

Natural Language Processing for e-Health

Pierre Zweigenbaum

LIMSI, CNRS, Université Paris-Saclay

DataIA-JST International Symposium on Data Science and AI
Paris, 2018/7/11



université
PARIS-SACLAY

- 1 Introduction: LIMSI, NLP for e-Health
- 2 Information Normalization: Coding Death Certificates for Health Statistics
- 3 Information Extraction from Social Networks for Pharmacovigilance
- 4 Human-Machine Dialogue to Train Medical Students

- 1 Introduction: LIMSI, NLP for e-Health
 - LIMSI, CNRS, Université Paris-Saclay
 - Biomedical Text Conveys Information and Knowledge
 - Ongoing NLP for e-Health Research at LIMSI
- 2 Information Normalization: Coding Death Certificates for Health Statistics
 - Finding Diagnoses in Text
- 3 Information Extraction from Social Networks for Pharmacovigilance
 - Finding Drugs, Signs and Symptoms in Text
- 4 Human-Machine Dialogue to Train Medical Students
 - A Virtual Patient to Train Medical Students

LIMSI: An Interdisciplinary Laboratory for Mechanics and Information Sciences

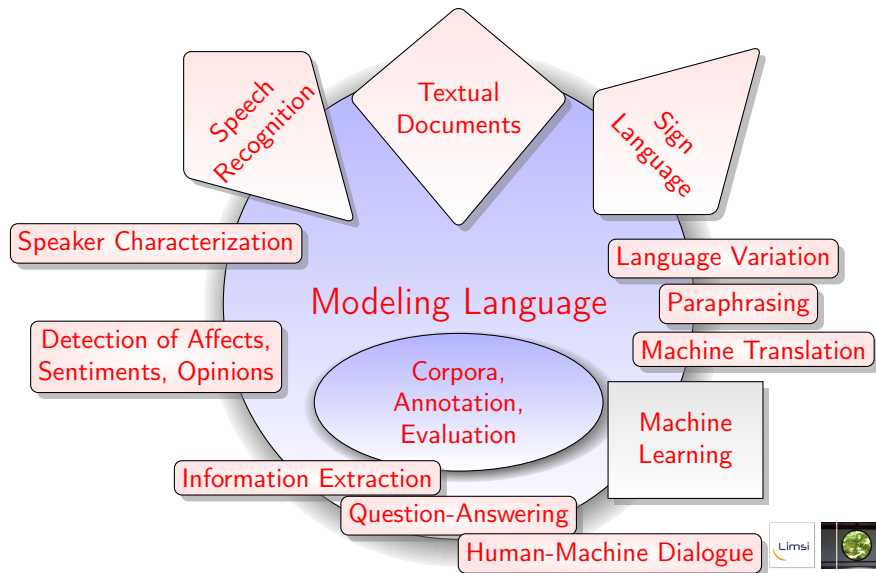
An Intramural **CNRS** Research Unit on the **Université Paris-Saclay** Campus

LIMSI's new building (currently moving in)



Human Language Processing at LIMSI

About 80 people including 30 permanent researchers



Biomedical texts convey information and knowledge

- **Natural Language Processing** of these texts unlocks that information and knowledge for further computer processing

General Bibliography

- Névéol A, Zweigenbaum P. Making sense of big textual data for health care: Findings from the section on clinical natural language processing. *Yearb Med Inform*, 26(01):228-233, 2017.
- Demner-Fushman D, Elhadad N. Aspiring to unintended consequences of natural language processing: A review of recent developments in clinical and consumer-generated text processing. *Yearb Med Inform*. 2016 Nov 10;(1):224-233.
- Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. *Yearb Med Inform*. 2008:128-44.
- Huang CC, Lu Z. Community challenges in biomedical text mining over 10 years: success, failure and the future. *Brief Bioinform*. 2016 Jan;17(1):132-44.
- Zweigenbaum P, Demner-Fushman D, Yu H, Cohen KB. Frontiers of biomedical text mining: current progress. *Brief Bioinform*. 2007 Sep;8(5):358-75. Epub 2007 Oct 30.



