How do Physiotherapists and Patients talk?

Developing the *RiMotivAzione* dialogue corpus.

Andrea Bolioli[1], **Francesca Alloatti**[1,2], **Mariafrancesca Guadalupi**[1], **Roberta Iolanda Lanzi**[1], **Giorgia Pregnolato**[3], **Andrea Turolla**[3] [1]

[1]CELI - Language Technology, Italy
[2]Department of Computer Science - Università degli Studi di Torino, Italy
[3]IRCCS Fondazione Ospedale San Camillo, Italy

Abstract

The research project *RiMotivAzione* aims at helping post-stroke patients who are following an arm and hand rehabilitation path. In this paper we present the *RiMotivAzione* corpus, the first collection of dialogues between physiotherapists and patients recorded in an Italian hospital and annotated following the RIAS annotation protocol. We describe the dataset, the methodologies applied and our first investigations on relevant features of the dialogue process. The corpus was the basis for the design of a conversational interface integrated with a wearable device for rehabilitation, to be used by the patient during the exercises that he or she may perform independently.[1]

1 Introduction

In recent years, computational linguistics and medical research have started to collaborate in order to analyze the communication in the healthcare domain, in particular between clinicians and patients. From a medical perspective, linguistic analysis and dialogue modeling can be used to better understand and potentially enhance communication in different healthcare settings (Sen et al., 2017; Chang et al., 2013; Marzuki et al., 2017), as well as to identify "preclinical" or "presymptomatic" diseases for specific ranges of patients, e.g. discovering early linguistic signs of cognitive decline (Beltrami et al., 2018).

Natural Language Processing (NLP) technologies are also used to develop new communicative tools, e.g. virtual assistants, to alleviate the burden on medical personnel or shift to a home-based patient-centered model of care. Through mHealth (mobile health), for example, people can receive assistance at home, and monitoring devices can check the well-being of a person (Sezgin et al., 2018). A recent review of scientific literature about Artificial Intelligence and IoT in healthcare can be found in (Shah and Chircu, 2018).

The research project *RiMotivAzione* aims at helping the patients who suffered from a stroke and are following an arm and hand rehabilitation path. The goal is to motivate the patients to follow the assigned exercises through the use of a new wearable device with motion sensors developed by the Istituto Italiano di Tecnologia (IIT), integrated with a visual App and a conversational interface. This last component guides the user through the therapeutic path proposing the exercises, giving advice and asking for feedback.

The implementation of voice technologies in the healthcare domain allows for patients with motor impairments to interact with devices through spoken language (Moore et al., 2018), while arm and hand are busy performing the assigned exercises. The interaction is seamless and spontaneous. The patient can keep up autonomously with the therapy thanks to the guidance provided by the voice assistant. The physiotherapist can monitor the patients at a distance, to evaluate their progress, and he can prevent a situation of therapy neglect by the patient, while the latter is motivated to stick to the path and he can reach his rehabilitation goals on time. Needless to say, these digital assistants are not meant to substitute the clinician.

2 Methodological Background and Related Work

As we described in the previous section, the study of communication and conversation in the medical domain is growing in the last years, as well as the introduction of conversational agents in the healthcare sector. A review of current applications and evaluation measures of conversational agents used for health-related purposes can be found, for example, in (Laranjo et al., 2018). Otherwise, there is no systematic review of scientific literature

concerning the linguistic analysis of dialogues in healthcare. Some scientific studies describe how communication can influence clinical outcomes in the rehabilitation setting, e.g. how patient satisfaction, decision-making, and stress level correlate with physicians' communicative acts (Hall and Roter, 2012). Some researchers propose methods to detect and track topics in psycho-therapeutic conversations (Chaoua et al., 2018). Other researchers conducted an analysis of actual communicative behaviors, including nonverbal ones, between physicians and patients in rehabilitation, using transcription and coding of utterances (Chang et al., 2013).

