The RiMotivAzione Dialogue Corpus - Analysing Medical Discourse to Model a Digital Physiotherapist

The *RiMotivAzione* dialogue corpus

Analysing Medical Discourse to Model a Digital Physiotherapist

Francesca Alloatti, Andrea Bolioli, Alessio Bosca, Mariafrancesca Guadalupi
CELI - Language Technology, Università degli Studi di Torino
Turin
{francesca.alloatti, andrea.bolioli, alessio.bosca, mariafrancesca.guadalupi}@celi.it

Abstract

The *RiMotivAzione* project aims at providing a digital assistant that guides patients in their physiotherapy sessions at home. To properly develop this assistant, we gathered a corpus of dialogues between patients and physiotherapists. In this paper we present a deep and extended analysis of this corpus over different levels of granularity. The linguistic features extracted from the medical discourse were employed to model the *RiMotivAzione* chatbot, which will be experimented with patients at San Camillo Hospital in Venice (Italy).

Keywords: dialogue, physiotherapy, chatbot

1. Introduction

In the recent years there has been a steep increase in the application of ICT technologies to the healthcare domain (Kafle and Huenerfauth, 2018; Mieskes and Stiegelmayr, 2018; Liao et al., 2019). Specifically, one of these technologies is chatbots, or conversational agents. Most of them are created to help users to better communicate with the clinicians, as well as to help the medical personnel to monitor their patients (Laranjo da Silva et al., 2018). These chatbots are developed with different frameworks involving various techniques (reinforcement learning, pattern matching, etc.) and they are deployed throughout a range of platforms (Montenegro et al., 2019). Despite the abundance of systems, little to no description is provided about the language employed by the chatbots. By "language" here we mean the combination of words, phrases, tone and pragmatic features employed while giving instructions or providing any kind of medical assistance. A lot of attention has been payed to the linguistic features clinicians apply when talking to patients (Ferguson, 2012), therefore it is logical that the same focus should be applied when chatbots are the ones conversing with the patients. A correct use of medical language has been proved to be essential to a positive outcome of the treatment path (Hull, 2016) and a conversational agent should use the same terminology used by doctors and nurses.

In the *RiMotivAzione*[1] research project, a conversational interface is integrated with a visual App and a wearable device equipped with motor sensors. The project aims at assisting elders who suffered from a stroke and are under treatment for upper limb motor rehabilitation (Bolioli et al., 2019). The chatbot works as a virtual physical therapist guiding the patients through the exercises, giving advice and asking for information about the person's well-being. Given the aforementioned importance of the use of correct language in the medical domain, the interface has been modeled after therapists' real linguistic behavior: a corpus of conversation between doctors and patients was collected, transcribed and studied to retrieve information about linguistic communication in the physical rehabilitation domain. Preliminary results can be found in Bolioli et al. (2019), while in this work we present the data in more detail. We conduct a deeper analysis at various levels of granularity and provide more insight into the features of the physiotherapist-patient communication.

To the best of our knowledge, this is the sole corpus that deals with linguistic features employed in a specific medical setting - physical therapy sessions - in Italian. The corpus is not publicly released due to privacy reasons. It can be obtained for research purposes by writing to the authors.

2. Related Work

An analysis of related work unveiled various studies that share similarities with this one. The most similar one gathered a corpus of conversations between therapists and patients and analyzed it (Chang et al., 2013), although the language used is Korean. The analysis highlights some interesting features that can also be found in our corpus, such as the imbalance between the number of patients' utterances and the doctors' ones. They have also taken into consideration non-verbal behavior to measure empathy. Their goal was in fact to improve empathetic communication, while ours is to model a chatbot. Chaoua et al. (2018) also concerns analysis of patients-therapists conversations, although their setting is a psychological one, and their goal is topic detection and extraction. In a similar way, Jin (2018) focus solely on the analysis of small talk. Mieskes and Stiegelmayr (2018) inspect data from psychotherapy sessions with the aim of identifying what constitutes a sign of cooperation between the two participants. For this reason, their analysis is mainly qualitative. Finally, Wang et al. (2018) gather data about conversations in the pediatric domain.

