Adapting the ISO 24617-2 Dialogue Act Annotation Scheme for Modelling Medical Consultations

Volha Petukhova[1] **and Harry Bunt**[2]
[1]Spoken Language Systems Group, Saarland University, Germany
[2]Department of Cognitive Science and Artificial Intelligence, Tilburg University, The Netherlands
v.petukhova}@lsv.uni-saarland.de; harry.bunt@uvt.nl

Abstract

Effective, professional and socially competent dialogue of health care providers with their patients is essential to best practice in medicine. To identify, categorize and quantify salient features of patient-provider communication, to model interactive processes in medical encounters and to design digital interactive medical services, two important instruments have been developed: (1) medical interaction analysis systems with the Roter Interaction Analysis System (RIAS) as the most widely used by medical practitioners and (2) dialogue act annotation schemes with ISO 24617-2 as a multidimensional taxonomy of interoperable semantic concepts widely used for corpus annotation and dialogue systems design. Neither instrument fits all purposes. In this paper, we perform a systematic comparative analysis of the categories defined in the RIAS and ISO taxonomies. Overcoming the deficiencies and gaps that were found, we propose a number of extensions to the ISO annotation scheme, making it a powerful analytical and modelling instrument for the analysis, modelling and assessment of medical communication.

Keywords: dialogue act, semantic annotations, medical interaction analysis

1. Introduction

The current call for cost-effective, accessible and user-friendly health care services, together with recent advances in interactive technologies, has triggered an enormous interest in digital medical applications. Many such services are provided online, e.g. ordering medicines, making doctor appointments, accessing medical records (Turgiss et al., 2011). Self-service healthcare is actively promoted. Interactive health screening kiosks are deployed where people can measure their vision, blood pressure, weight and body mass index, receive an overall health assessment, and access a database of local doctors (Bluth, 2009). Health care providers are sometimes replaced by virtual conversational agents (DeVault et al., 2014).

Of chief importance is that the quality of technology-enhanced and technology-mediated services is not significantly lower than conventional in person patient-provider encounters, but adopt a user-centred approach to achieve high effectiveness, relevance and quality. For successful designs and innovations, attention needs to be paid not only to technical possibilities but also very much to the social interactive environment in which these innovations may be placed. Consequently, it is important to understand how well a technical solution fits in with the activities and needs of the users in a proposed setting. Systematic and comprehensive interaction analysis and dialogue modelling methods are often used for obtaining a satisfactory degree of understanding of human interactive behaviour for the subsequent specification of mechanisms of human dialogue that need to be incorporated into a system. A multi-disciplinary analysis of user behavioural, physiological and functional data is required, with processes and results that are understandable by medical and non-medical experts, for staying close to the reality of doctors and patients, and for developing products that are well accepted by their users. The data analysis often involves annotation with dialogue act information. Annotation schemes have been constructed that are useful both for empirically-based studies of interactive and task-related phenomena, and for data-driven design of interactive systems.

A number of studies have proposed the use of a dialogue act taxonomy tailored to the medical domain (Sandvik et al., 2002; Miller and Nelson, 2005; Chang et al., 2013; Bolioli and others, 2019). Most of them are based on the RIAS scheme (Chang et al., 2013; Miller and Nelson, 2005; Bolioli and others, 2019), which has proved efficient for the analysis of various kinds of medical encounters[1], but which cannot be directly used for building a dialogue system or its components. The widely used domain-independent ISO 24617-2 dialogue act taxonomy, on the other hand, needs some adaptation to the medical domain, but is well suited for computational modelling and for dialogue system design. This study tests the assumption that the two schemes are in this sense complementary, and when combined together in a sensible way provide a unified model that supports the quantitative and qualitative analysis of observed behaviour in natural interactive medical settings, while also being useful for quality assessment of interactive and task-related performance of medical professionals, including technology-enhanced and technology-mediated interactions. Moreover, the combined taxonomy can facilitate user-based interactive data collection (real or simulated), as well as the design of conversational medical applications.

The paper is structured as follows. Section 2 specifies the use cases and discusses the related work performed in the analysis and modelling of medical encounters. Section 3 introduces the RIAS and ISO 24617-2 taxonomies. Section 4 presents annotation experiments performed to assess the compatibility of concepts defined in both taxonomies. We specify the corpus data and discuss the obtained results. Section 5 defines a mapping between the RIAS and ISO 24617-2 taxonomies, and proposes extensions to ISO

[1]For an overview see (Pires and Cavaco, 2014).

24617-2 in order to make it powerful and accurate, as required for the use cases of analysing and modelling medical interactions. Finally, Section 6 summarizes our findings and outlines directions for future research and development.

2. Use Cases

Dialogue occurs in almost all kinds of patient-provider encounters. It forms a foundation for diagnosis, examination, treatment and therapeutic management. Recording and automatic processing of patient-provider dialogues is desirable in many contexts. Large volumes of patient-related dialogue data can be useful for informed decision making by caregivers. Speech, language and interactive multimodal data has been used to detect dementia and related disorders (Chapman et al., 1998; Cuetos et al., 2007; Mirheidari et al., 2016), depressions and post-traumatic stress disorders (DeVault et al., 2013; Stepanov et al., 2018; Dham et al., 2017). Analysed dialogue data can enhance communication with patients by understanding their concerns and needs. Dialogue data also forms the source for design and training dialogue systems, personalized recommendations and interventions. We consider three important use cases: (1) medical interaction analysis; (2) quality assessment of technology-enhanced or technology-mediated interactions; and (3) dialogue system design.

2.1. Medical interaction analysis

An average physician conducts more than 200,000 consultations in his/her professional career (Silverman et al., 2016). The success of medical consultations relies heavily on how doctors respond to their patients' communicative actions (Langewitz et al., 2002; Conigliaro, 2001; Patel Kuehl, 2011). The principle characteristics of medical communication have been the subject of many studies. Asymmetries are observed where medical staff have the right to initiate and control the interaction and patients have limited initiative rights and responsive tasks, even though the patient has a foreground role in the interaction. Doctors provide and request information, give instructions, i.e. prohibiting or issuing commands, and patients respond 'submissively' to doctors' questions and rarely ask their own (Roter and Hall, 2006). Research on the effects of institutional frameworks on medical communication has been carried out using pragmalinguistic and discourse analyses (Bührig, 1996; Atkinson, 1999). With the current shift to a patient's autonomy in defining and following their medical treatments, interaction analysis can help healthcare providers to assess the degree of patient participation in medical encounters (Street Jr and Millay, 2001).

The success of interactive processes often depends not only on the medical competence of the doctor, but also on his/her linguistic, social and cultural competences (Suchman et al., 1997; Lindemann, 2015). Doctors can exercise several attitudes, e.g. active listening or empathetic silence, and use the emotional context of reassurance, support, and understanding (Kaplan et al., 1989; Lazare et al., 1995). Numerous studies have identified challenges related to cultural differences in language use in doctor-patient interactions (Schyve, 2007; Brach and Fraserirector, 2000; Collins et al., 2002). Studies on social factors affecting the outcome of medical consultations often focus on politeness and cooperativeness (Robins and Wolf, 1988; Adams, 2013). A considerable body of research has been carried out, with quantitative and qualitative studies reporting results on the number and types of questions asked by doctors and patients, on the use of indirect speech acts and social obligation acts, on the number of times a doctor interrupts a patient and vice versa, on the quantity of speech production repairs, etc., see e.g. (Aronsson and Sätterlund-Larsson, 1987; Ong et al., 1995; Kindler et al., 2005; Roter and Larson, 2002). Interaction analysis is useful to study how effectively caregivers talk to patients, how active patients are when they talk to their caregivers, and how the communicative behaviour of caregivers and that of their patients are related.

