

Reconnaissance vocale du discours spontané pour le domaine médical



L. Ormaechea Grijalba, P. Bouillon, J. Gerlach

Faculté de Traduction et d'Interprétation (FTI), Université de Genève

B. Lecouteux, D. Schwab

Université Grenoble Alpes (UGA)/Laboratoire d'Informatique de Grenoble (LIG)

H. Spechbach

Hôpitaux Universitaires de Genève (HUG)

Introduction (1/3)

1. Objectif du projet :

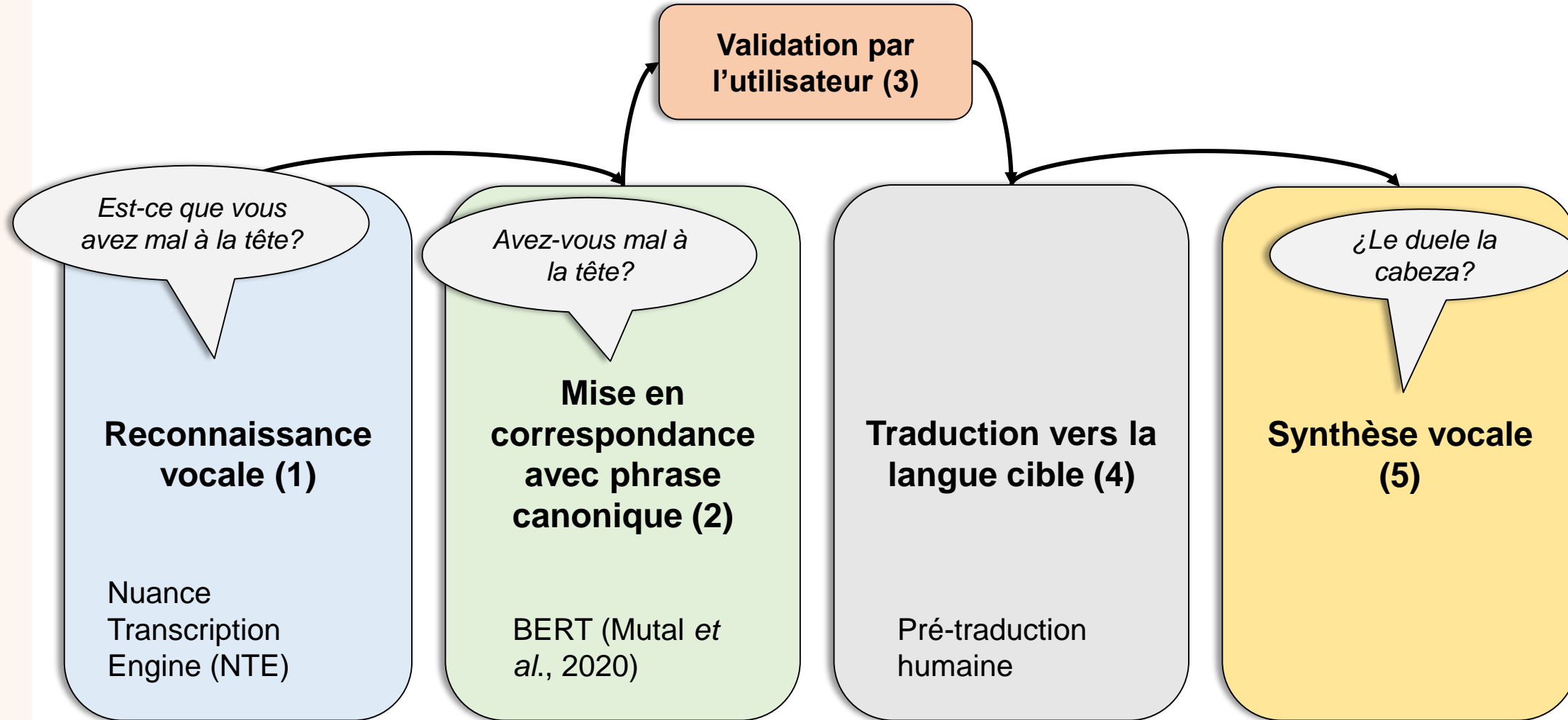
- **BabelDr** : outil de **traduction médicale** spécialisé.
- Utilisé au sein des unités d'urgences des **Hôpitaux Universitaires de Genève (HUG)**.
- Utilisant système de reconnaissance de la parole venant d'une **société privée** (Nuance).