Information extraction from clinical text

- ANR Cabernet — Building French resources for information extraction from medical text [Campillos et al., 2017](#)
- ANSM Vigi4med, PHARES — Information extraction, linking to specialized entities from lay language [Morlane-Hondère et al., 2016](#)
- CépiDc collaboration — ICD-10 coding with a large number of classes [Lavergne et al., 2016](#)
- De-identification — MEDINA software [Grouin & Névéol, 2014](#)
- DigiCosme PhD, with CEA — Temporal information extraction and event coreference [Tourille et al., 2017](#)

Information extraction from the scientific literature

- Paris-Saclay IDI PhD, with INRA — Hybrid embedding of words and concepts [Ferré et al., 2017](#)
- H2020 ITN MiRoR Joint PhDs with Amsterdam U — Spin detection; Text classification for systematic reviews [Koroleva & Paroubek, 2017](#); [Norman et al., 2018](#)
- ANR MIAM — Food and drug interaction [Hamon et al., 2017](#)



Question-answering and dialogue

- French-German ANR-DFG GoAsq (with LRI & TU Dresden) — Hybrid question-answering with text and ontologies **Kamath et al., 2018**
- FUI Patient Genesys, SATT PVdial — **Human-machine dialogue** with a very large concept and vocabulary space **Campillos et al., 2016**
- HEGP/Necker PhD — QA from clinical datawarehouse **A. Neuraz, A. Burgun, S. Rosset**

*Activities listed in red are further presented in the present talk.

Text simplification

- ANR CLEAR — Lay vocabulary and expert vocabulary Grabar & Hamon, 2016

Multilingual access

- Cochrane collaboration, with LIMSI/TLP — Machine translation of scientific abstracts, with a need for precision Ive et al., 2016
- Joint PhD with Sfax — Term detection, transfert to Arabic Neifar et al., 2016
- AP-HP/Inserm PhD — Multilingual terminology for interoperability Paris et al., 2018

- 1 Introduction: LIMSI, NLP for e-Health
 - LIMSI, CNRS, Université Paris-Saclay
 - Biomedical Text Conveys Information and Knowledge
 - Ongoing NLP for e-Health Research at LIMSI
- 2 Information Normalization: Coding Death Certificates for Health Statistics
 - Finding Diagnoses in Text
- 3 Information Extraction from Social Networks for Pharmacovigilance
 - Finding Drugs, Signs and Symptoms in Text
- 4 Human-Machine Dialogue to Train Medical Students
 - A Virtual Patient to Train Medical Students

Coding Death Certificates for Health Statistics

Text classification: From the text of a certificate to one or more labels (codes)

Needed in about every country for public health statistics

SEE INSTRUCTIONS ON OTHER SIDE

CAUSE OF DEATH

27. PART I. Enter the diseases, injuries, or complications that caused the death. Do not enter the mode of dying, such as cardiac or respiratory arrest, shock, or heart failure. List only one cause on each line.				Approximate Interval Between Onset and Death	
IMMEDIATE CAUSE (Final disease or condition resulting in death) →	a.	Rupture of myocardium DUE TO (OR AS A CONSEQUENCE OF):		Mins.	
	b.	Acute myocardial infarction DUE TO (OR AS A CONSEQUENCE OF):		6 days	
	c.	Chronic ischemic heart disease DUE TO (OR AS A CONSEQUENCE OF):		5 years	
	d.				
Sequentially list conditions, if any, leading to immediate cause. Enter UNDERLYING CAUSE (Disease or injury that initiated events resulting in death) LAST	PART II. Other significant conditions contributing to death but not resulting in the underlying cause given in Part I. Diabetes, Chronic obstructive pulmonary disease, smoking			28a. WAS AN AUTOPSY PERFORMED? (Yes or no) Yes	28b. WERE AUTOPSY FINDINGS AVAILABLE PRIOR TO COMPLETION OF CAUSE OF DEATH? (Yes or no) Yes
29. MANNER OF DEATH <input checked="" type="checkbox"/> Natural <input type="checkbox"/> Pending Investigation <input type="checkbox"/> Accident <input type="checkbox"/> Suicide <input type="checkbox"/> Could not be Determined <input type="checkbox"/> Homicide	30a. DATE OF INJURY (Month, Day, Year)	30b. TIME OF INJURY M	30c. INJURY AT WORK? (Yes or no)	30d. DESCRIBE HOW INJURY OCCURRED	
	30e. PLACE OF INJURY – At home, farm, street, factory, office building, etc. (Specify)		30f. LOCATION (Street and Number or Rural Route Number, City or Town, State)		