2.2. Quality assessment of technology-enhanced medical encounters

A growing body of research results demonstrates that the incorporation of health technologies can make health care more effective and efficient by electronically connecting clinicians to clinicians, patients to clinicians, and even patients to other patients (Clark et al., 2007; Kulshreshtha et al., 2010; Caiata-Zufferey et al., 2010; Weiss, 2004).

With many online and mobile applications now being developed, the effect of telemedicine and other digital health intervention systems on the quality of health provision is of particular concern. Many professionals argue the case for strict regulations, even discussing so-called 'Digital service prescription' of certified services (Murray et al., 2016). To assess the quality of these applications, a variety of evaluation frameworks has been proposed (Field and others, 1996; Grigsby et al., 1995). Although the majority of the research findings favour telemedicine, respondents have reported both positive (cost-effectiveness and accessibility) and negative results (e.g. relating to non-verbal behaviour and lack of touch) (Miller, 2001).

So far not a great deal of research has been devoted to the analysis of communication in technology-enhanced or -mediated consultations. While some attention has been paid to general communicative efficacy, the focus was more broadly on overall performance and satisfaction with the general (including technical) attributes of telemedicine and e-health (Bell, 2018). The impact of technology use on patient's and provider's task-related ('data-gathering' and 'education and counselling') behaviour and socio-emotional aspects ('building a relation' and 'activating and partnership building') is still understudied. Detailed interaction analysis is a useful instrument in the design of a successful technology-enhanced application. It enables the systematic identification, categorization, and quantification of salient features of doctor-patient communication, and when linked with a wide range of outcomes, including patient and provider satisfaction, adherence to treatment, health and clinical status, recall and understanding, and psychological well-being can serve the development of valid and efficient measurement/assessment systems.

Socio-emotional exchange	Task-focused exchange		Global affect ratings
personal remarks, social conversation	transition words		anger/irritation
laughs, tells jokes	gives orientation		anxiety/nervousness
shows (dis-)approval - direct	paraphrase/checks for understanding		depression/sadness
gives compliment - general	bid for repetition		emotional distress/upset
shows agreement or understanding	asks for understanding		dominance/ assertiveness
back-channel responses	asks for opinion		interest/ attentiveness
empathy	asks (open-/close-ended) questions	medical condition,	friendliness/warmth
shows concern/worry	gives information	therapeutic regimen,	responsiveness/engagement
reassures, encourages/shows optimism	counsels/directs behaviour	lifestyle,	sympathetic/empathetic
legitimizes		psychological feelings,	hurried/rushed
partnership		other	respectfulness
shows criticism	requests	services	
asks for reassurance		or medication	

Table 1: Taxonomy of the RIAS actions.

2.3. Dialogue system design

Multimodal dialogue (combinations of spoken and typed language, videos, pictures, facial expressions, haptic and other gestures) is not only the most natural and social form of interaction which is increasingly becoming the most attractive human-machine interface, but is proven to have positive effects in the treatment of certain cognitive impairments (Woods et al., 2012; Hughes et al., 2013), and in health self-management, (Luperfoy, 2004; Reid et al., 2018) patient education (Brixey et al., 2017; Wolf et al., 2019), health behaviour change (Petukhova et al., 2019), and mental and emotional well-being (Fitzpatrick et al., 2017; Inkster et al., 2018; DeVault et al., 2014).

The vast majority of existing dialogue systems make use of dialogue acts[2] as core semantic units to describe and model what is happening in dialogue. Dialogue data annotated with dialogue act information is used to train machine-learning algorithms for the automatic recognition and prediction of dialogue acts in a human-machine dialogue system. The dialogue act taxonomies used for these purposes vary from a simple list of mutually exclusive tags, modelling closed limited domains, to complex hierarchical multidimensional open-domain taxonomies, see (Petukhova, 2011) for an overview.

Currently, a steadily growing interest can be observed in data-driven modelling of dialogue phenomena and dialogue system design. Malchanau et al. (2018) proposed the Continuous Dialogue Corpus Creation (D3C) methodology, where a corpus is used as a shared repository for analysis and modelling of interactive dialogue behaviour, and for implementation, integration and evaluation of dialogue system components. The ISO 24617-2 standard data model is used to facilitate these purposes.

3. Semantic analysis: taxonomy of medical communicative actions

3.1. The Roter Interaction Analysis System

Interaction analysis has been employed in a wide variety of health care settings. The Roter Interaction Analysis System (RIAS) (Roter and Larson, 2002) is the most widely used analysis and evaluation system in medical communication. It was been designed to systematically study and assess medical dialogues in a variety of medical fields, including nursing, adult care, emergency medicine, pediatric primary care, oncology, etc. RIAS has also been used for training health care providers in communication skills.

RIAS views patient-provider communication as having at least three core functions in parallel: (1) to determine and monitor a medical problem; (2) to develop, maintain, or conclude a therapeutic relationship; and (3) to carry out patient education and implementation of treatment plans, see the 'three functions model' (Lazare et al., 1995). Thus, medical dialogue involves in the first place a **task-related** exchange, consisting of *question-asking* and *information-providing* actions in order to gather data, and *counselling* actions produced by a medical professional to educate a patient and direct/influence his future behaviour, motivating him to adhere to a treatment. Actions related to discussion (negotiation) and implementation of a treatment plan are not defined in RIAS. These actions have the purpose to determine areas/issues of differences (conflicts) between patient and provider, and negotiate to resolve them; communicate the diagnostic significance of the problem; negotiate and recommend appropriate diagnostic procedures and treatment; negotiate and recommend appropriate preventive measures and lifestyle changes; and enhance the coping ability by understanding and dealing with the social and psychological consequences of the disease and the treatment (Tuckett et al., 1985). Negotiation relevant actions such as offer, promise and acceptance/rejection of counselling acts, as well as modal operators for expressing importance, likelihood, desirability, possibility, necessity and ability are important in shared decision making for health behaviour change (Petukhova et al., 2019). In RIAS, task-focused actions also involve *activation strategies* that facilitate the expressions of partner's expectations, preferences and perspectives, such as asking for an opinion, understanding, paraphrasing and interpretation, and are important for a meaningful participation in treatment and decision making (Roter, 2000).

A second type of communicative actions is concerned with **therapeutic relation management**. This category comprises actions in order to (1) define the nature of the relationship; (2) communicate professional expertise; (3) communicate interest, respect, support and empathy; (4) recognize and resolve various relational barriers to patient-provider communication; and (5) elicit the patient's perspective (Lazare et al., 1995). Functions of type (5) are included in RIAS as task-focused actions as discussed above; functions of type (3) and (4) are defined in RIAS as *socio-emotional exchange* and are concerned mostly with **social and interpersonal relations management**. They comprise

[2]Many researchers use the traditional terminology of 'speech acts'. In designing conversational agents, designers refer to 'intent' rather than 'intention'.

expressions of worry and concern, optimism and reassurance, empathy and partnership building (social talk, jokes). Functions for (1) and (2) are partially covered in RIAS by the *give orientation* category.
Another important aspect concerns **affective** behaviour performed in order to *build an emotional relation* with the patient through the development of rapport and responsiveness to patient's emotions. The affective aspect includes expressions of and reactions to anger, anxiety, distress, sadness, dominance, etc.
Table 1 gives an overview of the RIAS analysis categories.

