

# A Spoken Drug Prescription Dataset in French for Spoken Language Understanding

Ali Can Kocabiyikoglu<sup>1,3</sup>, François Portet<sup>1</sup>, Prudence Gibert<sup>2</sup>, Hervé Blanchon<sup>1</sup>, Jean-Marc Babouchkine<sup>3</sup>, Gaëtan Gavazzi<sup>4</sup>

1. Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France, {ali-can.kocabiyikoglu, francois.portet, herve.blanchon}@univ-grenoble-alpes.fr
2. CHU Grenoble Alpes, Avenue Maquis-du-Grésivaudan, 38700 La Tronche, France, pgibert@chu-grenoble.fr
3. Calystene SA, 16 Rue Irène Joliot Curie, 38320 Eybens, France, jm.babouchkine@calystene.com
4. Clinique de médecine gériatrique, CHU Grenoble Alpes, Equipe Grépi, EA 7408, CS 10217, 38700, La Tronche, France, ggavazzi@chu-grenoble.fr



Our dataset is available at: <https://doi.org/10.5281/zenodo.6482586>

## 1. Introduction

We present PxSLU corpus a drug prescription dataset comprising around 4h of transcribed and annotated dialogues of drug prescriptions in French acquired through an experiment with 55 participants experts and non-experts in prescriptions.

- Corpus acquired from human-machine interactions using a prototype of a goal-oriented dialogue system
- Publicly available through a Attribution 4.0 International (CC BY-4.0) license.

## 2. Data Collection Protocol

We have established a simple protocol, inviting the participants to follow the following steps:

