COMMON GROUND, FRAMES AND SLOTS FOR COMPREHENSION IN DIALOGUE SYSTEMS

Philippe Blache and Matthis Houlès

Laboratoire Parole et Langage, CNRS, Aix-en-Provence, France

ABSTRACT

This paper presents a dialogue system for training doctors to break bad news. The originality of this work lies in its knowledge representation. All information known before the dialogue (the universe of discourse, the context, the scenario of the dialogue) as well as the knowledge transferred from the doctor to the patient during the conversation is represented in a shared knowledge structure called common ground, *that constitute the core of the system. The* Natural Language Understanding *and the* Natural Language Generation *modules of the system take advantage on this structure and we present in this paper different original techniques making it possible to implement them efficiently.*

KEYWORDS

Dialogue systems, common ground, natural language understanding.

1. INTRODUCTION

We present in this paper a multimodal dialogue system for training doctors to break bad news. This system consists in asking trainees, following a given scenario, to announce the patient a problem that occurred during a medical act [13]. Doctors are frequently faced with such a situation in real life, and official agencies (e.g. the French "*Haute Autorité de la Santé*") underline the fact that training communication skills for interacting with patients is of deep importance. A typical bad news is a damage associated to the care, consequence of an un-expected event that can be due to a medical complication, a dysfunction or a medical error. Experienced clinicians consider the task of announcing this type of information as difficult, daunting, and stressful. The problem is that training doctors in this perspective remains a complex task: it is organized by hospitals as workshops during which doctors interact with actors playing the role of patient [11]. Such training solution is difficult to implement, expensive and time-consuming. The ACORFORMed project has proposed to develop an immersive platform in virtual reality [13] with an *embodied conversational agent* simulating a patient interacting with the doctor. However, the first version of this platform was equipped with an efficient dialogue system, capable of understanding precisely doctor's productions and generating appropriate reactions. We describe in this paper a system fulfilling these requirements.

The main difficulty in dialogue systems concerns the understanding module and more generally semantic processing. This task still represents a challenge and a research question for open domain dialogues. Fortunately, task-oriented dialogue systems (and even more crucially training-purpose applications) correspond to a very specific situation in which semantics can be controlled precisely ad being restricted to the task itself. In our use case, the system (i.e. virtual patient) has a complete knowledge of the scenario and the context. Moreover, the user (i.e. the doctor to be trained) receives before the interaction a set of information with the patient's medical

David C. Wyld et al. (Eds): NATL, NETWORKS, Fuzzy - 2021
pp. 33-43, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.112003

folder, the context description that led to the problem (e.g. surgery, endoscopy, therapy, etc.), the description of the bad news and how (if possible) it should be fixed. He/she also receives several recommendations on the way to deliver the bad news, following official guidelines elaborated by national agencies. The user can freely talk with the virtual patient that generates in response multimodal verbal and non-verbal reactions. In such a context, the dialogue structure is very specific. Technically, this means that the semantic domain is closed and the system has the entire knowledge of the discourse universe (including the specific scenario associated to the task). Beside this characteristic, the system also have information about how the doctor has to announce the new. A last and very important feature of this specific training context is that the doctor remains the main speaker all along the interaction. On its side, the patient (played by the dialogue system) only reacts to the doctor's utterances, without taking the lead of the conversation.

These different characteristics deeply impact on the one hand the behaviour of the agent and on the other hand the technology to be used for the comprehension module. Moreover, as far as agent's behaviour is concerned, the most important feature of the system consists in how it reacts to doctor's utterances in particular by answering questions, producing feedbacks and asking for clarification. In terms of understanding techniques, thanks to the pre-defined knowledge of the discourse universe, the core of the architecture relies on a knowledge structure precisely defined before the conversation.

We propose in this paper a description of the main aspects of the comprehension module of our system. We use machine learning techniques when possible, but the main architecture remains symbolic, in particular because of lack of data in this domain. Moreover, many multimodal behaviours of the virtual agent are directly controlled by rules during the comprehension process. Finally, the system is designed to be used for training: the state of the common ground after the interaction is an important element of evaluation for the trainee. Deep learning approaches would not provide such a facility.

