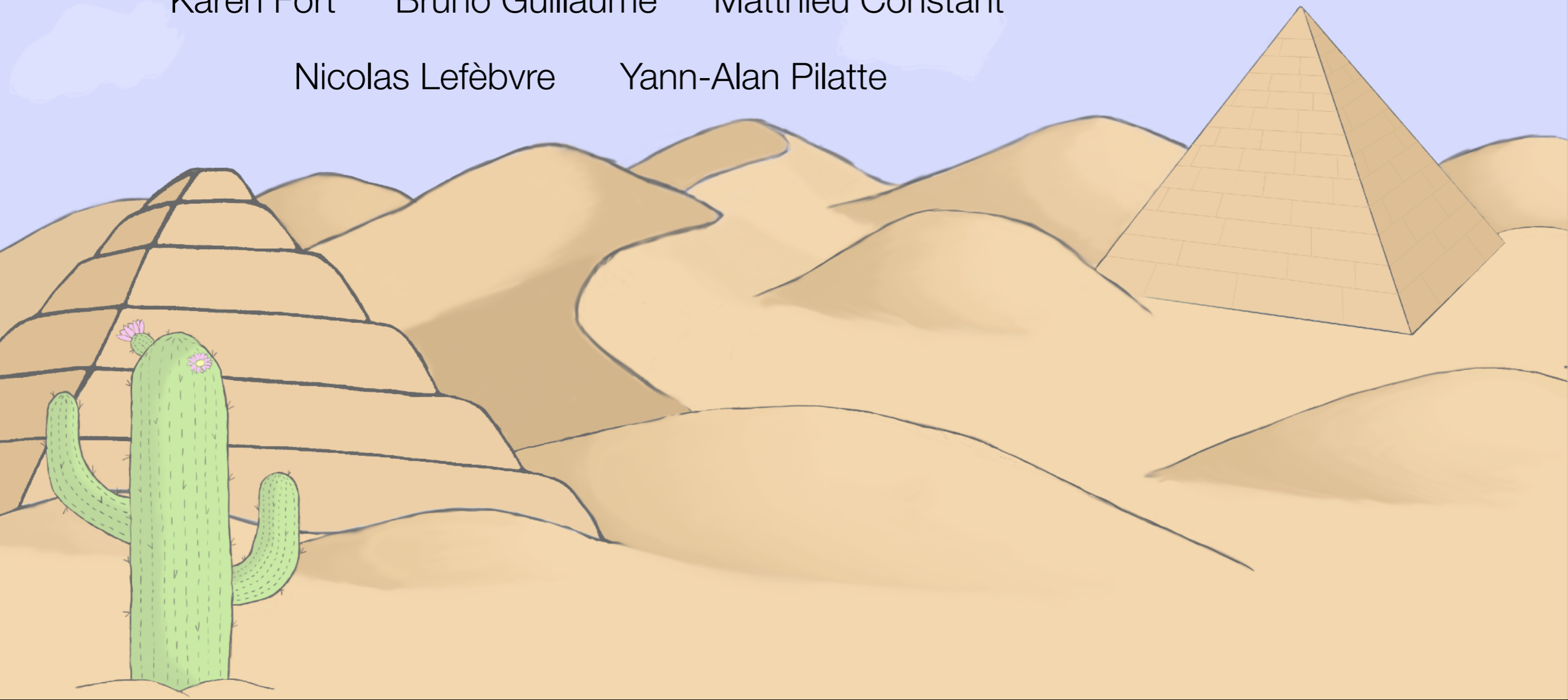


RIGORMORTIS

Evaluating Speakers' Identification of Multi-Word Expressions Using a Slightly Gamified Crowdsourcing Platform