Other works have different goals, such as producing a different metric to evaluate ASR system (Kafle and Huenerfauth, 2018), or even to model a dialog system (Gilmartin et al., 2018), although this last study does not focus on conversational interfaces in the healthcare domain.

[1] *RiMotivAzione* is a two-year Research and Innovation project supported by POR FESR 2014-2020 Regione Piemonte

Regarding the annotation and analysis of the corpus, we consulted the work by Shelley Staples (2016), in which certain linguistic features are extracted from corpora of medical discourse given their importance and ability to represent the quality of the exchange between doctors and patients. We analyzed our corpus identifying the features that we deemed relevant for the specific domain of physical rehabilitation. On the other hand, the annotation of speech acts poses a different kind of challenge. It has been tackled by various means and more recently through the employ of automatic systems (Basile and Novielli, 2018). However, no specific work focus on the automatic annotation of acts for the medical discourse, which may require a different set of tags and approach.

3. The *RiMotivAzione* Corpus

The *RiMotivAzione* corpus contains dialogues between a physician and a patient during the course of physiotherapy sessions. The people involved in the recording are two patients and three physiotherapists. Both patients are elderly (more than 60 years old) and males. Only one of the clinicians took care of both patients, while the other two were assigned to just one care recipient. The patients were selected by the research team at IRCSS San Camillo Hospital in Venice based on some preliminary tests. These tests aimed at identifying patients that could take part in the study by having certain characteristics: for instance, their speech needed to be sufficiently clear and they needed to be proficient in the Italian language. This was meant to exclude people who speak mainly in their own regional dialect, which is usually not intelligible from people from other parts of Italy.
Both patients signed an informed consent to be recorded and to have their data handled according to the current privacy laws.

3.1. Setting and Corpus Features

The sessions were recorded at IRCSS San Camillo Hospital. They were taped by means of a camera and the audio was extracted from the videos. This is due to some ambient conditions of the room where the sessions took place: high ceiling and the presence of temporary drywall generated a lot of noise and rumbling sounds from the rooms nearby. Professional recorders - that we employed at first - are very sensitive and captured each vibration, generating too much disturbance for the automatic transcription engine. Moreover, we did not want to use wearable microphones, since they would have disrupted the physiotherapy conditions.
The audio extracted from the video, in mono compression, has a lower quality but also presents less background noise. No additional information useful to the study could be deducted from the videos and they posed a privacy problem, therefore the visual track was not included in the corpus. The original files were transferred from the Hospital to the rest of the research team by mean of a private repository, in order to safely exchange data regarding the patients.
Each session is composed of three different stages, while only the last session contains four. The clinicians stopped the recording at each change of stage, so that the transition from one another was easier to understand even after the event. The stages are *Reception*, *Calibration* and *Therapy*. An addition *Screening* stage can be found mainly during the first or last session. The mean duration of each session is one hour.
During the transcription part each file pertaining to a different stage was joined in a chronological order, creating two main collections, one for each patient. The data available for each collection can be found in Table 1.

Data	**Format**
Unique ID per line	Integer
Transcription	String
Annotation	String
Timestamp	Date Time
Session stage	String
Session number	Integer
Name of the clinician	String
Patient anonymous ID	String
Age of the patient	Integer
Sex of the patient	String

Table 1: Corpus data and its format.

Additional information is available for each collection, such as the time span of the recordings, the total number of sessions and the total number of lines, whereas "lines" is used as a synonym for "turns". For the first patient collection:

- Time span: December the 3rd, 2018 to December the 20th, 2018
- Number of sessions: 14
- Number of lines: 3373

For the second patient collection:

- Time span: February the 25th, 2019 to April the 8th, 2019
- Number of sessions: 16
- Number of lines: 4293

