Ondrej Dušek and Filip Jurcıcek. 2016. A context-aware natural language generator for dialogue systems. In *17th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*. pages 185–190.

Heather Friedberg, Diane Litman, and Susannah BF Paletz. 2012. Lexical entrainment and success in student engineering groups. In *Spoken Language Technology Workshop (SLT)*. IEEE, pages 404–409.

Riccardo Fusaroli and Kristian Tyln. 2016. Investigating conversational dynamics: Interactive alignment, interpersonal synergy, and collective task performance. *Cognitive Science* 40(1):145–171.

Cindy Gallois, Tania Ogay, and Howard H. Giles. 2005. Communication accommodation theory: A look back and a look ahead. *W. Gudykunst (red.): Theorizing about intercultural communication. Thousand Oaks, CA: Sage* pages 121–148.

Jonathan Gratch, David DeVault, and Gale Lucas. 2016. The benefits of virtual humans for teaching negotiation. In *International Conference on Intelligent Virtual Agents (IVA)*. Springer, pages 283–294.

Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK.

Patrick GT Healey, Matthew Purver, and Christine Howes. 2014. Divergence in dialogue. *PloS one* 9(6):e98598.

Zhichao Hu, Gabrielle Halberg, Carolynn R Jimenez, and Marilyn A Walker. 2014. Entrainment in pedestrian direction giving: How many kinds of entrainment? In *In Proceedings of 5th International Workshop on Spoken Dialog System (IWSDS)*. Citeseer.

Harold H Kelley and Donald P Schenitzki. 1972. Bargaining. *Experimental Social Psychology. New York: Holt, Rinehart, and Winston* pages 298–337.

Ludovic Lebart, André Salem, and Lisette Berry. 1997. *Exploring textual data*, volume 4. Springer Science & Business Media.

José Lopes, Maxine Eskenazi, and Isabel Trancoso. 2015. From rule-based to data-driven lexical entrainment models in spoken dialog systems. *Computer Speech & Language* 31(1):87–112.

Carl H. Mooney and John F. Roddick. 2013. Sequential pattern mining – approaches and algorithms. *ACM Computing Surveys* 45(2):19:1–19:39.

Ani Nenkova, Agustin Gravano, and Julia Hirschberg. 2008. High frequency word entrainment in spoken dialogue. In *Proceedings of the 46th annual meeting of the association for computational linguistics on human language technologies (ACL-HLT): Short papers*. Association for Computational Linguistics, pages 169–172.

Kate G Niederhoffer and James W Pennebaker. 2002. Linguistic style matching in social interaction. *Journal of Language and Social Psychology* 21(4):337–360.

Gabriel Parent and Maxine Eskenazi. 2010. Lexical entrainment of real users in the let's go spoken dialog system. In *INTERSPEECH*. pages 3018–3021.

Martin J. Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences* 27(02):169–190.

David Reitter, Frank Keller, and Johanna D Moore. 2006. 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*. Association for Computational Linguistics, pages 121–124.

James Schenkein. 1980. A taxonomy for repeating action sequences in natural conversation. *Language production* 1:21–47.

Svetlana Stenchikova and Amanda Stent. 2007. Measuring adaptation between dialogs. In *8th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.

Svetlana Stoyanchev and Amanda Stent. 2009. 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*. Association for Computational Linguistics, pages 189–192.

Deborah Tannen. 2007. *Talking voices: Repetition, dialogue, and imagery in conversational discourse*, volume 26. Cambridge University Press.

Arthur Ward and Diane J Litman. 2007. Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora. In *Speech and Language Technology in Education (SLaTE2007)*. pages 57–60.

Zhou Yu, Leah Nicolich-Henkin, Alan W Black, and Alex I Rudnicky. 2016. 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)*. pages 55–63.