Objectif de notre travail → Créer un système de reconnaissance **alternatif**, basé sur des outils **open source**.



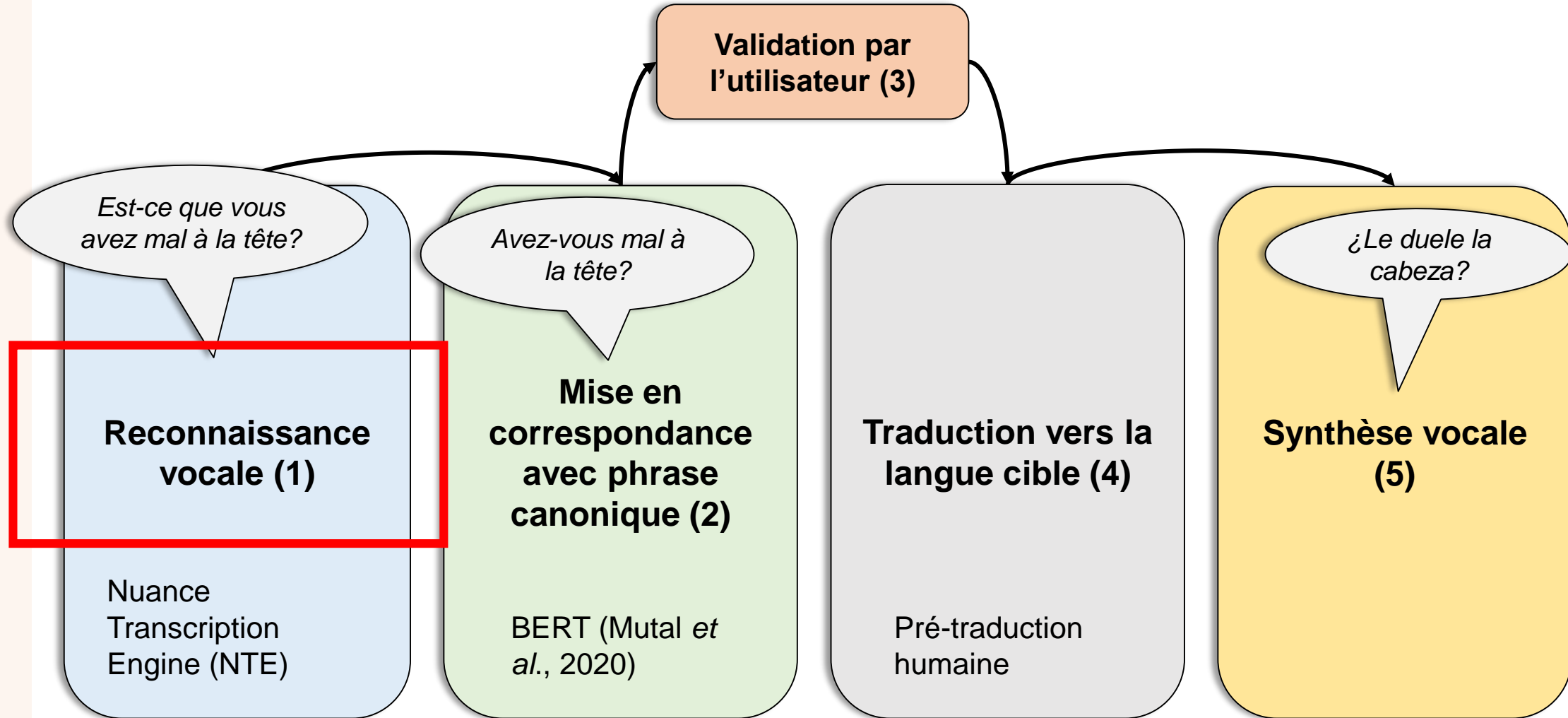
Introduction (2/3)

2. Contexte du projet – Architecture actuelle de BabelDr :



Introduction (2/3)