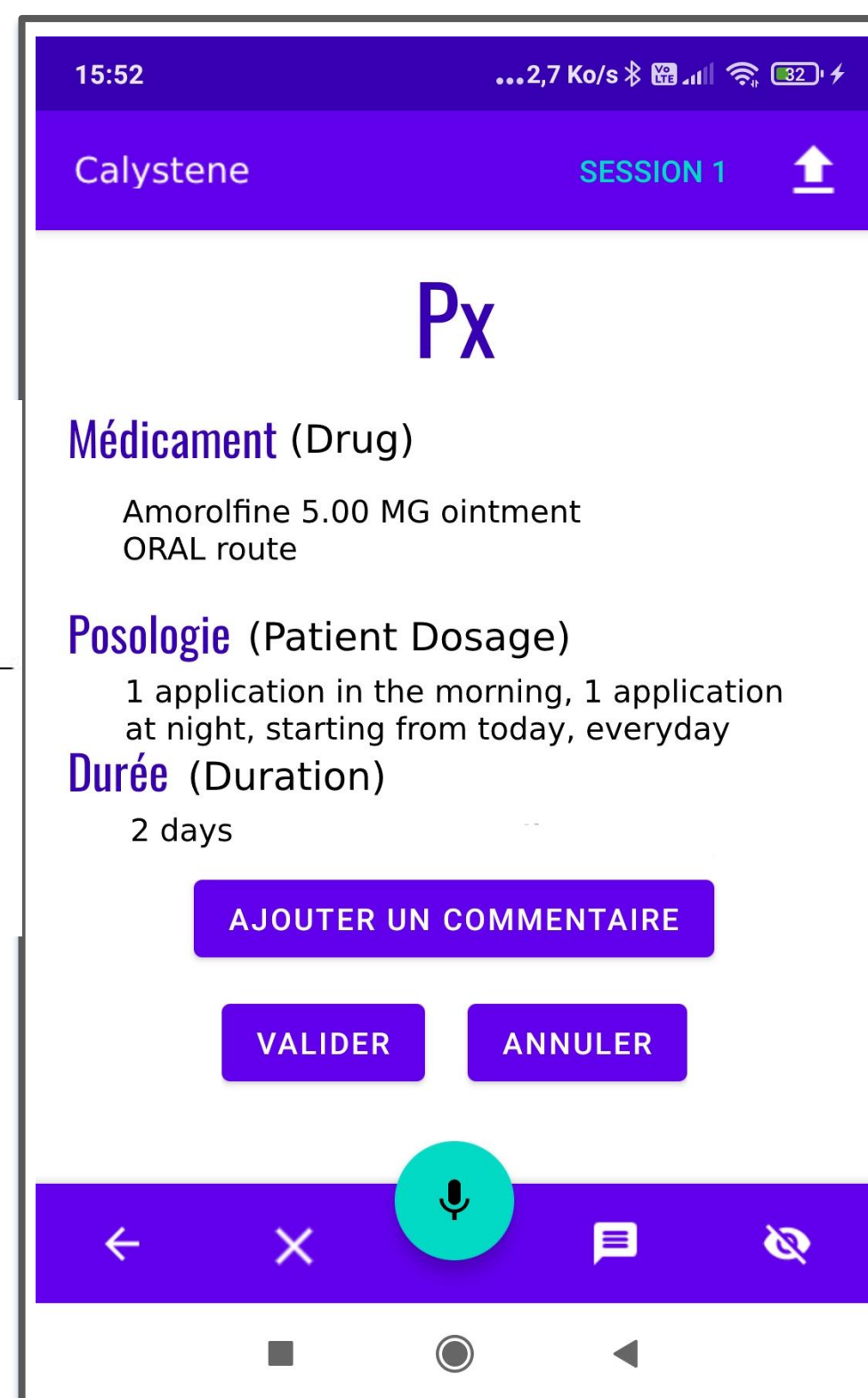
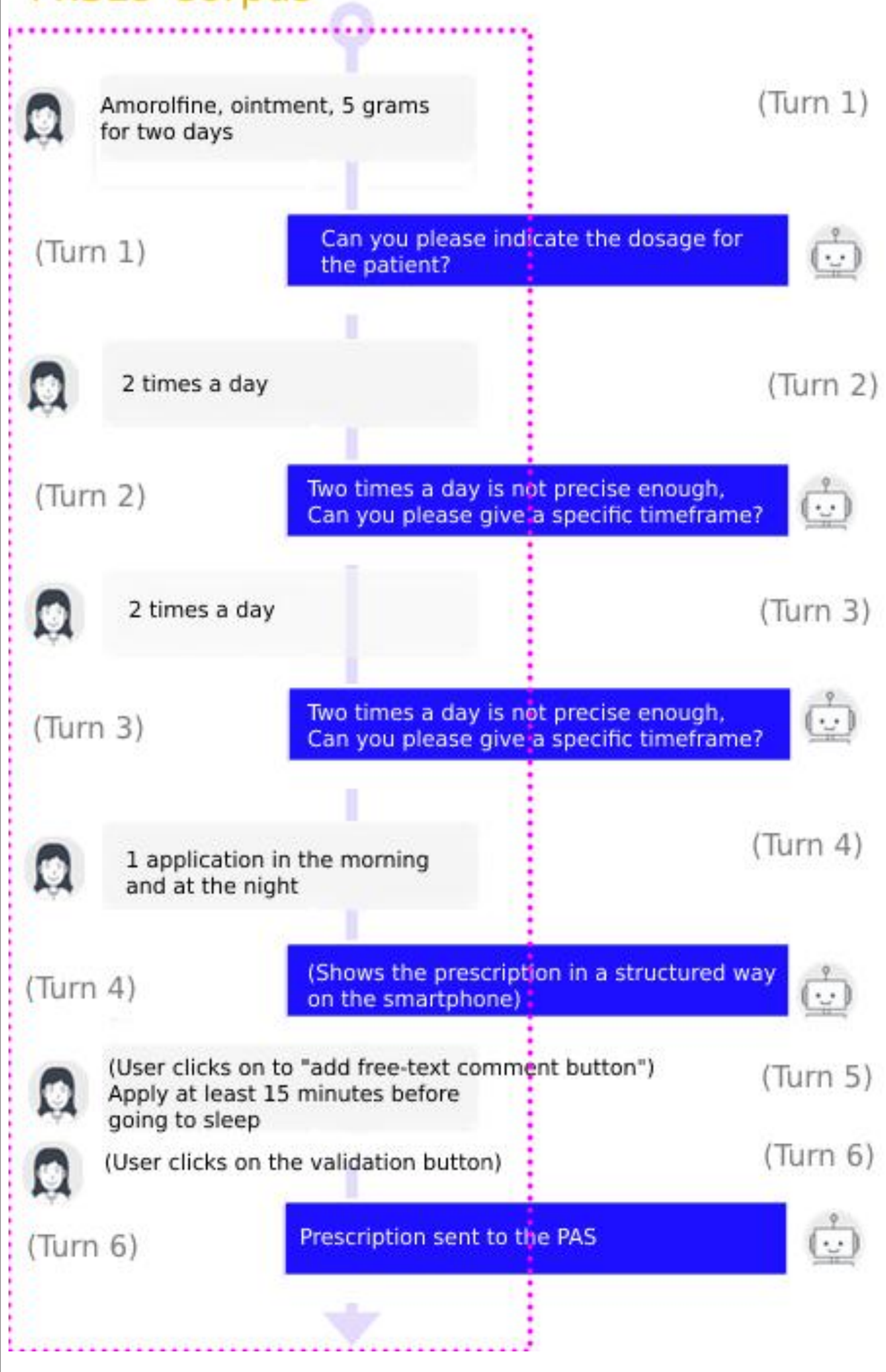
1. Registration on the form for requesting to participate in the experiment
2. Reception of the mobile application and follow the document explaining the installation and the course of the experiment
3. Reception of prescription examples depending on the audience
4. Filling the metadata survey, agree to the terms of use, and complete the experiment