3.2. ISO 24617-2 dialogue act annotation scheme

ISO 24617-2 (ISO, 2012) is not just a theoretically grounded and empirically tested inventory of dialogue acts with fine-grained distinctions, it presents a semantic framework for the systematic analysis and computational modelling of multimodal behaviour of dialogue participants. It takes a multidimensional view on dialogue in the sense that participation in a dialogue is viewed as performing several activities in parallel, such as pursuing the dialogue task or activity, providing and eliciting feedback, and taking turns. These activities in various 'dimensions' are called *dialogue acts* and are formally interpreted as update operations on the information states of the dialogue participants. Dialogue acts have two main components: a *semantic content* corresponds to what the utterance is about, e.g. objects, events, etc.; and a *communicative function*, which specifies how an addressee updates his information state with the semantic content when he understands the corresponding aspect of the meaning of a dialogue utterance. A communicative function captures beliefs and intentions of the speaker.
The ISO 24617-2 taxonomy distinguishes 9 dimensions, addressing information about: the domain or task (*Task*), feedback on communicative behaviour of the speaker (*Auto-feedback*) or other interlocutors (*Allo-feedback*), managing difficulties in the speaker's contributions (*Own-Communication Management*) or those of other interlocutors (*Partner Communication Management*), the speaker's need for time to continue the dialogue (*Time Management*), about who should have the next turn (*Turn Management*), the way the speaker is planning to structure the dialogue (*Dialogue Structuring*), and the information motivated by social conventions (*Social Obligations Management*). An updated version of the standard (Bunt et al., 2020) includes additionally the *Contact Management* dimension, adopted from the DIT^{++}[3] annotation scheme, for acts that serve to establish and manage contact and attention. Moreover, the *Task Management* dimension, known from the DAMSL annotation scheme, is defined as a possible extension, for dealing with discussion or explanation of a certain task or activity that is pursued through the dialogue (as opposed to performing that task/activity).
For each dimension, at most one communicative function can be assigned, which can occur either in this dimension alone (the function is *dimension specific*) or occur in all dimensions (the function is *general purpose*). For example, an utterance with the dimension-specific function Self Correction exclusively addresses the *Own Communication Management* dimension. Utterances with a *general purpose* function, such as Inform, can address any dimension (such as e.g. *Task* or *Discourse Structuring*).
The tagset contains 30 dimension-specific functions and 26 general-purpose functions, see Appendix A. When a unit addresses several dimensions simultaneously, multiple tags are assigned. To perform this systematically and accurately, ISO 24617-2 offers flexible segmentation strategies for identifying meaningful dialogue units in multiple dimensions, called 'functional segments', defined as the functionally relevant minimal stretches of communicative behaviour.
Speaker intentions may be complex and may be expressed with a particular attitude or emotion. Nuances concerning certainty, conditionality, or sentiment are captured by means of *qualifiers*. Moreover, dialogue acts are not produced in isolation, but various relations exists between them: functional dependence, feedback dependence and rhetorical relations, see (Bunt et al., 2018) for an updated view.
ISO 24617-2 includes the specification of the XML-based Dialogue Act Markup Language (DiAML) for the representation of dialogue act annotations (Bunt et al., 2012).

4. Applying the ISO 24617-2 standard to annotate medical dialogues

This section reports the results of small-scale annotation experiments, preformed with the aim to assess the applicability of the ISO 24617-2 dialogue act annotation standard to medical interactions and RIAS and ISO compatibility.

4.1. Corpus data

Unfortunately, publicly available dialogue corpora featuring real doctor-patient interactions are rare, primarily for ethical reasons concerning participants' privacy and data security. The corpus considered in this study is the **Distress Analysis Interview Corpus** (DAIC, Gratch et al., 2014), which contains clinical interviews to assist the detection of psychological disorders like anxiety, depression and post-traumatic stress disorder. The part of the corpus publicly released contains interviews collected in the Wizard-of-Oz setting (DAIC-WOZ corpus) where a virtual agent - Ellie - was controlled by humans playing the role of an interviewer who simulates standard protocols for identifying people at risk for post-traumatic stress disorder (PTSD) and depression based on the PTSD Checklist - Civilian Version (Blanchard et al., 1996). Wizards interact with humans who were (pre-)assessed by a professional therapist being either distressed or not-distressed. The DAIC corpus is a multimodal collection of semi-structured clinical interviews starting with neutral questions designed to build rapport and make the participant comfortable, progressing to more specific questions about symptoms and events related to depression and PTSD, and ending with a 'cool-down' phase to ensure that participants do not leave the interview in a distressed state of mind. The corpus contains audio, video, and depth sensor (Microsoft Kinect) recordings of 189 dialogues, and is used in a variety of studies, e.g. in the

[3] https://dit.uvt.nl/

Dimension	Functional segments (in %)		
	ALL	*from those*	
		Interviewer	Interviewee
Task/Activity	62.8	29.1	70.9
Auto Feedback	18.5	76.0	24.0
Allo Feedback	1.9	91.9	8.1
Discourse Structuring	1.1	100.0	0.0
Own Communication Man.	1.9	0.0	100.0
Social Obligations Man.	5.2	76.5	23.5
Turn Management	14.5	0.0	100.0
Time Management	9.4	0.0	100.0

Table 2: Distribution of functional segments across dimensions produced by the interviewer and an interviewee, in terms of relative frequency (in %).

analysis of verbal and non-verbal indicators of psychological distress (DeVault et al., 2013), in automatic depression and PTSD detection from multimodal behaviour (Stepanov et al., 2018; Dham et al., 2017), in the analysis of patient's (disclosure) behaviour when interacting with a virtual therapist, as well as in comparison to human-human interaction using the (unreleased) face-to-face dialogues of the DAIC dataset (Lucas et al., 2014) and the development of a virtual interviewer (DeVault et al., 2014).

To some extent, the DAIC-WOZ data covers all three use cases defined in Section 2. The dialogues are based on a real scenario, involving humans who are patients and humans who simulate medical interviewer behaviour in a role-playing setting. Previous research showed that open and closed role plays are effective for eliciting authentic interactive behaviour and for examining the impact of various factors on the participants' interactive behaviour (Kasper, 2000; Bardovi-Harlig and Hartford, 2005; Al-Gahtani and Roever, 2012). The role-playing method is commonly used in interactive dialogue data collection efforts (Brône and Oben, 2015), and underpins high-fidelity simulations of clinical cases and medical communication training (Kaplonyi et al., 2017; Ker and Bradley, 2013; McGaghie et al., 2010). The DAIC-WOZ dialogues feature technology-enhanced application in the domain of telemedicine, and form the basis for a dialogue system development - the SimSei Kiosk (DeVault et al., 2014).

4.2. Annotations

From the DAIC-WOZ corpus, 11 randomly selected dialogues were manually re-segmented and annotated with ISO 24617-2 dialogue acts and independently with RIAS categories. The selected dataset comprises 2,819 functional segments. The annotations were compared and mapped. If RIAS categories were more specific and captured the utterance meaning more accurately, or if they were not defined in the ISO taxonomy, they were proposed as elements for a future plug-in for as defined in (Bunt, 2019).

Annotations were performed using the ANVIL tool[4], which allows segmentation and annotation in multiple tiers so that for each participant all ISO dimensions and RIAS categories can be specified. Two randomly selected dialogues were annotated by two trained annotators who were not medical experts and were novice users of the RIAS scheme. Inter-coder agreement was measured in terms of Cohen's kappa for each tagset resulting in moderate agreement for RIAS (kappa of 0.52) and for ISO (kappa of 0.58) on average. Annotators disagreed the most when classifying social and feedback acts from both schemes, and the ISO sentiment qualifiers and the RIAS categories for global affect.

[4] www.anvil-software.de

4.3. Results

The analysis shows that the majority of the functional segments is assigned to the **Task** dimension where the Interviewee produced twice as many task-related acts as the Interviewer. The Interviewer is thus successful in achieving the goal to encourage the Interviewee to talk and disclose information and feelings. From the task-related acts about 27% are *questions*, mostly asked by the Interviewer. Following RIAS, questions where annotated as *closed-* (58.8%) and *open-question* (41.2%). It may be noticed here that the inter-annotator agreement assigning these two categories was rather low, measured as 0.47 in terms of kappa, which may be explained by the fact that the RIAS definitions are not very precise, leaving room for subjective interpretation, see (Sandvik et al., 2002; Roter and Larson, 2002) and our discussion in Section 5.

