

Déploiement de systèmes de traduction de la parole au Swiss AI Center

Alexis Allemann, Andrei Popescu-Belis, Bertil Chapuis (HEIG-VD / HES-SO)

Gopal Krishnamani, Keshav Krishnamani (Bhaasha SàRL)

Intérêt croissant du grand public pour la traduction de la parole (*speech-to-speech*)





Consommation

On en parle

Du lundi au vendredi de 8h36 à 10h sur la 1ère

[▶ DERNIÈRE ÉMISSION](#) [Partager](#) [Contacter l'émission](#)

Assurances, contrats, conso, nouvelles technologies, services en tout genre... "On en parle" vous oriente dans tout ce qui fait votre quotidien.



Quels outils de traduction pour une utilisation facile et efficace?

"Pour bien s'entendre, il faut se comprendre". Et la tâche s'est singulièrement simplifiée avec les progrès des outils de traduction. Gratuits ou payant on en trouve sur des sites web ou comme logiciels voire comme applications pour smartphone. Les traductions automatiques permettent de traiter aujourd'hui de grands volumes de textes. Comprendre le contexte d'une situation demande un apprentissage massif et les malentendus cocasses ou non restent nombreux. Avec le développement de l'IA, les métiers d'interprètes et de traductrices ou traducteurs s'adaptent mais sont loin de disparaître.

François Jeannet reçoit Andrei Popescu-Belis, professeur d'informatique et chercheur en traitements de langues et traduction automatique.

[Lire moins](#)

[▶ 9 min](#)

PLAY **RTS**
Émissions Directs Programme TV

Accueil >

A bon entendeur

Le magazine de la consommation.



Applications de traduction : gare aux malentendus !

Accueil > A bon entendeur > **ST** 20.06.2023 · 28 min

Elles sont un outil précieux sur la route des voyages : les applications de traduction qui permettent de communiquer dans le monde entier. La traduction vocale ouvre la voie de la discussion quelle que soit la langue de son interlocuteur, tandis que la simple photo d'un menu en coréen ou en mandarin permet de savoir ce qu'il y aura dans son assiette. Mais à quel point les traductions de ces applications sont-elles fiables ? ABE fait le test dans cette émission.

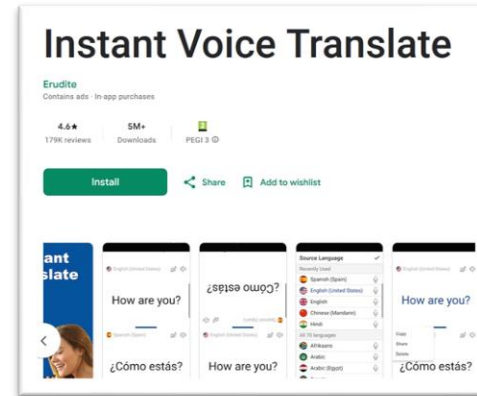
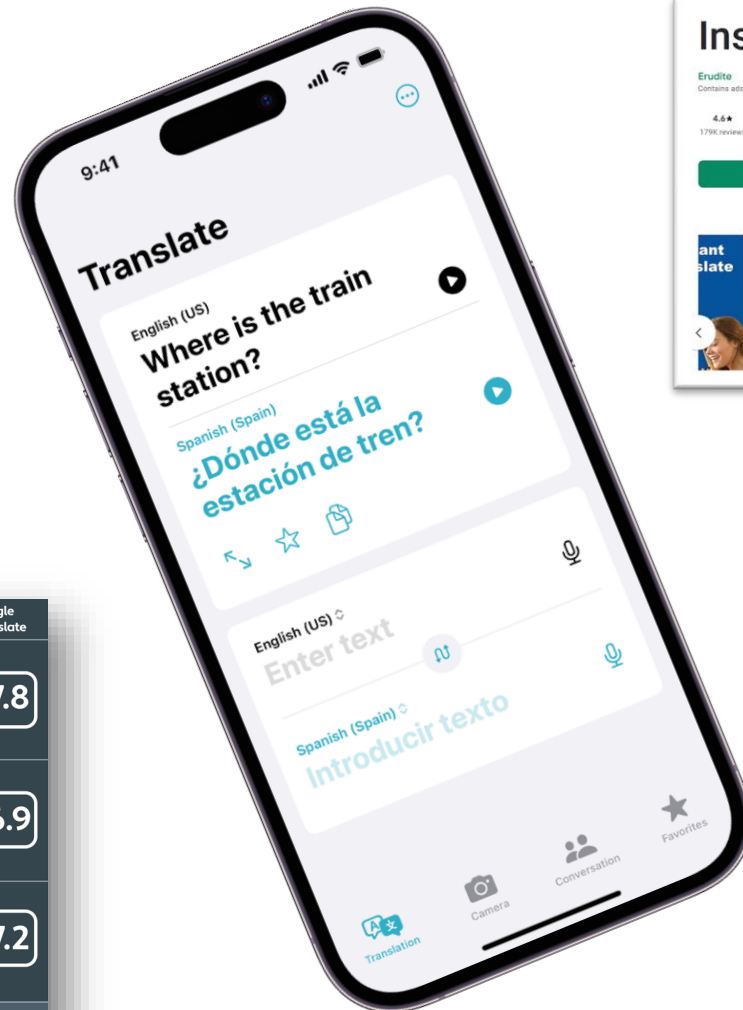
PLUS DE LIENS