2. Contexte du projet – Architecture actuelle de BabelDr :

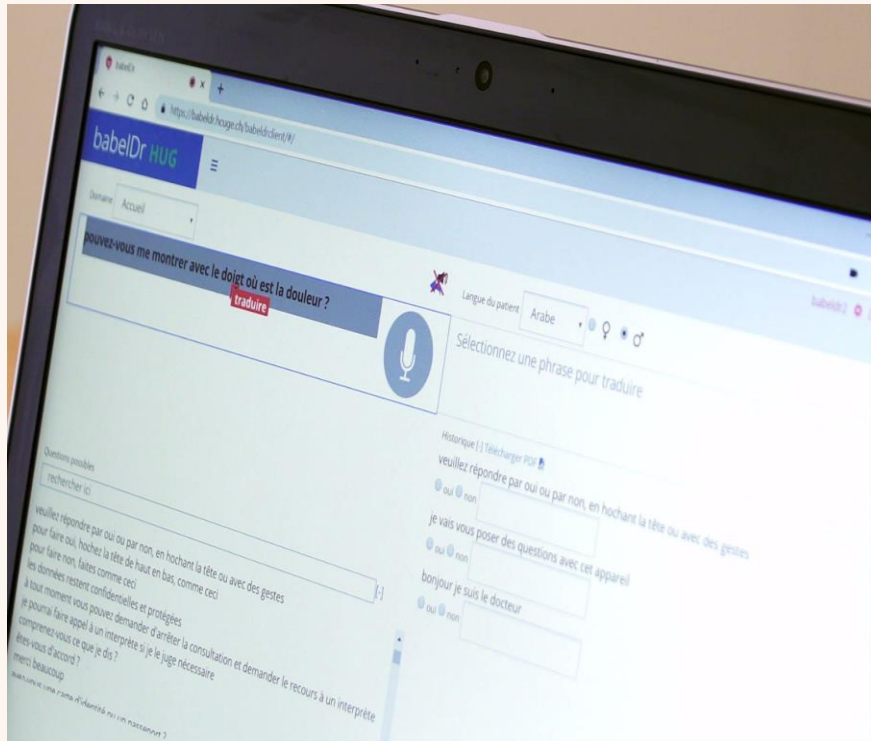




Introduction (3/3)

3. Questions de recherche :

- **Comment** peut-on construire un système de **reconnaissance vocale** pour un **domaine spécialisé** ?
- Quelle **performance** peut être atteinte avec un **système *open source*** ?

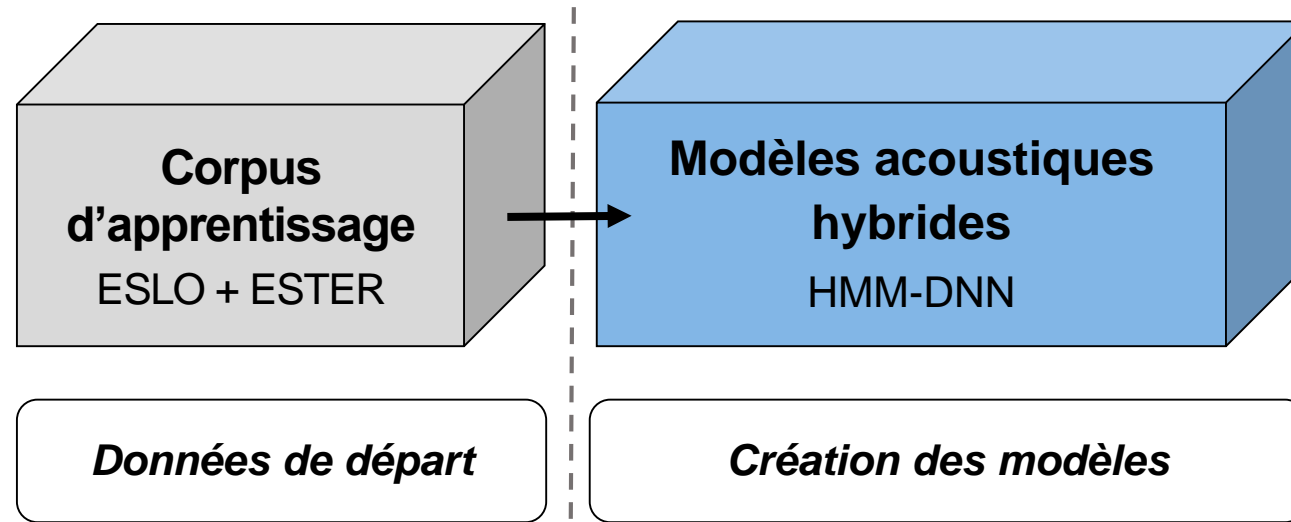


Méthodologie et évaluation (1/3)