## 3. Corpus Overview

The experiment was performed between January 2021 and October 2021. The data represents 262 minutes of recordings when all the participants are included.

	Sessions	Recordings	Time (m)
<b>Medical experts</b>	258	434	94.83
<b>Doctors</b>	230	570	105.21
<b>Non experts</b>	415	977	62.13
<b>Total</b>	903	1981	262.27

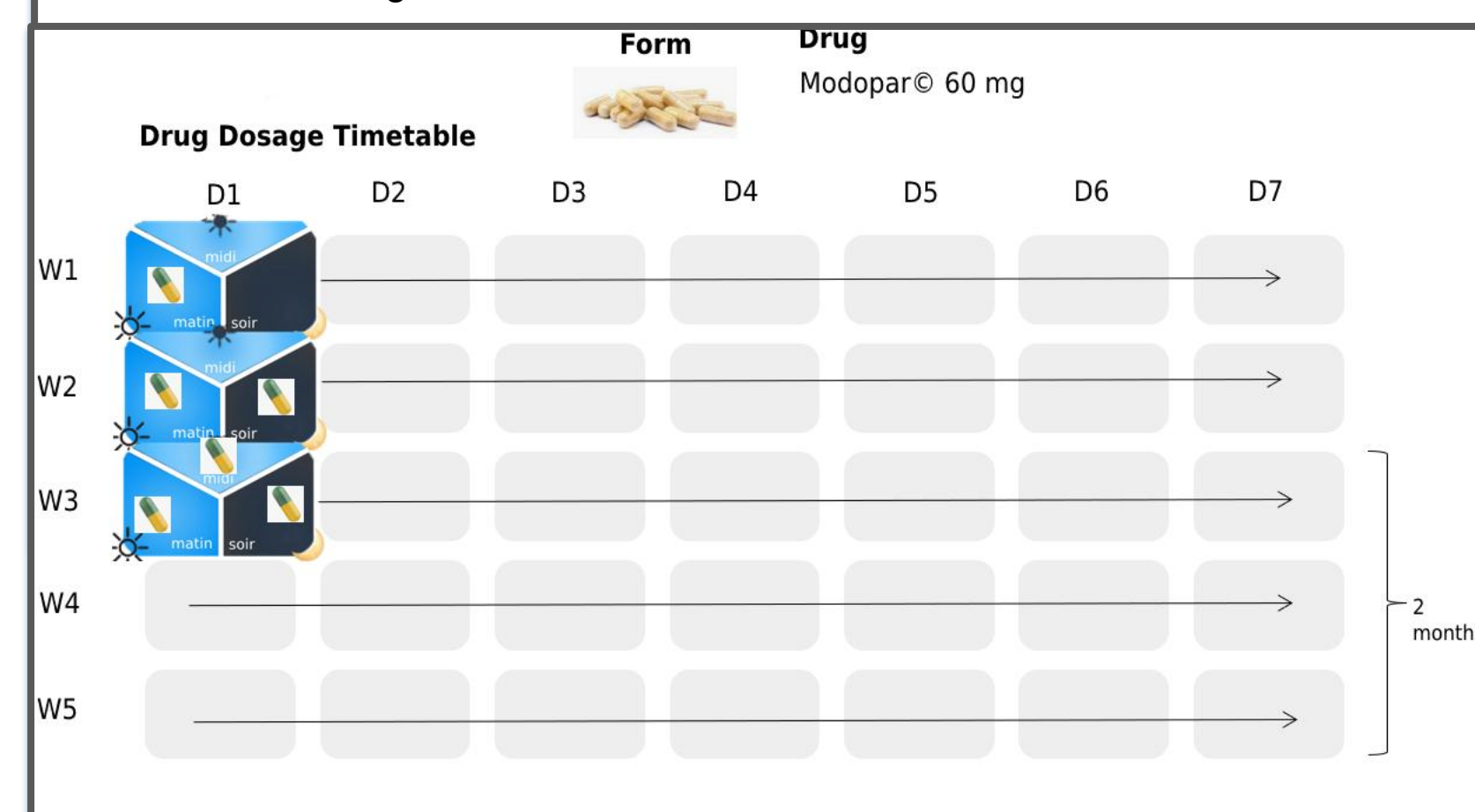
## PxSLU Corpus



## 4. Data Preparation

We targeted two type of participants:

- Medical practitioners (doctors, pharmacists, biomedical engineers, etc.)
  - 10 pictographs
  - 10 reading exercises
- Non-experts
  - 20 reading exercises

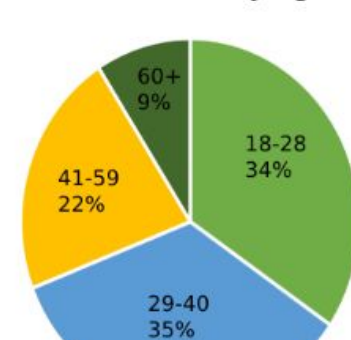


## 5. Distribution of participants

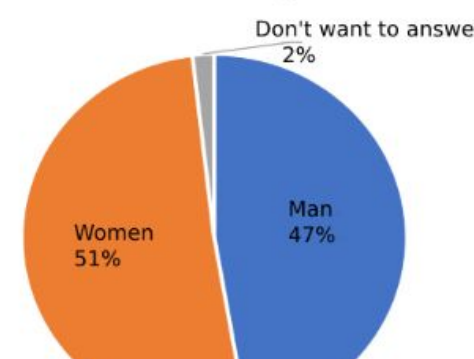
Distribution by category



Distribution by age



Distribution by sex



## 6. NLU Model Evaluation

Our model evaluation builds on previous work (Kocabiyikoglu et al., 2019) where we have presented initial NLU systems trained on artificial and textbook data:

- Baseline CRF model
- Triangular CRF extension (Jeong et Lee, 2008)
- Bi-RNN with attention (Liu et Lane, 2016)
- Flaubert pre-trained transformer LM (Le et al., 2019)

