Managing Linguistic and Terminological Variation in a Medical Dialogue System

Leonardo Campillos-Llanos Dhouha Bouamor Pierre Zweigenbaum Sophie Rosset
LIMSI, CNRS, Université Paris-Saclay, 91405 Orsay, France
firstname.lastname@limsi.fr

Abstract

We introduce a dialogue task between a virtual patient and a doctor where the dialogue system, playing the patient part in a simulated consultation, must reconcile a specialized level, to understand what the doctor says, and a lay level, to output realistic patient-language utterances. This increases the challenges in the analysis and generation phases of the dialogue. This paper proposes methods to manage linguistic and terminological variation in that situation and illustrates how they help produce realistic dialogues. Our system makes use of lexical resources for processing synonyms, inflectional and derivational variants, or pronoun/verb agreement. Specialized knowledge is used for processing medical roots and affixes, ontological relations and concept mapping, and for generating lay variants of terms according to the patient's non-expert discourse. We report the results of a evaluation of the non-contextual analysis module—which supports the Spoken Language Understanding step—after 11 users interacted with the system. The annotation of domain entities obtained 91.8% of Precision, 82.5% of Recall, 86.9% of F-measure, 19.0% of Slot Error Rate, and 32.9% of Sentence Error Rate.

Keywords: medical terminology, natural language understanding, virtual patient consultation

1. Introduction and Related Work

Terminology management is a core component in medical informatics applications. While this need has long been identified for health professionals (Cimino, 1998), the needs of patients and lay people have only been addressed recently (McCray et al., 2000; Zeng-Treitler et al., 2007).

Virtual patients (VP) are interactive systems and require managing terms—e.g. by formalizing ontological concepts (Nirenburg et al., 2008)—and a Natural Language Understanding (NLU) module. The NLU component may rely on text meaning representations for resolving paraphrases (Nirenburg et al., 2009) or a corpus of questions and answers curated by an expert (Kenny et al., 2008).

We are developing a conversational agent to be used in a simulated consultation with a VP, where the system aims at training medical doctors (Campillos-Llanos et al., 2015). Users (medical students or doctors) interact with the VP to collect information that allows them to provide a correct diagnosis. Medical trainers define each e-learning case beforehand by entering the VP profile data in a clinical record (e.g. symptoms or medical history).

Managing linguistic and terminological variation is crucial to match a user's question to a term in the clinical record and to select suitable terms for answer generation. This paper gives an overview of the difficulties (Section 2.) and strategies applied in both analysis (Section 3.) and generation (Section 4.). We also report the results from an evaluation (Section 5.) and conclusions (Section 6.). Although the system currently only supports French, the challenges found might be raised regardless of users' language.

2. General Overview

Each turn in a dialogue system includes steps of analysis, dialogue management, and generation. Additionally, a virtual patient must have a model of its (health) state, which is here provided by the contents of its clinical record.

The analysis involves, firstly, a *non-contextual analysis* (NCA) step (i.e. analysing the input without context information). In the NCA step, the question terms, standard entities and medical entities in the input are detected and annotated semantically (e.g. *paracetamol* is a DRUG). The current version of the system manages 139 entity types: 100 domain entities (71.9%), 24 (17.3%) miscellanea tags—e.g. general question types such as *quand* ('when') or *pourquoi* ('why')—and 15 labels (10.8%) for managing the dialogue—e.g. salutations such as *bonjour* ('hello'). To annotate entities from the medical domain, we use gazetteers/lists (Table 1) and semantic rules.

The second step of the analysis concerns matching entities against data in the clinical record. This processing poses difficulties caused by the variability of medical terms[1]. A concept may be referred to by a variety of acronyms and jargon terms (e.g., *tonsillectomie*, 'tonsillectomy' and *extraction des amygdales*, 'removal of tonsils') and lay variants from other registers (e.g., *opération des amygdales*, 'tonsils operation'). Another challenge concerns semantic modeling: the system should know that *essoufflement* ('breathless') is a symptom and is related to a physiological function (*respirer*, 'to breathe'). In the generation step, the VP should reply coherently, as a patient, with lay terms.

