

Comparative Analysis of verbal alignment in human-human and human-agent interactions

Sabrina Campano, Jessica Durand, Chloé Clavel

Institut Mines-Télécom, Télécom ParisTech, CNRS LTCI
37-39, rue Dareau, 75014 Paris, France

sabrina.campano@telecom-paristech.fr, jessica.durand@telecom-paristech.fr, chloe.clavel@telecom-paristech.fr

Abstract

Engagement is an important feature in human-human and human-agent interaction. In this paper, we investigate lexical alignment as a cue of engagement, relying on two different corpora : CID and SEMAINE. Our final goal is to build a virtual conversational character that could use alignment strategies to maintain user's engagement. To do so, we investigate two alignment processes : shared vocabulary and other-repetitions. A quantitative and qualitative approach is proposed to characterize these aspects in human-human (CID) and human-operator (SEMAINE) interactions. Our results show that these processes are observable in both corpora, indicating a stable pattern that can be further modelled in conversational agents.

Keywords: engagement, human-agent interaction, corpus analysis

1. Introduction

Engagement can be considered as “the value that a participant in an interaction attributes to the goal of [...] continuing the interaction” (Poggi, 2007 in Peters et al., 2005). It is an important feature in human-human interaction, as well as in human-agent interaction (between a human and a virtual character). In this paper, we aim at characterizing verbal alignment processes that can be considered as cues of engagement: shared vocabulary and other-repetitions. Our final goal is to build a virtual conversational character able to detect and display such cues, in order to maintain user's engagement.

Alignment refers to the endorsement of an activity in progress (Gorisch et al., 2012). It can be defined as the convergence of two participants on several aspects. Low-level aspects include lexical repetitions (Ward and Litman, 2007) (Bertrand et al., 2013), linguistic style (Niederhoffer and Pennebaker, 2002), or speech activity (Campbell and Scherer, 2010). High-level aspects include concepts (Brennan and Clark, 1996), or inter-comprehension (Shockley et al., 2009). Low-level and high-level aspects are interdependent, and often analyzed simultaneously.

Participants use lexical repetition to show their involvement (Tannen, 1992), a concept that seems very close to engagement. As engagement is necessarily a two-participant process, in this paper we focus on *other-repetitions* (or allo-repetitions). Other-repetition occurs when a speaker repeats one or several words said previously by another speaker. Other-repetition has several functions, as conveying the understanding of an emotional stance (Svennevig, 2004), the receipt of a message (Perrin et al., 2003) (Tannen, 1992) (Bazzanella, 2011), or the contribution to the topic of the conversation (Halliday, 1967) cited in (Tannen, 1992). These functions allow us to assess whether participants are involved in the interaction process. Hence, other-repetitions could be used in human-agent interaction to foster user's engagement. This is why we need to study their features, and how they are related to user's engagement.

In this paper, we investigate : (i) the quantitative evolution of shared vocabulary between two speakers, and (ii) the detailed occurrences of other-repetitions, that is, which words are other-repeated and when. Shared vocabulary is a measure related to other-repetitions, allowing us to identify general trends over time. Our study is grounded on two corpora: the SEMAINE corpus (McKeown et al., 2010), containing human-agent interactions, and the CID corpus (Bertrand et al., 2008), containing human-human interactions. The outline of the paper is as follows: the corpora are described and compared in Section 2.. Section 3. proposes statistics on shared vocabulary, giving insights on some alignment processes. Methods for visualising shared vocabulary and other-repetitions are proposed in Section 4., and finally an in-depth analysis of the resulting figures is provided in Section 5..

2. Corpora Overview

2.1. SEMAINE and CID

The SEMAINE database is a corpus of emotionally coloured conversations, taking place in a machine-human scenario called the Sensitive Artificial Listener (McKeown et al., 2010). We worked on the subset of SEMAINE database which is currently available. In this subset, human users interact with human operators playing the role of a virtual agent. Each session involves two speakers. Operators' responses were restricted to sentences from a script. An operator can endorse four different roles, or characters, corresponding to four different emotional styles: Prudence, even-tempered and sensible, Poppy, happy and outgoing, Spike, angry and confrontational, and Obadiah, depressive. The conversational goal of the operator is to shift the user towards one of these emotional states. To our knowledge, operators were not told to align with the user, nor was it a strategy included in the scenario. The SEMAINE corpus is relevant for the design of conversational virtual characters, since the experimental setting is close to human-agent interaction. Studying SEMAINE allows us to catch some

specificities of verbal alignment in this context.

