Alignment in a Multimodal Interlingual Computer-Mediated Map Task Corpus

Justine Reverdy, Hayakawa Akira, Carl Vogel

ADAPT Centre

Computational Linguistics Group

School of Computer Science and Statistics, Trinity College Dublin, Ireland

{reverdyj, campbeak, vogel}@tcd.ie

Abstract

This work presents an assessment of interlocutor alignment using a semi-automated method in the context of multimodal interlingual (English-Portuguese) computer-mediated interactions. We study the adaptation phenomenon (also known as convergence behaviour and alignment behaviour) by looking at verbal repetition at different levels of linguistic representation. Since alignment behaviour has already been analysed in direct human-to-human and in human-to-agent dialogues, one may wonder whether the same behaviour is observed in interlingual computer-mediated communication. First, we compare repetitions patterns in task-oriented dialogues of human-to-human communication (HCRC Edinburgh Map Task corpus) and interlingual computer-mediated human-to-human communication (ILMT-s2s corpus), for eye-contact and no eye-contact scenarios. Secondly, we study the relation between the cognitive state of the subject, and the alignment process in interlingual computer-mediated communication settings. Results show that above chance repetitions, signalling verbal alignment, are present in both direct human-to-human communication and interlingual computer-mediated interactions, and that interlingual computer-mediated setting yields significantly more self-repetitions than direct human-to-human interactions. Also, in interlingual computer-mediated communication, a lack of alignment cues for long sequences correlated with a high amount of negative cognitive states in the eye-contact setting, implying a potential lack of mutual understanding.

Keywords: alignment, mutual understanding, task-oriented, computer-mediated, interlingual communication

1. Introduction

Interlocutor alignment (repetition of linguistic choices) is said to be an important part of human-to-human communication. In particular, the Interactive Alignment Model (Pickering and Garrod, 2004) has been taken as the basis of various works exploring this phenomenon. Different levels of linguistic representation reflect this alignment, for example in lexical choices and syntactic structures (Branigan et al., 2000; Reitter and Moore, 2007; Garrod and Anderson, 1987) or prosodic features (Giles et al., 1991).

The results of those studies show evidence that interlocutors tend to align their representation of the world to establish mutual understanding throughout conversation (Turnbull, 2003) that is sufficient for the purpose of the exchange (Newlands et al., 2003, p. 327). The achievement of mutual understanding is never entirely certain; however, interlocutors can achieve a state in which they lack direct evidence of misunderstanding (Taylor, 1992), i.e., achieve a level of understanding that is adequate to accomplish a given task (depending if the task type requires this achievement) (Brown et al., 1985). Repetition mechanisms are central in the alignment process and hold multiple functions (Tannen, 2007). They can signal understanding and by contrast, in other cases, can express a misunderstanding that will induce repair. The presence of repetitions is also an indicator of involvement or engagement in an interaction.

In two previous studies we conducted (Reverdy and Vogel, 2017a; Reverdy and Vogel, 2017b) using the data of the HCRC Edinburgh Map Task corpus (Anderson et al., 1991), we reported that in task-based interaction, repetitions that occur ‘above chance’ have an impact on task-success, results which are consistent with other findings in direct human-to-human communication (Branigan et al., 2000; Reitter and Moore, 2007; Nenkova et al., 2008).

Using the map-task setting, a study comparing face-to-face and video-mediated interactions (O’Malley et al., 1996, p. 177) suggested that “when speakers are not physically co-present, they are less confident in general that they have mutual understanding [...], and therefore over-compensate by increasing the level of both verbal and non-verbal information”. Other studies about the alignment process with a virtual agent reported evidence of exaggerated alignment when the speakers thought they were talking to a machine (Branigan et al., 2010; Dubuisson Duplessis et al., 2017). Previous experiments have also found that dialogue acts used by the subjects during task-oriented computer-mediated communication differ substantially from direct human-to-human communication, with backchannel utterances (acknowledging understanding) reduced significantly in computer-mediated interlingual communication (Hayakawa et al., 2016b).