The analysis of speech acts and conversational interaction can play a relevant role in dialogue modeling for healthcare thanks to the classification of utterances, the analysis of dialogue turns and threads, the discovery of recurrent patterns. Speech acts have been investigated in linguistics and computational linguistics for long. Specifically, the task of automatic speech act recognition has been addressed leveraging both supervised and unsupervised approaches (Basile and Novielli, 2018). Otherwise, in the healthcare domain there is still much room for investigation.

In the *RiMotivAzione* project, we deal with physiotherapy sessions in a hospital. The task is to collect and analyze para-linguistic and linguistic data, according to the aforementioned goal of the research project. In this specific setting, i.e. conversational analysis of physician-patient discourse, the most widely used method is the Roter Interaction Analysis System (RIAS). RIAS was developed as a tagset for coding medical dialogue since 1991 by Debra Roter et al. (Roter, 1991; Roter and Larson, 2002) and it has been constructed as to be viable for all kind of sessions, e.g. conversations in the oncological setting (2017), between patients and psychotherapists or even patients and pharmacists. Moreover, RIAS was originally developed to annotate audio, while we transcribed the speech and annotated the transcriptions. This is motivated by the NLP analysis we wanted to perform on the text, e.g. syntactic and semantic analysis, machine learning, automatic dialogue act classification. Other dialogue annotation schemes exist, namely (Bunt et al., 2017; Serban et al., 2017; Stolcke et al., 2000), that includes rich taxonomies of communicative functions. The ISO 24617-2 standard, for example, includes the specification of the Dialogue Act Markup Language (DiAML), used in many annotated corpora. In *RiMotivAzione* project, we deemed RIAS as the most useful one for its specific focus on medical conversation. Even though RIAS is the closest domain tagset to annotate our corpus, some problems still emerged and they will be presented in next section.

3 Corpus Annotation

The *RiMotivAzione* corpus includes two complete cycles of physiotherapy sessions with two patients in post-stroke rehabilitation (namely, P1 and P2) and three physiotherapists (T1, T2, T3). The interviews were video recorded in IRCCS Fondazione Ospedale "San Camillo" in Venice. Each session lasted about 1 hour. The physiotherapy cycle for patient P1 included 14 sessions, while P2 took 16 sessions. Therefore the total duration of recordings is about 30 hours.

The patients were carefully selected by the doctors, since they must present some features. Above all, they had to agree to be part of the experimentation and they needed to talk in Italian. In an environment where dialect is still strong, their ability to speak Italian was not to be treated lightly. Moreover, the patients did not have to present any issues related to aphasia. These requirements restrained the viable options to two candidates.

Both speakers were encouraged to talk freely about any topic that may have emerged. Their only constraint was the use of Italian; when people slipped into dialectal terminology (in this case, Venetian), it was explicitly marked with the `<dialect>` tag in the corpus. The audio tracks were transcribed and annotated following Savy's (2005) guidelines for orthographic transcription for spoken Italian, where applicable. As a pre-processing, we used two Automatic Speech Recognition (ASR) systems, i.e. Google Speech-to-Text and Nuance Transcription Engine. Automatic transcriptions were corrected manually and anonymized. Video and audio tracks have been separately saved for future projects.

Overlapping between the two speakers and pauses were not marked, as it was not relevant to our study. Similarly, any intervention in the dialogue from a third party was not transcribed since our interest was solely in the doctor and patient's linguistic behaviours. Each dialogue turn of the corpus was annotated by two different annotators

following the RIAS guidelines. All the annotators have a background in linguistics and a specific education about linguistic corpora. As a single dialogue turn may contain more than one sentence and more than one speech act, the tags assigned to each turn may be more than one.

RIAS tagset includes 29 categories divided in four macro-categories called Medical Interview Functions (MIF) that cover the majority of the exchanges between a doctor and a patient: Data Gathering, Information Exchange, Emotional Expression and Responsiveness, Partnership Building and Activation. Table 1 contains the list of categories occurring at least 200 times in the corpus, together with examples.