We focus in the remaining of the paper on these two aspects: knowledge representation for understanding and generating patient's behaviours.

2. KNOWLEDGE REPRESENTATION: THE *COMMON GROUND*

As underline above, the context of dialogue systems for training is very specific from many respects. First, the universe of discourse (i.e. the semantic domain)is fully specified both for the knowledge it concerns (in our case the context of the damage, and all the medical aspects) but also in the way the information has to be delivered, according to certain requirements and recommendations [17]. The doctor's discourse is organized around three main phases: *greetings, damage description and remediation, closing* [12]. Moreover, and this is of great importance for knowledge representation, both the doctor and the patient have a complete knowledge of the context, the degree of severity, the risk, etc.

In terms of interaction theories [15], information updating consists in building a shared knowledge between the speakers, called "*common ground*" [18], made of what is supposed to be known by both participants. The task consists in adding step by step during the conversation new information in relation to an item of the common ground. In this common knowledge base, many information is also presupposed or can be inferred automatically depending on the instantiated knowledge. What is specific to the common ground is that all participants suppose the others also have access to the same knowledge. In the case of a training dialogue environment, the context, the scenario and the recommendations are already known by the system before the interaction (this fact is hidden to the trainee). The evolution of the information transfer from the doctor to the

patient consists in specifying in the knowledge base what has been transferred. Moreover, in the situation of a task-oriented dialogue, the system knows at anytime not only what has been updated, but also what remains to be instantiated.

Formally, the common ground is made of a set of frames in the sense of frame semantics [8], defined as attribute-value matrices gathering different pieces of in-formation, called slots as illustrated in figure 1. A slot value can be atomic (e.g. values of the slots *Name*, *Age*, *Gravity*, etc.) or refer to another frame (e.g. *Person*, *Pathology*). Moreover, each slot can be associated with different control information [2]. First, each slot may be weighted, in a 3-value scale: *mandatory*, *important*, *optional*. Second, an information may depend from another frame or slot value. For example, the doctor cannot describe any remediation before having presented the pathology: in this case, we say that the frame `Remediation` depends on the frame `Pathology` description. Finally, the last important control concerns slots: depending on certain values, particular agent's reactions may be triggered. For example, if the value *high* is instantiated to the slot *Severity*, then an emotional feedback may be generated by the agent. All this information is to be encoded by specific constraints associated to the frame or the slot description, each slot bearing the following features: *type*, *weight*, *dependent-values*, *inference*.

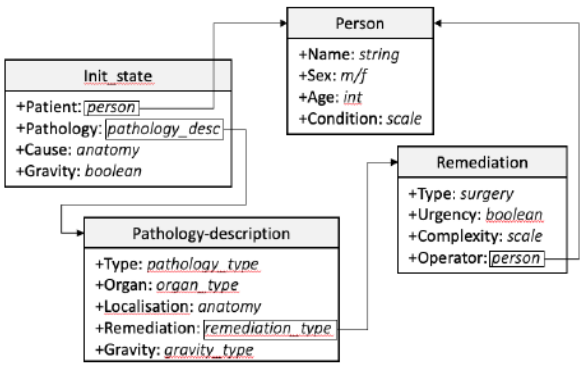

Figure 1: Example of frames and slots

3. UNDERSTANDING WITH COMMON GROUND: FRAMES AND SLOT INSTANTIATION

As sketched in the previous section, the understanding mechanism of our dialogue system only relies on common ground instantiation. This knowledge structure being based on a frame lattice (each frame being made of slots), the first step of the comprehension module consists then in identifying such frames from the doctor's speech. In this perspective, it is interesting to note the correspondence between dialogue acts [7, 6] and frames required by the common ground representation (for example the different phases of the dialogue). The mechanism for frame identification relies on this correspondence, and we propose to use a dialogue act classification technique for that.

3.1. Dataset

We collected a corpus of training sessions, in French, organized between doctors (the trainees) and patients (played by human actors). The corpus is made of 7 sessions each lasting around 15mns. The audio input (representing 37,000 words) has been transcribed and manually

corrected. The corpus has been automatically segmented into inter-pausal units (with pauses higher than 250ms).