The Corpus of Interactional Data (CID) is an audiovisual recording of 8 hours of French human-human conversational dialogs (Bertrand et al., 2008). The corpus provides eight sessions of two speakers interacting, with different participants in each session. Participants were asked to tell about conflicts or unusual events of their personal lives. They were free to negotiate their roles as listeners or tellers. According to the authors, this kind of task is known to promote the occurrence of reported speech, a type of aligned response. The CID corpus is relevant for the study of verbal alignment and other-repetitions, as shown by previous research work (Bertrand et al., 2013), (Guardiola and Bertrand, 2013), (Bigi et al., 2014). It will help us to draw a comparison with SEMAINE, and determine whether some alignment processes can be found in both corpora.

SEMAINE and CID corpora have several differences as well as common features. SEMAINE is oriented by operator's task (shift the user to a given emotional state) and sentences, while in CID corpus the goal was to obtain a free and natural conversation. The common feature between the two corpora is that participants regularly engage in a storytelling activity (e.g. telling how was the day, or telling an unusual event), and each session involves two speakers.

In the following sections, we compare the two corpora in terms of speech activity and vocabulary.

2.2. Speech Activity

Speakers' speech activity is an important feature for the study of alignment. The display of speech activity has been previously used to analyze dominance changes in a conversation (Campbell and Scherer, 2010). In our study, we use the display of speech activity to compare the distribution of conversational turns between CID and SEMAINE corpora. We computed the speech activity distribution for each participant per session¹, and we used the LISA tool², in order to visualise this distribution (see Figure 1).

Figure 1 shows a difference between the two corpora: speech activity is delimited in big blocks in SEMAINE, while it is more interleaved in CID. It is likely because in CID, participants spontaneously take turns, while in SEMAINE, a participant waits for the operator to ask a question/make a commentary, and then takes turn.

Speech activity is evenly distributed between the two speakers in CID, while it is not in SEMAINE (the user occupies 65.5% of the total speech time). It is likely because in SEMAINE, only the user is a teller, whereas in CID the two participants endorse the role of a teller. This distribution also varies according to operator's role in SEMAINE. We computed the percentage of users' speech time for all sessions corresponding to a specific operator's role. The minimal percentage is obtained for Obadiah (60.6%), and the maximal is obtained for Prudence (70.4%). This may be due to the respective personalities of these agents (played by human operators): the even-tempered sensible nature of

Prudence may lead the user to talk more, while the depressive mood of Obadiah may lead the user to talk less.

2.3. Specificity of Speakers' Vocabulary

Preprocessing In order to be analyzed, the corpora were pre-processed according to the following procedure. For each transcript of SEMAINE, sentences were lemmatized and POS-tagged with Python NLTK³. We built supplementary data where function words were removed⁴, as well as words belonging to the english stopwords list resource from Université de Neuchâtel⁵ (571 words). This list was established on the guideline proposed by Fox (1989). This manipulation allows us to discard the most frequently used English words, which is useful in the study of word repetitions. The same procedure was applied to CID, except lemmatization and POS-tagging were performed by the Tree Tagger tool⁶, and the french stopwords list from Université de Neuchâtel was used (463 words) (Savoy, 1999).

A major difference between the two corpora is the duration of an interaction: 5 minutes in SEMAINE, and 1 hour in CID. Consequently, the mean vocabulary size per session for CID ($\mu = 1744$, $\sigma = 244.18^7$) is higher than for SEMAINE ($\mu = 257$, $\sigma = 69.78$).

In order to know whether participants use a specific vocabulary in each corpus, we used TF-IDF statistic (Term Frequency-Inverse Document Frequency). TF-IDF was computed on lemmatized data with stopwords removed. All the words said by a specific speaker is considered as a document in terms of TF-IDF. In SEMAINE, operators playing a specific role, eg Poppy, are considered as a unique speaker. This means that all the words said by Poppy in various sessions represent one document. Measures were performed separately for each corpus.

TF-IDF results show that each participant, either in CID or SEMAINE, effectively uses a specific vocabulary. This indicates that they contribute to specific topics and possess their own linguistic style. Table 1 includes the words with the highest TF-IDF score for each corpus. Regarding SEMAINE, words with the highest TF-IDF scores for each operator role are highly representative for their affective style. As an example, the 20 top-ranked words for Obadiah (depressive) include "*miserable, suffering, disappointed, hurt, darkness, sad, terrible, worse, depressed*", whereas those for Poppy (happy) include "*excellent, aha, cool, exciting, fun, hapiness, holiday*". For SEMAINE users, the top-ranked words are related to topics or user's lexical style. A speaker's lexical style can be characterized by the frequent occurrences of specific adverbs or interjections, that are overly used by the speaker compared to other speakers. The 10 top-ranked words obtained for one of the users include "*ya, yeah*" (user's lexical style) and "*shower, dress, weekend, holiday*" (topics).