Information-providing acts constitute about 69% of all functional segments and are produced mainly by an Interviewee. The fact that there are almost twice as many information-providing than information-seeking acts can be explained by the fact that the Interviewees' answers were very elaborate. Directives account for about 4% of the observed dialogue acts, mostly in the form of Interviewer *requests* to provide more information or to give examples. These directive dialogue acts cannot be directly mapped to RIAS *counselling* acts since the latter mostly concern medical actions to be undertaken by the patient, described in the semantic content. We mapped them to requests, however, applied to a broader range of semantic content categories than originally specified in RIAS.

As for the semantic content of task-related acts, this concerns discussion of social and demographic conditions including relationships with partners, family and friends plus living conditions and employment details (48.3%), feelings and emotions (27.4%), life style issues such as diets, habits, holidays and exercise/sporting activities (12.4%). and medical conditions which include own and family medical history, illnesses and hospitalizations, recent/current symptoms; and tests and references to diagnostic and prognostic issues (11.4%).

The second large category of dialogue acts is formed by those that report about the speaker's and addressee's processing achievements. This category comprises positive and negative **Auto-** and **Allo-Feedback** acts, and is rather heterogeneous when taking *sentiment qualifiers* into account. While on a binary classification (positive vs negative) almost perfect inter-annotator agreement was achieved (kappa of 0.83), the assignment of qualifiers posed a problem. ISO does not provide a fixed set of sentiment qualifiers. The W3C recommendation EmotionML does not provide a single repository of emotion descriptors, and the available alternative emotion vocabularies it provides are rather general. We used the RIAS categories for 'socio-emotional exchange' and 'global affect'. Our analysis

RIAS category	Segments (in %)	
	Interviewer	Interviewee
task-focused exchange	16.3	83.3
- Open-ended questions	6.2	0.0
- Closed-ended questions	7.8	0.4
- Gives information	0.3	66.3
- Requests	2.5	0.0
- Transition words	0.0	14.5
- Gives orientation	0.4	0.0
- Bid for repetition	0.0	0.4
- Checks for understanding	0.0	1.2
socio-emotional exchange	66.0	34.0
- Show approval	37.1	0.0
- Asks approval	2.6	10.2
- Back-channel responses	12.9	0.0
- Show understanding	17.1	0.0
- Laughs	2.9	0.0
- Personal remarks, social talk	14.3	0.0
- Gives compliment	2.9	0.0
global affect	76.2	23.8
- engagement/responsiveness	4.8	4.8
- interest/attentiveness	9.5	0.0
- anger/irritation	0.0	19.0
- friendliness/warmth	52.4	4.8
- respectfulness	4.7	0.0

Table 3: Distribution of interviewer and interviewee communication categories according to the Roter Interaction Analysis System (RIAS), in terms of relative frequency (in %).

shows that some RIAS socio-emotional acts address participants' processing of own or partner(-s) previous communicative behaviour, such as *back-channel responses*, others are performed for a slightly different purpose (although having that meaning as well), namely to establish and maintain a respectful interpersonal relationship between interlocutors, aiming at a form of partnership and trust for triggering self-disclosure acts and making participants comfortable. We therefore propose to add an additional dimension for social activities management called **Interpersonal Relation Management**. These acts are different from *Social Obligations Management* acts since they are not motivated by social conventions and norms. For example, the utterance 'I'm sorry' in the analysed medical dialogues is not produced with the purpose to apologize for mistakes, dispreferred reactions, misunderstandings or any other infelicitous behaviour, but to express empathy and compassion with the situation the addressee (mostly the patient) is experiencing.

Global affect categories are annotated as ISO sentiment qualifiers. Since the terminology related to emotion, mood, attitude, and sentiment can be rather confusing, we suggest to adopt 'affect' as a general term which denotes a concept used in psychology to describe the experiencing of feeling or emotion, and 'affective states' that are psychophysiological constructs which connect mental and physical processes (Hogg and Abrams, 2007). For ISO 24617-2 plug-ins for affective state qualifiers in medical discourse see the next section.

Table 2 shows the distribution of annotated dialogue acts across ISO dimensions, indicating also the percentage of identified functional segments per dimension produced by different speakers, i.e. by Interviewer and an Interviewee.

It is interesting to observe that certain behaviour is performed exclusively by Interviewer, like for the purpose to structure the discourse. This is not surprising, since medical care providers are those who as experts have the power to make decisions concerning what will be discussed. Other dialogue acts, on the other hand, are produced exclusively by an Interviewee like Own Communication, Turn and Time Management acts. This is however assumed to be an artefact of this corpus. In real patient-doctor interaction, it is highly likely that doctors exhibit such behaviour as well since it is very human to stall for time, edit one's own speech, and regulate turn allocations. Virtual conversational agents can improve if they generate these types of dialogue acts as well. What types, where and how frequently, should be estimated when analysed real face-to-face interactions.

Table 3 summarizes the results of annotation performed with the RIAS scheme.

5. RIAS inspired plug-ins for ISO 24617-2

The latest revised version of the ISO 24617-2 dialogue act annotation standard[5] defines **ten** core dimension. RIAS clusters medical actions into **three** categories as discussed above. Even if not explicitly defined, a systematic mapping of RIAS acts to ISO 24617-2 dialogue acts shows that the majority of ISO dimensions is addressed in RIAS and shows a one-to-one correspondence. Other RIAS acts are domain- or use-case dependent, are not defined in ISO but represent a useful extension of the latter in the form of plug-ins. The ISO scheme makes several extensions possible provided they meet certain requirements and formal constraints specified in (Bunt et al., 2018; Bunt, 2019) and summarized in DIT^{++} Release 5.2[6]. Figure 1 provides an overview of the resulting high-level categorization, where ISO dimensions are highlighted in grey boxes and the extensions obtained from the mapping to RIAS are marked in red.

A top-level distinction is made between communicative actions advancing the underlying **task** and *managing the task*, such as instructions, questions, and answers, and actions that *control the dialogue* (see Bunt, 1994). Dialogue control acts are concerned with cognitive processing (*feedback*) of previously produced behaviour, *interaction management* and *social activities management*.

5.1. Task-focused actions

Medical interactions may be motivated by various purposes, however, the majority of them involve question-answering parts, e.g. for medical history taking, to collect complaints, and to survey problems. For medical professionals, mastering interviewing skills is very important for mature decision-making and action-taking.

RIAS differentiates between more directed focussed questions (*closed-ended*) and more open questions (*open-ended*) that allow greater respondent discretion and a more

[5] A proposed second edition is submitted to ISO for circulation and reviewing by ISO member bodies and their experts in September 2019.

[6] https://dit.uvt.nl/#Release5.2

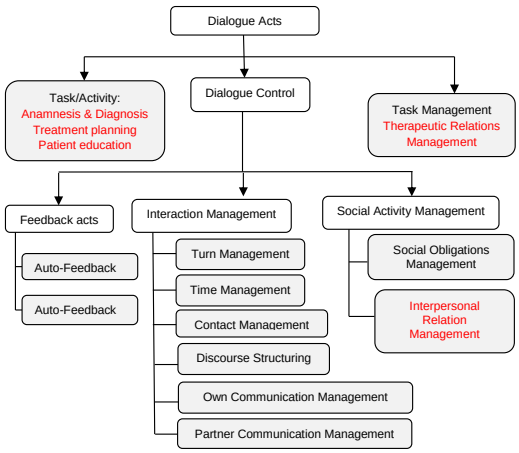

Figure 1: ISO 24617-2 modified to fit medical interaction use cases based on mapping to the RIAS categories. Dimensions are given in grey boxes; optional dimensions in dotted grey boxes; and the modifications obtained from RIAS are marked red.

detailed response. In our annotation experiments, annotation of question forms was found to be complicated but important, as Roter and Larson (2002) also noted. For instance, it has been observed that it is good to start a medical interview with an open-ended question which convey an interest in listening (Patel Kuehl, 2011), whereas an early pursuit of closed questioning may prevent doctors from discovering all the issues and even lead to an incorrect diagnosis (Silverman et al., 2016). RIAS suggests that closed-ended questions produce focused and curtailed responses, while open inquiries and exploratory, investigative or unspecific probing is indicative for the open nature of open-ended questions. Questions where the speaker wants to obtain the truth of a proposition or where the speaker wants to know some or all of the elements of a certain set, thus requiring a specific answer, are closed-ended questions. An open-ended question, as its name suggests, does not seek a specific answer at all, see also (Dhillon et al., 2004).