To the best of the authors' knowledge, the RIAS system has never been used to annotate sessions of physiotherapy until now. This means that not all of the tags applied completely to the situation, or that some tags may be under-represented compared to other studies: for instance, the tag `Concerns` was applied to few sentences, since patients in physiotherapy sessions may inherently express less concern than oncological patients.

All the categories defined in Roter et al. (2017) were used. Moreover, two more tags were added to include all the exchanges: `Unclear` and `Technical problems`. The first applied to incomplete sentences, unintelligible ones (also marked with the `<unclear>` tag), or even in cases where the sentence referred to the physical context, making the general meaning impossible to retrieve for the annotator. The second tag applied to situations where the wearable device wasn't working properly, therefore resulting in some technical issue out of the scope of the therapy.

Another issue concerns the use of irony. Specifically, Patient 2 heavily employed irony while talking to the therapist, even when the dialogue concerned his health and well-being. Irony is hard to interpret, resulting in the difficulty to assign correctly a tag to those sentences. Tag `Jokes` was used in this case, and where inappropriate, a discussion between the annotators oriented the choice.

As the annotation task was difficult and it was inherently affected by subjectivity, we measured the resulting inter-annotator agreement and we put in place strategies to solve the disagreement, in order to annotate all the dialogue turns. The agreement calculated at this stage, according to the Cohen's score, was promising (k = 0.63). In case of disagreement (about 25% of the data), the process was followed by reconciliation or a final decision by a super annotator, where the two annotators could not overcome the disagreement.

The *RiMotivAzione* corpus has been built and archived according to GDPR norms. It is not publicly available but it can be requested to the authors for research purposes.

4 Corpus Analysis

The *RiMotivAzione* corpus contains about 98778 tokens. The total number of dialogue turns is 7670: 3377 dialogue turns in P1 sessions, 4293 in P2 sessions.

In Table 2 and Table 3 we reported the number of types, tokens, the ratio between types and tokens (the Lexical Richness Index) and the number of questions for the two patients.

It is worth noticing that Lexical Richness Index ranges from 0 to 1 and it is closer to 0 in the doctors' speech, meaning that medical personnel employ a poorer vocabulary while talking to a patient. This is due to the fact that a therapist needs to stick to a protocol and cannot digress over a certain limit. On the other hand, the patient talks quantitatively less: he pronounces fewer words, and most of the time those words are simple answers to the questions posed by the clinician. The patient talks less but he can wander more across conversation topics: he may disclose some personal detail about his life or just chit chat. This behavior is actually encouraged by the therapist, since it makes the therapy session less dull and more spontaneous for both the participants (Delany et al., 2010; Edwards et al., 2004). To sum up, the doctor needs to talk a lot to instruct the patients about the exercise they need to fulfill, as well as to ask questions (mainly regarding general well-being and inquiries about the therapy itself). Meanwhile, the patient may talk less because most of the time he just has to answer short questions (such as *"Does it hurt?"*); or, when he talks more, it is about some external topic which generates an increment in the vocabulary richness index.

As the main goal of the study is to replicate the clinician's communicative style onto a conversational interface, the major interest is on how the therapists talk, rather than the patients. Patients' manner of speaking is taken into consideration when imagining all the orders or phrases

Specific RIAS code	Examples
Social talk	non vedevo l'ora di venirla a trovare.
Directions	per scendere chiudo, per salire apro la mano.
Agreements	esatto, perché lo abbiamo registrato proprio cosí.
Medical condition	un po', poco, fastidio piú che male.
Approvals	bravissimo.
Unclear	[dialect] vara!
Therapeutic regimen	venerdí faremo la parte clinica ti faró io la scala di valutazione.
Jokes and laughter	ci vediamo domani, è piú una minaccia che un invito.
Asking for understanding	vorrei portarla cosí, hai capito?
Checking for understanding	chiudo le dita. cosí?
Concerns	sei sicura che funziona?
CeQ Medical condition	a fare gli esercizi non ha dolore?