In both steps, lexical resources provide synonyms, inflectional and derivational variants, or pronoun/verb agreement. Specialized knowledge is used for concept mapping, ontological relations and medical roots and affixes. Table 2 breaks down the number of variants, minimun, maximum and mean values per word entry or per CUI, and number of word entries or CUIs in each resource (for relations between CUIs, the number of related pairs is reported).

Figure 1 shows a sample dialogue with all the types of variation that we detail in the next sections (examples are in English for the sake of understandability).

[1] We do not address here the variation related to spelling errors, for which we have tested two spelling correctors.

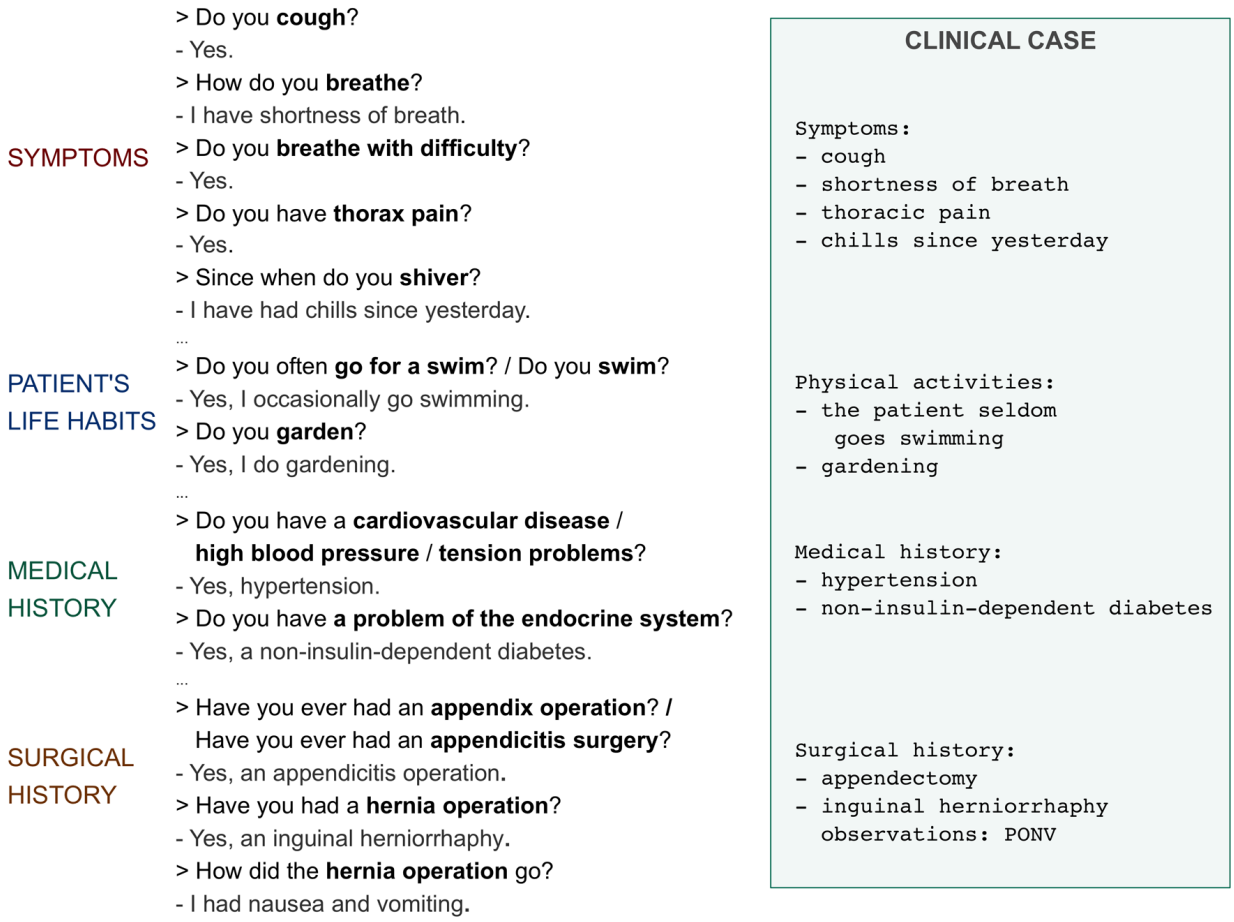

Figure 1: Sample dialogue and clinical record

3. Analysis Step

3.1. Linguistic Variation

Linguistic variation between the input and the contents of the clinical record is managed through inflectional and derivational variants and synonyms. Inflectional variants (e.g., *jardinez* and *jardiner*, 'to garden') are obtained from a general-language dictionary (Courtois, 1990). Deverbal nouns (e.g., *jardiner*, 'to garden', and *jardinage*, 'gardening') are obtained from VerbAction (Hathout et al., 2002). Derivational variants of medical terms (e.g., *thorax* with *thoracique* ['thoracic']) come from the UMLF lexicon (Zweigenbaum et al., 2005). Additional synonyms (e.g., *nage* and *natation*, 'swimming') are obtained from a synonym dictionary (Rosset et al., 2008).

3.2. Terminological Variation

Medical vocabulary is mainly processed through the UMLS® (Bodenreider, 2004) Metathesaurus terms and relations. Medical roots and affixes and auxiliary lists of terms not found in the UMLS complement these strategies.