Another study examined alignment in machine-translated communication, but in a de-contextualized setting (Schneider and Luz, 2011), including a Wizard-of-Oz experiment where participants were asked to answer machine-translated questions. Half of the questions contained translation mistakes resembling ones an MT system would produce. Their results indicate that people align their answer and reproduce the obvious errors (translation mistakes), assuming that the speech-to-speech machine translation (S2S-MT) system would understand them better. To the best of our knowledge, alignment has yet to be studied where the communication is mediated by an S2S-MT system, between two people who are aware that they are interacting with each other, in particular in the context of a map-task, where specific lexical items need to be transmitted in order to achieve a common goal.

Therefore, we see a need to extend these studies to

computer-mediated communication to verify how alignment through repetition changes in this new communication style. In addition, the speaker's cognitive state could be an identifier of smooth or problematic communication. For example, results of a study in the context of call centres show that customers' frustration, irritation or surprise (that one could define as negative cognitive states) , have a negative impact on communication. The call centre staff would try to reduce the customers' negative emotional attitudes to ease the interaction and resolve the customers' issues (Botherel and Maffiolo, 2006, p. 3).
In this paper, we exploit two multimodal corpora to observe repetition of linguistic choices as cues of an alignment process. (i) We first compare direct human-to-human communication with interlingual computer-mediated communication to verify if alignment is exaggerated in computer-mediated interlingual communication. The results show that the method detected equivalent cues of alignment in both direct human-to-human and interlingual computer-mediated communication settings, with the latter displaying significantly more self-repetition than direct human-to-human communication. (ii) Secondly, we emphasize the possible role of repetitions in relation with the cognitive states of the subjects within computer-mediated interlingual communication. In those settings, we found that the lack of alignment cues for long sequences correlated with high amounts of negative cognitive states, pointing to possible communication problems (lack of mutual understanding).

2. Data Set

Data from two multimodal corpora that use the Map Task technique to elicit spontaneous communicative behaviour was used. For the direct human-to-human communication, we used a subset of 16 dialogues from the HCRC Map Task corpus (Anderson et al., 1991), and for the computer mediated interlingual communication, we used all 15 dialogues from the ILMT-s2s corpus (Hayakawa et al., 2016c), see Table 1. The subjects were assigned the role of Information Giver (IG) or Information Follower (IF) and each given a map containing similar landmarks. The IG had a map with a route drawn along the landmarks with a START and a FINISH indicated, and was tasked with guiding the IF through a map not displaying FINISH.

	HCRC (Subset)	ILMT-s2s	
Language	English	English	Portuguese
Tokens	22,106	13,761	12,671
Turns	3,790	2,310	2,236
SELF REP	2,448	3,877	2,306
OTHER REP	2,653	2,407	1,107

Table 1: HCRC Map Task and ILMT-s2s Corpora Summary; SELF REP and OTHER REP (see definition § 3.) are given for the linguistic representation level token only.

2.1. The HCRC Map Task corpus

The HCRC Map Task corpus consists of 128 English dialogues of direct human-to-human task based interactions. The recordings were split in two settings, with half the subjects being able to see their interlocutor's face (i.e., with eye-contact), while the other half had screens placed between them (i.e., without eye-contact). To standardise the data, only dialogues that used the same maps (maps 1 & 7) as those used in the ILMT-s2s corpus (§ 2.2.) were kept for this study, resulting in a total of 16 out of the 128 (half male, half female in both the main corpus and the subset).