³<http://nltk.org/>

⁴articles, prepositions, conjunctions, pronouns

⁵<http://members.unine.ch/jacques.savoy/clef/index.html>

⁶<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

⁷the mean is denoted as μ , and the standard deviation as σ .

¹we performed our measures on the sessions for which transcription files are provided

²developed by computer science laboratory of Avignon <http://lia.univ-avignon.fr/>

CID ($\alpha = 0.009$)		
word	score	speaker
globalement (on the whole)	0.00958	NH
bijou (jewel)	0.00928	NH
gamin (kid)	0.00950	AC
mage (sorcerer)	0.01605	SR
rembourseur (repay)	0.00963	SR
buffet (buffet)	0.01755	LL
conflit (conflict)	0.01699	LL
dessin (drawing)	0.01318	LJ
fouille (excavations)	0.01065	LJ
violet (purple)	0.00930	MG
conflit (conflict)	0.01273	EB
doudou (cuddly toy)	0.01481	ML
couche (diaper)	0.00931	ML

SEMAINE ($\alpha = 0.025$)		
word	score	operator role
miserable	0.02561	Obadiah
excellent	0.02579	Prudence
fool	0.03250	Spike
annoyed	0.03123	Spike
excellent	0.03417	Poppy
aha	0.03379	Poppy

word	score	user
bloody	0.03224	2
beautiful	0.029622	9
shipped	0.02506	11
nope	0.02506	11
hang	0.02506	11
as	0.02506	11
language	0.03233	12
room	0.04196	16

Table 1: TF-IDF measure: top-ranked words (score $> \alpha$) in SEMAINE and CID corpora

In CID, we also find words related to topics and speakers’ lexical style at the top of the list. The 10 top-ranked words for one of the speakers include “*bijou (jewel)*, *liqueur (liquorice)*” (topics) and “*globalement (on the whole)*” (speaker’s lexical style). There are also some high score words related to affect, but they are less numerous than in SEMAINE, and not ranked at the top of the list. This shows that when speakers are not encouraged to talk about affect, they do not use many affect words. However, speakers in CID regularly express their opinion, corresponding to some high-score words: “*ignoble*” (awful, 25th rank for speaker NH when NH words are ordered by TF-IDF score) “*injuste*” (unfair, 170th rank for NH), “*sympathique*” (nice, 54th rank for AB), “*extraordinaire*” (extraordinary, 64th rank for AB), “*terrible*” (terrible, 68th rank for AB). These words corresponds to appraisal expressions, as defined by Martin and White (2005) (see Langlet and Clavel (2014) for more details on definitions of affect, judgment and appreciation and studies on polarity accordance in the SEMAINE corpus).

3. Statistics on Shared Vocabulary

Shared vocabulary between speakers is a dimension of verbal alignment which is related to other-repetitions. It provides a more general overview than word occurrences. In this section, we propose a formal definition of shared vocabulary, and present main statistics on shared vocabulary in the two corpora.

3.1. Definition

Given a time interval I in an interaction, and two participants p_1 and $p_2 \neq p_1$, the *shared vocabulary* is the set containing each word used by both p_1 and p_2 in I . We define $voc_I(p)$ as the vocabulary used by p in I . Then, the shared vocabulary between p_1 and p_2 in I is defined as:

$$SV_I(p_1, p_2) \in [0, 1] = \frac{card(voc_I(p_1) \cap voc_I(p_2))}{card(voc_I(p_1) \cup voc_I(p_2))}$$

The higher is $SV_I(p_1, p_2)$, the more vocabulary is shared between p_1 and p_2 in I .

For each participant $p \in \{p_1, p_2\}$, we define a supplementary measure representing the vocabulary that p shares relating only to its own vocabulary. This allows us to compute whether a user shares more of its vocabulary than another. This measure is defined as :

$$SV_I(p_1) \in [0, 1] = \frac{card(voc_{p_1} \cap voc_{p_2})}{card(voc_{p_1})}$$

The higher is $SV_I(p_1)$, the more p_1 shares its own vocabulary with p_2 in I .

3.2. Main statistics

Notations: in the following results, the mean is denoted as μ , and the standard deviation as σ . To test the significance of the results we used the Student’s t-test (denoted as t).

3.2.1. Shared Vocabulary and Speech Activity

For each corpus, we computed the shared vocabulary mean score as:

$$\frac{1}{N} \sum_{s \in S} SV_{I^s}(p_1^s, p_2^s)$$

where S is the set of corpus sessions, I^s is the session time interval, p_1^s, p_2^s the two speakers in session s , and N the number of sessions in the corpus.