[Quelle application de traduction choisir pour vos vacances ? ↗](#)







[Traducteurs vocaux : quelle fiabilité ? ↗](#)

Augmentation de l'offre mais qualité variable

1. Best Overall: Google Translate
2. Best for Accuracy: DeepL Translate
3. Best for Business: Microsoft Translator
4. Best for Extra Features: iTranslate
5. Best for Live Conversations: SayHi
6. Best for Messaging: Language Translator
7. Best for FaceTime Calls: Translate Now
8. Best for Phrases: Travel Phrasebook
9. Best for Offline Use: Translator with Speech
10. Best for Simplicity: Voice Translator Free



Similar apps →

-  Talking Translator
idealappcenter
4.7 ★
-  DeepL Translate
DeepL SE
4.7 ★
-  Transcribe Speech to Text
Transkriptor
3.1 ★
-  Microsoft Translator
Microsoft Corporation
4.6 ★
-  Camera Translator
Sleeping Music
2.3 ★
-  Reverso Translate and Learn
Reverso Technologies Inc.
4.8 ★

		i Translate	Microsoft Translator	Yandex	Talk & Translate	Google Translate
Portugais	Note	6.1	7.3	7.7	7.1	7.8
Grec	Note	4.9	5.9	6.6	6.7	6.9
Thailandais	Note	4.0	3.5	5.1	5.9	7.2
Prix		5.50 / mois	gratuit	gratuit	71.00 / an	gratuit

Automatiser l'interprétation communautaire

- Traduction de la parole = « interprétation »
- Interprétation communautaire = situations de dialogue informel entre des personnes qui ne se comprennent pas (≠ conférences, négociations, ...)

Bhaasha⁺ offre une plateforme reliant les interprètes humains à leurs clients; considère l'ajout d'un service automatique d'interprétation.

- **Les apps actuelles ne sont pas utilisables...**
 - pour des langues « rares » (peu de données d'entraînement)
 - dans des situations spécifiques : cabinet médical, administrations, parents d'élèves, etc.
 - lorsque la confidentialité doit être assurée
- **Combiner des modules performants** basés sur l'IA dans un système propriétaire

Contributions de la présentation

1. Un système pour combiner divers modules en cascade
 - reconnaissance vocale + traduction écrite + synthèse vocale
 - services/API ou modèles prêts-à-l'emploi ou modèles entraînés
2. Utilisation de l'infrastructure *cloud* du Swiss AI Center
3. Sélection de la **meilleure combinaison** de modules
 - sur un jeu de données de développement turc-français
4. Extension du système pour la récolte des données

Plate-forme Swiss AI Center (CSIA-PME)

Reconnaissance vocale → traduction écrite → synthèse vocale

speech-to-text (STT) → machine translation (MT) → text-to-speech (TTS)
(or automatic speech recognition, ASR) *(or speech synthesis)*

Interface utilisateur : plusieurs modes

traduction parlée | traduction + évaluation | traduction + évaluation + correction

Speech translator

1. Record or upload speech



2. Select the source and target languages

Source language: French ▾

Target language: English ▾



3. Translate and play speech

A green rounded rectangular button with a white right-pointing arrow and the text "Translate and play".