Each inter-pausal unit forms an utterance, 1,822 such utterances have been produced throughout the 7 dialogues by the doctors. The corpus has been annotated automatically when possible (tokenization, POS tagging) and manually as for the dialogue acts associate to each utterance (5 annotators among which two experts).

3.2. Step 1: Frame identification

A dialogue in our type of discourse is structured into different phases, on top of classical opening and closing: the description of the patient's initial state (the cause of the hospitalization), the bad news description (typically an incident during a surgery) and the patient's current state. Moreover, the doctor also gives explications, asks questions, reassures the patient and have different types of social interactions. These different actions correspond to different dialogue acts to be annotated: *Opening, Init state, Init remediation, Bad news state, Bad news remediation, Current state, Current remediation, Reassurance, Explication, Social interaction, Discourse, Question, Closing*.

In our approach, each dialogue act corresponds to a frame in the common ground. The first step of the comprehension module is then to identify these frames that can be associated to each doctor's utterance during the dialogue. This problem corresponds to a classification one in which the predictive variables are extracted from the utterance. We propose to implement this classification task on the basis of different linguistic features that already have been shown to be effective [4]:

Classical features: We first use a set of features classically involved in DA identification. It consists in combining TF-IDF principles with word and character n-grams. Applying a principal component analysis, we extracted 4 combinations to be tested:

- f-TFIDF (TFIDF on word n-grams from 1 to 3 words keeping the 250 best, and character n-grams (from 3 to 5 chars, keeping the 250 best)
- s-TFIDF The f-TFIDF features, filtered with a singular value decomposition in order to obtain a better representation density
- w-TFIDF TFIDF only based on the word n-grams, keeping the 500 best
- l-TFIDF TFIDF based on the lemmas n-grams, keeping the 500 best

Morpho-syntactic features: We also involve in the model low-level morpho-syntactic features, based on POS tags: number of discourse markers in the utterance, number of filled-pauses, number of tokens.

Lexical features: A dictionary specific to our domain has been created, containing medical words in which we distinguished pathological terms vs. others. Moreover, we annotated the data with a specific label tagging the medical words depending on they appear for the first time in the dialogue or not (corresponding to the given/new distinction used in discourse analysis).

Context features: As proposed in several works [5, 16], context (i.e. the labels of the preceding dialogue acts) is taken into account. We implemented three different context representations, in a 1 to 5 window: one hot encoding of the preceding DAs, bag-of-words(encoding the number of times the DA appears in the context of the utterance), n-grams of words (up to 0.5% frequency).

Syntactic features: High-level syntactic information can play a role in the characterization of certain classes. In particular, dialogue sequences corresponding to a description or an explanation are usually associated to more complex structures, with more modifiers (adjectives and adverbs) and more complex clauses (subordinates, relatives, prepositional phrases). We propose two features for a simple approximation of these characteristics: the ratio of the number of adjectives and adverbs to the total number of tokens in the utterance $\frac{nbAdj + nbAdv}{\sum tokens}$ and the ratio of the number of conjunctions, pronouns and prepositions to the total number of tokens $\frac{nbConj + nbPrep + nbPro}{\sum tokens}$.

A hierarchical top-down classification, limited to two levels (DA meta-classes and leaf classes) which consists in training a multi-class classifier for each level[10] has been implemented. We keep as first-level classes the meta-classes specified in the ISO 24617-2 scheme: `Opening`, `Discourse`, `Inform`, `Question`, `Closing`. These classes are not only easier to identify, they also correspond to different agent's reactions (standardized reaction or feedbacks in association with `Opening`, `Discourse` and `Closing` and appropriate answers (see next section) with `Questions`. The dialogue act `Inform` correspond to the majority class, that we separate into 8 subclasses: `Init_state`, `Init_remediation`, `Bad_new_state`, `Bad_new_remediation`,`Current_state`, `Current_remediation`, `Social_interaction`, `Explication`.

Different algorithms and feature combinations have been tried for training the classifiers at both levels. The best result for the first level classification has been obtained using a linear regression classifier with an Anova to select the k-best features. As expected, the accuracy is very high, reaching 94% (89% of balanced accuracy). Note that the 1st-level classes are relatively stable and easy to recognize, the context feature did not bring any improvement there.