3.2. Transcription Methodology

The corpus was produced by means of a semi automatic approach; we manually revised the textual output created by an automatic transcription engine in order to correct the problems emerged during the transcription and to obtain a dialogue corpus with a high degree of accuracy.
The automatic transcription was carried out with a transcription engine developed for commercial purposes. To adapt it to our need, the engine was fine tuned to a portion of the corpus data. This pre-processing was essential to improve the final performance of the system. However, the outcome still presented a significant Word Error Rate over the entire corpus, such that ex post processing (e.g. post transcription fine-tuning) did not produce meaningful results.
The rationale of this poor performance is to be found in the intrinsic nature of such data - dialogues in a real setting

- which is inherently more difficult than standard corpora. To this matter, we tested various ASR engines and obtained similar results. Moreover, the patients spoke Italian with a heavy accent, and even though they were asked not to use dialect, sometimes they slipped some words in their spoken flow without realizing it. None of these difficulties could be addressed automatically, so the entire corpus was manually revised.

3.3. Manual Revision

The output of the system was manually revised and corrected following Savy's guidelines (2005) for transcription of spoken Italian. We added proper punctuation to help interpret the meaning of the sentences and marked with a specific tag `Unclear` the parts that were either unintelligible or in dialect. Since the `Unclear` tag could be applied to single words or to entire sentences, the Word Error Rate proved to be an unreliable metric: some words could not be understood because of the dialect, or entire sentences were muffled by background noise such that even a human transcriber could not understand them. For the first patient, the `Unclear` tag appears 238 times, while for the second one 145 times. Proper names of patients and doctors were anonymized to preserve privacy.
Overlapping contents between the two speakers and pauses were not specifically marked or tagged, as it was not relevant to our study.

4. Corpus Analysis

Each collection properly assembled and corrected was analyzed with the goal of obtaining objective measurements of the physiotherapist linguistic behaviour. The features extracted were to help model the chatbot ability to efficiently communicate with the patients.
Even though the major focus is on the physiotherapist's part of the dialogue, we also analyzed and discussed the patient's speech. The goal was to highlight how he reacted to certain linguistic stimuli given by the doctor, if there were certain words he did not understand, what were his expectations - if he ever expressed any, etc. Since a chatbot is inherently less smart than a human therapist, we needed to predict any possible difficulty conveyed by the patient so to address it properly through an efficient conversational design.
The analysis was carried out on two levels of granularity: in the more detailed one, we considered the single token up to its morphemes, as well as the dependencies in a single sentence. This analysis was conducted with open source *StanfordNLP* library for Python [2]. On a broader level, we annotated each turn pertaining to the patient or the therapist with a dialogue act tag (ore more than one, if necessary). We employed the RIAS tagset, which is specific to the domain of medical discourse and thus allowed for a more precise definition of each dialogue act. RIAS was developed for encoding conversation in the medical domain in 1991 by Debra Roter et al. (1991) (2002) and it has been applied to various settings, e.g. to annotate exchanges between doctors and oncological patients, for psychotherapy sessions or even when the dialogue takes place between clients and pharmacists (Roter et al., 2017).

4.1. The RIAS tagset

The RIAS tagset has been designed to cover all the speech acts that could appear in the medical discourse. It contains 29 categories grouped in four macro-categories called Medical Interview Functions (MIF). These macro-categories are *Data Gathering, Information Exchange, Emotional Expression and Responsiveness, Partnership Building* and *Activation*. Table 2 contains an excerpt of the complete list of categories. For brevity reasons, we present only the ones that occur at least 200 times in the corpus, together with real examples taken from the dialogues. The examples are translated for the purpose of this paper and are selected for their clarity with respect to the category.

All the categories defined in Roter et al. (2017) were employed in the annotation. Nonetheless, not all of the tags always applied completely to the situation, or some tags were under-represented in this corpus compared to other studies: for instance, the tag `Concerns` was assigned to fewer turns, since these patients did not present a critical clinical situation and their chance of recovery was good (in contrast to other situation, such as an oncological one). Two additional tags were included to cover the entirety of speech acts in the dialogues: `Unclear` and `Technical problems`. The first one was used to tag incomplete sentences, the ones where the original audio was too compromised to understand the words, or the patients talked in their dialect. The `Unclear` tag was also employed in cases where the speech referred to the context in real time, making the general meaning impossible to retrieve for the annotator just by listening to the audio. The video track did not provide any help in resolving these matter. On the other hand, the `Technical problems` tag applied where the armband device used by the patients presented some issue. The two speakers may then discuss the subject of technological devices, which went beyond the tags presented in the RIAS tagset.