2.2. The ILMT-s2s corpus

As with the HCRC Map Task corpus, the dialogues use the map task technique, but with a difference that the subjects are located in different rooms, speak different languages to each other and communicate via a Speech-to-Speech Machine Translation (S2S-MT) system — the ILMT-s2s System. The ILMT-s2s corpus consists of fifteen dialogues between fifteen English, and fifteen Portuguese subjects (16 females, 14 males). The maps that are used are the same as the HCRC Map Task corpus, in their original version for the English speakers, and translated for the Portuguese speaking subjects. The ILMT-s2s System is a rapidly built system that uses off-the-shelf components — the Google Speech API for Automatic Speech Recognition (ASR), the Microsoft Bing translation service for Machine Translation (MT), and the Apple system voices provided with Mac OS X computers for Text-to-Speech synthesis (TTS) — to perform the S2S-MT. The corpus was annotated for the cognitive states of Frustration, Amusement,[1] and Surprise, for each speaker in all the dialogues, with the assessment made through video and audio modalities. The inter-coder agreement for the labels was calculated[2] and the results are well above .6. A user survey was also conducted to collect the user's appreciation of the system. Each question follows a 7 point Likert scale ranging from '1 – Strongly disagree' to '7 – Strongly agree', designed in such a way that the more they agreed to the statement, the more positive their experience was. Due to the push-to-talk activation method of the system, subjects did not only talk to their interlocutor (*On-Talk*), but also spoke out loud to themselves and other people in the room (*Off-Talk*) (Hayakawa et al., 2016a). To standardise the data between corpora, only *On-Talk* was used for the analysis.

3. Method

We counted the repetition of tokens of a contribution and the immediately preceding contribution, that we assimilated as a dialogue turn of each speaker (Vogel and Behan, 2012; Vogel, 2013). A REGISTER is created for each participant, containing her or his most recent contribution. For each dialogue turn, the REGISTER is populated with counts of each repetition of a token, for other-repetitions (repetition of a token uttered by the other participant — OTHERSHARED) and self-repetitions (SELFSHARED). Tokens are counted as n-grams, up to $n = 5$. The n-grams length was divided into three length types — N: n = All ($n = 5$); N1: $n = 1$; N2+: $n > 1$ (from 2 to 5, long sequences). In each dialogue, the turns are then randomly re-ordered

[1]We note that Amusement was considered negative for English speaking subjects, as it was a reaction to high word error rate utterances output (Hayakawa et al., 2017).

[2]Using the modified kappa feature of ELAN (Wittenburg et al., 2006) version 4.9.0's "Inter-Annotator Reliability…" function.

ten times. This resulted in ten randomly ordered dialogues where other and self-repetitions were counted again. In the direct human-to-human dialogues, the count was carried out between the utterances of the two human subjects. However, for the computer-mediated dialogues, the count was carried out within the same language — the utterances from the English speakers are coupled with the English translation of the Portuguese speakers utterances and vice-versa, which created two fully monolingual dialogues.

A pre-process labelling, designed to measure five different levels of linguistic repetition types, was applied: (i) Token, (ii) Lemma, (iii) Part-Of-Speech (POS), (iv) a combination of Lemma with POS, and (v) a combination of Token with POS. Data from the HCRC Map Task corpus and the English dialogues of the ILMT-s2s corpus were labelled with the TreeTagger English training set (Schmid, 1994), while the Portuguese dialogues of the ILMT-s2s corpus were labelled using the TreeTagger tagset proposed by Pablo Gamallo (Gamallo and Garcia, 2013). The aim is to observe if a significant difference is identified between the actual dialogues and the randomized dialogues, using the statistical test described below.