3.2.1. Terms Referring to the Same Concept

The UMLS Metathesaurus records term variants associated to the same concept through a common Concept Unique Identifier (CUI). For example, *pression artérielle élevée* ('high blood pressure', input) is mapped to *hypertension* (clinical record) thanks to their common CUI (C0020538).

However, not all terms are recorded in the UMLS. This is the case of most verbs referring to symptoms (e.g., *tousser*, 'to cough'), which miss a link to the corresponding nouns (e.g., *toux*, 'cough', C0010200). We created lists to cluster them, including single- and multi-word verbs/idioms. *Lemmatized forms* are obtained for multi-words: e.g., *[vous] respirez avec difficulté*, reduced to *respirer avec difficulté*, maps to *difficulté à respirer* ('difficulty breathing').

An auxiliary list is used for additional lay variants (e.g., *problèmes de tension* 'tension problems' maps to *hypertension* (C0020538)).

More approximate designations are sometimes used: for example, although there is no direct UMLS relation between *appendicectomie* ('appendicectomy', C0003611) and *appendicite* ('appendicitis', C0003615), they may be related in a dialogue. To cope with this when other methods fail, we rely on lists of medical affixes and roots: e.g., *appendic-* in the previous examples, to match the terms *chirurgie de l'appendicite* ('appendicitis surgery', input) and *appendicectomie* ('appendicectomy', clinical record). To build these lists, we selected neoclassical compounds in the Specialist lexicon® (McCray et al., 1994) and adapted them to French morphology according to (Namer and Zweigenbaum, 2004).

3.2.2. Using Hierarchical Relationships

Hierarchical relationships are needed to cope with a variety of contexts involving disorders or surgical procedures.