Table 1: Tags and examples of categories occurring at least 200 times in the corpus.

Parameters	Patient 1	Therapist
Types	2065	3017
Tokens	10533	39305
Lexical Richness Index	0,19	0,07
Questions	40	667

Table 2: Patient 1 corpus.

Parameters	Patient 2	Therapist
Types	2451	2406
Tokens	18233	30707
Lexical Richness Index	0,13	0,07
Questions	380	805

Table 3: Patient 2 corpus.

Word	Frequency
vai	1166
apri	432
rilassa	400
bravissimo	353
mantieni	314
bravo	288
lascia	199
fare	187
prova	156
ottimo	153

Table 4: Most frequent Verbs and Adjectives used by therapist 1.

that the user could say to the voice assistant to express his needs. Table 4 and Table 5 list the most frequent Verbs and Adjectives pronounced by the physiotherapists. Apart from "Okay", which is the most frequent word for both therapists (1231 and 1019 occurrences), both therapists often use adjectives of positive value: *bravissimo, bravo, ottimo, buono*. Other frequent words are mainly verbs expressed at the first plural person, such as *we do, we'll try,* or equivalent expressions (*let's relax*). The use of the "we" is a communication element that aims at putting on the same level the clinician and the patient; the goal is to make the patient feel more comfortable and therefore enhancing the probability of therapy adherence. At the same time, adjectives such as "good" and "very good" praise the patient's efforts, underlining the progress he is making. The psychological component is of paramount importance during physiotherapy, especially for patients that suffered a stroke (Palma and Sidoti, 2019).

The quantitative analysis operated over the annotated corpus confirms the qualitative remarks made so far. In Figure 1 we present the distribution of dialogue tags, both for patients and therapists, i.e. the distribution of utterance type according to RIAS categories. We plotted on a logarithmic scale the frequencies of the tags.

Sentences annotated as `Social talk` were abundant, while those marked as `Concerns` were copious just for a patient, because he was frustrated about his health situation and the difficulties to manage the physiotherapy. During the sessions with Patient 1, the physiotherapist was able to engage a conversation about a hobby of his (motorcycles); even though this discussion topic is not relevant to the therapy, the fact that they were talking about something interesting for the patient contributed to the improvement of his med-