To verify if there was a difference in the subject repetition patterns in the two corpora, the single-step Tukey HSD multiple comparison test was performed using a general linear model with a binomial error family (Bretz et al., 2016). The null hypothesis for the test was as follows:

$H_0 : Random.Speaker.Level.N - Actual.Speaker.Level.N \geq 0$

The null hypothesis (H_0) states that the difference between the amount of repetitions in the randomized dialogues and the actual dialogues should equal (or exceed) zero if repetitions are simply due to chance. If rejected, the assumption is that a potential role in the communication could be accepted. For each dialogue, the model was computed and dialogues with repetitions 'above chance' or not were identified: (i) per speaker (IG: Information Giver, IF: Information Follower), (ii) per n-gram (All n-grams [up to length 5]; N1: $n = 1$ [length 1]; N2+: $n > 1$ [length 2 to 5]), (iii) per type of repetition (OTHERSHARED and SELFSHARED), and (iv) per linguistic Level: TOKEN (L1), LEMMA (L2), LEMMA+POS (L3), POS (L4), TOKEN+POS (L5). This allowed us to observe a rate of H_0 rejection, defined as the "number of actual rejections of the null hypothesis" over the "number of possible rejections of the null hypothesis" in each categories. We compared the rates of rejection of H_0 in the two corpora, and the combinations of those tests is the basis of our meta-analysis. Since the two corpora contained dialogues with and without eye-contact, and the ILMT-s2s corpus is annotated for cognitive states and two languages, we observed the rates of rejections in relation with those conditions.

4. Results

4.1. Human-to-Human vs Computer-Mediated

The null hypothesis (H_0), with the threshold of $p \geq 0.05$, was rejected 233 times out of 300 for OTHERSHARED and 273 times out of 300 for SELFSHARED in the ILMT-s2s corpus across all linguistic levels while in the data from the HCRC Map Task, OTHERSHARED was rejected 111 times out of 160 and SELFSHARED was rejected 25 times out of 160 (Table 2). This reveals a considerable difference in the rejection rate for SELFSHARED repetitions between the direct human-to-human dialogues of the HCRC Map Task corpus (25/160 $= 0.15$) and those of the ILMT-s2s corpus (273/300 $= 0.91$), with SELFSHARED repetitions happening 'above chance' more often in the computer-mediated corpus. A Mann-Whitney-Wilcoxons test found that across all linguistic levels, the number of SELFSHARED repetitions is significantly different ($p = 2.686e - 06$) between the HCRC Map Task (with an average rejection of $\overline{x} = 2.5$) and the ILMT-s2s corpus (with an average rejection of $\overline{x} = 13.65$). However, no significant difference ($p = 0.9636$) was found between the two corpora concerning OTHERSHARED repetitions at level n-grams = All, both corpora showing a high rate of rejection of H_0. No significant difference was found between the two corpora in terms of speaker role, language spoken, and eye-contact modality at level n-grams = All.

Lng	SHARED	Role	L1	L2	L3	L4	L5	*M*
ILMT-s2s English n-grams = All								
Eng	OTHER	IG	12	12	12	11	12	11.8
Eng	OTHER	IF	12	12	13	9	13	11.8
Eng	SELF	IG	14	14	14	13	14	13.8
Eng	SELF	IF	14	14	14	11	14	13.4
H_0 rejection: 254 / 300 (OTHER: 118 / 150, SELF: 136 / 150)								
ILMT-s2s Portuguese n-grams = All								
Por	OTHER	IG	13	12	13	10	13	12.2
Por	OTHER	IF	12	12	12	6	12	10.8
Por	SELF	IG	14	15	15	14	14	14.4
Por	SELF	IF	14	14	14	9	14	13
H_0 rejection: 233 / 300 (OTHER: 115 / 150, SELF: 137 / 150)								
HCRC Map Task n-grams = All								
Eng	OTHER	IG	11	12	10	4	6	8.6
Eng	OTHER	IF	15	14	14	10	15	13.6
Eng	SELF	IG	**2**	**2**	**3**	**0**	**2**	**1.8**
Eng	SELF	IF	**4**	**2**	**4**	**2**	**4**	**3.2**
H_0 rejection: 136 / 320 (OTHER: 111 / 160, **SELF: 25 / 160**)								

Table 2: Rejection count of H_0 for levels L1 to L5 and mean (*M*) values in the ILMT-s2s corpus and HCRC Map Task corpus for all n-grams. For each dialogue at each level, the number of possible H_0 rejection is 15 in the ILMT-s2s corpus, and 16 in the HCRC Map Task corpus.