Clinical record section	Type of semantic entity	Example	Count
Lifestyle	Addictive substances	*marijuana*	70
	Alcoholic beverages	*vin* ('wine')	35
	Daily activities and acts	*monter des escaliers* ('to climb stairs')	111
	Diets	*régime* ('diet')	69
	Food	*viande* ('meat')	1491
	Recreational activities	*nager* ('to swim'), *natation* ('swimming')	327
History/symptoms	Allergies	*allergie au latex* ('allergy to latex')	227
	Anatomy	*thorax, thoracique* ('thoracic")	24369
	Anesthesias	*péridurale* ('epidural')	499
	Circumstances related to conditions	*effort physique* ('physical effort')	206
	Disorders	*hypertension*	147058
	Findings	*logement humide* ('damp housing')	134
	Medical devices	*pacemaker*	1758
	Medical doctors and specialists	*cardiologue* ('cardiologist')	105
	Obstetric/gynecological history	*césarienne* ('cesarean')	1020
	Physiological functions	*respirer* ('to breathe'), *digestion*	96
	Surgical procedures	*appendicectomie* ('appendectomy')	5269
	Transfusions	*autotransfusion*	64
	Vaccines	*vaccin antigrippal* ('antigripal vaccine')	142
	Symptoms	*saigner* ('to bleed'), *hémorragie* ('bleeding')	9140
	Descriptions of signs/symptoms:		
	Changes in symptom/condition	*aggravé* ('aggravated')	152
	Colours	*jaune* ('yellow')	33
	External characteristics	*sanglant* ('bloody')	33
	Intensity	*violent*	62
	Irradiation of pain	*irradier* ('to irradiate')	13
	Onset type of symptom/condition	*progressif* ('progressive')	37
	Other features (e.g. type of pain)	*lancinant* ('stabbing')	124
	Volume	*épais* ('thick')	21
Treatments	Galenic form	*comprimé* ('pill')	96
	Medical drugs	*paracétamol*	43222
	Method of administration	*par voie orale* ('orally')	130
	Treatments	*dialyse* ('dialysis')	2404
Examinations/analyses	Analyses/diagnostic procedures	*radiographie* ('radiography')	4899
	Examinations involving surgical proc.	*coloscopie* ('colonoscopy')	39
	Laboratory and clinical tests	*hémogramme* ('blood count')	5841
Miscellanea	Adverbs and expressions of manner	*anormalement* ('abnormally')	31
	Adverbs and expressions of quantity	*beaucoup* ('many')	10
	Expressions of duration	*constamment* ('constantly')	50
	Expressions of frequency	*souvent* ('often')	115
	Relative position	*à droite* ('to the right'), *inférieur* ('lower')	34
Total			249541

Table 1: Lists in the resources for NCA analysis in the current version

For example, the doctor might ask whether the patient has a type of disorder (e.g., *maladie cardiovasculaire* 'cardiovascular disease') when the clinical record mentions a specific disorder (e.g., *hypertension*). UMLS *child of* (CHD) relationships are used for this purpose.

Some terms referring to classes of disorders follow the pattern *disease* + ANATOMY: e.g., *maladie* + *de* + ANATOMY (*maladie des yeux*, 'eye disease'). However, term variants in the UMLS do not always match this pattern exactly. For example, concept C0015397 has term *trouble de l'oeil*, but not *maladie des yeux*. Fortunately, most disorder terms are related to their anatomical site by SNOMED CT relation *has finding site* found in the UMLS: e.g., 'non-insulin-dependent diabetes' *has finding site* 'endocrine system', from which we obtain that *diabète non insulinodépendant* ('non-insulin-dependent diabetes') is a kind of *affection du système endocrinien* ('problem of the endocrine system').

Some symptoms or disorders are related to physiological functions: e.g., *essoufflement* ('breathlessness') and *respirer* ('to breathe'). A list of correspondences between those types of entities is used to match them. Data were extracted from UMLS terminologies and their relations: namely, ICD10, MeSH and SNOMED. Hierarchical relationships between concepts referring to symptoms or disorders (especially, *is a*) were also used. For example,

Step	Resource	Variants	Min	Max	Mean	Entries/CUIs
Generation	Verb/pron. correspondences	48429	1	1	1.00	24829
	Scient./lay term corr. (with CUIs)	22	1	9	5.50	4
	Scient./lay term corr. (without CUIs)	60	1	1	1.00	36
Analysis	Inflection	631035	1	61	7.96	91571
	Synonyms	18663	1	143	13.50	15049
	Derivational variants	20043	1	9	2.56	8008
	Terms with CUIs:					
	Anatomy	7861	1	29	3.15	18177
	Disorders/Symptoms	106387	1	34	2.86	369846
	Surg./therap. procedures	33741	1	24	2.63	130685
	Terms without CUIs:					
	Symptoms (vbs./idioms)	707	1	36	14.42	50
	Other terms	122	1	22	9.54	13
	Roots/affixes	681	1	12	2.14	318
	Relations between CUIs:	# Pairs of concepts (CUIs)				
	Child of	170571				
	Procedure - Disorder	11854				
	Procedure - Anatomy	95744				
	Disorder - Phys. function	8144				

Table 2: Resources for managing linguistic and terminological variation in the current version

disorders and symptoms related to *respirer* ('to breathe') were extracted using, among others, ICD10 class R06 ('Abnormalities of breathing') and MeSH subtree C23.888.852 ('Signs and symptoms, respiratory').