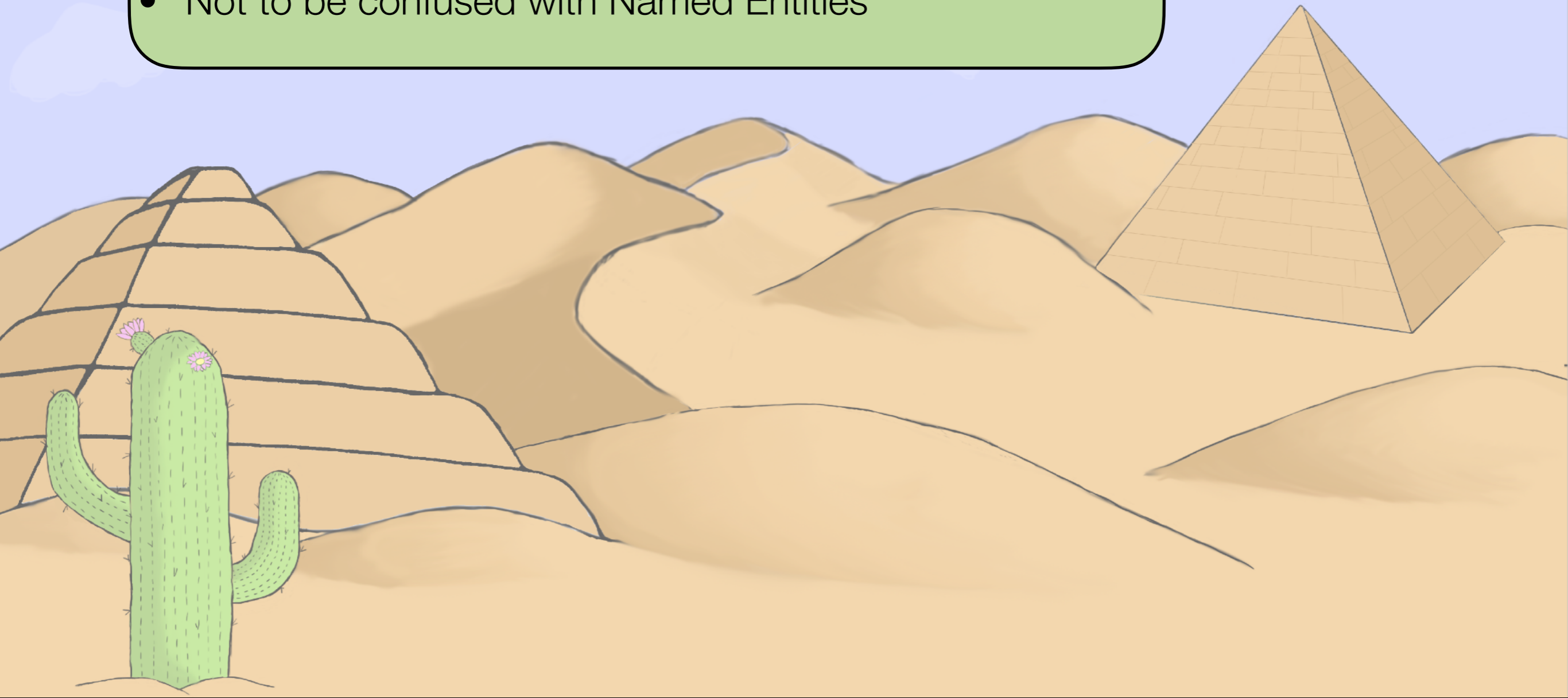
Karën Fort Bruno Guillaume Matthieu Constant

Nicolas Lefèbvre Yann-Alan Pilatte



Multi-Word Expression (MWE):

- A more or less fixed association of several words
- The meaning of the whole is not predictable from parts
- Not to be confused with Named Entities



Rigor Mortis:

- Experiment Crowdsourcing to build MWE annotated corpus
- Crowdsourcing:
 - Intuitive task (no training needed): image labelling
 - Complex task (training required): syntax annotation
- MWEs are in the midway:
 - Each native speaker has an intuitive knowledge of MWEs
 - MWE annotation is a complex task, even for experts (PARSEME)



Experiment the intuition part:

- No prior training: informal description and a couple of examples
- No feedback before the end (avoir biaisés)
- Publicized on social networks and French NLP mailing lists

Reference Corpus

- Annotated by experts (PARSEME)
 - 10 sentences (1 without MWE)
 - 16 MWEs
- Only fixed lexical components are annotated
⇒ approximate matching



Phase débloquée

Test de l'intuition des explorateurs.