I23.3 Rupture of cardiac wall without haemopericardium as current complication following acute myocardial infarction

(急性心筋梗塞合併症心膜血腫伴心(壁)破裂)

I21.9 Acute myocardial infarction, unspecified (急性心筋梗塞，詳細不明)

I25.9 Chronic ischaemic heart disease, unspecified (慢性血性心疾患，詳細不明)



NLP Challenge: Variation and Ambiguity

Study of *pneumectomie* in death certificates

= removing lung

pneumectomie
pneumectomie
poumon unique
lobectomie pulmonaire
résection pulmonaire
pneumectomisé
PNEUMECTOMISEE

cancer du **poumon** traité chimio
et **lobectomie**

carcinome **bronchique** lobaire
supérieur gauche **lobectomisé**
adénocarcinome **bronchique**
(récidive métastatique
controlatérale) avec **lobectomie**
primitive

⇒ ICD Z90.2:
Acquired absence of lung [part of]

PNEUMECTONIE DROITE

BPCO SEVERE /
LOBECTOMIE MOYENNE
ET SUPERIEURE DROITE
POUR ADENOCARCINOME
lobectomie supérieure gauche



NLP Challenge: Variation

Terms for *myocardial infarction* in the UMLS Metathesaurus

Terme	Libellé (VARIANTE)
L0027051	<i>Myocardial Infarction</i> (PF)
L0027051	MYOCARDIAL INFARCTION (VC)
L0027051	<i>Myocardial infarction</i> (VC)
L0027051	<i>myocardial infarction</i> (VC)
L0027051	<i>Myocardial Infarctions</i> (VP)
L0027051	<i>Infarctions (Myocardial)</i> (VWP)
L0027051	<i>Infarctions, Myocardial</i> (VWP)
L0027051	<i>Infarction, myocardial</i> (VCW)
L0027051	<i>Infarction;myocardial</i> (VCW)
L0027051	<i>Infarction, Myocardial</i> (VW)
L0027051	<i>Myocardial infarction, NOS</i> (VO)
L0284112	<i>Heart attack, NOS</i> (PF)
L0284112	HEART ATTACK (VO)
L0284112	<i>Heart Attacks</i> (VO)
L0284112	<i>Heart attack</i> (VO)
L0284112	<i>heart attack</i> (VO)
L0284112	ATTACK HEART (NOS) (VCW)
L0284112	<i>Attack heart (NOS)</i> (VCW)
L0306107	<i>Infarction of heart, NOS</i> (PF)
L0306107	HEART INFARCTION (VO)
L0306107	<i>Infarction, heart</i> (VO)
L0306107	<i>Infarction;heart</i> (VO)
L0379717	<i>Cardiac infarction, NOS</i> (PF)