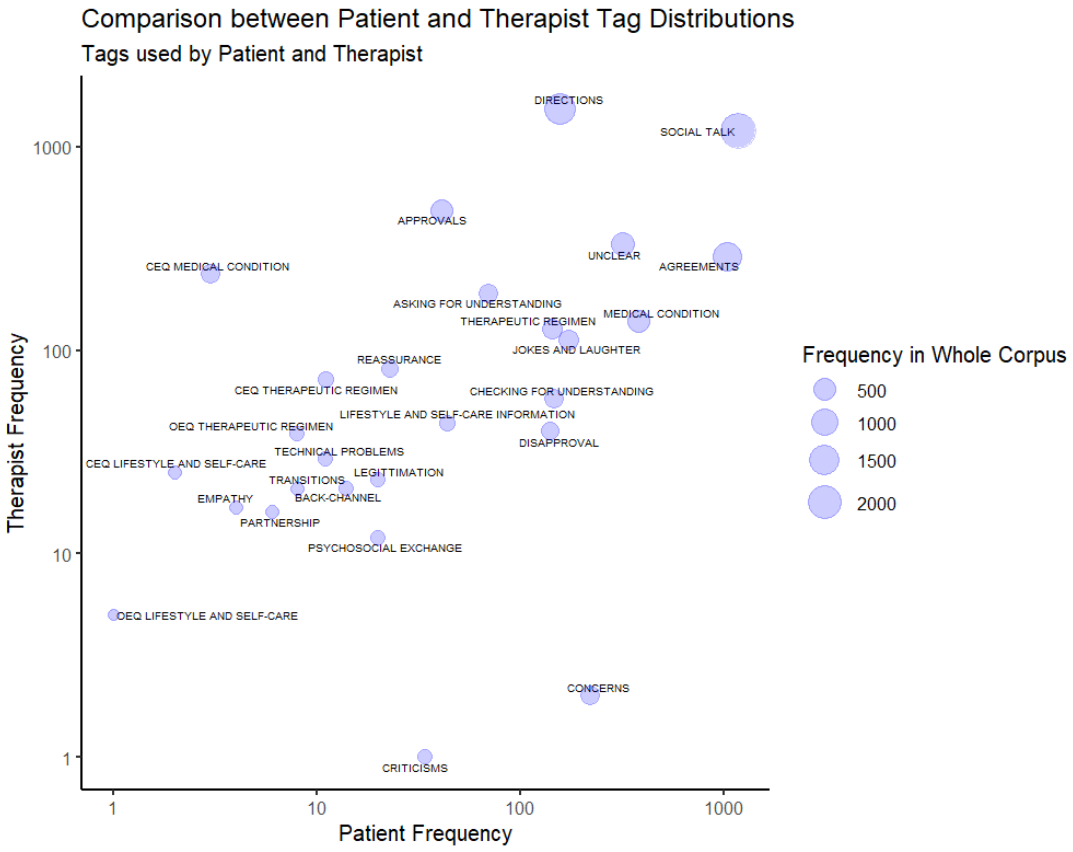

Figure 1: Distribution of dialogue tags in *RiMotivAzione* corpus

Word	Frequency
vai	340
proviamo	199
apro	198
pronto	174
facciamo	134
attento	124
andare	123
scendere	120
vediamo	115
fare	111

Table 5: Most frequent Verbs and Adjectives used by therapist 2.

ical condition (Gard and Gyllenstein, 2000).

All of these conversational elements are put in place willingly by the clinician and, even more, it is the style patients are used to. In the voice assistant design we try to mirror these strategies, providing praises when appropriate and asking questions to constantly monitor the user's well-being. The data extracted from the transcription and the annotation represents the most frequent linguistic behaviors emerged during the conversations. These patterns were used to build the conversational style and infrastructure of the dialogue system.

5 Conclusions and Next Steps

We created a corpus of conversations between patients and clinicians, in Italian, and we annotated the dialogue turns according to the Roter Interaction Analysis System (RIAS). This corpus was the first step in the design of a conversational interface integrated with a smart wearable device, to guide and assist the patients through the exercises assigned by the physiotherapist.

The first step in the future work will be to deepen the linguistic analysis conducted on the corpus, especially regarding the tagged dialogue acts. A stronger qualitative investigation over the data will be carried out. The second step will be to enrich the dataset: unfortunately, only two patients were deemed appropriate for the experimentation, while a corpus should contain dialogues from more speakers.

The *RiMotivAzione* corpus can be requested to

the authors for research purposes.

The system prototype will be tested in San Camillo Hospital by a set of stroke patients, following the clinical trial procedures. Thanks to the results of the test, we will produce experimental data to investigate if and how a voice assistant integrated with a wearable device can increase the effectiveness of the therapy.

6 Acknowledgments

RiMotivAzione is a two-year Research and Innovation project supported by POR FESR 2014-2020 Regione Piemonte. The partners are Koiné Sistemi, CELI, IRCCS Fondazione Ospedale San Camillo, Synesthesia, Istituto Italiano di Tecnologia (IIT) and Morecognition. We are thankful to our colleagues and project partners, in particular Paolo Ariano and Nicoló Celadon.