The second step of the classification consists in applying a new classifier to the sequences labeled `Inform` by the 1st-level classifier. For this step, the best results have been obtained using random forests with the complete set of features and reaching an accuracy of 77.2% (71% balanced accuracy). More details on this stage can be found in [4].

3.3. Step2: Slot filling

The second step in the common ground instantiation concerns slot filling. At each doctor's utterance, we identify a frame (corresponding to the dialogue act)thanks to the classifier. Several approaches have proposed to process at the same time frame classification and slot filling in a unique mechanism [9]. In our case, the dialogue act classifier returns the frame: thanks to the common ground, we know then the set of slots to be instantiated. This task consists in identifying the value and the slot to be instantiated. More precisely, two steps can be specified:

1. Extracting from the utterance the possible values for the different slots
2. Selecting the slot, verifying the type compatibility, instantiating the value

The fact that the list of slots prone to instantiation is very small opens the possibility to adopt a specific mechanism. Instead of trying to identify before-hand the possible slot values from the utterance and then to look for the slot taking into account the compatibility of the type of its value, we propose to implement a reverse mechanism based on semantic similarity. Instead of an abstract type, each slot is associated with a prototypical value. For example, the slot "*specialty*"

of the frame "*doctor*" takes as value "*surgeon*". Then, for each term of the utterance, the semantic distance with the prototypical term of the slots is calculated. Above a given similarity threshold, the slot is then instantiated taking as value the term itself. In our case, the similarity is calculated with the *Gensim* package (*https://radimrehurek.com/gensim/*).

4. QUESTIONS, CLARIFICATION, FEEDBACKS

In a task oriented-dialogue, both the semantic domain and the task to be filled are completely known by the system before the interaction. Moreover, in the case of a medical conversation (typically for breaking bad news), the main speaker remains the user (i.e. the doctor). The main task of the dialogue system understanding module is to update the knowledge transferred by the doctor, in other words to instantiate the common ground. The virtual patient, on its side, has only few information to transfer and its main activity consists in reacting to doctor's messages and behaviours. This is of deep importance for agent's naturalness and credibility. Three main reactions have to be implemented in priority: answering doctor's questions, generating conversational feedbacks, and asking for clarification. This section presents different solutions addressing these specific problems.

4.1. Answering questions

Questions are identified by the dialogue act classifier. A distinction is done between open-ended questions (wh-questions) and closed-ended questions (yes-no questions).

Yes-no questions: This type of questions focuses on the patient's condition ("*Are you in pain this morning*?", "*Did they bring you pain medication*?"), under-standing ("*Do you have any questions?*"), or social aspects ("*Do you want us to call your son?*"). The answers to these questions depend on a scenario or user profile. It does not provide information used by the dialog system. The choice of the type of answer, "*yes*" or "*no*" (or any other positive or negative rephrasing) is left to the system and not based on any particular semantic processing.

Wh-questions: In this case, it is necessary to identify the type of questioning and the informational *focus* of the question. Generally speaking, an open-ended question is made up of an interrogative particle giving the type of question followed by a description of the focus of the question, which is a specific property of the object or event to which the question relates. An open-ended question can therefore be represented by the doublet <*question type; focus property*>. *Question type*: We traditionally distinguish 8 types of open questions, corresponding to different forms of the interrogative particle: *who, what, when, where, why, how, what, to whom*. We propose to associate a generic type for the answer of each of these question types:

Table 1. Interrogative particles and their types.

Interrogative particule	*Question type*
who; to whom	person
what	object
when	time
where	location
why	event
how	condition

Focus of the question: Depending on the type of question, the associated characteristic is identified, describing a property of the object of the answer. For example, if the type of question is a location, the focus will generally be an action (or possibly a state), represented by a verb phrase (the head of which being an action verb). Table 2 summarizes the prototypical focus for each type of question.

Generating the answer: Knowing the expected answer type and the focus of the question makes it possible to generate straightforwardly an answer, based on generic patterns. All questions refer necessarily to the patient's state or personal data. This information is then already encoded in the common ground, as part of the scenario. Generating the answer consists then in looking into the common ground for a slot value corresponding to the expected answer type, together with the focus as a key for identifying the associated frame.