Création du système de reconnaissance automatique de la parole :

- Basé sur la boîte à outils *open source* **Kaldi** (Povey *et al.*, 2011).
- Modification et ajustement de **2 des composants de son architecture typique** :

❑ Modèles acoustiques :



Méthodologie et évaluation (2/3)

❑ Modèles de langue ou grammaires :

```

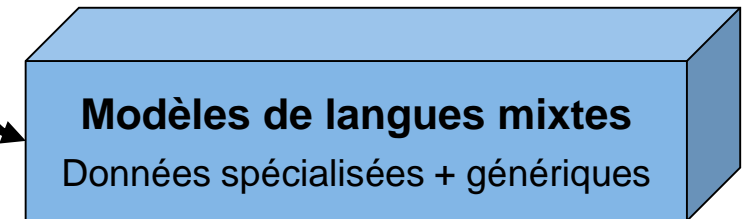
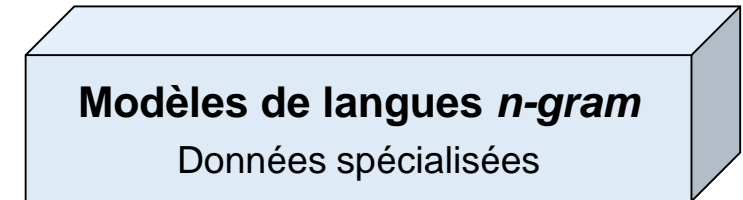
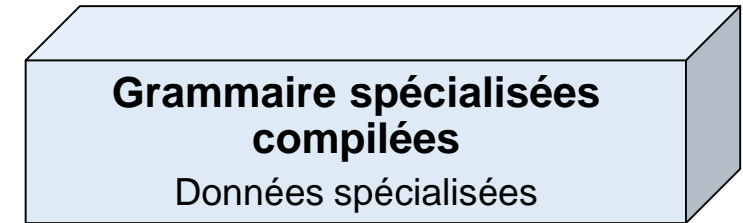
Utterance
Source $avez_vous ( mal | des
douleurs ) quelque part
EndUtterance

TrPhrase $avez_vous
Source ( avez-vous | vous avez )
EndTrPhrase
    
```

Extrait des grammaires spécialisées.



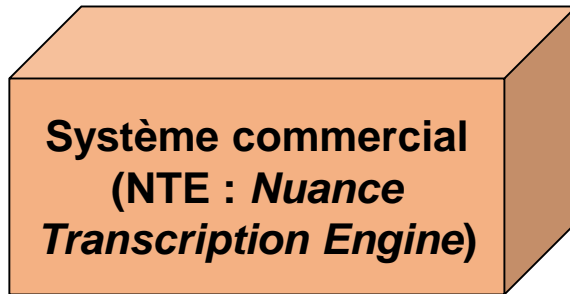
Données de départ



Création des modèles

+
Rigidité des modèles de langue
-

Méthodologie et évaluation (3/3)



Transcription de données

Corpus oral :

- Collecté aux **HUG** avec des médecins via **BabelDr**.
- Varié du point de vue **inter-locuteur** + **conditions acoustiques** diverses.

Corpus	Mots	Phrases	Locuteurs	Parole
<i>Dev</i>	16758	2864	24	2h15
<i>Test</i>	15166	2708	14	2h19

Évaluation

Résultats (1/3)

1. Résultats du corpus test en termes de *Word Error Rate* (WER) :

Kaldi [Nuance-NTE]					
Corpus test	Mots reconnus	Inser.	Suppr.	Subst.	WER (%)
<i>dc1</i>	2364/2981	88	213	316	20,70 [45,28]
<i>dc2</i>	5959/6797	195	182	461	12,33 [16,62]
<i>exp2</i>	4663/5388	151	262	312	13,46 [15,83]
Total	12986/15166	WER total ponderé (%)			14,37 [22,93]

- Réduction globale d'environ 9% WER avec le système Kaldi.