We found that speakers share vocabulary in either SEMAINE or CID. However, the mean $SV_I(p_1, p_2)$ score for CID is slightly higher than the score for SEMAINE (CID : $\mu = 0.29, \sigma = 0.02$; SEMAINE : $\mu = 0.258, \sigma = 0.066$). This is likely because in SEMAINE operators speak less than users (as previously shown in section 2.2.). Hence, the intersection between user vocabulary and operator vocabulary is smaller. In CID, the speech activity of each speaker is more evenly distributed.

In order to know whether a user shares more of its own vocabulary than an operator, we also computed the shared vocabulary mean score for users as:

$$\frac{1}{N} \sum_{s \in S} SV_{I^s}(user_s)$$

where $user_s$ is the user of session s , and N the number of sessions in the corpus. We also computed the shared vocabulary mean score for operators (same formula).

Results indicate that the mean $SV_I(p_1)$ scores for operators and for users is different (operators : $\mu = 0.567, \sigma = 0.115$; users $\mu = 0.344, \sigma = 0.119$). This difference is significant ($t = 11.204, p < 0.001$). This means that an operator shares more of its own vocabulary with a user than a user

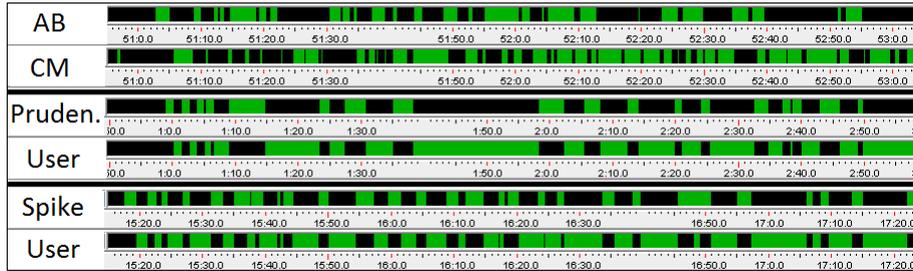


Figure 1: Showing speech activity per participant on 3 minutes (LISA tool) - 1st session : CID, 2nd and 3rd sessions : SEMAINE

shares its vocabulary with an operator. This could be explained by several factors : (i) the user speaks more than the operator, hence uses more vocabulary, and (ii) the operator has to play a role, hence uses a restrained vocabulary specific to its role. Then, in human-agent interaction, when a user has a low $SV_I(p_1)$ score compared to an agent, it should not be interpreted as a low alignment of the user towards the agent.

3.2.2. Shared Vocabulary and Operator’s Role

In SEMAINE, we also measured whether the affective style of each operator (Poppy, Prudence, Obadiah and Spike) is correlated with the $SV_I(p_1, p_2)$ score. For each role, we computed the mean $SV_I(p_1, p_2)$ score for all the sessions where the operator is playing this role. We found that the scores are different between roles (Obadiah : $\mu = 0.274$, $\sigma = 0.080$; Poppy : $\mu = 0.248$, $\sigma = 0.049$; Prudence : $\mu = 0.233$, $\sigma = 0.051$; Spike : $\mu = 0.279$, $\sigma = 0.072$). The shared vocabulary between operator and user is higher for Spike and Obadiah than for Poppy and Prudence. Hence, negative emotions could be associated with more shared vocabulary. However, these differences are not statistically significant. We have to be cautious with this hypothesis that needs to be tested on more samples. We only observe a significant result for the mean $SV_I(p_1)$ score for users on the pair *Prudence/Spike* ($t = -2.948, p < 0.01$). This means that a user shares more of its own vocabulary with Spike ($\mu = 0.40$, $\sigma = 0.13$) than with Prudence ($\mu = 0.29$, $\sigma = 0.08$). This may be because when an operator playing Spike is offending the user, the user sometimes reacts in repeating operator’s words. On the contrary, as Prudence is the most even-tempered of the four roles, it may trigger less affective alignment. The correlation between anger and alignment has been mentioned by other research work (Niederhoffer and Pennebaker, 2002).

4. Visualizing Lexical Alignment

4.1. Evolution of Shared Vocabulary

We computed the evolution of shared vocabulary measure on sliding time windows, and displayed it on figures (one figure per session). This procedure allows us to visually identify increases and decreases of shared vocabulary through a session, and to figure out where these variations could come from. One example can be seen in Figure 2. The goal is to provide a qualitative analysis of these curve variations, and to establish whether they are linked to engagement cues.