Other terms referring to surgical procedures follow the pattern *operation/intervention* + ANATOMY (e.g. *intervention cardiaque*, 'heart intervention'). However, term variants in the UMLS do not always instantiate this pattern: e.g., concept C0003611 is designated by the term *appendicectomie*, but not *opération de l'appendice*. Again, to detect the equivalence of these two terms, SNOMED CT relationships such as 'appendicectomy' *has procedure site* 'appendix' were extracted from the UMLS.

Entities referring to surgeries may also have the structure DISORDER + *surgery* (e.g., *opération de hernie*, 'hernia surgery'). Entities with this pattern are not always matched to UMLS variants. For example, concept C0021446 has the term *cure de hernie inguinale*, but not *opération de hernie inguinale*. Here, SNOMED CT relations *has procedure morphology* and *has direct morphology* were obtained from the UMLS to link surgery procedure terms to related diseases: e.g., *opération d'hernie* ('hernia operation') and *herniorraphie inguinale* ('inguinal herniorrhaphy').

4. Generation Step

Two main constraints are addressed during this step. First, data in the clinical record include personal pronouns and verbs referring to the patient in the third person, which require to be changed for the virtual patient to reply in the first person. Regular expressions and a list of pronoun and verb transformations are applied before output. In the record of Figure 1, the string *le patient fait de la natation rarement* ('the patient seldom goes swimming') is changed into *je fais de la natation rarement* ('I seldom go swimming').

Second, the virtual patient should favor lay terms over more technical terms. For this purpose, each set of terms sharing the same UMLS CUI is sorted by degree of technicality: e.g., for concept C0036973 ('shiver'), *grelottements* is the less technical term, and *frissonnement* is the most technical. This degree of technicality was computed by comparing the probabilities of a term according to two language models respectively trained on a technical corpus of medical articles (CRTT)[2] and on a non-technical corpus of online medical forums[3]. The degree of technicality of a term is computed as the likelihood ratio of these two probabilities (Bouamor et al., 2016). To generate a lay variant of a term, its CUI is determined and the least technical term for this CUI is chosen.

Additionally, a manually created list of {technical, lay} term pairs is used for terms lacking a UMLS CUI, or terms for which no degree of technicality could be computed because they were unseen in our training corpora: e.g., *nausées et vomissements* ('nausea and vomiting') refers to *NVPO* ('PONV', 'Postoperative Nausea and Vomiting').

5. Results and Evaluation

The system has been tested on three patient cases with project partners and during public demonstrations, and a first evaluation has currently been carried out. The interface of the prototype to test the cases is available online[4].

Here we report the results from the evaluation of the *non-contextual analysis* (NCA) step. We have thus evaluated the ability of the system to *understand* user's input (i.e. the *Spoken Language Understanding* component). The evaluation we explain just focuses on domain entities; out-of-domain entities such as conversational acts (e.g. salutations) will not be considered in the results here presented.

[2] http://crtt.univ-lyon2.fr/les-corpus-medicaux-du-crtt-613310.kjsp

[3] http://www.atoute.org

[4] www.audiosurf.net/pg5c/select_case.php

HYP: À <Qquantite> combien </Qquantite> était votre <Qsymptome> <symptome> fièvre </symptome> </Qsymptome> ?
REF: <Qtemperature> À combien était votre <symptome> fièvre </symptome> </Qtemperature> ?
('How high was your fever ?')