To elaborate on the response framing power of questions, it has been observed that minor changes in question wording can have a major impact on responses (Schwarz and Oyserman, 2001), and can easily and inadvertently direct the patient away from self-disclosure (Beckman and Frankel, 1984). Linguistic and psychological studies revealed that questions may shape answers (Kellermann, 2007), e.g. suggesting ones and excluding others by open, focused choice, leading, confirmatory questions (De Dreu and Van Kleef, 2004); by carrying assumptions (Zillmann, 1972); and inviting (dis-)agreement, openness or evasion, and threat or comfort (Schuman and Presser, 1996). The ISO 24617-2:2012 set of question types can be further extended to model these differences.

Medical encounters also involve counselling, where doctors *direct* behaviour of their patients expressing their wish that the patient performs or avoids a certain action, possibly dependent on a condition. Different types of directives carry different strengths of the speaker's assumptions about the ability and willingness of the addressee to perform an action (Bunt, 2011).

As noticed above, *commissives* acts are not covered by RIAS, although they play an important role in medical negotiations. For adequate modelling, we need to take into account that negotiators may perform several types of dialogue acts expressing various levels of commitment, but also qualified ('modalised') actions expressing participants' attitudes and preferences, and negotiation strategies (Petukhova et al., 2016; Petukhova et al., 2017; Petukhova et al., 2019).

5.2. Task Management

There is still an ongoing discussion whether Task Management should be included as a separate dimension in ISO 24617-2, as is the case in the DAMSL multidimensional annotation scheme to define acts that discuss the problem solving process or experimental scenario ((Allen and Core, 1997). To model DBOX games (Petukhova et al., 2014) and Metalogue multi-issue bargaining and debate dialogues (Petukhova et al., 2016; Petukhova et al., 2018), Task Management acts were introduced as an ISO 24617-2 extension to address aspects related to game, debate or negotiation processes, phases and procedures. In dialogues in institutional settings, as in a court room or doctor-patient dialogues, task management acts may occur rather frequently, since there is often a clear power relation between the participants. We leave this set largely unspecified for the time being, however, we propose two communicative functions to illustrate this dimension: (1) *Give Orientation* for statements and directives related to an examination or clinical visits, e.g. 'The signal is faint, please speak louder'; and (2) *Discuss Expertise* related to participant roles and areas of expertise, e.g. 'I am your cardiologist'. Other communicative functions can be defined specifically to scenario and/or therapeutic regime management and other arrangements.

5.3. Feedback

In medical interactions, it is important for the doctor not only to signal active listening but also to show a genuine interest and understanding of the patient's behaviour by repeating the information revealed, rephrasing the previously asked questions or provided instructions, confirming or checking for understanding, consistency and validation of the information revealed. Doctors also need to encourage patients to ask questions, express their attitudes, preferences, concerns, fears and opinions. In RIAS, these acts are defined as *activation* strategies.

5.4. Interaction Management

Concerning the Interaction Management functions, only **Turn Management** together with **Time Management** and **Discourse Structuring** are defined in RIAS. However, our annotation experiments indicate that medical interaction analysis will benefit from inclusion of **Contact Management** acts, in particular when applying to telemedicine; of **Own Communication Management** acts when analysing specific patient speech production behaviour; and of specifying **Partner Communication Management** acts to analyse the abilities of participants to detect difficulties and errors in a partner's communicative behaviour.

5.5. Social Activity Management

5.5.1. Social Obligations Management

Participating in a dialogue is a social activity, where one is supposed to do certain things and not to do other things, and to act in accordance with the norms and conventions for social behaviour. A dialogue participant has besides functional also ethical tasks and obligations, and performs social obligation management acts to fulfil these. Social Obligations Management acts are not just 'social', they also improve the transparency of the dialogue. For example, people greet each other also for establishing their presence, and say good-bye also to close the conversation. Such acts, defined in ISO 24617-2, are not covered by RIAS.

5.5.2. Interpersonal Relation Management

A goal in any medical encounter is to establish and maintain a kind of partnership between doctor and patient. Utterances produced for this purpose are not so much meant to exchange information or influence each other's behaviour, but to establish a certain bond between the dialogue participants. Successful partnership building actions promote better cooperation. Here, statements are important that convey the doctor's alliance with the patient in terms of health and support, decision-making, or the development of a therapeutic plan. Patients are often anxious about their medical condition, express concerns or worry, and seek reassurance or special attention. RIAS is particularly explicit concerning these acts, which can be clustered in a separate ISO dimension - **Interpersonal Relation Management** addressing information about the process of patient-provider relationship building, which is important to improve patient satisfaction and health outcomes (Lucas et al., 2014). This is a reason to incorporate types of relational (but also emotional see below) communicative behaviour into the analysis and further modelling.

5.6. Affect

Doctors must be aware of the patient's feelings, motivations, insecurities, engagement and reasons for whether they want to do certain things or not. In ISO 24617-2 this information can be annotated using sentiment qualifiers for which the standard does not specify any specific set of tags. In the revised 2nd edition of the ISO scheme, it is recommended to look to EmotionML (Burkhardt and Schröder, 2008) for specifications of possible sets of emotion and attitude values, and for more sophisticated annotation of the affective aspects of dialogue behaviour. RIAS defines a set of *global affects* that can be used in an ISO 24617-2 plug-in for the specification of participants' attitudes (such as responsiveness, attentiveness, friendliness) and local affective states relating to dialogue acts (such as anger, irritation).

5.7. Semantic Content

ISO 24617-2 focuses on the functional meaning of dialogue acts and does not annotate the semantic content. In the 2nd edition of ISO24617-2, plug-ins are introduced for extending annotations of the functional meaning of dialogue acts with information about their semantic content. it is shown that the degree of detail in which the semantic content of a dialogue act is appropriately represented depends on the application domain (Bunt, 2019). For some domains a simple representation as a list of attribute-value pairs may be adequate. For others a representation in terms of events with their participants, time and place may be more appropriate, and again for more advanced applications it may be necessary to take general aspects of natural language utterance meaning into account, including quantification and modification phenomena.

RIAS supports a high-level specification of the semantic content of medical actions. Task-focused actions are about *medical conditions*, *therapeutic regime*, *lifestyle*, *psychological feelings*, *services*, *medication* and *other* content. Miller and Nelson (2005) define a semantic content category related to *technology* used in medical dialogue. An alternative medical interaction analysis system, Medical Interaction Process System (MIPS) (Ford et al., 2000), defines additional semantic content categories, such as *tests*, *side-effects*, *drugs*, *social/demographic circumstances* and *administrative/practical details*.

The table in Appendix A gives an overview of ISO 24617-2 dimensions (in bold) and communicative functions (black), and proposed extensions for medical interaction analysis and modelling in terms of dimensions, communicative functions, sentiment qualifiers and high-level semantic content (in red).