Shared vocabulary measure was computed on sliding windows with varying lengths. The principle of this approach is to fragment temporal data (uttered words) in several overlapping time frames, and to compute the shared vocabulary $SV_I(p_1, p_2)$ mean value for each time frame (see section 3.1. for details on the computation of $SV_I(p_1, p_2)$). Then, the values can be displayed on a figure. The window length l_w can be expressed according two units of duration: the number of conversational turns and the time in milliseconds. As shared vocabulary depends on a conversational process, we focused on turns. The overlap between the frames was set at 50%. We performed this computation with two different window lengths: 2 and 16 turns. 2 allows us to identify when a speaker repeated other speaker’s words in the immediate successive turn. This is useful to localize short-distance other-repetitions. After an analysis of the transcripts, we found out that 16 seems to be the maximum number of turns during which a topic is discussed in SEMAINE. According to Langlet and Clavel (2014), the average number of high-level topics per SEMAINE session is 4, which tends to confirm our observations. A window length of 16 is then useful to localize the evolution of shared vocabulary that depends on topic change.

4.2. Other-Repeated Words

For each session, we displayed on a figure the word occurrences for each speaker along a time axis. Our aim is to visually identify which words are other-repeated, and when. We recall that an *other-repetition* occurs when a speaker repeats a word previously said by another speaker (Guardiola and Bertrand, 2013).

We draw one figure for each interaction. The x axis represents time units, whereas the y axis represents words. Time units are expressed in turns, and in milliseconds (different figures were generated for each time unit). An example can be seen on figure 3. In these figures, it is possible to identify what is the distance between a word occurrence and its other-repetition. Distance is an important feature to distinguish other-repetition functions (Perrin et al., 2003).

4.3. Selected Pre-processing

The computation of figures was performed on the four types of pre-preprocessing (see Section 2.3.). We analyzed the figures, and found out that the most relevant pre-processing seems to be the lemmatized form, where function words and stopwords are removed. If such words are not removed, we observe increases in shared vocabulary figures that are

not relevant, and it alters the visibility of relevant data. As an example, we found on a figure an increase corresponding to the fact that the two participants used both the words “I”, “am” and “you” in the same time window. If such words are removed, the remaining increases are more relevant. It is then possible to clearly identify when two speakers are talking about the same topic (“sports”), or express the same emotion words (“happy”). The same principle applies to figures with other-repeated words. If function words are not removed, then a lot of irrelevant words are remaining, and identifying interesting data is more difficult. The counterpart of this approach is that some relevant displays of words could be missing.

4.4. Additional Pre-processing for CID

The CID corpus is originally segmented in InterPausal Units (IPUs). IPUs are blocks of speech bounded by silent pauses over 200 ms. In order to compare CID with SEMAINE, we converted IPUs in turns. To do so, we aggregated the IPUs of a same speaker that are not separated by the intervention of the second speaker.

5. In-depth Analysis

In this section, we present an analysis of shared vocabulary and word occurrences figures generated from SEMAINE and CID corpora (as explained in Section 4.). In order to illustrate alignment processes appearing in these figures, we present a review of the analysis of 4 SEMAINE sessions (~20 minutes), and 1 CID session (~one hour). We selected randomly 4 SEMAINE sessions with the same user and 4 different operator roles (Prudence, Poppy, Obadiah and Spike), and one CID session.

5.1. Methodology

Our analysis of each session is grounded on the following methodology. First, we analyzed the curve variations in the *shared vocabulary figure*. If an important increase or decrease was found at a given turn (e.g., turn 32), we analyzed the corresponding phase in corpus transcripts and video recordings (e.g., if a decrease is found at turn 32, the corresponding phase is from turn 24 to turn 40 because the window length is 16). Doing so, we try to find an explanation for the observed curve variations. Second, we analyzed the *word occurrence figure* by looking at the words occurring in the corresponding phase. Our aim is to know which type of words are other-repeated and if other-repetitions are related to a specific lexical field (e.g., opinion, topic).

5.2. Review of observed phenomena

We give here a list of the main phenomena observed during the analysis illustrated by examples.

User’s interest in the topic and user’s emotional involvement in SEMAINE sessions We identify two interesting phenomena that seem to be associated to the peaks of shared vocabulary curve: the user’s interest in the topic and the user’s emotional involvement.

An example of the first phenomenon (user’s interest) can be found in Session 70 (Prudence). The figure of shared vocabulary (window length = 16 turns) shows two peaks, at turns 24 and 56 (see Figure 2) corresponding to phases

[16-32] and [48-64]. During these two phases, we observe a special interest by the user in the current topic (“holidays” in [19-32] turns and “trip” in [50-64] turns). Between these two phases, there is a gap in the shared vocabulary curve: the user is effectively not pleased about the topic of “work” and at turn 50, says that the topic is boring.

