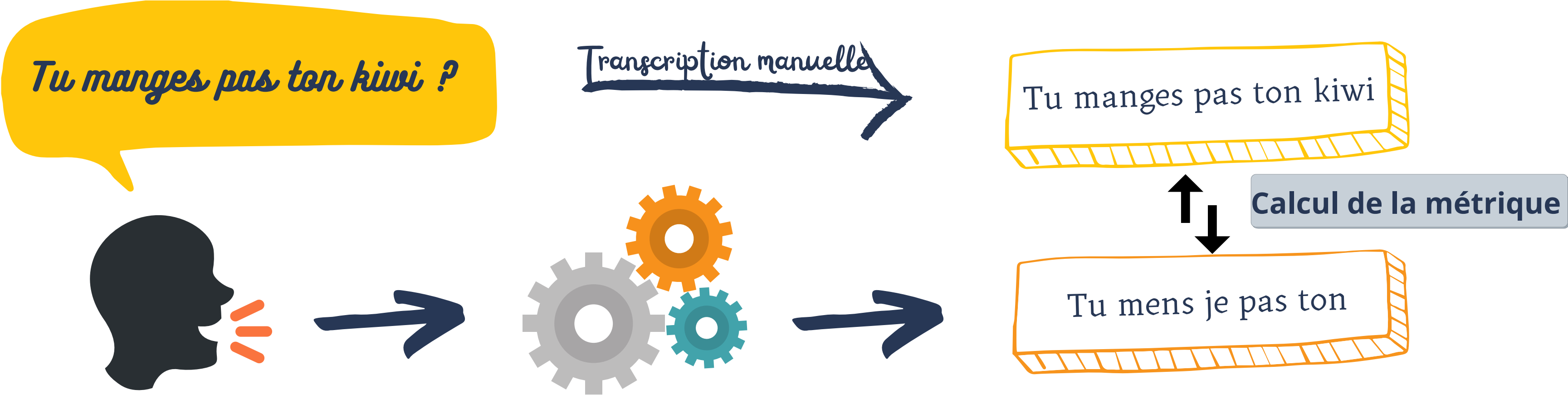




Qualitative Evaluation of Language Model Rescoring in Automatic Speech Recognition

Automatic Speech Recognition



UN PEU DE CONTEXTE...

Le Word Error Rate (WER) ou Taux d'Erreur Mot (TER) est LA métrique utilisé en RAP

UN PEU DE CONTEXTE...

Le Word Error Rate (WER) ou Taux d'Erreur Mot (TER) est LA métrique utilisé en RAP

le WER se calcule alors comme suit :

$$WER = \frac{\#Sub + \#Ins + \#Del}{\#mots\ de\ la\ référence}$$

UN PEU DE CONTEXTE...

Le Word Error Rate (WER) ou Taux d'Erreur Mot (TER) est LA métrique utilisé en RAP

le WER se calcule alors comme suit :

$$WER = \frac{\#Sub + \#Ins + \#Del}{\#mots\ de\ la\ référence}$$

Référence	Tu	ne	manges		pas	ton	kiwi
	=	=	<i>Sub</i>	<i>Ins</i>	=	<i>Sub</i>	<i>Del</i>
Hypothèse	Tu	ne	mens	je	pas	toi	

UN PEU DE CONTEXTE...

Le Word Error Rate (WER) ou Taux d'Erreur Mot (TER) est LA métrique utilisé en RAP

le WER se calcule alors comme suit :

$$WER = \frac{\#Sub + \#Ins + \#Del}{\#mots\ de\ la\ référence}$$

Référence	Tu	ne	manges	pas	ton	kiwi
	=	=	<i>Sub</i>	<i>Ins</i>	=	<i>Sub</i> <i>Del</i>
Hypothèse	Tu	ne	mens	je	pas	toi

"mange" (sans 's') est une erreur aussi grave que "mens"

UN PEU DE CONTEXTE...

Quelle est l'objectif ?