Figure 2: Sample of NCA annotation

The evaluation procedure was as follows: 11 non medical professionals (computer science researchers and engineers) interacted with the three cases available online. They were told to interact freely with the virtual patient, but some instructions were given regarding the types of dialogue acts the system can cope with (e.g. questions related to medical history or lifestyle, but not instructions or out-of-domain requests). From December 2015 to February 2016, 349 utterances[5] were collected. We rejected a total of 51 lines (14.6%) with spelling or grammar errors, or expressing dialogue acts the system is not designed to process (e.g. instructions or prescriptions). For the evaluation presented in this work, we were interested in evaluating how the system processed only domain entities when users requested data from the clinical record. That is why we also ruled out 56 utterances with dialogue acts unrelated to the record, and deleted out-of-domain entities. We finally used 242 utterances of the initial data collected (1356 words; Table 3).

Users	Tests	Utterances			Words		
		N	M	SD	N	M	SD
11	20	242	12.1	8.6	1356	5.6	2.6

Table 3: Details of the data for this evaluation (NCA step)

We shall recall that the NCA step in our system relies on the annotation of domain-specific entities in user's utterances. Figure 2 is an example of annotation for the utterance *À combien était votre fièvre ?* ('How high was your fever?'). The system hypothesis appears above, and the correct reference, below; the utterance is incorrectly tagged with one substitution and one insertion. We have therefore used standard metrics of named entity recognition systems for this evaluation, namely Precision, Recall and F-measure, and the Slot Error Rate (Makhoul et al., 1999). These measures are computed with the following counts:

- C: Correct tags in the hypothesis (true positives)
- I: Inserted entities in the hypothesis (false positives)
- D: Deleted entities in the hypothesis (false negatives)
- S: Substituted entities in the hypothesis
- Hyp: Total of correct and wrong tags in the hypothesis
- Ref: Total of correct and wrong tags in the reference

Precision (P) is the ratio between the correct annotations and all annotated entities:

$$P = \frac{C}{Hyp}$$

Recall (R) is the ratio between the correct annotations and the entities found in the reference:

$$R = \frac{C}{Ref}$$

The F-measure (F) is the harmonic mean between P and R, which is normally balanced with $\beta = 1$:

$$F = \frac{(1+\beta^2)PR}{\beta^2 P + R} = \frac{2PR}{P+R}$$

Finally, the Slot Error Rate (SER) is calculated as follows:

$$SER = \frac{S+D+I}{Ref}$$

We have also computed a commonly used metric in dialogue and natural language processing tasks: the Sentence Error Rate (SeER), which is the ratio between the sentences with at least one error and all of the correct sentences:

$$SeER = \frac{\#Wrong\ sentences}{\#Correct\ sentences}$$

All of these measures are expressed as percentages. Broadly speaking, the lower SER and SeER values, the better the system annotates; conversely, higher P, R and F-measures reflect a better performance.
Table 4 breaks down the results of the evaluation. The current version of the SLU module lacks enough coverage of domain entities, although the detected items tend to be annotated with high precision.

Ents. hypothesis: 390		Ents. reference: 434		
C	I	D	S	Errors
358 (82.5%)	15 (3.5%)	59 (13.6%)	17 (3.9%)	91 (21.0%)
Precision	Recall	F-measure	SER	SeER
91.8	82.5	86.9	19.0	32.9

Table 4: Results of the evaluation (NCA step)

An analysis of the evaluation data showed interesting results. Firstly, we looked at the types of domain entities annotated in users' interventions. We considered the groups related to each clinical record section (see Table 1) [6]. Figure 3 shows that most entities annotated in users' input (system hypothesis) were related to patient's history or complaints (48.5%), especially symptoms (23.8%). 20.8% of annotated entities were miscellanea items unrelated to any clinical record section (e.g. entities expressing frequency or quantity). Then, 15.1% of entities addressed sections related to patients' lifestyle (e.g. recreational activities, 4.8%, or smoking habits, 3.6%). Entities annotating personal data

[5] We denote by the term *utterance* any user's turn; system replies are excluded here. The NCA module may tag any utterance with more than one entity, or without any entity at all.