An example of the second phenomenon (user’s emotional involvement) can be found in Session 73 (Spike). There are two peaks of shared vocabulary in this session at turns 24 and 64 (figure not shown). During phase [56-72], the user tells that he is often mistaken for an american, whereas he is canadian and seems to be annoyed by this confusion. Operator empathizes with the user, using expressions “*that’s a lot*”, “*piss me off*”. These expressions are other-repeated by the user almost identically at the following turn. As the role of Spike is to make the user angry, these expressions are employed as a strategy to reinforce user annoyance. A function of other-repetitions is to express **emotional stances** (Svennevig, 2004). We also found in Section 3.2.2. that verbal alignment is stronger when operators play Spike.

Other-repetition of appraisal expressions There are numerous examples of appraisal expressions in SEMAINE and CID. We focus on 3 extracts found in the corpora thanks to word occurrence and shared vocabulary figures. In Session 70 (Prudence), during the “trip” topic, two interesting words are repeated several times by the user and the operator: the words “excessive” and “absurd” (see Figure 3). They are other-repeated in order to display the same stance. According to Bazzanella (2011), other-repetitions can be used to show that ones **share attitudes and opinions**.

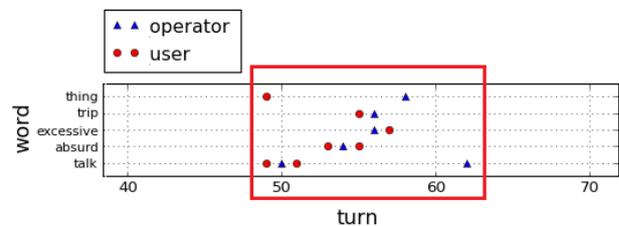


Figure 3: Word occurrences for user and operator (cut on x and y axes), session 70, SEMAINE corpus.

In session 72 (Obadiah), we found many appraisal expressions. This seems related with Obadiah affective style. In this role, the operator expresses attitudes about life and about the user, which triggers short-distance repetitions. Here are two examples:

37 - Obadiah: “*Life is hard sometimes.*”

38 - User: “(Nods). *Life can suck sometimes. I agree.*”

41 - Obadiah: “*Yeah. But you can’t be cheery all the time.*”

42 - User: “(Shakes head). *Oh God I’m not cheery (laughs). [...]*”

A speaker’s evaluation could be similar (as in turns 37-38) or opposed (as in turns 41-42) to the other speaker’s evaluation. In the word occurrence figure of this session, we also

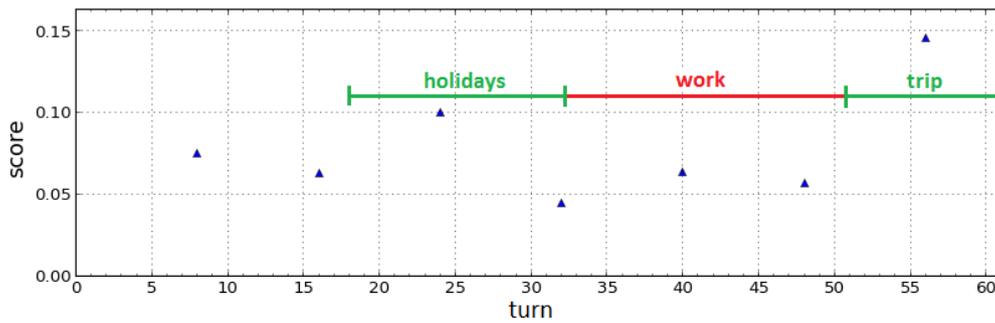


Figure 2: Evolution of shared vocabulary score, session 70, SEMAINE corpus (window length = 16 turns)

observe a large number of other-repetitions related to affect, as “happy”, “feel”, “sad”, “bored”, “interesting”. According to Martin and White (2005), these words express affect evaluations.

In CID, we observed some occurrences of the expression of surprise during the narration of a story. As an example, a peak of shared vocabulary is found when AB talks about an old woman who stayed all the night in a bedroom with an opened window and snow coming through it. CM repeats large parts of other participant sentences, showing astonishment and empathy for the old woman. This peak is shown on Figure 4. As previously said, **expressing emotional stances** as surprise is a function of other-repetitions (Svennevig, 2004).

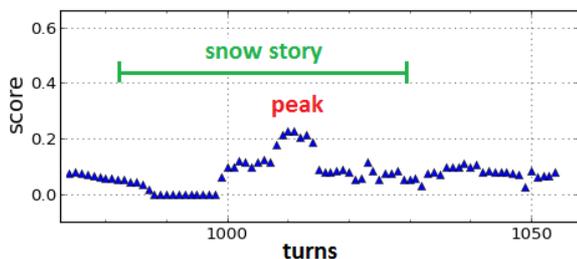


Figure 4: Peak of shared vocabulary in CID corpus

Other-repetitions and listening behavior A hearer can express that he is listening to the speaker in several ways. He may use short verbal or non verbal feedbacks, or repeat the words said by speaker. This function of other repetition is defined as **Taking Into Account function (TIA)** by Per-rin et al., 2003). We give examples of these phenomena in CID and SEMAINE.