- Mettre en évidence le besoin d'une métrique plus proche de la perception humaine
- Évaluer différents aspects (grammaire, sémantique, mot, caractère, intérêt pour une tâche en aval, etc)
- Remettre en cause le taux d'erreur mot

PLAN

PRÉSENTATION RAPIDE DU SYSTÈME ÉVALUÉ

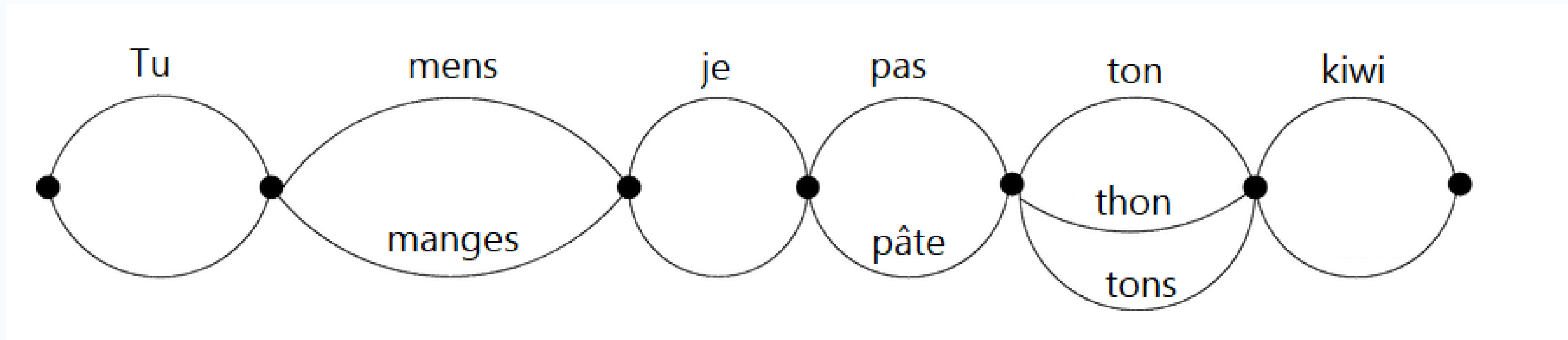
MÉTRIQUES UTILISÉS

RÉSULTATS ET ANALYSES DES MÉTRIQUES

Systemes de reconnaissance de la Parole utilisé

- Système basé sur une recette existante de Kaldi
- Système acoustique neuronal
- Trois modèles de langage utilisé
- Entraîné sur ESTER 1/2, EPAC, ETAPE, REPERE
- **Rescoring** (ou réordonnancement a posteriori des hypothèses) appliqué grâce à deux modèles de langages pour améliorer les prédictions des systèmes

RESCORING



Permet de sélectionner des phrases plus probables

Aperçu des métriques d'évaluation pour la RAP

- WER (mot)
- CER (caractère)
- POSER (part-of-speech)
- LER (lemme)
- EmbER (embeddings/mot)
- SemDist
- BERTScore (similarité max entre embeddings)

WER & CER

DISTANCE DE LEVENSHTTEIN

- Sur les mots et sur les caractères

POSER (Part-Of-Speech Error Rate)

WER appliqué sur les POS

- POS obtenues grâce au tagger POET utilisant des POS étendues

POSER (Part-Of-Speech Error Rate)

WER appliqué sur les POS

- POS obtenues grâce au tagger POET utilisant des POS étendues

Tu	PPER2S	ne	ADV	manges	VERB	€	pas	ADV	ton	DET	kiwi	NMS
=		=		=		I	=		=		D	
Tu	PPER2S	ne	ADV	mens	VERB	je	PPER1S	pas	ADV	ton	DET	€

Variantes :

- dPOSER (avec les POS étendues)
- uPOSER (avec les POS universelles)

POSER (Part-Of-Speech Error Rate)

WER appliqué sur les POS

- POS obtenues grâce au tagger POET utilisant des POS étendues

Tu	PPER2S	ne	ADV	manges	VERB	€	pas	ADV	ton	DET	kiwi	NMS
=	=	=		I	=	=					D	
Tu	PPER2S	ne	ADV	mens	VERB	je	PPER1S	pas	ADV	ton	DET	€

LER(Lemma Error Rate)