4.2. Within Computer-Mediated Interactions

No impact of 'above chance' repetition in relation to the cognitive states of the participants was found at n-grams length n = All (count listed in Table 3). However, differences appeared for OTHERSHARED repetitions of Portuguese (IF) at n-gram length n >1 (N2+) in "Eye-Contact" conditions (Table 4). While in all other settings the rate of rejections of H_0 remains high, the Portuguese IF speakers did not repeat the English speakers' token in the same proportion in the "Eye-Contact" condition.

This relation is highlighted with Pearson's standardized residuals from log-linear models in Figure 1. For long sequences of n-gram repetitions (N2+), we observe that when there is Eye-Contact, the Portuguese speakers show higher levels of negative cognitive states than expected when they are at the same time not repeating the English speaker.

Role	IF			IG			Total
Cog.	Fru	Sur	Amu	Fru	Sur	Amu	
Eng	67	57	220	103	54	263	764
Por	290	137	113	210	105	184	1039
Total	884			919			1803

Table 3: Number of Cognitive States per Subject Role (Information Follower, Information Giver), Spoken Languages (English, Portuguese) and Cognitive State Type (Frustrated, Surprised, Amused) in the ILMT-s2s corpus

Lng	SHARED	Role	L1	L2	L3	L4	L5	M
With Eye-Contact n >1 (N2+)								
Eng	OTHER	IG	6	6	6	6	6	6.0
Eng	OTHER	IF	6	6	5	5	5	5.4
Eng	SELF	IG	7	7	7	7	7	7.0
Eng	SELF	IF	8	8	8	6	6	7.2
Por	OTHER	IG	5	4	5	4	5	4.6
Por	OTHER	IF	**3**	**4**	**4**	**3**	**2**	**3.2**
Por	SELF	IG	7	7	7	7	7	7.0
Por	SELF	IF	7	7	6	5	6	6.2

Table 4: Rejection count of H_0 for levels L1 to L5 and mean (M) values. In each case the number of possible H_0 rejection is 8 (modality: eye-contact).

Meanwhile they show less frustration than expected if they repeat the English speaker for long sequences (N2+).

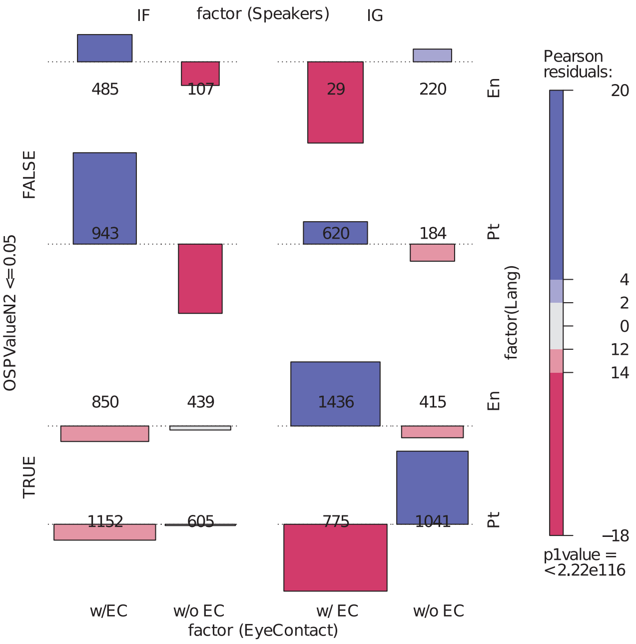

Figure 1: Association Plot of significant OTHERSHARED residuals (TRUE: p <=0.05 — FALSE: p > 0.05) for n-gram>1 (N2+), Subject Role (IG: Information Giver—IF: Information Follower), Eye-Contact (w/ EC: with Eye-Contact—w/o EC: without Eye-Contact), and Language Spoken (En: English—Pt: Portuguese)