6. Conclusions and Future Efforts

In this paper we proposed a number of extensions to the standard dialogue act annotation scheme, ISO 24617-2, to make it a powerful analytical and modelling instrument for medical interactive data analysis and design of digital services/applications. We started from the assumption that the ISO 24617-2 dialogue act taxonomy and the de-facto standard system for medical interactions analysis, RIAS, would have complementary strengths and weaknesses. Derived from social-exchange theories related to interpersonal influence, problem-solving and empowerment, RIAS has been proven to be useful in in-depth studies of communication dynamics and its relationship to outcomes of patient-provider encounters. Taking the complexities of natural human dialogue into account, ISO 24617-2 adopts a multidimensional view on communication, which has been recognized to be empirically well motivated and to allow accurate modelling of theoretical distinctions. The multidimensional nature of the ISO taxonomy also enables various extensions and offers the opportunity to tailor it to specific applications and domains.

We considered a number of use cases related to medical interaction analysis, quality assessment and dialogue system design. We presented significant findings of communication research performed in face-to-face, technology-enhanced and technology-mediated interaction between healthcare providers and their patients. In addition, we performed a mapping between the ISO and RIAS schemes. Dialogues from the DAIC-WOZ corpus were annotated according to each of them, and the correspondences between assigned tags were analysed. In this way, systematic differences and correspondences between schemes, and their strengths and weaknesses became apparent.

The research reported here has some practical limitations. First of all, the corpus used in this study comprises dialogues in the mental healthcare domain. In such interactions, rapport and trust building is essential for patient's self-disclosure. In other types of medical dialogue, different doctor and patient actions, their sequences and distributions may be observed. However, since the RIAS scheme has been applied for many medical domains and is commonly acknowledged as a generic medical scheme, we do not expect that important aspects (dimensions) are missing in our analysis. Nevertheless, other schemes will be explored which are specific to a particular type of interaction, e.g. the ISBAR scheme for medical handover communication analysis, see e.g. (Spooner et al., 2018), OPTION5 and OPTION12 for shared decision making (Elwyn et al., 2003), or specific to an element of communicative behaviour such as emotions, e.g. the Verona Coding Definition of Emotional Sequences (VR-CoDES) (Del Piccolo et al., 2011).

Not all the labels defined in the two schemes are present in our annotations since the available corpus we used was not large and not very specific. Another limitation is that we could not find multiple annotators equally experienced with both schemes to apply them reliably. Nevertheless, meaningful extensions for medical dialogues were identified that can be converted to plug-ins for the general ISO scheme and can be applied in the future on larger datasets. The plug-ins need to be tested for their usability and coverage in manual and automatic annotation.

Future efforts will be also directed towards larger collections of simulated and real patient-provider dialogue data, with the perspective to enrich task-focused, relationship-building and effective verbal and non-verbal communication strategies for multimodal dialogue systems in healthcare settings and medical training applications.

7. Bibliographical References

Adams, R. L. (2013). *Politeness strategies in decision-making between GPs and patients*. Ph.D. thesis, University of Birmingham.

Al-Gahtani, S. and Roever, C. (2012). Role-playing l2 requests: Head acts and sequential organization. *Applied Linguistics*, 33(1):42–65.

Allen, J. and Core, M. (1997). Draft of damsl: Dialog act markup in several layers. Available at `http://www.cs.rochester.edu/research/cisd/resources/damsl/`.

Aronsson, K. and Sätterlund-Larsson, U. (1987). Politeness strategies and doctor-patient communication. on the social choreography of collaborative thinking. *Journal of Language and Social Psychology*, 6(1):1–27.

Atkinson, P. (1999). Medical discourse, evidentiality and the construction of professional responsibility. *Talk, work and institutional order: Discourse in medical, mediation and management settings*, 1.

Bardovi-Harlig, K. and Hartford, B. S. (2005). *Interlanguage pragmatics: Exploring institutional talk*. Routledge.

Beckman, H. B. and Frankel, R. M. (1984). The effect of physician behavior on the collection of data. *Annals of Internal medicine*, 101(5):692–696.

Bell, T. (2018). Can telemedicine be both cost efficient and high quality. *US News and World Report*.

Blanchard, E. B., Jones-Alexander, J., Buckley, T. C., and Forneris, C. A. (1996). Psychometric properties of the ptsd checklist (pcl). *Behaviour research and therapy*, 34(8):669–673.

Bluth, C. P. (2009). Community based managed health kiosk system for soliciting medical testing and health study participants, September 24. US Patent App. 12/407,677.

Bolioli, A. et al. (2019). How do physiotherapists and patients talk? developing and annotating rimotivazione dialogue corpus. In *Proceedings of the Sixth Italian Conference on Computational Linguistics*, Bari, Italy.

Brach, C. and Fraserirector, I. (2000). Can cultural competency reduce racial and ethnic health disparities? a review and conceptual model. *Medical Care Research and Review*, 57(1_suppl):181–217.

Brixey, J., Hoegen, R., Lan, W., Rusow, J., Singla, K., Yin, X., Artstein, R., and Leuski, A. (2017). Shihbot: A facebook chatbot for sexual health information on hiv/aids. In *Proceedings of the 18th annual SIGdial meeting on discourse and dialogue*, pages 370–373.

Brône, G. and Oben, B. (2015). Insight interaction: a multimodal and multifocal dialogue corpus. *Language resources and evaluation*, 49(1):195–214.

Bührig, K. (1996). *Reformulierende Handlungen: Zur Analyse sprachlicher Adaptierungsprozesse in institutioneller Kommunikation*, volume 23. Gunter Narr Verlag.

Bunt, H., Alexandersson, J., Choe, J.-W., Fang, A. C., Hasida, K., Petukhova, V., Popescu-Belis, A., and Traum, D. R. (2012). ISO 24617-2: A semantically-based standard for dialogue annotation. In *LREC*, pages 430–437. Citeseer.

Bunt, H., Gilmartin, E., Keizer, S., Pelachaud, C., Petukhova, V., Prévot, L., and Theune, M. (2018). Downward compatible revision of dialogue annotation. In *Proceedings 14th Joint ACL-ISO Workshop on Interoperable Semantic Annotation*, pages 21–34.

Bunt, H., Petukhova, V., Gilmartin, E., Pelachaud, C., Fang, A., Keizer, S., and Prévot, L. (2020). The ISO Standard for Ddialogue Act Aannotation, Ssecond Eedition. In *Proceedings 12th International Conference on Language Resources and Evaluation (LREC(2020)*, Paris. ELRA.

Bunt, H. (2011). The semantics of dialogue acts. In *Proceedings of the Ninth International Conference on Computational Semantics*, pages 1–13. Association for Computational Linguistics.

Bunt, H. (2019). Plug-ins for content annotation of dialogue acts. In *Workshop on Interoperable Semantic Annotation (ISA-15)*, page 33.

Burkhardt, F. and Schröder, M. (2008). Emotion markup language: Requirements with priorities. *W3C Incubator Group report*.

Caiata-Zufferey, M., Abraham, A., Sommerhalder, K., and Schulz, P. J. (2010). Online health information seeking

in the context of the medical consultation in switzerland. *Qualitative health research*, 20(8):1050–1061.

Chang, C. L., Park, B. K., and Kim, S. S. (2013). Conversational analysis of medical discourse in rehabilitation: A study in korea. *The journal of spinal cord medicine*, 36(1):24–30.

Chapman, S. B., Highley, A. P., and Thompson, J. L. (1998). Discourse in fluent aphasia and alzheimer's disease: Linguistic and pragmatic considerations. *Journal of Neurolinguistics*, 11(1-2):55–78.

Clark, R. A., Inglis, S. C., McAlister, F. A., Cleland, J. G., and Stewart, S. (2007). Telemonitoring or structured telephone support programmes for patients with chronic heart failure: systematic review and meta-analysis. *Bmj*, 334(7600):942.

Collins, K. S., Hughes, D. L., Doty, M. M., Ives, B. L., Edwards, J. N., and Tenney, K. (2002). *Diverse communities, common concerns: assessing health care quality for minority Americans*. Commonwealth Fund New York.