WER appliqué sur les lemmes

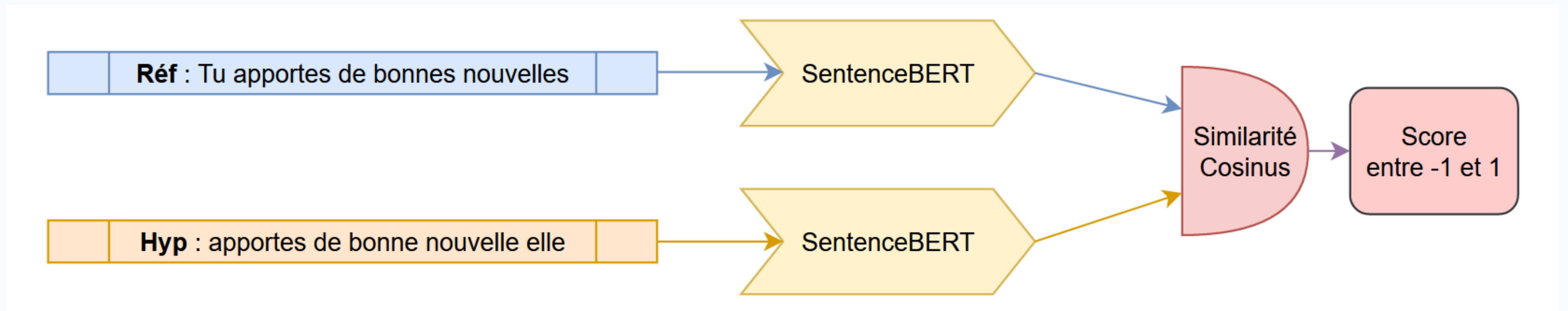
- Utilisation d'un lemmatizer issu de Spacy pour le français

EmbER (Embedding Error Rate)

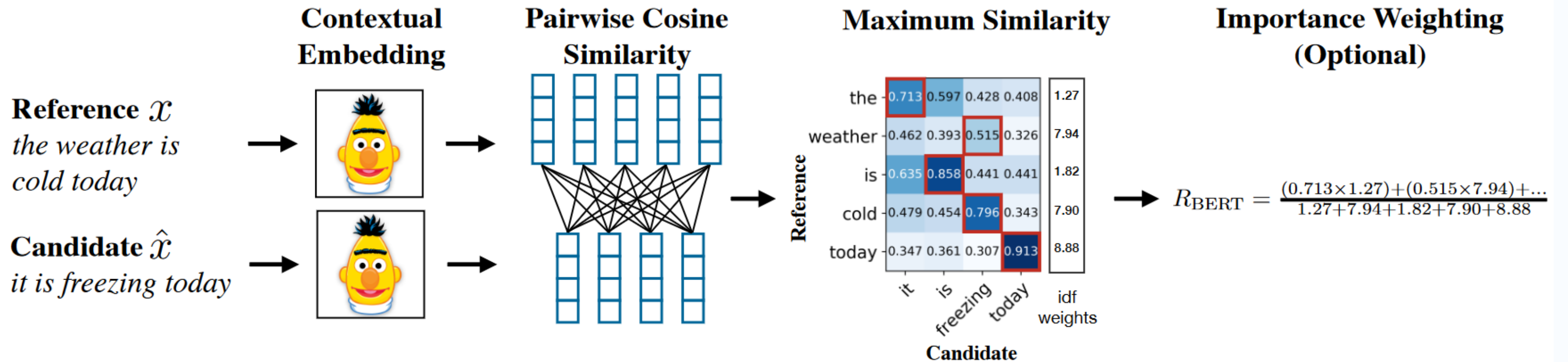
WER pondéré par la distance sémantique des substitutions

- On calcule la similarité cosinus lors d'une substitution et applique une erreur en fonction