4.2. The Annotation Process

While the detailed analysis was conducted automatically with the help of *StanfordNLP* library, the speech act annotation was carried out manually. Three annotators took part in the work: one of them also served the purpose of super-annotator in case of disagreement. All the annotators have a formal education in Linguistics and they are aware of standards and annotation procedure regarding linguistic corpora. Each dialogue turn may contain more than one sentence and it may express more than one speech act. Therefore, a single turn can be tagged with two or more tags.

Inter-annotator agreement between two of the workers was calculated at (k = 0.63) according to Cohen's score. In case of disagreement, which happened in about 25% of the data, the super-annotator worked as a conciliator until all the annotators agreed to a final decision.

[2] `https://stanfordnlp.github.io/stanfordnlp/`

RIAS code	Example from the corpus
Social talk	non vedevo l'ora di venirla a trovare. *I couldn't wait to come and meet you.*
Directions	per scendere chiudo, per salire apro la mano. *to lower the hand I close it, to lift it up I open it.*
Agreements	esatto, perché lo abbiamo registrato proprio cosí. *exactly, we set it this way.*
Medical condition	un po', poco, fastidio piú che male. *just a bit, it bothers me rather than hurting me.*
Approvals	bravissimo. *very good.*
Unclear	[dialetto veneto] vara! *[venetian dialect] look!*
Therapeutic regimen	venerdí faremo la parte clinica ti faró io la scala di valutazione. *Friday we will do the clinical part I'll make the evaluation scale for you.*
Jokes and laughter	ci vediamo domani, è piú una minaccia che un invito. *see you tomorrow, sounds more like a threat than an invitation.*
Asking for understanding	vorrei portarla cosí, hai capito? *I'd like to bring it down like this, you see?*
Checking for understanding	chiudo le dita, cosí? *do I close my fingers, like this?*
Concerns	sei sicura che funziona? *are you sure it works?*
CeQ Medical condition	a fare gli esercizi non ha dolore? *do you feel pain while you do the exercises?*

Table 2: Frequent RIAS codes. Each code is presented with an explanatory example taken from the corpus.

5. Results

The complete *RiMotivAzione* corpus contains about 98778 tokens. The total number of dialogue turns is 7670: 3377 lines for Patient 1, and 4293 lines for Patient 2. To have a first overview on the exchanges between physiotherapists and patients, we present the number of types, tokens, and ratio between types and tokens (defined as the Lexical Richness Index), as well as the amount of questions for the two parts of the corpus (Table 3 and Table 4). Although the two patients do not present the same behavior regarding the number of questions, it can be noticed that the Lexical Richness Index ranges from 0 to 1 in both cases and it has a lower value for the physiotherapists' discourse. This means that doctors do not deploy a large and differentiated terminology, instead they rather stick to a certain script (which is usually an official one that has been validated by the hospital). On the other hand, patients may chat more about personal subjects since they do not need to comply to official clinical procedures.

From a stricter quantitative perspective, the patient produces less words. If we cross this information with the Lexical Richness Index data, we can infer that the patients may talk less but he can roam more freely from one topic to another. In fact, he may chat about some interest of his or about his personal life. This behavior is not only allowed but also encouraged, because it serves as a conversation enhancer and it produces health benefits for the patient, as mentioned in (Delany et al., 2010; Edwards et al., 2004) and in contrast with other findings in literature (Maynard and Hudak, 2008).

Data	Patient 1	Clinician
Types	2065	3017
Tokens	10533	39305
Lexical Richness Index	0.19	0.07
Questions	40	667

Table 3: Data from Patient 1 sessions.

Data	Patient 2	Clinician
Types	2451	2406
Tokens	18233	30707
Lexical Richness Index	0.13	0.07
Questions	380	805

Table 4: Data from Patient 2 sessions.