Conigliaro, R. L. (2001). Communicating with today's patient: Essentials to save time, decrease risk, and increase patient compliance. *Jama: The Journal of the American Medical Association*, 286(6):725.

Cuetos, F., Arango-Lasprilla, J. C., Uribe, C., Valencia, C., and Lopera, F. (2007). Linguistic changes in verbal expression: a preclinical marker of alzheimer's disease. *Journal of the International Neuropsychological Society*, 13(3):433–439.

De Dreu, C. K. and Van Kleef, G. A. (2004). The influence of power on the information search, impression formation, and demands in negotiation. *Journal of Experimental Social Psychology*, 40(3):303–319.

Del Piccolo, L., De Haes, H., Heaven, C., Jansen, J., Verheul, W., Bensing, J., Bergvik, S., Deveugele, M., Eide, H., Fletcher, I., et al. (2011). Development of the verona coding definitions of emotional sequences to code health providers' responses (vr-codes-p) to patient cues and concerns. *Patient education and counseling*, 82(2):149–155.

DeVault, D., Georgila, K., Artstein, R., Morbini, F., Traum, D., Scherer, S., Morency, L.-P., et al. (2013). Verbal indicators of psychological distress in interactive dialogue with a virtual human. In *Proceedings of the SIGDIAL 2013 Conference*, pages 193–202.

DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., et al. (2014). Simsensei kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 1061–1068. International Foundation for Autonomous Agents and Multiagent Systems.

Dham, S., Sharma, A., and Dhall, A. (2017). Depression scale recognition from audio, visual and text analysis. *arXiv preprint arXiv:1709.05865*.

Dhillon, R., Bhagat, S., Carvey, H., and Shriberg, E. (2004). Meeting recorder project: Dialog act labeling guide. Technical report, INTERNATIONAL COMPUTER SCIENCE INST BERKELEY CA.

Elwyn, G., Edwards, A., Wensing, M., Hood, K., Atwell, C., and Grol, R. (2003). Shared decision making: developing the option scale for measuring patient involvement. *BMJ Quality & Safety*, 12(2):93–99.

Field, M. J. et al. (1996). *Telemedicine: A guide to assessing telecommunications for health care*. National Academies Press.

Fitzpatrick, K. K., Darcy, A., and Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): a randomized controlled trial. *JMIR mental health*, 4(2):e19.

Ford, S., Hall, A., Ratcliffe, D., and Fallowfield, L. (2000). The medical interaction process system (mips): an instrument for analysing interviews of oncologists and patients with cancer. *Social Science & Medicine*, 50(4):553–566.

Grigsby, J., Schlenker, R. E., Kaehny, M. M., Shaughnessy, P. W., and Sandberg, E. J. (1995). Analytic framework for evaluation of telemedicine. *Telemedicine Journal*, 1(1):31–39.

Hogg, M. A. and Abrams, D. (2007). Social cognition and attitudes.

Hughes, T., Flatt, J., Fu, B., Chang, C.-C., and Ganguli, M. (2013). Engagement in social activities and progression from mild to severe cognitive impairment: the myhat study. *International psychogeriatrics*, 25(04):587–595.

Inkster, B., Sarda, S., and Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth*, 6(11):e12106.

ISO. (2012). *Language resource management – Semantic annotation framework – Part 2: Dialogue acts. ISO 24617-2*. ISO Central Secretariat, Geneva.

Kaplan, S. H., Greenfield, S., and Ware Jr, J. E. (1989). Assessing the effects of physician-patient interactions on the outcomes of chronic disease. *Medical care*, pages S110–S127.

Kaplonyi, J., Bowles, K.-A., Nestel, D., Kiegaldie, D., Maloney, S., Haines, T., and Williams, C. (2017). Understanding the impact of simulated patients on health care learners' communication skills: a systematic review. *Medical education*, 51(12):1209–1219.

Kasper, G. (2000). Data collection in pragmatics research. *Culturally speaking: Managing rapport through talk across cultures*, 316341.

Kellermann, K. (2007). Persuasive question asking: how question wording influences answers. In *Annual Meeting of the State Bar Association of California, Anaheim, CA*.

Ker, J. and Bradley, P. (2013). Simulation in medical education. *Understanding medical education: Evidence, theory and practice*, pages 175–192.

Kindler, C., Szirt, L., Sommer, D., Häusler, R., and Langewitz, W. (2005). A quantitative analysis of anaesthetist–patient communication during the pre-operative visit. *Anaesthesia*, 60(1):53–59.

Kulshreshtha, A., Kvedar, J. C., Goyal, A., Halpern, E. F., and Watson, A. J. (2010). Use of remote monitoring to

improve outcomes in patients with heart failure: a pilot trial. *International journal of telemedicine and applications*, 2010:3.

Langewitz, W., Conen, D., Nübling, M., and Weber, H. (2002). Communication matters–deficits in hospital care from the patients' perspective. *Psychotherapie, Psychosomatik, medizinische Psychologie*, 52(8):348–354.

Lazare, A., Putnam, S. M., and Lipkin, M. (1995). Three functions of the medical interview. In *The medical interview*, pages 3–19. Springer.

Lindemann, K. (2015). Emotionen in medizinischer kommunikation. *Handbuch Sprache in der Medizin*, 11:154.

Lucas, G. M., Gratch, J., King, A., and Morency, L.-P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37:94–100.

Luperfoy, S. (2004). Retrofitting synthetic dialog agents to game characters for lifestyle risk training. In *AAAI Fall Symposium on Dialogue Systems for Health Communication*.

Malchanau, A., Petukhova, V., and Bunt, H. (2018). Towards continuous dialogue corpus creation: Writing to corpus and generating from it. In *Proceedings of the 11th International Conference on Language Resources and Evaluation(LREC 2018)*, Myiazaki, Japan.

McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., and Scalese, R. J. (2010). A critical review of simulation-based medical education research: 2003–2009. *Medical education*, 44(1):50–63.

Miller, E. A. and Nelson, E.-L. (2005). Modifying the roter interaction analysis system to study provider–patient communication in telemedicine: promises, pitfalls, insights, and recommendations. *Telemedicine Journal & e-Health*, 11(1):44–55.

Miller, E. A. (2001). Telemedicine and doctor-patient communication: an analytical survey of the literature. *Journal of telemedicine and telecare*, 7(1):1–17.

Mirheidari, B., Blackburn, D., Reuber, M., Walker, T., and Christensen, H. (2016). Diagnosing people with dementia using automatic conversation analysis. In *Proceedings of interspeech*, pages 1220–1224. ISCA.

Murray, E., Hekler, E. B., Andersson, G., Collins, L. M., Doherty, A., Hollis, C., Rivera, D. E., West, R., and Wyatt, J. C. (2016). Evaluating digital health interventions: key questions and approaches.

Ong, L. M., De Haes, J. C., Hoos, A. M., and Lammes, F. B. (1995). Doctor-patient communication: a review of the literature. *Social science & medicine*, 40(7):903–918.

Patel Kuehl, S. (2011). Communication tools for the modern doctor bag. physician patient communication part 1: beginning of a medical interview. *Journal of community hospital internal medicine perspectives*, 1(3):8428.

Petukhova, V., Gropp, M., Klakow, D., Schmidt, A., Eigner, G., Topf, M., Srb, S., Motlicek, P., Potard, B., Dines, J., et al. (2014). The DBOX corpus collection of spoken human-human and human-machine dialogues. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 14)*. European Language Resources Association (ELRA).

Petukhova, V., Stevens, C. A., de Weerd, H., Taatgen, N., Cnossen, F., and Malchanau, A. (2016). Modelling multi-issue bargaining dialogues: Data collection, annotation design and corpus. In *LREC*.