Terme	Libellé (VARIANTE)
L0308108	<i>Myocardial Infarct</i> (PF)
L0308108	MYOCARDIAL INFARCT (VC)
L0308108	<i>Myocardial infarct</i> (VC)
L0308108	<i>Myocardial Infarcts</i> (VP)
L0308108	<i>Infarct, Myocardial</i> (VW)
L0308108	<i>Infarcts, Myocardial</i> (VWP)
L0308108	INFARCT MYOCARDIAL (VCW)
L0308108	<i>Infarct myocardial</i> (VCW)
L0873038	<i>Myocardial necrosis</i> (PF)
L1007490	<i>Attack coronary</i> (PF)
L1007490	ATTACK CORONARY (VC)
L1022045	<i>Myocardial infarction syndrome</i> (PF)
L1024662	<i>Necrosis myocardium</i> (PF)
L1024662	NECROSIS MYOCARDIUM (VC)
L1024664	<i>Myocardial necrosis syndrome</i> (PF)
L1088741	CT - Coronary thrombosis (PF)
L0586860	MI - Myocardial infarction (PF)
L0026809	<i>mi <1></i> (PF)
L0026809	MI (VO)
L0026809	<i>mi</i> (VO)
L1217656	AMI <3> (PF)
L1217656	AMI (VO)
L1374411	<i>Cardiopathy necrotic</i> (PF)

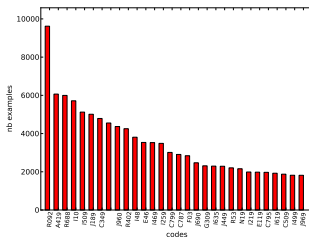
NLP Challenge: Contextual Coding

- The same input may express two different overall phenomena

Certificate (FR)	<i>choc hémorragique</i>	R571
Certificate (FR)	<i>hémorragie digestive</i>	K922
Certificate (FR)	<i>Cirrhose hépatique alcoolique</i>	K703
English gloss	hemorrhagic shock	R571
English gloss	gastrointestinal bleeding	K922
English gloss	alcoholic liver cirrhosis	K703
Certificate (FR)	<i>choc hémorragique</i>	T794
Certificate (FR)	<i>polytraumatisme</i>	T07
Certificate (FR)	<i>accident de la voie publique</i>	V892
English gloss	hemorrhagic shock	T794
English gloss	polytrauma	T07
English gloss	traffic accident	V892

- R571 *Hypovolaemic shock* < R57 *Shock, not elsewhere classified*
- T79.4 *Traumatic shock* < T79 *Certain early complications of trauma, not elsewhere classified*

- Large number of classes + Zipf's law
 - >3,000 disease codes in dataset



→ use both supervised learning (frequent) and dictionary (rare)



First Normalize Text as Much as Possible

First transform input string to reduce variation

- Downcase characters
 - AVC hémorragique → avc hémorragique
- * Split line into 'words' (*tokens*) (avc, hémorragique)
- Remove 'stop words' (of the and...)
- Remove diacritics
 - hémorragique → hemorrhagique
 - Behçet sévère → Behcet severe
- Fix some spelling errors based on training corpus and dictionaries
 - syndrome:4526 ← sydrome:15
 - hemorragique:2573 ← hemorragqiue:3
- Stem each token
 - hemorrhagie, digestive → hemorrhag, digest
 - avc, hemorrhagique → avc, hemorrhag

Then Extract Text Features for Supervised Classifier

Features computed from the normalized text of a death certificate

avc hemorrag

Encoding a sequence of words as a set of...

- Word N-grams

- **u**: unigrams: 'avc', 'hemorrag'
- **b**: bigrams: 'avc hemorrag'

longer n-grams are more precise

- Character N-grams

- bigrams: _a, av, vc, c_, _h, he, em, mo, or, rr, ra, ag, g_
- trigrams: _av, avc, vc_, _he, hem, emo, ... rra, rag, ag_

character n-grams are more robust

Dictionary Text Features

Take advantage of a priori knowledge

[colon perforation leading to stercoral peritonitis]

colon perfor lead stercor peritoniti

Use dictionary to generate features for classifier

- Dictionary entries (pre-processed with same normalization)
 - colon perfor/K631 (colonic perforation/K631)
 - peritoniti/K659 (peritonitis/K659)
- Dictionary lookup
 - Features: p:colon perfor, c:K631, p:peritoniti, c:K659