4.2. Conversational feedbacks

We propose an approach making it possible to generate feedbacks on the basis of different cues that can be identified in real time from doctor's behaviour. In our feedback model, besides low-level classical cues (such as breaks, turn length, POS, etc.), we also integrate higher level semantic or discourse-level cues [3].

Table 2. Expected answers depending on the type of question.

Question type	Expected Answer	Focus	Example
person	Proper name, professional category, family category	Action VP	*Who brought you a painkiller? Who did you talk to?*
object	Common Name	Generic object	*What medication did you take?*
time	Temporal NP	Action VP	*When were you brought the medication?*
location	Spatial NP	Action/state VP	*Where does it hurt? Where do you put your glasses?*
state	Adv, PP	Action/state VP	*How are you feeling this morning? How did you get to the hospital?*

The dialogue systems mainly have two input streams: the audio signal and its transcription. Prosodic features (silent pauses, pitch, IPU duration, etc.) are extracted from the audio stream. Temporal features, also coming from this stream, are kept updated, in particular the duration since the last feedback, the indication of the current state of the production (speech or silent pause), the duration of the speech since the last pause, the duration of the pause, etc. On their side, linguistic features can be acquired from the transcription stream: morphosyntax (POS n-grams), lexicon (some terms can trigger specific feedbacks), but also at a higher level the information structure (the introduction of a new referent) or the discourse organization (transition between phases) can also be associated with specific listener's reactions [1]. Finally, semantics plays a central role in generating feedbacks: many listener's reactions are triggered upon instantiation of the common ground.

We propose to implement a semantic-based feedback generation by associating CG slots to specific feedbacks (for example, a feed-back expressing fear is triggered instantiation of the slot "urgency"). Note that in the case of a task-oriented dialogue, most feedbacks are triggered by linguistic cues. As a consequence, when the doctor speaks, we first look at linguistic cues whereas during a pause, the feedback generator is mainly based on pause duration. The different cues extracted from the analysis of the input streams serve as input to a feedback type identification function, based on a set of rules, as illustrated in figure 3. Given the feedback type to be generated and the current mode (pause or speech), the last step consists in generating the

feedback itself, by choosing among a list of possible candidates. This list is in a probability space which also depends on the current state (e.g. visual or bimodal feedback will be preferred during speech where verbal feedbacks will be favoured during pauses).

4.3. Clarification questions

Clarification questions plays an important role in dialogue not only for the verification of the common ground construction, but also (even to a greater extent) for the naturalness of the virtual agent: such questions show very efficiently that the agent understands and follows the conversation. Different conditions can trigger such questions.

As explained in the description of the common ground structure, frames and slots can be associated with different controls. First, some frames or slots can be instantiated only when other frames or slots are already instantiated. Such values form a pre-requisite and are called *dependent values*. For example a doctor cannot present a diagnostic (i.e. the system cannot instantiate a `diagnostic` frame)before having presented the symptoms (resp. instantiation of a `symptom` frame. In the same way, in the case of our use case, the bad news frame cannot be created before having developed the `init_state` one. Such relations make it possible to implement both sequentiality and semantic dependencies. A dependent value conflict is detected when a slot is about to be instantiated with a dependent value still free.

Table 3. Feedback generation rules

Level	*Cue*	FB type	Description
Prosody	elapsed_time_pause > 200ms	generic	
Discourse	new_referent	specific	Use of a new term
Syntax	POS == [V,N | V,Adv]	generic	The last previous POS
Semantics	medical_term	generic	When using a medical term
Semantics	positive_emotion	agreement	a positive emotion term triggers an agreement
Semantics	negative_emotion	disagreement	negative emotions trigger disagreement
Discourse	DA == bad news	fear	When the phase becomes "bad news"
Discourse	DA == social interaction	generic	After a social interaction