The distributions of negative cognitive states was found significantly different between 'above chance' and non-'above chance' OTHERSHARED repetitions for the Portuguese IF speakers at n-gram>1 level (W = 883, p-value = 0.027).
The low rate of N2+ repetitions detected is echoed in the user survey conducted in the ILMT-s2s corpus. The Portuguese speakers (IF) in "Eye-Contact" conditions showed the lowest appreciation of the system (Median score = 3.0; Overall Median score = 5.0), which correlates with a high amount of negative cognitive states for those speakers.

5. Discussion

The high rate of 'above chance' OTHERSHARED repetition in the computer mediated dialogues of the ILMT-s2s corpus indicates that alignment occurs in at least the same proportion as in direct human-to-human communication. We did not find evidence of its' exaggeration with the method, as it detected equally high alignment cues in direct human-to-human communication. However, 'above chance' repetitions occurred at all linguistic levels at a high rate in the ILMT-s2s corpus, for both OTHERSHARED and SELFSHARED. This is different from the direct human-to-human dialogues where 'above chance' SELFSHARED repetitions occurred at a much lower rate. This high rate of SELFSHARED repetition could be attributed to the perceived difficulty for the speakers to have their utterance properly recognized by the ASR and correctly translated to their interlocutor, hence their tendencies to repeat themselves more. The high rate of repetition, in both types (OTHER and SELF), in this interlingual computer-mediated corpus, follows past findings that suggest strong alignment in human-computer interaction. To the best of our knowledge, this is the first time that a method of assessing alignment, by counting repetition, has been applied to dialogues of interlingual computer-mediated task-based communication.
Secondly, a relation emerged within the computer-mediated dialogues, between negative cognitive states and low 'above chance' repetitions of long sequences. Portuguese speakers in eye-contact conditions had a higher than expected negative cognitive states which also related to their low appreciation of the system. Previous work suggested that exaggerated alignment toward a system was detrimental to the interaction since the subjects also repeated translation errors (Schneider and Luz, 2011). Our findings show that the lack of alignment of long token sequences in video conditions indicates problematic interactions.

6. Conclusion

We note that even if the small size of the two corpora prevents us from making too broad a statement, the repetitions patterns detected by the automatic method present S2S-MT software design cues that constitute another step toward aiding human-to-human communication when interacting through machine translation. One might wonder if the reason that differences appeared between English and Portuguese speakers could be interpreted as a cultural difference. This could be examined in the future by comparing other language pairs and/or larger data sets.

7. Acknowledgements

This research is supported by Science Foundation Ireland through the Research Centres Programme (Grant 13/RC/2106) in the ADAPT Centre (www.adaptcentre.ie) at Trinity College Dublin. The ADAPT Centre for Digital Content Technology is co-funded under the European Regional Development Fund.

8. Bibliographical References

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., and Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech*, 34(4):351–366.

Botherel, V. and Maffiolo, V. (2006). Regulation of emotional attitudes for a better interaction: Field study in call centres. In *Proceedings of 20th International Symposium on Human Factors in Telecommunication*, Sophia Antipolis, France.

Branigan, H. P., Pickering, M. J., and Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75(2):B13–B25.

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

Bretz, F., Hothorn, T., and Westfall, P. (2016). *Multiple comparisons using R*. CRC Press.

Brown, G., Anderson, A., Shillcock, R., and Yule, G. (1985). *Teaching talk: Strategies for production and assessment*. Cambridge University Press.

Dubuisson Duplessis, G., Clavel, C., and Landragin, F. (2017). Automatic measures to characterise verbal alignment in human-agent interaction. In *Proceedings of SIGDIAL 2017*, pages 71–81.

Gamallo, P. and Garcia, M. (2013). FreeLing e TreeTagger: um estudo comparativo no âmbito do Português. Technical report, Universidade de Santiago de Compostela.