5.1. Part-of-Speech Analysis

A deeper analysis was conducted with respect to the part-of-speech of each token. Complete results can be found in Figure 1. For each patient, two physiotherapists conducted the sessions, according to their availability. The therapist number two intervened for both patients.

We focus on the most abundant PoS tag: verbs. Verbs are indeed the core of a sentence in a language such as Italian, and they express the essence of the action. In detail, verbs at the plural form were deemed to be particularly significant, in the light of their abundance. Table 5 and Table 6 highlight the usage of such verbs from both patients and clinicians. Although the values for Patient 2 are higher, in

	P1	T1	T2	T1+T2	P2	T2	T3	T2+T3	TOT_PATIENTS	TOT_THERAPISTS
ADJ	3403	2931	2849	5780	2172	2674	2579	5253	5575	11033
ADP	5227	4156	4026	8182	3499	3535	3742	7277	87726	15459
ADV	6556	5313	5290	10603	5122	5720	5357	11077	11678	28236
AUX	2182	1496	1445	2941	2018	1469	1470	2939	4200	5880
CCONJ	1288	1063	1065	2128	1255	1059	1001	2060	2543	4188
DET	6207	4990	4661	9651	3948	3948	4024	7972	10155	17623
INTJ	1711	833	975	1808	1288	1432	1436	2868	2999	4676
NOUN	8478	6942	6522	13464	5847	5751	5953	11704	14325	25168
NUM	496	305	288	593	381	247	356	603	877	1196
PRON	3696	2656	2571	5227	3574	2443	2536	4979	7270	10206
PROPN	268	121	127	248	295	124	167	291	563	539
SCONJ	1008	803	790	1593	1166	811	812	1623	2174	3216
VERB	10419	8431	8696	17127	7363	7250	7023	14273	17782	31400
X	2352	1662	1605	3267	2120	1552	1557	3109	4472	6376

Figure 1: Complete results of the PoS tagging analysis. *P* refers to Patient, whil *T* to the Physiotherapist.

both cases the one who largely employs verbs is the doctor. This is coherent with the greater use of nouns by the patient. Verbs are often in the indicative mood, which means that most sentences are main clauses. Main clauses are clearer, easier to process from a neurological point of view and they would serve better in the medical domain, where clarity is of paramount importance (Fengler et al., 2016). To corroborate these considerations about the doctors' manner of speaking, we cross this data with the analysis conducted on the dialogue acts. Most verbs in the indicative mood from the physiotherapists' discourse are embedded in turns tagged as `Directions`, where the clinician explains to the patient what to do in order to perform an exercise properly. The use of indicative can be expected while giving directions, since it allows for a clear discourse without any nested subordinate, but at the same time it is more polite than the imperative mood.

Verbs at the plural form	1185
Indicative mood	1019
Patient	182
Physiotherapists	1003
Embedded in `Directions`	846

Table 5: Verbs in Patient 1 sessions.

Verbs at the plural form	1381
Indicative mood	1292
Patient	492
Physiotherapists	1168
Embedded in `Directions`	969

Table 6: Verbs in Patient 2 sessions.

The second most frequent PoS tag are nouns. However, to the purpose of this study they did not represent an interesting area of analysis. Nouns may pertain to a broad variety of subjects, even some unrelated to the physiotherapy session. Patient 1, for instance, chatted about a hobby of his (motorcycles), therefore some nouns pertained to that area, which is not useful when analysing a medical discourse. This chatty behavior is quite frequent in the elderly, since they tend to talk about a variety of subjects even if they are not related to the context (Kallirroi et al., 2010).

On the other hand, the adjectives, especially the ones employed by the doctors, proved to be an interesting feature. Table 7 lists the the ten most frequent adjectives used by the physiotherapists. The frequency is computed over the totality of the corpus.

Rank	Frequency	ADJ
1	376	bravo *good*
2	366	bravissimo *very good*
3	271	pronto *ready*
4	244	esatto *exact*
5	159	ottimo *great*
6	126	attento *careful*
7	120	alto *high*
8	120	fermo *still*
9	102	giusto *right*
10	99	rilassato *relaxed*

Table 7: Adjectives employed by the physiotherapists.