References

P. Basile and N. . Novielli. 2018. Overview of the evalita 2018 italian speech act labeling (ilisten) task. In *Proceedings of the Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2018) co-located with the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018)*.

D. Beltrami, G. Gagliardi, Rossini Favretti, E. R., Ghidoni, F. Tamburini, and L. Calzá. 2018. Speech analysis by natural language processing techniques: a possible tool for very early detection of cognitive decline? *Frontiers in Aging Neuroscience*, 10:369.

Harry Bunt, Volha Petukhova, David Traum, and Jan Alexandersson. 2017. Dialogue act annotation with the iso24617-2 standard. *Multimodal Interaction with W3C Standards*.

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

I. Chaoua, D. R. Recupero, S. Consoli, A. Harma, and R. Helaoui. 2018. Detecting and tracking ongoing topics in psychotherapeutic conversations. *AIH@ IJCAI*, pages 97–108.

C.M. Delany, I. Edwards, G.M. Jensen, and E. Skinner. 2010. Closing the gap between ethics knowledge and practice through active engagement: an applied model of physical therapy ethics. *Physical Therapy*, 90(7):1068–1078.

I. Edwards, M. Jones, J. Carr, A. Braunack-Mayer, and G.M. Jensen. 2004. Clinical reasoning strategies in physical therapy. *Physical Therapy*, 84(4):312–330.

G. Gard and A. L. Gyllenstein. 2000. The importance of emotions in physiotherapeutic practice. *Physical Therapy Reviews*, 5(3):155–160.

J. A. Hall and D. L. Roter. 2012. Physician-patient communication. In H. A. Friedman, editor, *The Oxford Handbook of Health Psychology*. Oxford University Press.

L. Laranjo, A.G. Dunn, H.L. Tong, A.B. Kocaballi, and al. 2018. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, 25(9):1248–1258.

E. Marzuki, C. Cummins, H. Rohde, H. Branigan, and G. Clegg. 2017. Resuscitation procedures as multiparty dialogue. In *Proc. SEMDIAL 2017 (SaarDial) Workshop on the Semantics and Pragmatics of Dialogue*, pages 60–69.

R.J. Moore, M.H. Szymanski, R. Arar, and G. J. Ren. 2018. *Studies in Conversational UX Design*. Springer.

S. Palma and E Sidoti. 2019. La comunicazione nei processi di cura. *COMUNIT IMPERFET*, 46(4):243–251.

D. Roter. 1991. *The Roter method of interaction process analysis (RIAS manual)*. The Johns Hopkins University, Baltimore.

D. Roter, S. Isenberg, and L. Czaplicki. 2017. The roter interaction analysis system: Applicability within the context of cancer and palliative care. *Oxford Textbook of Communication in Oncology and Palliative Care*, pages 717–726.

D. Roter and S. Larson. 2002. The roter interaction analysis system (rias): utility and flexibility for analysis of medical interactions. *Patient education and counseling*, pages 128–132.

R. Savy. 2005. *Specifiche per la trascrizione ortografica annotata dei testi in Italiano Parlato. Analisi di un dialogo*. Liguori, Napoli.

T. Sen, M.R. Ali, M.E. Hoque, R. Epstein, and P. Duberstein. 2017. Modeling doctor-patient communication with affective text analysis. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 170–177.

Iulian Vlad Serban, Ryan Lowe, Peter Henderson, and Joelle Pineau Laurent Charli and. 2017. A survey of available corpora for building data-driven dialogue systems. *arXiv:1512.05742*.

E. Sezgin, S. Yildirim, S. Ozkan-Yildirim, and E. Sumuer. 2018. *Current and Emerging MHealth Technologies: Adoption, Implementation, and Use*. Springer.

R. Shah and A. Chircu. 2018. Iot and ai in healthcare: A systematic literature review. *Issues in Information Systems*, 19(3):33–41.

Andreas Stolcke, Klaus Ries, and Elizabeth Shriberg Noah Coccaro. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3).