Also Extract Features from Metadata

Features obtained from metadata provided with a death certificate

- **Year** of coding (e.g., 2011)
- **Age** of the deceased (e.g., 85)

Encoding a point in an ordered set as a set of intervals

- **y: Coding year:** coding rules change over the years
 - 2×9 features instantiated for $y \in [2006 \dots 2014]$
 - $2011 \rightarrow >2006, >2007, \dots >2010, \leq 2011, \dots \leq 2014$
- **a: Patient's age**
 - 2×11 features instantiated for $a \in [0 \dots 100]$
 - $85 \rightarrow >0, >5, >10, >15, >20, \dots >80, \leq 85, \leq 90, \leq 95, \leq 100$

Top results on CLEF eHealth 2017 data

- US death certificates
 - P=90, R=81, F=85
- French death certificates
 - P=89, R=77, F=83
- French death certificates, line-aligned
 - P=87, R=87, F=87

Collaboration with users for take-up

- **CepiDc**: Epidemiology center on medical causes of death (WHO coding center)
 - Insert into coding workflow
- **Santé Publique France**: National public health agency
 - Detect “syndromic groups” for yearly outbreak surveillance and public health monitoring



Coding Causes of Death


Acknowledgements


- Classification: with Thomas Lavergne (LIMSI, Univ. Paris-Sud)
- CLEF eHealth 2016–2018 Shared Tasks with
 - Cyril Grouin, Thomas Lavergne, Aurélie Névéol (LIMSI)
 - Aude Robert, Grégoire Rey (INSERM, French CépiDc data)
 - Robert Anderson (US CDC data)
 - Francesco Grippo and Chiara Orsi (Italian data)
 - László Pelikán (Hungarian data)

- 1 Introduction: LIMSI, NLP for e-Health
 - LIMSI, CNRS, Université Paris-Saclay
 - Biomedical Text Conveys Information and Knowledge
 - Ongoing NLP for e-Health Research at LIMSI
- 2 Information Normalization: Coding Death Certificates for Health Statistics
 - Finding Diagnoses in Text
- 3 Information Extraction from Social Networks for Pharmacovigilance
 - Finding Drugs, Signs and Symptoms in Text
- 4 Human-Machine Dialogue to Train Medical Students
 - A Virtual Patient to Train Medical Students

Social Networks

New sources of information on health

 **cabergoline side effects** [Watch this](#)



pharmgirl24

[View More Posts](#)




[Ignore](#)

Maybe there's already a [discussion](#) out there for this, but I thought I would start a new one to see if we could compile some specifics. I just started [cabergoline](#) (3 doses down already) and wanted to hear about experiences people have had on it. I know I was terrified (partly because I know too much already & partly from what I've read online) and really wanted to know what to expect. What sort of side effects have you had? How long before they go away? Do certain things never go away? What do you do or take to combat those side effects?

So far I've done pretty well. First dose gave me horrible headache & brain fog for a day. After that I've really just had mild dizzy spells, headaches off and on (never had a headache from tumor) & insomnia. I'll fall asleep fine, then toss & turn a few hours later & wake up several hours before I usually do. Then I'm wide awake (at 4am). Then of course I'm tired all day. Not much nausea but I find that I don't want to eat much & get full right away. Almost that "take it or [leave it](#)" kind of attitude. I love that part! Definitely would love to lose 10lbs!

I don't want to take anything for the insomnia because I don't want to mess with too many brain chemicals all at once. I'm hoping it will eventually go away. Nothing really seems to help the headaches either but I'm thinking that will go away too?