In the case of a slot dependency, the conflict is directly identified by verifying whether the dependent slot is already instantiated or not. If not, the slot name and its value type are passed to the generation module which select a question pattern filled with this information. As for frames, the dependent value conflict requires a more complex process. The problem consists in identifying whether a frame is instantiated or not: in most of the cases, only part has been informed, the frame remaining incomplete. The problem is then to evaluate whether the frame can be considered as complete or not. As presented with the common ground, we have seen that each slot is associated to a weight (3-value scale). When a frame A is dependent from a frame B, the system verifies whether all mandatory slot values of B are already instantiated1. If not, then a clarification question is generated, using the same generation mechanism as for slots. The second situation triggering a clarification question occurs when no slot can be instantiated in spite of the identification of a frame by the classifier. In this case, none of the terms belonging to the utterance is similar enough (i.e. reaches a sufficient similarity threshold) with one of the prototypical values of the different slots. In this case, we can say that there is a type mismatch between the term used by the doctor for a slot value and the expected value. The clarification question generated is general, simply indicating an incomprehension of the system. The last case processed by our system concerns general questions that can be asked by the system at the end of the interaction or when the doctor asks the agent whether he/she has some questions. The mechanism consists in selecting either a non saturated frame or the frame with the more non-

instantiated slots weighted as important. This frame being chosen, the system selects arbitrarily one of the non instantiated important slots and generates a question.

5. GENERATING A MULTIMODAL ANSWER

We briefly sketch in this section the multimodal aspects of the dialogue system. Concerning the input signal, our goal being to develop a generic application, the first constraint concerns the user's equipment: we want to remain totally free of any instrument or specific sensor, the user freely speaking (or writing) to the agent. Our second goal is to offer the possibility of having a written output (in the situation where the system only generates written sentences), an audio (the system speaks to the user) or a multimodal one. In this last case, the system generates the code for a complete behaviour (speech and gestures) of an embodied conversational agent.

Concerning the input stream, at this stage of development, only the audio modality is processed. Automatic speech recognition is applied to the doctor's production, providing its transcription plus some prosodic information, in particular concerning pauses. The transcription is then segmented into discourse units delimited by discourse markers (*then, because, and, but,* etc.) that indicate approximately a change between different discourse units. Such segmentation makes it possible to identify homogeneous semantic units that can be associated to only one or two frames by the classifier. At this stage, the video signal that could be used for detecting head movement, smiles, etc. is not already processed by the system. As a consequence, we mainly use as input unimodal information based on the transcription only. On the contrary, concerning the output, the system generates multimodal behaviours played by an ECA implemented in the Virtual Interactive Behaviour platform [14]. This platform includes an XML dialect called FML encoding instructions for generating the agent's verbal and non-verbal behaviour. The mechanism consists then for our system to generate (dynamically when necessary) this FML code for each answer or reaction of the agent.

Figure 2. The embodied virtual patient of the ACORFORMed platform

We implemented two different ways for generating such multimodal behaviours depending on whether they include flexible verbal material or not. In the last case (typically feedbacks), the list of possible behaviours is closed and rather small. Moreover, they are very canonical and the verbal material very limited, fixed or even absent. It is then possible (and preferable in terms of efficiency)to create a specific gesture library corresponding to the required feedbacks and then to encode a predefined list of FML files for these prototypical behaviours.
Each feedback may have different realizations that can be chose randomly in order to introduce variability. The generation of flexible behaviours including verbal material requires a three-step

mechanism. First, the verbal part is generated following the methods described above. The verbal utterance to produce is then passed to the FML generator. This module generates a first version of the FML code my using the standard VIB function (FMLAnnotator) and then refine this code by completing or adding different instructions (e.g. the prosody).

6. CONCLUSION

Using dialogue techniques for training doctors represent a seminal use case both in the perspective of developing new dialogue techniques, but also in terms of application: training doctor's social skills is known to be of crucial importance in the therapy process. Such dialogue systems require at the same time to be very precise, reactive and to allow doctors to interact freely, with spontaneous speech: this correspond to the most difficult challenges for dialogue technology. We have presented in this paper an approach fulfilling these requirements by taking advantage of the particularities of this type of application. Our approach relies on a precise knowledge representation, the common ground, which constitutes the core of the understanding architecture. The frame-based representation first offer the possibility to use classification techniques identifying directly the frame to be instantiated. Thanks to this first step, we have proposed an original slot filling method, based on the common ground and distributional semantics information. The generation of the agent's reactions and its adaptation to the doctor's speech directly takes advantage of our common ground representation.