Viens tester ton intuition sur des expressions multi-mots.

Voir mes résultats

Phase débloquée

Formation des explorateurs

Couloir Khéops

Couloir Khéphren

Couloir Mykérinos

Couloir Khentkaous

Couloir Sphinx

3^e phase

Exploration libre.

Lance-toi à la chasse aux expressions multi-mots !

Jouer

Classement

1 - Wellington	57555
2 - Methos31	46655
3 - NicoZombi	32830
4 - YeonWoo	31865
5 - Matadora	26445
6 - iougz	25955
7 - fanfan69	25060
8 - Lyco	21530
9 - manoste	18100
10 - Karen	15775
11 - Pr. Indianaperrier	10000
12 - Helice	8110
13 - lyly	7925
14 - bruno	7725
15 - Gwendo66	7275
16 - mhlay	7030
17 - FATMA MELLAK	5550

Trouve les expressions multi-mots présentes dans la phrase


Besoin
d'aide ?



- 25 avril 2001 : le juge Éric Halphen se **déclare incompetent** pour instruire le cas du président Chirac, **en raison** de son **immunité présidentielle**.

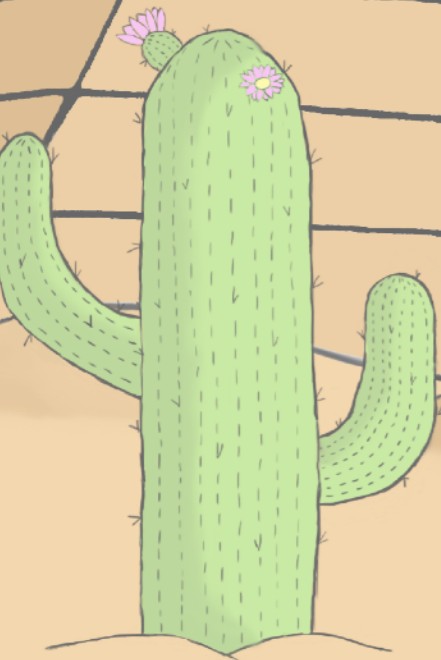
immunité présidentielle



Valider 

déclare incompetent

en raison



- 25 avril 2001 : le juge Éric Halphen se **déclare incompetent** pour instruire le cas du président Chirac, **en raison** de son **immunité présidentielle**.

en **raison** Réponse exacte !

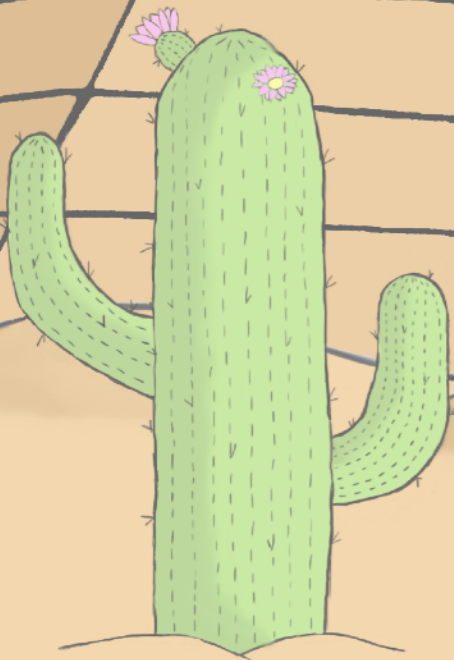
déclare **incompétent** Ce n'est pas une expression multi-mots !

immunité **présidentielle** Ce n'est pas une expression multi-mots !

On parle **entre autre** d'un métis qui dès son arrivée à Brazzaville a été autorisé à sortir de la **file d'attente**.

file **d'** **attente** Réponse exacte !

entre **autre** Tu as oublié cette expression multi-mots !