Status



Speech-to-text (STT)



Machine translation (MT)



Text-to-speech (TTS)

Results

STT	<p>French transcription</p> <p>Bonjour et bienvenue aux journées de l'intelligence artificielle à Lausanne.</p>	Rating Unrated ▾
MT	<p>Translated text</p> <p>Hello and welcome to the artificial intelligence days in Lausanne</p>	Rating Unrated ▾
TTS	A dark grey audio player with a play button, a progress bar, a 0:00 / 0:04 timer, and a volume icon.	Rating Unrated ▾

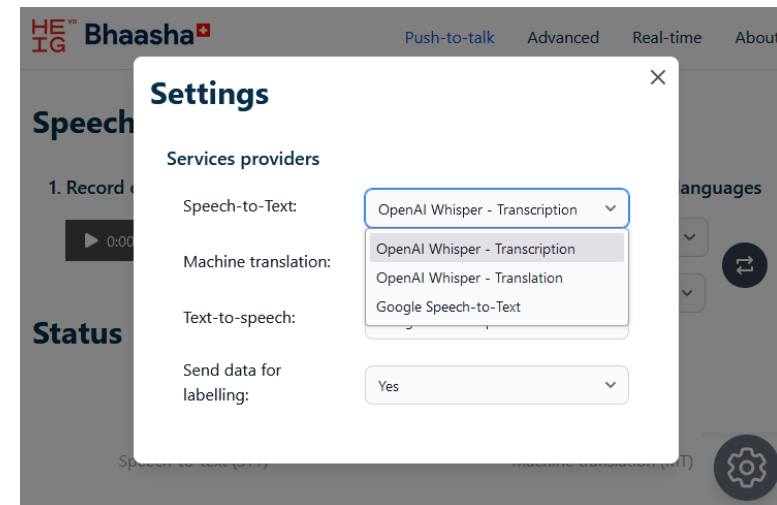
Complexité potentielle du système

Modules disponibles

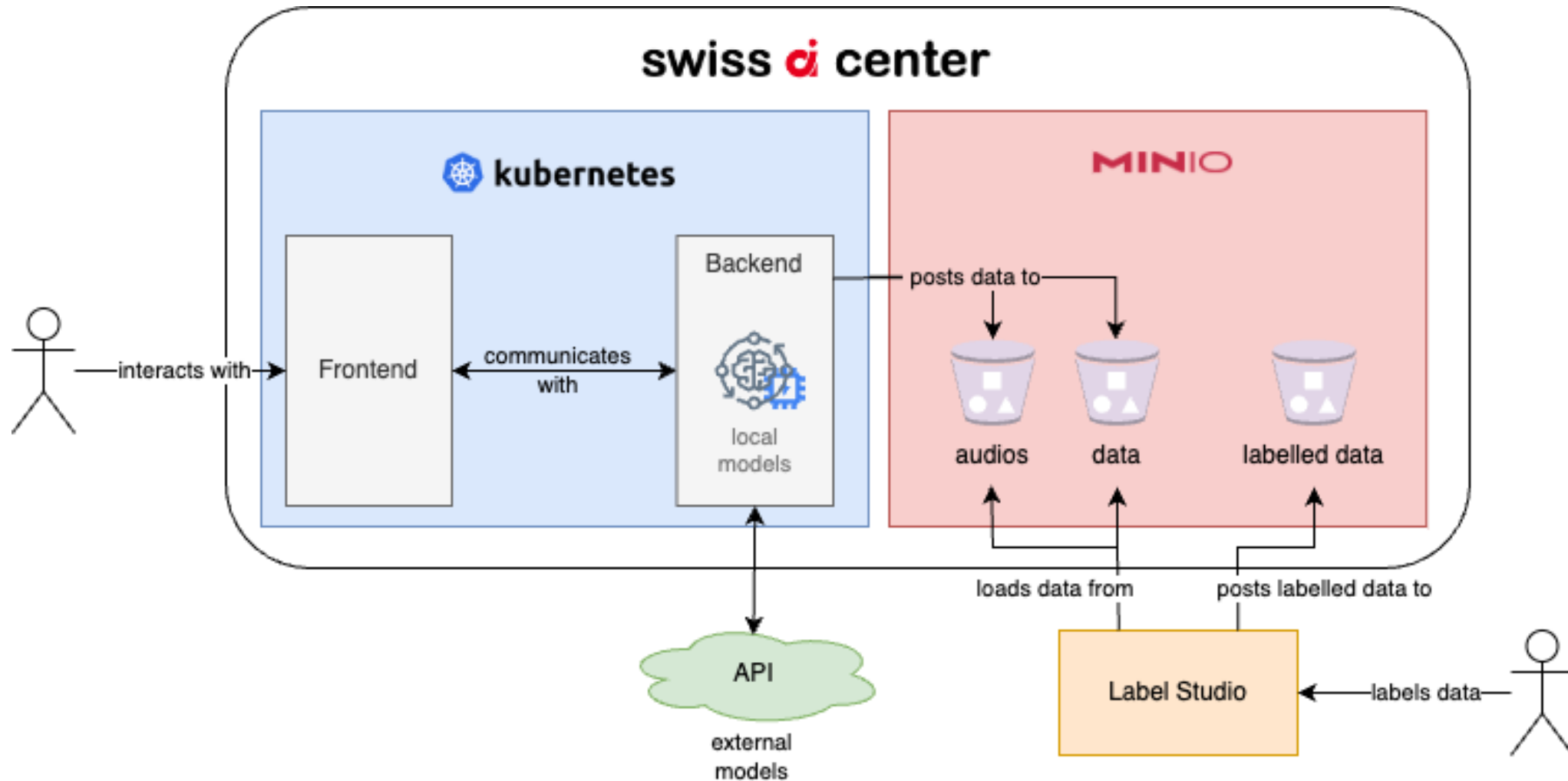
- STT
 - OpenAI Whisper (local)
 - Google API (service)
- STT + MT combiné
 - OpenAI Whisper (local)
- MT
 - Google API (service)
 - MarianNMT entraîné sur OpusMT par HelsinkiNLP (local)
- TTS
 - Google API (service)