The Figure 5 represents the evolution of shared vocabulary during the first half of a story telling in CID (turns [180-250]). AB is narrating an unusual story to CM. On this figure, we observe a peak of shared vocabulary, followed by a flat phase (low shared vocabulary). At the beginning of this story, CM repeats AB words in order to show that she takes them into account: “*feu d’artifice*” (fireworks), “*bonne soeur*” (nun). This short sharing activity is then followed by a long flat phase, where CM is narrating a story. During this phase, AB simply provides non verbal or short verbal feedbacks as “*mh mh*” or “*ouais*” (yeah).

In SEMAINE, we also observe these two types of listening behavior. In Session 71 (Poppy), there is a gap in

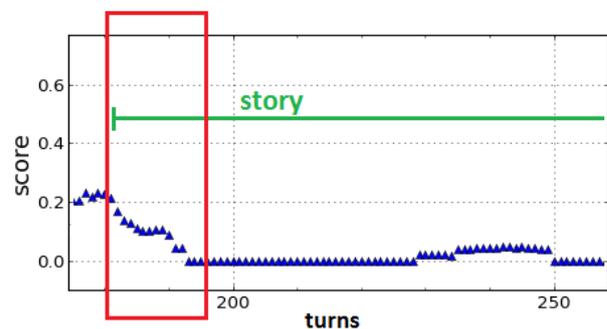


Figure 5: Evolution of shared vocabulary score, session AB-CM, extract from CID corpus (window length = 16 turns)

shared vocabulary corresponding to phase [24-32]. During this phase, the operator provides feedbacks such as “*hmh*”, “*yeah*”, “*ok*”, “*ah*”, but does not repeat user’s words. At other moments, the operator repeats what the user said, adding an evaluative stance besides the TIA function, as in “*Australia will be cool.*” after the user has talked about a trip in Australia.

Other-repetitions and definition of a word in CID In CID, participants are given the instructions to talk about unusual stories. Figure 6 shows that at the beginning of the session, both speakers use the word “*insolite*” (unusual) several times in the phase [0-50]. During this phase, speakers are discussing about the meaning of the word “*insolite*” (unusual). As shown by the figure, the repetition of this word occurs at other moments in the interaction. This indicates that speakers are trying to find unusual stories to tell about.

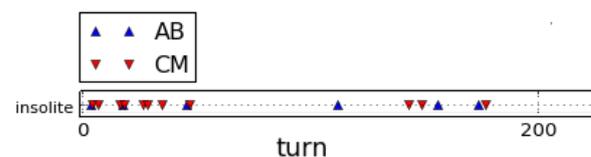


Figure 6: Occurrences of the word “insolite” (figure cut on y axis) - session AB-CM, SEMAINE corpus

5.3. Alignment Functions Found in Corpora

The previous analysis on the generated figures allowed us to identify several functions of other-repetitions:

- *express an emotional stance*
- *show that ones share attitudes, opinions*
- *take into account (TIA) what the other says*
- *clarify a concept*

Expressing an *emotional stance* and *sharing attitudes* are functions that are widely used by speakers in both corpora. According to Svennevig (2004), other-repetitions can be used to display an emotional stance, as surprise, interest, approval, etc. It is also a way of receiving information. For Bazzanella (2011), other-repetitions can be used to show that ones share attitudes and opinions.

The TIA function, as described by Perrin (Perrin et al., 2003), is the process “by which a speaker indicates that what was just said by the interlocutor has been heard and interpreted”. It is also referred to a “ratifying listenership” by Tannen (Tannen, 1992).

Expressing an appraisal and TIA are functions closely related to speakers’ engagement in dialogue: the first provides evidence of an affective involvement, and the second is a cue of speakers’ involvement in the dialogue process. We can draw the conclusion that user’s engagement could be detected according to the presence or absence of these alignment cues, and that we can model an conversational character able to produce such cues in order to show its engagement.

6. Conclusion

In this paper, we presented an analysis of verbal alignment processes related to speaker’s engagement. We focused on shared vocabulary and other-repetitions. Our aim was to determine whether such cues can be effectively found in real dialogues (CID and SEMAINE corpora), and how they are related to speakers’ engagement. Our final goal is to use this work to build a virtual agent able to detect such cues, and to produce them. We provided main statistics on the two corpora, as well as figures showing the speakers’ word occurrences, and the evolution of shared vocabulary between speakers over time. We found out that sharing vocabulary depends in part on user’s engagement towards a specific topic. In particular, we distinguished engagement as a stance, where a speaker expresses an attitude towards a target (person, event), and engagement as the contribution to the conversation, where a speaker shows that he takes into account what the other says. Future work will focus on a rule-based characterization of these two types of alignment processes. This model will be used to help the automatic detection of user engagement, and to build a virtual conversational agent able to take this engagement into account in order to select appropriate strategies.