Global results:

- Sentences played by 66.8 participants on average
- 59% correct annotation of the sentence without MWE

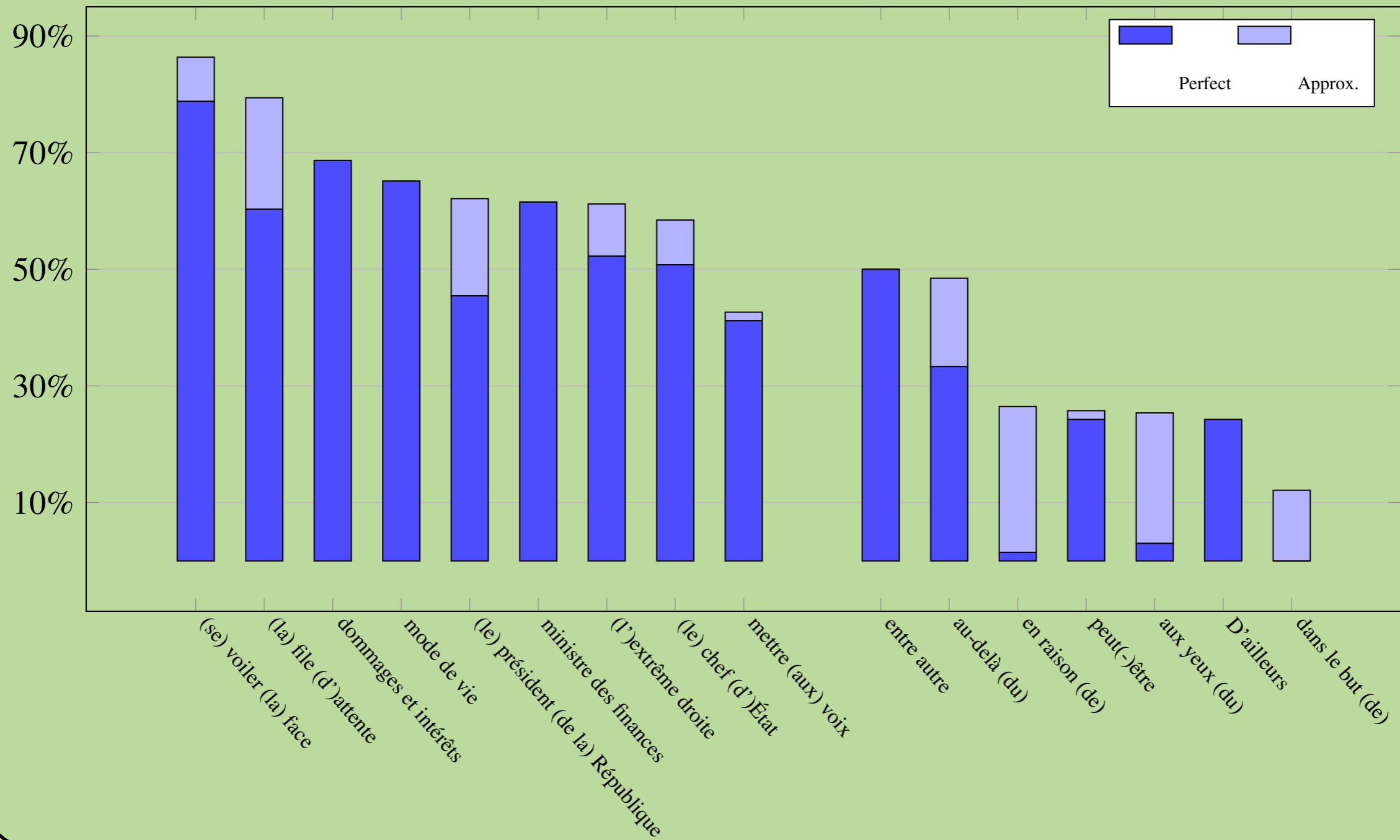
	Precision	Recall	F-measure
Perfect	48.13	41.22	44.41
Approximate	58.22	49.86	53.72