[6] Note that this classification only fits our project needs; some entity types could be classified in other groups.

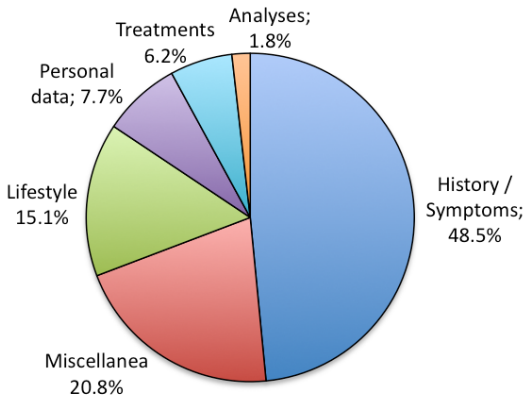

Figure 3: Entity types in users' input w.r.t. clinical record

represented a 7.7%, and a few proportion of entities were related to treatments (6.2%) or analyses (1.8%).

Secondly, an error analysis of the data showed the entities causing a lower performance. Table 5 shows some examples of entities that deserve commenting on. Note that we only recount entities occurring at least five times in the samples[7] (figures need to be put in perspective due to the scarcity of our data). The entity types with poorer recall (i.e. the system did not annotate them) were those related to ambiguous items that can be both a disease or symptom (e.g. *tension*) or entities for detecting physiological functions (e.g. *respirer*, 'to breathe'). Rules for processing questions on the aim of the consultation had also a poor recall. Regarding symptoms, the list of noun (N) entities (e.g. *toux*, 'cough') showed lower precision and higher recall, whereas the list of verb (V) entities (e.g. *tousser*, 'to cough') had lower recall but higher precision. Conversely, lists of surgery and disease entities had high recall but low precision. Large lists of these types of entities increased recall but caused false positives: e.g. *rouge* ('red') in *viande rouge* ('red meat', FOOD) was annotated as DISEASE.

Entity type	Example	P	R	F
Ambiguous (symp./dis.)	*tension*	100.0%	14.3%	25.0%
Phys. funct.	*respirer* ('breathe')	100.0%	33.3%	50.0%
Symptom (N)	*toux* ('cough')	90.6%	100.0%	95.1%
Symptom (V)	*tousser* ('to cough')	100.0%	80.0%	88.9%
Surgery	*greffe* ('transplant')	60.0%	100.0%	75.0%
Disease	*diabète* 'diabetes'	35.7%	100.0%	52.6%

Table 5: Results of the annotation of some domain entities

Our data are insufficient to fully evaluate how input entities are matched with concepts in the clinical record. Nevertheless, in Table 6, we show some examples of users' terms that succesfully matched concepts in the patient's record (UMLS CUIs are included to identify concepts). None of these terms used were unmatched, and the virtual patient replied accurately with the requested data.

CUI	Token	Count
C0020538	*hypertension*	2
C0005823	*tension*	1
C0795691	*problèmes cardiaques* ('heart problems')	1
C0030193	*mal* ('pain')	11
	douleurs	7
	douleur	1
C0010200	*toux* ('cough')	2
	toussez ('to cough')	1
C0850149	*toux sèche* ('dry cough')	1
C0042963,	*vomissement* ('vomiting')	1
C0042965	*vomissez* ('you vomit')	1

Table 6: Examples of successfully matched concepts

The scarcity of the data for evaluating our system sets limits for generalising results. Nonetheless, the annotation of domain entities has been fairly accurate, although some entity types need higher recall. Precision rates dropped mainly due to large lists of entities causing false positives.

The next stage of the evaluation will focus on usability and medicine professionals will be the users to test the system. Comparing this evaluation with the results here presented will be stimulating to understand term variation as well as the classes of entities used by different types of users.