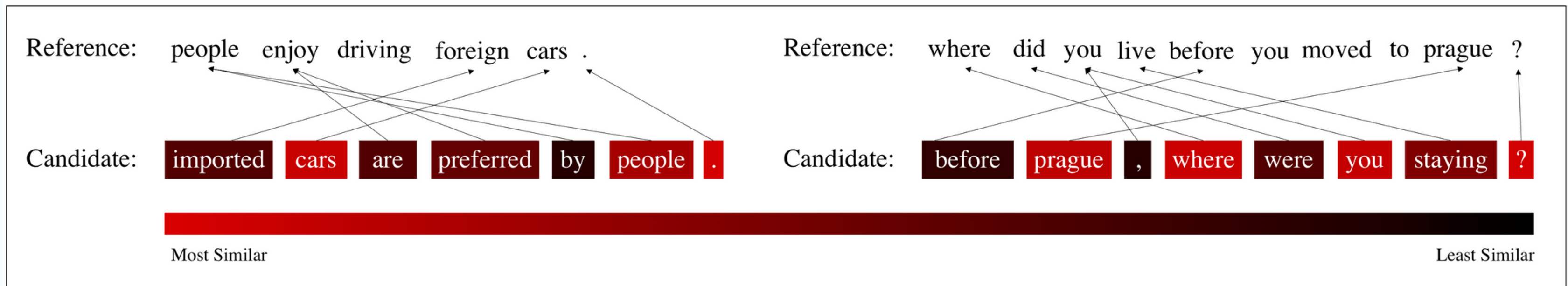
SemDist



BERTScore



BERTScore



Résultats

- Impact du Rescoring
- Corrélation inter-métriques
- Gain sémantique par POS
- Impact local du Rescoring

Corrélation inter-métriques

	WER	CER	LER	LCER	dPER	uPER	EmbER	SemDist	FBERT
WER									
CER	89.34								
LER	88.08	88.49							
LCER	87.10	98.31	91.40						
dPER	92.96	90.02	92.70	89.51					
uPER	90.40	90.58	93.69	90.81	97.95				
EmbER	96.51	91.51	86.57	88.78	91.00	88.98			
SemDist	71.81	64.78	62.22	62.60	65.33	64.13	75.73		
FBERT	74.63	74.27	72.60	73.00	74.09	74.25	84.51	63.35	

Impact du Rescoring

System	WER	CER	dPOSER	uPOSER	LER	LCER	SemDist	BERTScore	EmbER
Base	15.45	8.57	14.59	12.22	14.35	8.78	7.89	9.12	12.33
Rescoring	13.24	7.70	12.51	10.79	12.08	8.00	7.18	8.38	10.79
<i>Reduction</i>	-14.3 %	-10.2 %	-14.3 %	-11.7 %	-15.8 %	-8.8 %	-9.0 %	-8.1 %	-12.5 %

Gain sémantique par POS

	Base	Rescoring	Reduction
INTJ	14.07	10.45	-3.63
CCONJ	9.83	6.82	-3.01
VERB	6.10	4.20	-1.90
ADJ	5.08	3.41	-1.67
AUX	4.66	3.27	-1.39
PRON	5.37	4.12	-1.25
SCONJ	3.51	2.43	-1.08
PROPN	6.72	5.82	-0.90
NOUN	3.34	2.57	-0.77
ADV	3.23	2.49	-0.74
ADP	2.90	2.25	-0.65
DET	2.95	2.42	-0.53
NUM	2.96	2.62	-0.34

Impact local du rescoring

- 50% des phrases sont améliorées
- 37% des phrases sont dégradées
- 13% ne sont pas impactées

Impact local du rescoring

- 50% des phrases sont améliorées
- 37% des phrases sont dégradées
- 13% ne sont pas impactées



Impact local du rescoring

- 50% des phrases sont améliorées
- 37% des phrases sont dégradées
- 13% ne sont pas impactées



Hypothèse : Le rescoring normalise la parole spontanée vers le domaine de l'écrit. Les méta-informations du corpus REPERE indiquent 1,23 fois plus de disfluences en moyenne dans les phrases qui sont dégradées que dans les phrases qui sont améliorées.

A SUIVRE

EVALUER LA CORRELATION DES MÉTRIQUES AVEC LA
PERCEPTION HUMAINE

MERCI DE VOTRE ATTENTION