	non-functional	functional
Perfect	58.19	19.48
Approximate	65.05	30.41



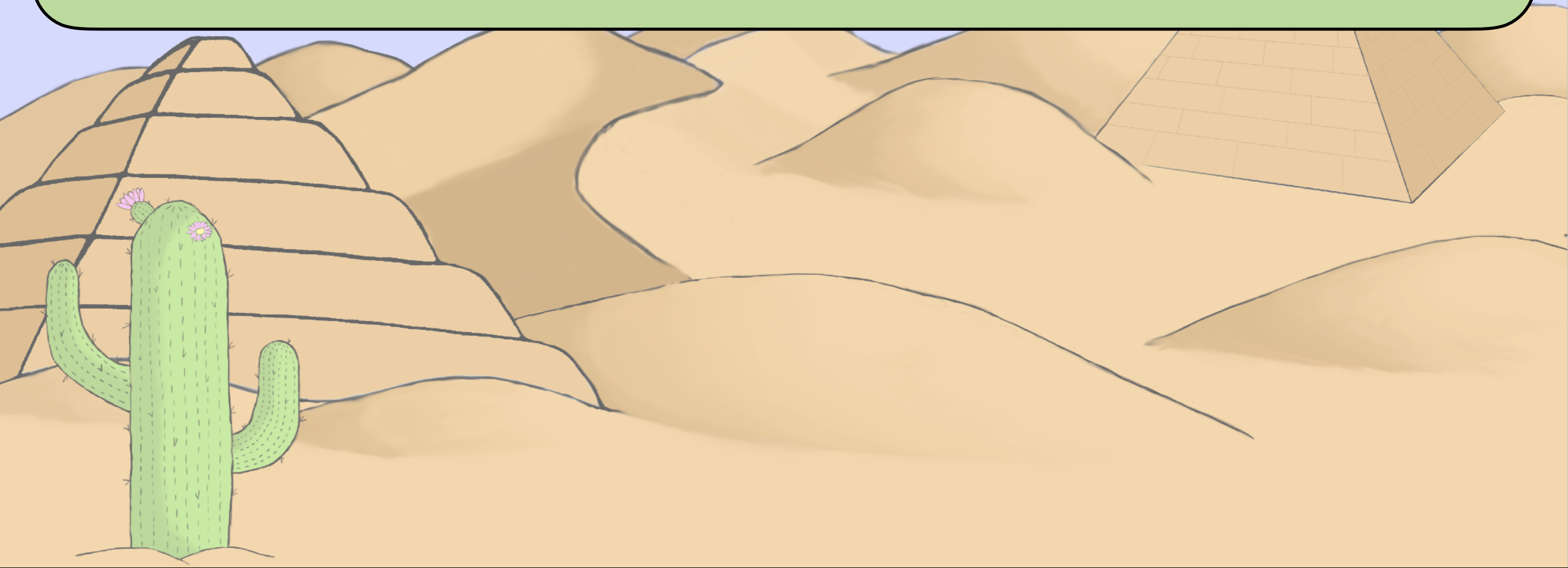
Recall of the players on each MWE

Left: non-functional Right: functional



Noise

Noisy expression (fr)	Noisy expression (en)	% of pl.	Comment
immunité présidentielle	presidential immunity	55,88 %	collocation (not MWE)
élection présidentielle	presidential election	34,85 %	collocation (not MWE)
aux yeux du public	to the eyes of the public	28,36 %	boundary error
destin tragique	tragic faith	23,88 %	collocation (not MWE)
Monsieur le président	Mister Predisent	19,70 %	common civilities



Discussion and perspectives

- The evaluation corpus is small
- Crowdsourcing helps finding non-functional MWEs
- Complementary experiment to evaluate how participant can be trained

Resources:

- Game available for French:

<http://rigor-mortis.org>

- Code freely available:

<https://github.com/gwaps4nlp/rigor-mortis>