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

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

Hayakawa, A., Haider, F., Luz, S., Cerrato, L., and Campbell, N. (2016a). Talking to a system and oneself: A study from a Speech-to-Speech, Machine Translation mediated Map Task. In *Proceedings of Speech Prosody 2016*, pages 776–780, Boston, MA, USA. ISCA.

Hayakawa, A., Luz, S., and Campbell, N. (2016b). Talking to a System and Talking to a Human: A Study from a Speech-to-Speech, Machine Translation Mediated Map Task. In *Proceedings of INTERSPEECH'16*, pages 1422–1426, San Francisco, CA, USA. ISCA.

Hayakawa, A., Luz, S., Cerrato, L., and Campbell, N. (2016c). The ILMT-s2s Corpus — A Multimodal Interlingual Map Task Corpus. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of LREC 2016*, pages 605–612, Portorož, Slovenia. ELRA.

Hayakawa, A., Vogel, C., Luz, S., and Campbell, N. (2017). Perception Changes With and Without the Video Channel: A Study from a Speech-to-Speech, Machine Translation Mediated Map Task. In *Proceedings of CogInfoCom 2017*, pages 401–406, Debrecen, Hungary. IEEE.

Nenkova, A., Gravano, A., and Hirschberg, J. (2008). High frequency word entrainment in spoken dialogue. In *Proceedings of ACL HLT 2008*, pages 169–172. Association for Computational Linguistics.

Newlands, A., Anderson, A. H., and Mullin, J. (2003). Adapting communicative strategies to computer-mediated communication: an analysis of task performance and dialogue structure. *Applied Cognitive Psychology*, 17(3):325–348.

O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G., and Bruce, V. (1996). Comparison of face-to-face and video-mediated interaction. *Interacting with computers*, 8(2):177–192.

Pickering, M. J. and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02):169–190.

Reitter, D. and Moore, J. D. (2007). Predicting Success in Dialogue. In *Proceedings of ACL 2007*, pages 808–815, Prague, Czech Republic. Association for Computational Linguistics.

Reverdy, J. and Vogel, C. (2017a). Linguistic Repetitions, Task-based Experience and A Proxy Measure of Mutual Understanding. In *Proceedings of CogInfoCom 2017*, pages 395–400, Debrecen, Hungary. IEEE.

Reverdy, J. and Vogel, C. (2017b). Measuring Synchrony in Task-Based Dialogues. In *Proceedings of INTERSPEECH'17*, pages 1701–1705, Stockholm, Sweden. ISCA.

Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of the International Conference on New Methods in Language Processing*, pages 154–164, Manchester, UK.

Schneider, A. and Luz, S. (2011). Speaker Alignment in Synthesised, Machine Translated Communication. In *Proceedings of IWSLT 2011*, pages 254–260, San Francisco, California, USA. ISCA.

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

Taylor, T. J. (1992). *Mutual Misunderstanding: Scepticism and the theorizing of language and interpretation*. Duke University Press.

Turnbull, W. (2003). *Language in action: Psychological models of conversation*. Psychology Press.

Vogel, C. and Behan, L. (2012). Measuring Synchrony in Dialog Transcripts. In *Cognitive Behavioural Systems. Lecture Notes in Computer Science*, volume 7403, pages 73–88. Springer Berlin Heidelberg.

Vogel, C. (2013). Attribution of Mutual Understanding. In *Journal of Law and Policy*, volume 21.2, pages 377–420.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. In *Proceedings of LREC'06*, pages 1556–1559, Genoa, Italy. ELRA.

9. Language Resource References

Hayakawa, Akira and Luz, Saturnino and Cerrato, Loredana and Campbell, Nick. (2015). *The ILMT-s2s Corpus*. CNGL Programme, distributed via ELRA, Trinity College Dublin, 1.0, ISLRN 100-610-774-625-0.