Langues disponibles

- Anglais, français, russe, turc
 - en entrée et en sortie ⇒ directions
- Tous les modules ne sont pas disponibles pour toutes les langues
 - interface ⇒ choix de la combinaison



Utilisation de l'infrastructure *cloud* du Swiss AI Center



Label Studio Projects / Bhaasha / Labeling

audio_url aud sc III Lk #33 + AL alexis.allemand #8WL7y 2 minutes ago

<input type="checkbox"/>	<input type="button" value="Click to load"/>	tr
<input type="checkbox"/>	<input type="button" value="Click to load"/>	tr
<input type="checkbox"/>	<input type="button" value="Click to load"/>	tr
<input type="checkbox"/>	<input type="button" value="Click to load"/>	en
<input type="checkbox"/>	<input type="button" value="Click to load"/>	tr
<input type="checkbox"/>	<input type="button" value="Click to load"/>	tr
<input type="checkbox"/>	<input type="button" value="Click to load"/>	tr

Source audio

00:00:00:000 00:00:00:000

Source language : tr Target language : fr

Transcription

Generated transcription

Suriye sınırındaki mayınların temizlenmesine dair karar, siyasetçilerin gündeminde olduğu kadar basın da gündeminde önemli yer tuttu.

Evaluation

☆☆☆☆☆ 1=very poor, 2=poor, 3=acceptable, 4=good, 5=very good

Reference transcription

Suriye sınırındaki mayınların temizlenmesine dair karar, siyasetçilerin gündeminde olduğu kadar basın da gündeminde önemli yer tuttu.

Add

Translation

Generated translation

La décision de déminer la frontière syrienne occupait une place importante dans l'agenda de la presse ainsi que dans celui des hommes politiques.

Evaluation

☆☆☆☆☆ 1=very poor, 2=poor, 3=acceptable, 4=good, 5=very good

Reference translations

La décision de déminer la frontière syrienne a occupé une place importante non seulement dans l'agenda politique, mais aussi dans l'agenda de la presse.

Add

Submit

Label Studio

Intégration avec le *Core Engine* du Swiss AI Center ?

- 3 services

Speech-to-Text

STT

Transcribes spoken words into written text, enabling users to convert audio speech into readable content.

VIEW

Machine Translation

MT

Translates text from one language to another, facilitating communication across linguistic barriers.

VIEW

Text-to-Speech

TTS

Converts written text into spoken words, allowing users to listen to the synthesized audio output.

VIEW

- 1 pipeline

Speech Translation

STT MT TTS

Translates spoken sentences from one language to another, enabling communication between speakers of different languages.

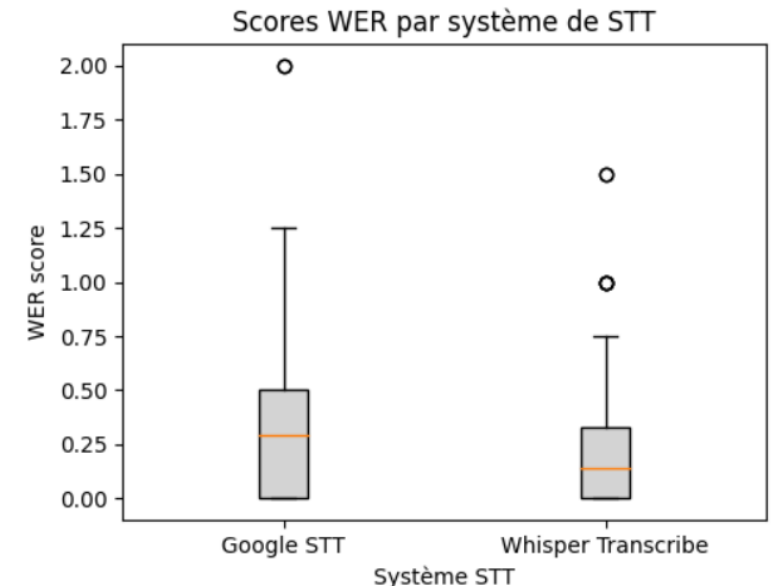
VIEW

Avantages

- Implémentation facilitée
- Une seule requête à la *pipeline*
- Gestion simplifiée des erreurs
- Services gérés par le Swiss AI Center
- Infrastructure externalisée
- Réutilisation des services