Petukhova, V., Bunt, H., and Malchanau, A. (2017). Computing negotiation update semantics in multi-issue bargaining dialogues. In *Proceedings of the SemDial 2017 (SaarDial) Workshop on the Semantics and Pragmatics of Dialogue*, Saarbrücken, Germany.

Petukhova, V., Malchanau, A., Oualil, Y., Klakow, D., Luz, S., Haider, F., Campbell, N., Koryzis, D., Spiliotopoulos, D., Albert, P., Linz, N., and Alexandersson, J. (2018). The metalogue debate trainee corpus: Data collection and annotations. In *Proceedings of the 11th International Conference on Language Resources and Evaluation*. Paris: ELRA.

Petukhova, V., Sharifullaeva, F., and Klakow, D. (2019). Modelling shared decision making in medical negotiations: Interactive training with cognitive agents. In *International Conference on Principles and Practice of Multi-Agent Systems*, pages 251–270. Springer.

Petukhova, V. (2011). *Multidimensional Dialogue Modelling. PhD dissertation*. Tilburg University, The Netherlands.

Pires, C. M. and Cavaco, A. M. (2014). Communication between health professionals and patients: review of studies using the rias (roter interaction analysis system) method. *Revista da Associação Médica Brasileira*, 60(2):156–172.

Reid, M., Walsh, C., Raubenheimer, J., Bradshaw, T., Pienaar, M., Hassan, C., Nyoni, C., and Le Roux, M. (2018). Development of a health dialogue model for patients with diabetes: A complex intervention in a low-/middle income country. *International journal of Africa nursing sciences*, 8:122–131.

Robins, L. S. and Wolf, F. M. (1988). *Confrontation and politeness strategies in physician-patient interactions*. Elsevier.

Roter, D. and Hall, J. A. (2006). *Doctors talking with patients/patients talking with doctors: improving communication in medical visits*. Greenwood Publishing Group.

Roter, D. and Larson, S. (2002). The roter interaction analysis system (rias): utility and flexibility for analysis of medical interactions. *Patient education and counseling*, 46(4):243–251.

Roter, D. (2000). The medical visit context of treatment decision-making and the therapeutic relationship. *Health Expectations*, 3(1):17–25.

Sandvik, M., Eide, H., Lind, M., Graugaard, P. K., Torper, J., and Finset, A. (2002). Analyzing medical dialogues: strength and weakness of roter's interaction analysis system (rias). *Patient education and counseling*, 46(4):235–241.

Schuman, H. and Presser, S. (1996). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. Sage.

Schwarz, N. and Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and ques-

tionnaire construction. *The American Journal of Evaluation*, 22(2):127–160.

Schyve, P. M. (2007). Language differences as a barrier to quality and safety in health care: the joint commission perspective. *Journal of general internal medicine*, 22(2):360–361.

Silverman, J., Kurtz, S., and Draper, J. (2016). *Skills for communicating with patients*. CRC Press.

Spooner, A. J., Aitken, L. M., Corley, A., and Chaboyer, W. (2018). Developing a minimum dataset for nursing team leader handover in the intensive care unit: A focus group study. *Australian Critical Care*, 31(1):47–52.

Stepanov, E. A., Lathuiliere, S., Chowdhury, S. A., Ghosh, A., Vieriu, R.-L., Sebe, N., and Riccardi, G. (2018). Depression severity estimation from multiple modalities. In *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pages 1–6. IEEE.

Street Jr, R. L. and Millay, B. (2001). Analyzing patient participation in medical encounters. *Health communication*, 13(1):61–73.

Suchman, A. L., Markakis, K., Beckman, H. B., and Frankel, R. (1997). A model of empathic communication in the medical interview. *Jama*, 277(8):678–682.

Tuckett, D., Boulton, M., Olson, C., and Williams, A. (1985). *Meetings Between Experts. An Approach to Sharing Ideas in Medical Consultations*. London: Tavistock.

Turgiss, J. L., Boylan, R. M., and Harrison, G. R. (2011). Interactive, internet supported health and fitness management system, September 27. US Patent 8,027,822.

Weiss, N. (2004). E-mail consultation: clinical, financial, legal, and ethical implications. *Surgical neurology*, 61(5):455–459.

Wolf, M., Petukhova, V., and Klakow, D. (2019). Term-based extraction of medical information: Pre-operative patient education use case. In *Proceedings of the International Conference on Recent Advances in Natural Language Procesing (RANLP 2019)*, Varna, Bulgaria.

Woods, B., Aguirre, E., Spector, A., and Orrell, M. (2012). Cognitive stimulation to improve cognitive functioning in people with dementia. *Cochrane Database Syst Rev*, 2(2).

Zillmann, D. (1972). Rhetorical elicitation of agreement in persuasion. *Journal of personality and social psychology*, 21(2):159.

8. Language Resource References

Gratch et al. 2014 The distress analysis interview corpus of human and computer interviews. In Proceedings of LREC 2014, pages 3123–3128

Appendix A ISO 24617-2:2012 dimensions and communicative functions in extended with RIAS acts, semantic content and sentiment qualifiers

General-Purpose Communicative Functions	Semantic Content	Dimension-Specific Communicative Functions		Sentiment Qualifiers
		Function	Dimension	
Open-ended Question	medical conditions	Give orientation	**Task Management**	anger/irritation
- Open-ended Set-Question	- symptoms	Discuss Expertise		anxiety/nervousness
- Open-ended Propositional Question	- diagnosis	AutoPositive	**Auto-Feedback**	depression/sadness
Close-ended Question	- prognosis	AutoNegative		emotional distress/upset
- Close-ended Set-Question	- history	AlloPositive	**Allo-Feedback**	dominance/ assertiveness
- Close-ended Propositional Question	therapeutic regimen	AlloNegative		interest/ attentiveness
- - Check-Question	- tests	FeedbackElicitation		friendliness/warmth
- Choice-Question	- medication	- Elicit Understanding/Approval		responsiveness/engagement
Inform	- treatment	- Elicit Opinion		sympathetic/empathetic
- Agreement	psychological feeling	Stalling	**Time Management**	hurried/rushed
- Disagreement	- dreams	Pausing		respectfulness
- - Correction	- memories	Turn Take	**Turn Management**	
- Answer	- thoughts	Turn Grab		
- - Confirm	- images	Turn Accept		
- - Disconfirm	lifestyle	Turn Keep		
Request	- habits	Turn Give		
- Instruct	- diet	Turn Release		
- Counsel	- hobby	Self-Error	**Own Communication Man.**	
Address Offer	- occupation	- Retraction		
- Accept Offer	- sport	- - Self-Correction		
- Decline Offer	soc./demographic circumstances	Completion	**Partner Communication Man.**	
Suggest/Advise	- family	Correct Misspeaking		
Address Suggestion	- partners	Interaction Structuring	**Discourse Structuring**	
- Accept Suggestion	- friends	- Opening		
- Decline Suggestion	- living conditions	Init-Greeting	**Social Obligations Man.**	
Offer	- education	Return Greeting		
- Promise	- employment	Init-Self-Introduction		
Address Request	administrative details	Return Self-Introduction		
- Accept Request	- GP contact	Apology		
- Decline Request	- appointments	Accept Apology		
	- med.forms	Thanking		
	- other arrangements	Accept Thanking		
	services	Init-Goodbye		
	- collection medication	Return Goodbye		
	- transport	Compliment	**Interpersonal Relation Man.**	
	- calling up	Empathy		
	- home visits	Concern/worry		
	- shopping	Reassurance		
	- cleaning	Legitimize		
	other	Criticism		
		Compassion		
		Self-disclosure		
		Jokes		
		Small talk		

Table 4: ISO 24617-2 dimensions (in bold) and communicative functions (black), and proposed RIAS extensions for medical interaction analysis and modelling in terms of dimensions, communicative functions, sentiment qualifiers and high-level semantic content (in red).