Most adjectives express a positive sentiment, while the rest concerns technical aspects of the therapy (such as "high", employed while giving instructions to the patients on how to position the wrist). When modeling the chatbot such considerations are fundamental, because if a patients expect to be praised and encouraged during the sessions through the use of certain words, a digital assistant should behave the same way.

5.2. Analysis of the Dialogue Acts

The first analysis carried out on the dialogue acts aimed at identifying the quantity of each tag in the totality of the corpus. The distribution of the tags was plotted on a logarithmic scale for patients and physiotherapist (Figure 2). The utterances tagged as `Social talk` are abundant for both speakers, followed by the `Directions` sentences. Even though we could expect tags related to the physiotherapy sessions to be more copious, only the latter is directly connected to the therapy. `Social talk` is very present because it serves its purpose during the sessions: talking about personal matters, doing small talk has a positive ef-

fect on the medical outcome (Gard and Gyllenstein, 2000). Unfortunately, the tag `Unclear` is also quite abundant, although it must be considered that a single turn may have multiple tags, and the `Unclear` may refer to just a word or a part of that line, not the entire exchange.
Given the goal of the study - to model a chatbot after the doctors' way of talking - we focus on the tags that appear more in the physiotherapists' discourse. The great quantity of `Approval` tags is coherent with the analysis conducted on adjectives from the previous subsection: praises such as "alright, that was great" (Patient 2, line 224) or "very good, now close your hand in a fist" (Patient 2, line 115) contain the aforementioned adjectives and are indeed tagged as `Approval`.
The tag `CEQ - Medical conditions` refers to questions through which the physiotherapist checks on the patient's well-being. There are numerous utterances under this tag, which means that the clinicians often check on the status of their care recipient (e.g. "do you feel comfortable?" in Patient 1, line 1). The same considerations can be made for the `Asking for understanding` tag, where the doctor makes sure the patient is onboard with the therapy. `Concerns` are particularly abundant for the patient, as it is to be expected. No specific trend of tags could be found across the corpus, which means that the `Concerns` are distributed all over the sessions, and they do not increase nor decrease along the therapy path.

6. Discussion

The analysis over the physiotherapists' discourse revealed some interesting features. First of all, the great presence of verbs in a plural form. Most of these verbs are used together with the *we* pronoun, which suggests a cooperation between the patient and the doctor. Empathy is a fundamental component during the sessions which allows for a quicker healing process (Palma and Sidoti, 2019). The most frequent adjectives highlighted by the analysis are functional to the same pattern of action. The physiotherapists praise the patient's effort and employ a communication strategy that puts the two of them on the same level, eliminating any hierarchy that may cause discomfort.
Some dialogue acts also comply to this strategy: the abundance of `Social talk` and `Approvals` tags imply that digressing from the strict subject of the therapy serves a purpose in the medical discourse. If a patient is chatting and his efforts are reckoned, he may relax more, feel less pain and therefore find the physiotherapy session less hard.
All of these linguistic features represent valuable instruction on how a digital assistant for physiotherapy should be developed. However, not all the information from the dialogue can be mapped in the chatbot. The `Jokes and laughter` tag, for instance, refers to the use of irony (particularly heavy for Patient 2) and other jokes made from both speakers. Given the contextual nature of laughing matters, it would be unwise for a digital system to mimic such linguistic behavior. We said before that some of the `Unclear` tags are used when the patient is speaking in dialect. Such a feature, although very interesting from a sociolinguistic point of view, cannot be used when predicting the possible input. Dialect does not get properly transcribed by the ASR systems and it can be tricky to interpret even afterwards.
The *RiMotivAzione* chatbot should be clear when giving instructions but not stiff. It needs to check quite often on the patients well-being, making sure what is their level of pain and how are they handling it. It has to be able to correct the patient when he is performing the exercise wrong, but it should also praise him when he is getting good results. It has to be able to conduct small talk, but not to make jokes or comments that may result inappropriate out of context.