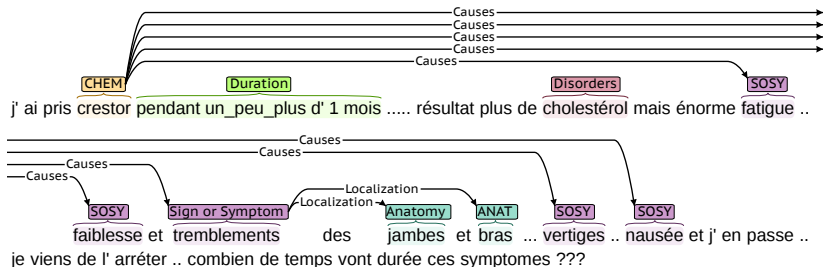
Posted on 02/21/11, 07:46 am

 Hug pharm...  [Send to a Friend](#)  [Share This](#)



Objectives and Methods

Spotting adverse drug event signals: pharmacovigilance, needed for drug safety



NLP challenges: non-standard vocabulary, variability, ambiguity

Methods: supervised information extraction; collect non-standard variants

Robust Construction of Semantic Classes

Collecting standard and non-standard term variants

- Large terminology resources: e.g., drug names
 - UMLS semantic class
- Bootstrapping: e.g., drug-taking verbs
 - I **take** **aspirin** every day
 - prendre, prescrire, être sous, donner, avoir. . .
 - Drug names absent from our lists
 - e.g., food complements, Canadian drug names
 - Intentional variants
 - Lévothyrox → lévo
 - Duphaston → dudu
 - pilule du lendemain → pdl
 - Spelling errors
- Unsupervised distributed representations, word clusters
 - gynécologue, gynécolgue, gynécologue, gynéco, gynécho, génico, gyné, gygy, etc.



Outcome: Baclofen study (Lillo-Le Louet et al., ISOP 2017)

Baclofen safety and its use in social media: a preliminary pharmacovigilance study

- 61,256,905 posts processed
- 18,570 posts mentioning **baclofen** and a **clinical finding** retrieved from NLP output database
 - Exclude alcohol dependence → 2621 posts
- Reviewed by two evaluators from pharmacovigilance center
 - 782 (29.2%) mentioned potential adverse drug reactions
- Compared to cases in French National Database
 - 717 ADR cases with baclofen
- Main findings
 - Information about baclofen dosage **more frequent in posts** (27.7% vs 9.2%)
 - 75% of the patients had a daily **dosage over 90 mg, higher** than the maximum recommended daily dosage
 - Demographic information (**gender, age**) **scarce** in posts



Pharmacovigilance from Social Media

Acknowledgements

Cyril Grouin, François Morlane-Hondère (LIMSI)

Funding: ANSM: National Agency for Drug Safety

Vigi4med (2014–2017)

- PI: Dr Cédric Bousquet
 - St-Étienne Medical School
 - LIMICS, Inserm, Paris
- 2 pharmacovigilance centers
- Detect adverse drug reactions
- Proof of concept, batch processing

PHARES (2017–2019)

- PI: Dr Marie-Christine Jaulent
 - LIMICS, Inserm, Paris
- 1 pharmacovigilance center
- Strengthen causal link detection
- Explore drug misuse
- On-line processing

LIMSI: NLP



- 1 Introduction: LIMSI, NLP for e-Health
 - LIMSI, CNRS, Université Paris-Saclay
 - Biomedical Text Conveys Information and Knowledge
 - Ongoing NLP for e-Health Research at LIMSI
- 2 Information Normalization: Coding Death Certificates for Health Statistics
 - Finding Diagnoses in Text
- 3 Information Extraction from Social Networks for Pharmacovigilance
 - Finding Drugs, Signs and Symptoms in Text
- 4 Human-Machine Dialogue to Train Medical Students
 - A Virtual Patient to Train Medical Students

A Virtual Patient to Train Medical Students

Je viens consulter pour une fatigue importante.
Je suis épuisé

depuis quand ?

Je suis très fatigué depuis 7 mois

avez-vous d'autres symptômes ?

Plusieurs problèmes, j'ai des insomnies, je suis anxieux et j'ai mal au dos

dites m'en plus

Maintenant, j'ai perdu du poids et je n'ai plus de motivation

avez-vous été opéré ?

J'ai eu une opération de l'appendicite

quand ?

J'ai eu une opération de l'appendicite a 10 ans.

cela s'est bien passé ?