7. Acknowledgments

The authors thank Catherine Pelachaud, Mohamed Chetouani and Candy Sidner for valuable insights and

suggestions. This work has been supported by the French collaborative project A1:1, and performed within the Labex SMART (ANR-11-LABX-65) supported by French state funds managed by the ANR within the Investissements d’Avenir programme under reference ANR-11-IDEX-0004-02.

8. References

- Carla Bazzanella. 2011. Redundancy, repetition, and intensity in discourse. *Language Sciences*, 33(2):243–254.
- Roxane Bertrand, Philippe Blache, Robert Espesser, Gaëlle Ferré, Christine Meunier, Béatrice Priego-Valverde, Stéphane Rauzy, et al. 2008. Le cid-corpus of interactional data-annotation et exploitation multimodale de parole conversationnelle. *Traitement Automatique des Langues*, 49(3):1–30.
- Roxane Bertrand, Gaëlle Ferré, Mathilde Guardiola, et al. 2013. French face-to-face interaction: repetition as a multimodal resource. *Coverbal Synchrony in Human-Machine Interaction*, page 141.
- Brigitte Bigi, Roxane Bertrand, and Mathilde Guardiola. 2014. Automatic detection of other-repetition occurrences: application to french conversational speech. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*.
- 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.
- Nick Campbell and Stefan Scherer. 2010. Comparing measures of synchrony and alignment in dialogue speech timing with respect to turn-taking activity. In *INTER-SPEECH*, pages 2546–2549.
- Christopher Fox. 1989. A stop list for general text. In *ACM SIGIR Forum*, volume 24, pages 19–21. ACM.
- Jan Gorisch, Bill Wells, and Guy J Brown. 2012. Pitch contour matching and interactional alignment across turns: An acoustic investigation. *Language and Speech*, 55(1):57–76.
- Mathilde Guardiola and Roxane Bertrand. 2013. Interactional convergence in conversational storytelling: when reported speech is a cue of alignment and/or affiliation. *Frontiers in Psychology*, 4:705.
- Michael AK Halliday. 1967. Notes on transitivity and theme in english, part 2. *Journal of linguistics*, 3(2):199–244.
- Caroline Langlet and Chloé Clavel. 2014. Modelling user’s attitudinal reactions to the agent utterances: focus on the verbal content. In *5th International Workshop on Emotion, Social Signals, Sentiment and Linked Open Data held at the 9th International Conference on Language Resources and Evaluation (LREC 2014)*.
- James R Martin and Peter RR White. 2005. *The language of evaluation*. Palgrave Macmillan Basingstoke and New York.
- Gary McKeown, Michel François Valstar, Roderick Cowie, and Maja Pantic. 2010. The semaine corpus of emotionally coloured character interactions. In *Multimedia and Expo (ICME), 2010 IEEE International Conference on*, pages 1079–1084. IEEE.

- Kate G Niederhoffer and James W Pennebaker. 2002. Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, 21(4):337–360.
- Laurent Perrin, Denise Deshaies, and Claude Paradis. 2003. Pragmatic functions of local diaphonic repetitions in conversation. *Journal of Pragmatics*, 35(12):1843 – 1860.
- Christopher Peters, Catherine Pelachaud, Elisabetta Bevacqua, Maurizio Mancini, Isabella Poggi, and Università Roma Tre. 2005. Engagement capabilities for ecas. In *AAMAS05 workshop Creating Bonds with ECAs*.
- Isabella Poggi. 2007. *Mind, hands, face and body: a goal and belief view of multimodal communication*. Weidler.
- Jacques Savoy. 1999. A stemming procedure and stopword list for general french corpora. *JASIS*, 50(10):944–952.
- Kevin Shockley, Daniel C Richardson, and Rick Dale. 2009. Conversation and coordinative structures. *Topics in Cognitive Science*, 1(2):305–319.
- Jan Svennevig. 2004. Other-repetition as display of hearing, understanding and emotional stance. *Discourse studies*, 6(4):489–516.
- Deborah Tannen. 1992. *Talking voices: Repetition, dialogue, and imagery in conversational discourse*, volume 6. Cambridge University Press.
- Arthur Ward and Diane Litman. 2007. Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora. In *Proceedings of the SLATE Workshop on Speech and Language Technology in Education*.