7. Conclusions

We gathered a corpus of dialogues between patients and physiotherapists recorded during real therapy sessions. The aim is to analyse the medical discourse and to extract relevant linguistic features at different levels of granularity. We first considered the single words and explored the value of the most frequent parts-of-speech: verbs and adjectives. Nouns were deemed not to be useful. Then, we annotated the dialogues with the RIAS tagset, a group of tags created to annotate medical discourse. The annotation and the subsequent analysis produced interesting results: the most frequent tags do not strictly concern the therapy, they rather serve as a psychological support for the patient. The analysis was expanded and deepened with respect to previous work (Bolioli et al., 2019).
All of these features have been incorporated in the development of the chatbot. The *RiMotivAzione* digital assistant is able to explain the exercises and provide praises when they are executed correctly. It can also check on the patient status and gather feedback about his level of pain. The chatbot, together with the smart wristband and the app will be experimented in San Camillo Hospital. After the experimentation, patients will be able to provide validation over various aspects of the project, including the language employed by the chatbot.

7.1. Future Work

In the future we plan on expanding the corpus. Unfortunately, only two patients could be enrolled in the present study, while we would like to add supplementary contributions to make the corpus more robust. More data could also be useful to conduct tasks such as automatic annotation and analysis of the tags distribution.
Future work will also embody the results from the experimentation with the patients in San Camillo Hospital, as well as more details about the interaction between the chatbot, the app and the smart wristband.

8. Acknowledgements

RiMotivAzione is a two-year Research and Innovation project supported by POR FESR 20142020 Regione Piemonte. The partners of the project are Koiné Sistemi, CELI, IRCCS Fondazione Ospedale San Camillo, Synesthesia, Istituto Italiano di Tecnologia (IIT) and Morecognition. The smart wristband was developed by Morecognition while Synesthesia developed the app that coordinates the wristband with the chatbot. We thank Andrea Turolla and Giorgia Pregnolato who enrolled the patients and will conduct the experimentation.