Non, il n'y a eu rien de particulier.

Prenez-vous un traitement ?



A Virtual Patient to Train Medical Students



D: What brings you here?

P: I am seeing you about persistent backache

D: When did it start?

P: I don't remember

D: Where is the pain exactly?

P: I have pain in the lower back, the pain is located in the lower back and sometimes down the legs

D: Can you sleep well?

P: The pain often makes it difficult to fall asleep

D: Do you suffer from any disease?

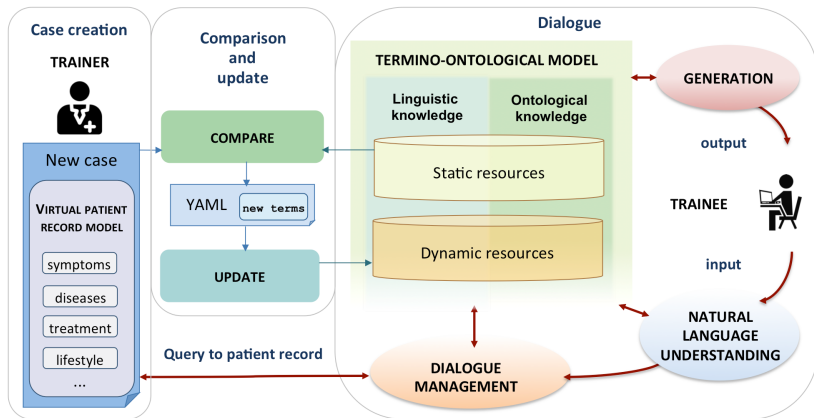
P: Chronic stomach ulcers

D: Do you follow any treatment for that?

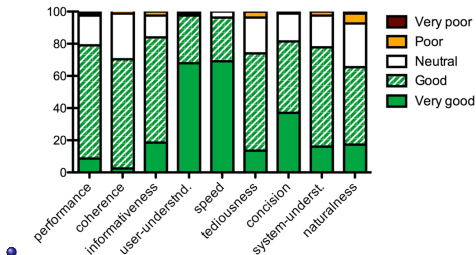
P: Omeprazole for the chronic stomach ulcers. Maybe there is more information in my file

- Match user vocabulary to patient case vocabulary
- Large concept and vocabulary space
- Need for precision
- Includes question-answering
- Dialogue management, instantaneous answer

System Architecture



- Functional system
 - Good user evaluations



- Robust across medical specialties
- Robust on unseen patient cases
- Ready for transfer to company



A Virtual Patient to Train Medical Students

Acknowledgements

- Sophie Rosset, Éric Bilinski
- Leonardo Campillos Llanos, Catherine Thomas, Dhouha Bouamor
- LIMSI, CNRS
- FUI / BPI ; SATT Paris-Saclay
- Interaction Healthcare / SimForHealth, CHU d'Angers, Voxygen, Vidal

- 1 Introduction: LIMSI, NLP for e-Health
 - LIMSI, CNRS, Université Paris-Saclay
 - Biomedical Text Conveys Information and Knowledge
 - Ongoing NLP for e-Health Research at LIMSI
- 2 Information Normalization: Coding Death Certificates for Health Statistics
 - Finding Diagnoses in Text
- 3 Information Extraction from Social Networks for Pharmacovigilance
 - Finding Drugs, Signs and Symptoms in Text
- 4 Human-Machine Dialogue to Train Medical Students
 - A Virtual Patient to Train Medical Students

- Multiple applications of NLP to biomedical text processing
- Key role of entity normalization / entity linking
- Relative importance of human and data-driven knowledge
- Remaining challenges
 - Variability of data and **generalizability**
 - Need for **annotations** for supervised learning
 - **Access to clinical text** to train systems

Thank You!