Model	Intent (acc)	Micro Avg			Macro Avg		
		P	R	F1	P	R	F1
CRF	92%	0.81	0.80	0.80	0.60	0.57	0.56
Tri-CRF	91%	0.83	0.82	0.82	0.64	0.57	0.59
Att-RNN	93%	0.83	0.87	0.85	0.55	0.55	0.53
Flaubert	94%	0.89	0.91	0.90	0.69	0.74	0.70

- Flaubert model gives the best results on both micro and macro averages
- Other models have comparable performance

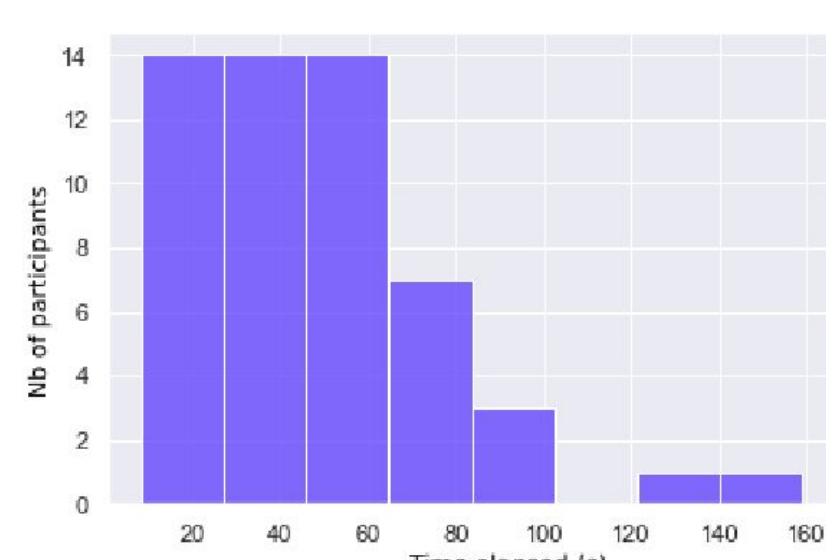
## 7. K-Fold (K=5) Cross Validation on PxSLU

Fine-tuning Flaubert model using PxSLU (flaubert-base-cased)

K #	Micro Average			Macro Average		
	P	R	F1	P	R	F1
1	0.93	0.93	0.93	0.79	0.75	0.75
2	0.93	0.94	0.94	0.76	0.80	0.77
3	0.89	0.90	0.90	0.54	0.48	0.50
4	0.92	0.91	0.91	0.68	0.67	0.66
5	0.95	0.95	0.95	0.73	0.74	0.72
avg	0.92	0.92	0.92	0.70	0.68	0.68
SD	0.02	0.02	0.02	0.10	0.12	0.10

- Similar performances with model trained with artificial+textbook data
- Data choice impacts the macro performance and the coverage of slots

## 8. Collected Dialogue Analysis



- Histogram of average time of session by number of participants
  - Most of the participants completed the task in less than 1 minutes
  - Medical practitioners on average less than 40 seconds
  - Doctors have spent more time on dialogue sessions (avg 40-100 seconds)

## References

Jeong, M., & Lee, G. G. (2008). Triangular-chain conditional random fields. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(7), 1287-1302.

Le, H., Vial, L., Frej, J., Segonne, V., Coavoux, M., Lecouteux, B., ... & Schwab, D. (2019). Flaubert: Unsupervised language model pre-training for french. *arXiv preprint arXiv:1912.05372*.

Liu, B., & Lane, I. (2016). Attention-based recurrent neural network models for joint intent detection and slot filling. *arXiv preprint arXiv:1609.01454*.

Kocabiyikoglu, A. C., Portet, F., Blanchon, H., & Babouchkine, J. M. (2019, October). Towards spoken medical prescription understanding. In *2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)* (pp. 1-8). IEEE.