ACKNOWLEDGEMENTS

This work has benefited from support from Institut Convergence ILCB (ANR-16-CONV-0002).

REFERENCES

[1] Bertrand, R., Espesser, R. (2017) Co-narration in French conversation storytelling: a quantitative insight. In Journal of Pragmatics111
[2] Blache, P. (2017) Dialogue management in task-oriented dialogue systems. In: International Workshop on Investigating Social Interactions with Artificial Agents
[3] Blache, P., Abderrahmane, M., Rauzy, S., Bertrand, R. (2020) An integrated model for predicting backchannel feedbacks. In: ACM International Conference on IntelligentVirtual Agents
[4] Blache, P., Abderrahmane, M., Rauzy, S., Ochs, M., Oufaida, H. (2020) Two-level classification for dialogue act recognition in task-oriented dialogues. In: COLING'20
[5] Bothe, C., Weber, C., Magg, S., Wermter, S. (2018) A context-based approach for di-alogue act recognition using simple recurrent neural networks. In: LREC 2018
[6] Bunt, H., Fang, A.C., Petukhova, V. (2017) Revisiting the iso standard for dialogue act annotation. In: Proceedings of the 13th Joint ISO-ACL Workshop on InteroperableSemantic Annotation (ISA-13)
[7] Core, M., Allen, J. (1997) Coding dialogs with the damsl annotation scheme. In: Working Notes of AAAI Fall Symposium on Communicative Action in Humans and Machines. pp. 28–35
[8] Fillmore, C.J., Baker, C. (2009) A frames approach to semantic analysis. In: Heine, B.,Narrog, H. (eds.) The Oxford Handbook of Linguistic Analysis. Oxford UniversityPress
[9] Firdaus, M., Golchha, H., Ekbal, A., Bhattacharyya, P. (2020) A deep multi-task model for dialogue act classification, intent detection and slot fill-ing. Cognitive Computation https://doi.org/10.1007/s12559-020-09718-4,https://doi.org/10.1007/s12559-020-09718-4
[10] Freitas, A., de Carvalho, A. (2008) A tutorial on hierarchical classification with applications in bioinformatics. In Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications pp. 119–145
[11] Granry, J.C., Moll, M. (2012) Rapport de mission, état de l'art en matière de pratiques de simulation dans le domaine de la santé. Tech. rep., HAS 2012
[12] Ochs, M., Blache, P., Montcheuil, G., Pergandi, J.M., Bertrand, R., Saubesty, J.,Francon, D., Mestre, D. (2018) The Acorformed corpus: Investigating multimodality inhuman-human and human-virtual patient interactions. In: CLARIN Annual Conference 2018. p. 16

[13] Ochs, M., Mestre, D., de Montcheuil, G., Pergandi, J.M., Saubesty, J., Lombardo, E., Francon, D., Blache, P. (2018) Training doctors' social skills to break bad news: Evaluation of the impact of virtual environment displays on the sense of presence. In Journal on Multimodal User Interfaces1
[14] Pelachaud, C. (2009) Studies on gesture expressivity for a virtual agent. In Speech Communication51(7), 630–639
[15] Pickering, M., Garrod, S. (2013) An integrated theory of language production and com-prehension. In Behavioral and Brain Sciences36(04), 329–347
[16] Raheja, V., Tetreault, J. (2019) Dialogue act classification with context-aware self-attention. In: NAACL-2019
[17] Schnebelen, C., Pothier, F., Furney, M.: Annonce d'un dommage associé aux soins. (2011) Tech. rep., Haute Autorité de Santé
[18] Stalnaker, R. (2002)Common ground. In Linguistics and Philosophy 25(5), 701–721

AUTHORS

Philippe Blache is senior researcher at the CNRS, France (Laboratoire Parole & Langage, Aix-en-Provence). He is the founder and the former director of the ILCB (Institute of Language, Communication and the Brain)

Matthis Houlès is student at the Engineer school of Nantes