6. Conclusion

We described the lexical and terminological resources used in a dialogue system simulating a virtual patient to train medical students. Results from a first evaluation have been reported, although our data are scarce and more evaluation tests are needed. Term ambiguity raises challenges we are still addressing in the project (e.g., *tension* can refer to either 'hypertension' or 'mental tension'). We would like to highlight that, through the use of comparable resources, most strategies presented here should be portable to other languages (e.g., Spanish or English).

7. Acknowledgments

BPI funded partly this work through the FUI Project PatientGenesys (F1310002-P).

8. References

Bodenreider, O. (2004). The Unified Medical Language System (UMLS): Integrating biomedical terminology. *Nucleic Acids Research*, 32(Database issue):D267–270.

Bouamor, D., Campillos-Llanos, L., Ligozat, A.-L., Rosset, S., and Zweigenbaum, P. (2016). Transfer-based learning-to-rank assessment of medical term technicality. In *LREC*.

[7]For example, the system did not tag and process specific questions related patient's hospitalisation, emergency care or medication compliance; however, these occured just once in the data.

Campillos-Llanos, L., Bouamor, D., Bilinsky, E., Ligozat, A.-L., Zweigenbaum, P., and Rosset, S. (2015). Description of the PatientGenesys dialogue system. In *Proc. of 16th SIGDIAL*, pages 438–440.

Cimino, J. J. (1998). Desiderata for controlled medical vocabularies in the twenty-first century. *Methods of Information in Medicine*, 37(4–5):394–403.

Courtois, B. (1990). Un système de dictionnaires électroniques pour les mots simples du français. *Langue française*, 87:11–22.

Hathout, N., Namer, F., and Dal, G. (2002). An experimental constructional database: the MorTAL project. *Many morphologies*, pages 178–209.

Kenny, P., Parsons, T. D., Gratch, J., and Rizzo, A. (2008). Evaluation of Justina: a virtual patient with PTSD. In *Intelligent virtual agents*, pages 394–408. Springer.

Makhoul, J., Kubala, F., Schwartz, R., and Weischedel, R. (1999). Performance measures for information extraction. In *Proceedings of DARPA Broadcast News Workshop*, pages 249–252.

McCray, A. T., Srinivasan, S., and Browne, A. C. (1994). Lexical methods for managing variation in biomedical terminologies. In *Proc. of Annual Symposium Computer Applic. Medical Care*, page 235. AMIA.

McCray, A. T., Dorfman, E., Ripple, A., Ide, N. C., Jha, M., Katz, D. G., Loane, R. F., and Tse, T. (2000). Usability issues in developing a web-based consumer health site. In *Proc AMIA Symp*, pages 556–560.

Namer, F. and Zweigenbaum, P. (2004). Acquiring meaning for french medical terminology: contribution of morphosemantics. *Eleventh MEDINFO International Conference*, pages 535–539.

Nirenburg, S., Beale, S., McShane, M., Jarrell, B., and Fantry, G. (2008). Language understanding in Maryland virtual patient. In *COLING 2008: Proceedings of the workshop on Speech Processing for Safety Critical Translation and Pervasive Applications*, pages 36–39, August.

Nirenburg, S., McShane, M., and Beale, S. (2009). A unified ontological-semantic substrate for physiological simulation and cognitive modeling. In *Proceedings of the first international conference on biomedical ontology (ICBO-2009)*, page 139.

Rosset, S., Galibert, O., Adda, G., and Bilinski, E. (2008). The LIMSI participation in the QAst track. In *Advances in Multilingual and Multimodal Information Retrieval*, pages 414–423. Springer-Verlag, Berlin, Heidelberg.

Zeng-Treitler, Q., Goryachev, S., Kim, H., Keselman, A., and Rosendale, D. (2007). Making texts in electronic health records comprehensible to consumers: a prototype translator. In *AMIA Annu Symp Proc*, pages 846–850.

Zweigenbaum, P., Baud, R. H., Burgun, A., Namer, F., Jarrousse, É., Grabar, N., Ruch, P., Le Duff, F., Forget, J.-F., Douyère, M., and Darmoni, S. (2005). A unified medical lexicon for French. *International Journal of Medical Informatics*, 74(2–4):119–124.