Sélection de la meilleure combinaison de modules, pour une paire de langue (1/2)

- Données de développement **turc-français** (préliminaires)
 - corpus CoVoST-2 / turc-anglais / dev : 1614 phrases
 - audio + transcription correcte + traduction correcte
 - nous avons traduit **127 phrases** vers le français
- Qualité de la reconnaissance vocale (STT) →
 - mesurée par la WER : *word error rate*
- Conclusion
 - **Whisper** (gratuit et exécuté localement) est perceptiblement meilleur que **GoogleSTT** (service payant accessible via API)



Sélection de la meilleure combinaison de modules, pour une paire de langue (2/2)

3 > 1

4 > 2

- Passer de Whisper STT à GoogleSTT dégrade le score BLEU de ca. 20%

1 >> 2

3 >> 4

- Passer de GoogleMT au Modèle local tr-fr dégrade le score BLEU de ca. 45-50%
⇒ on doit améliorer notre modèle

3 >> 5

4 ≈ 6

5 > 6

- La combinaison de deux traductions (Whisper STT+MT tr-en puis MT en-fr) a une qualité plus faible ou égale aux précédentes (modèle local en-fr un peu inférieur à GoogleMT)

	Système STT	Système MT	BLEU	COMET
1.	GoogleSTT	GoogleMT tr-fr	45.9	80.0
2.	GoogleSTT	Modèle local tr-fr	22.8	73.4
3.	WhisperSTT	GoogleMT tr-fr	60.9	89.0
4.	WhisperSTT	Modèle local tr-fr	28.4	81.8
5.	Whisper STT+MT (en)	GoogleMT en-fr	37.8	82.4
6.	Whisper STT+MT (en)	Modèle local en-fr	31.5	81.4

Extension du système pour la récolte des données : mode avancé

- Comment évaluer de manière fiable, pour diverses paires de langues ?
- Nécessité d'une collecte efficace des données
 - avec l'interface en mode avancé, ou avec LabelStudio
- *Workflows* possibles pour obtenir des données
 - A. Lire à haute voix des phrases écrites, dans les deux langues
 - provenant d'un corpus parallèle (traduit) représentatif du domaine/usage souhaité
 - B. Prononcer des phrases spontanées et les interpréter
 - représentatives du domaine/usage, plus naturelles
 - nécessite ensuite les transcriptions correctes

1. Record or upload speech



2. Select the source and target languages

Source language: Turkish

Target language: French



3. Translate and play speech

Optional: Upload an evaluation file

reference2.txt



Milliyet gazetesi yazarı Fikret Bila yirmi dokuz Mayıs iki bin dokuz tarihli yazısında mayın işinde AKP gurubu da isteksiz başlığını kullandı.

Status



Speech-to-text (STT)



Machine translation (MT)



Text-to-speech (TTS)

Results

Turkish transcription		WER	Rating	
STT	Miniyet Gazetesi yazarı Fikret Bila, 29 Mayıs 2009 tarihli yazısında Mayın işinde AKP grubu da isteksiz başlığını kullandı.	33.3	5 - Very good	
Translated text		BLEU	COMET	Rating
MT	Dans son article du 29 mai 2009, Fikret Bila, rédacteur du journal Miniyet, a utilisé le titre Le groupe AKP est également réticent à s'engager dans le secteur minier.	31.6	73.4	4 - Good
TTS			Unrated	



Perspective : projet INTERCOM (HES-SO)

Collecte de données

- Enregistrement audio de phrases source
- Transcription correcte
- Traduction correcte
- Enregistrement audio de phrases cible
- **Transcription système** (avec son score WER et son score humain)
- **Traduction système** (avec ses scores BLEU/COMET et son score humain)
- **Génération audio automatique** (avec son score humain)

Systèmes et leur évaluation

- Évaluer automatiquement des pipelines avec WER, BLEU/COMET
- **Estimer la fiabilité de l'évaluation automatique**
 - en comparant les scores humains avec ceux de WER, BLEU/COMET, etc.
- **adapter ou entraîner des systèmes**