Résultats (2/3)

2. Juste une évaluation WER ? :

Limitations :

- *Word Error Rate* → **Mesure d'évaluation globale.**
- Dans un domaine aussi **sensible** que le **médical** → Besoin de **plus de contrôle** sur la transcription.

REF	pouvez-vous	***	montrer	où
HYP	pouvez-vous	me	montrer	où
	C	I	C	C

Exemple 1.

REF	avez-vous	des	problèmes	avec	le	foie
HYP	avez-vous	des	problèmes	avec	le	froid
	C	C	C	C	C	S

Exemple 2.

Face à ça :

- Besoin d'une évaluation de **type sémantique** → *SemER (Semantic Error Rate)*.

Résultats (3/3)

3. Résultats du corpus test en termes de *Semantic Error Rate* (SemER) :

- **SemER** → % phrases reconnues pour lesquelles la phrase canonique résultante n'est pas identique à la phrase canonique correcte.
- Visant à vérifier → **sens(phraseReconnue) == sens(phraseCanoniqueCorrecte)**

Ex.: *Est-ce que vous avez de la température ? == Avez-vous de la fièvre ?*

Corpus test	Kaldi		Nuance-NTE	
	Phr. reconnues	SemER(%)	Phr. reconnues	SemER(%)
<i>dc1</i>	505/622	18,67	370/622	40,42
<i>dc2</i>	1002/1158	13,32	579/1158	49,91
<i>exp2</i>	802/928	13,20	797/928	13,74
SemER total ponderé (%)		14,73	SemER total ponderé (%)	35,52

Conclusions et perspectives

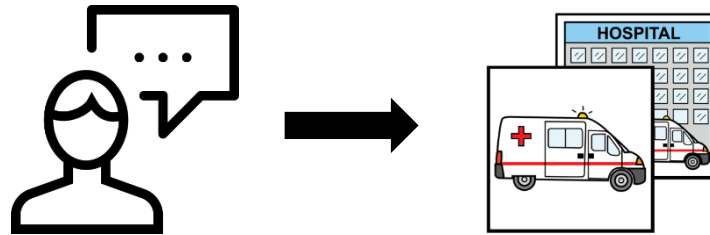
Récapitulatif :

- **Amélioration des résultats Kaldi** par rapport aux systèmes Nuance.
- Possible de créer un système de reconnaissance **performant** pour un **domaine spécialisé**.

Poursuites :

- **Court-terme :**
 - Réentraînement de nos modèles acoustiques (incorporation de *Common Voice*).
 - Création d'un modèle de langue plus flexible.
- **Long terme :**
 - Utilisation de ce système dans le cadre du projet **ANR/FNS PROPICTO** (*PRojection du langage Oral vers des unités PICTOgraphiques*) → [<https://propicto.unige.ch/>]

Développement de **systèmes de traduction parole-pictogrammes**





Merci de votre attention !

Références bibliographiques

- [Motal *et al.*, 2020] MUTAL, J., GERLACH, J., BOUILLON, P., & SPECHBACH, H. (2020). Ellipsis Translation for a Medical Speech to Speech Translation System. *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, 281-290.
- [Povey *et al.*, 2011] POVEY, D., GHOSHAL, A., BOULIANNE, G., BURGET, L., GLEMBEK, O., GOEL, N., HANNEMANN, M., MOTLICEK, P., QIAN, Y., SCHWARZ, P., SILOVSKY, J., STEMMER, G., & VESELY, K. (2011). *The Kaldi Speech Recognition Toolkit*. IEEE 2011 Workshop on Automatic Speech Recognition and Understanding.
- [Rayner *et al.*, 2016] RAYNER, M., BOUILLON, P., TSOURAKIS, N., SPECHBACH, H., & GERLACH, J. (2018). Handling Ellipsis in a Spoken Medical Phraselator. In *Statistical Language and Speech Processing* (Vol. 11171, p. 140-152). Springer International Publishing.