Welcome to MEDINFO 2019
International Conference on Medical Informatics
Lyon, France, August 2019



- Campillos, L., L. Deléger, C. Grouin, T. Hamon, A.-L. Ligozat, and A. Névéol. "A French clinical corpus with comprehensive semantic annotations: development of the Medical Entity and Relation LIMS I annotated Text corpus (MERLOT)". In: Language Resources and Evaluation (2017), pp. 1–31.
- Campillos, L., D. Bouamor, P. Zweigenbaum, and S. Rosset. "Managing Linguistic and Terminological Variation in a Medical Dialogue System". In: International Conference on Language Resources and Evaluation. Portorož SI, 2016, pp. 3167–3173.
- Ferré, A., P. Zweigenbaum, and C. Nédellec. "Representation of complex terms in a vector space structured by an ontology for a normalization task". In: BioNLP workshop, ACL. Vancouver CA, 2017, pp. 99–106.
- Grabar, N. and T. Hamon. "Exploitation de la morphologie pour l'extraction automatique de paraphrases grand public des termes médicaux". In: TAL (Traitement Automatique des Langues) 57.1 (2016), pp. 85–109.
- Grouin, C. and A. Névéol. "De-identification of clinical notes in French: towards a protocol for reference corpus development". In: Journal of Biomedical Informatics 50 (2014), pp. 151–161.
- Hamon, T., T. Vincent, F. Mouglin, N. Grabar, and F. Thiessard. "POMELO: Medline corpus with manually annotated food-drug interactions". In: Proceedings of the Biomedical NLP Workshop, Recent Advances in Natural Language Processing (RANLP). Varna BG, 2017, pp. 73–80.
- Ive, J., A. Max, F. Yvon, and P. Ravaud. "Diagnosing High-Quality Statistical Machine Translation Using Traces of Post-Edit Operations". In: International Conference on Language Resources and Evaluation -Workshop on Translation Evaluation: From Fragmented Tools and Data Sets to an Integrated Ecosystem. Portorož SI, 2016, 8p.
- Kamath, S., B. Grau, and Y. Ma. "Verification of the expected answer type for biomedical question answering". In: First International Workshop on Hybrid Question Answering with Structured and Unstructured Knowledge, The Web Conference. Lyon FR, 2018, pp. 1093–1097.
- Koroleva, A. and P. Paroubek. "On the contribution of specific entity detection and comparative construction to automatic spin detection in biomedical scientific publications". In: Workshop on Processing Emotions, Decisions and Opinions. Poznan PL, 2017, 5p.
- Lavergne, T., A. Névéol, A. Robert, C. Grouin, G. Rey, and P. Zweigenbaum. "A dataset for ICD-10 coding of death certificates: Creation and usage". In: Proceedings of the Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM2016, COLING). Osaka, JP, 2016, pp. 60–69.
- Morlane-Hondère, F., C. Grouin, and P. Zweigenbaum. "Identification of Drug-Related Medical Conditions in Social Media". In: 10th International Conference on Language Resources and Evaluation. Portorož SI, 2016, pp. 2022–2028.
- Neifar, W., T. Hamon, P. Zweigenbaum, M. Ellouze Khemakhem, and L. Hadrich Belguith. "Adaptation of a Term Extractor to Arabic Specialised Texts: First Experiments and Limits". In: International Conference on Intelligent Text Processing and Computational Linguistics. Konya TR, 2016, 12p.
- Norman, C., M. Leeftang, P. Zweigenbaum, and A. Névéol. "Automating Document Discovery in the Systematic Review Process: How to Use Chaff to Extract Wheat". In: International Conference on Language Resources and Evaluation. Miyazaki JP, 2018, 7p.
- Paris, N., M. Mendis, C. Daniel, S. Murphy, X. Tannier, and P. Zweigenbaum. "i2b2 implemented over SMART-on-FHIR". In: AMIA Jt Summits Transl Sci Proc. San Francisco US, 2018, pp. 369–378.
- Tourille, J., O. Ferret, X. Tannier, and A. Névéol. "Temporal information extraction from clinical text". In: Conference of the European Chapter of the Association for Computational Linguistics. Valencia ES, 2017, pp. 739–745.