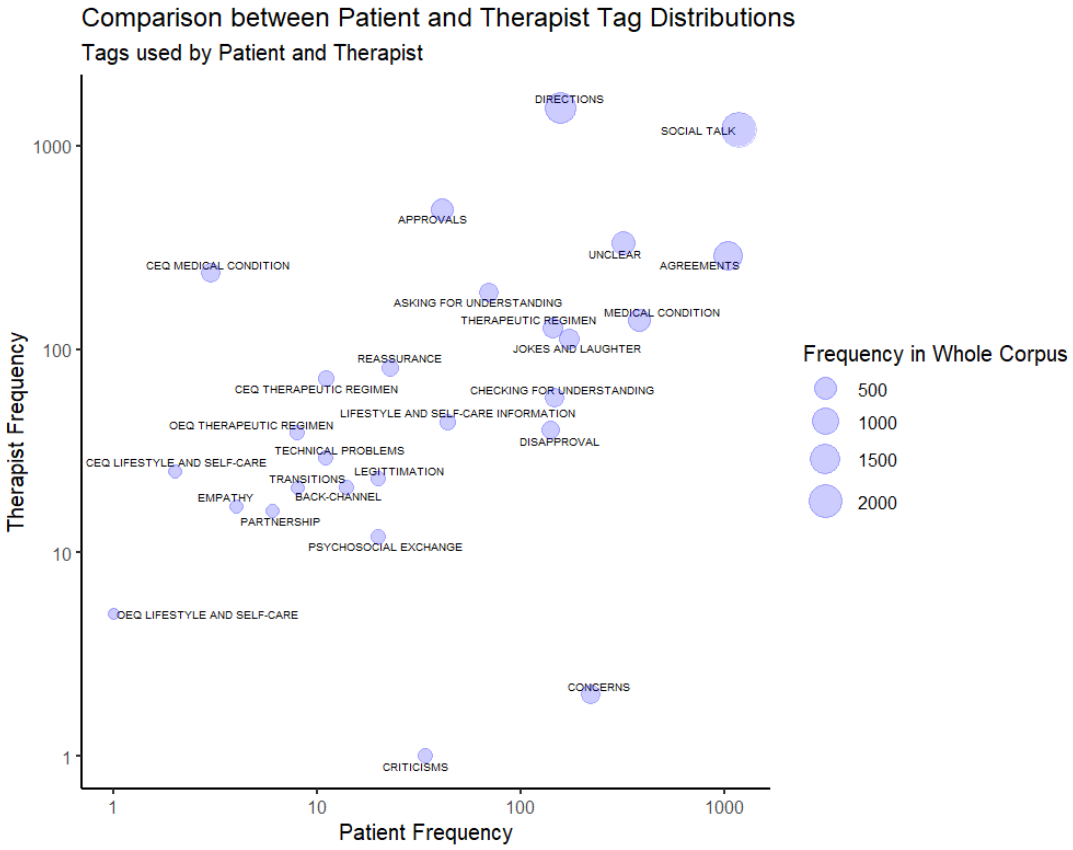

Figure 2: Distribution of dialogue tags across the entire corpus.

9. Bibliographical References

Basile, P. and Novielli, N. . (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)*.

Bolioli, A., Alloatti, F., Guadalupi, M., Lanzi, R. I., Pregnolato, G., and Turolla, A. (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, November 12-15, 2019.

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

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

Delany, C., Edwards, I., Jensen, G., and Skinner, E. (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.

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

Fengler, A., Meyer, L., and Freiderici, A. D. (2016). How the brain attunes to sentence processing: Relating behavior, structure, and function. *Neuroimage*, 129:268–278.

Ferguson, G., (2012). *English for Medical Purposes*, pages 243–261. John Wiley and Sons.

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

Gilmartin, E., Vogel, C., and Campbell, N. (2018). Chats and Chunks: Annotation and Analysis of Multiparty Long Casual Conversations. In Nicoletta Calzolari (Conference chair), et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12, 2018. European Language Resources Association (ELRA).

Hull, M. (2016). Medical language proficiency: A discussion of interprofessional language competencies and potential for patient risk. *International Journal of Nursing Studies*, 54:158–172.

Jin, Y. (2018). Small talk in medical conversations: Data from china. *Journal of Pragmatics*, 134:31–44.

Kafle, S. and Huenerfauth, M. (2018). A Corpus for Modeling Word Importance in Spoken Dialogue Transcripts. In Nicoletta Calzolari (Conference chair), et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*,

Miyazaki, Japan, May 7-12, 2018. European Language Resources Association (ELRA).

Kallirroi, G., Wolters, M., Moore, J. D., and Logie, R. H. (2010). The match corpus: A corpus of older and younger users' interactions with spoken dialogue systems. *Language Resources and Evaluation*, 44(3):221–261.

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

Liao, Q. V., Wang, Y.-C., Bickmore, T., Fung, P., Grudin, J., Yu, Z., and Zhou, M. (2019). Human-agent communication: Connecting research and development in hci and ai. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, pages 122–126, New York, NY, USA. Association for Computing Machinery.

Maynard, D. and Hudak, P. (2008). Small talk, high stakes: Interactional disattentiveness in the context of prosocial doctor-patient interaction. *Language in Society*, 37:661–668.

Mieskes, M. and Stiegelmayr, A. (2018). Preparing Data from Psychotherapy for Natural Language Processing. In Nicoletta Calzolari (Conference chair), et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12, 2018. European Language Resources Association (ELRA).

Montenegro, J. L. Z., da Costa, C. A., and da Rosa Righi, R. (2019). Survey of conversational agents in health. *Expert Systems with Applications*, 129:56–67.

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

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

Roter, D., Isenberg, S., and Czaplicki, L. (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.

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

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

Staples, S., (2016). *Identifying Linguistic Features of Medical Interactions: A Register Analysis*. Palgrave Macmillan.

Wang, N., Song, Y., and Xia, F. (2018). Constructing a Chinese Medical Conversation Corpus Annotated with Conversational Structures and Actions. In Nicoletta Calzolari (Conference chair), et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12, 2018. European Language Resources Association (ELRA).