

L. Besacier, B.
Lecouteux & NQ.
Luong

Utilisation de mesures de confiance pour améliorer le décodage en traduction de la parole

Laurent Besacier, Benjamin Lecouteux & NQ. Luong

Équipe GETALP, Laboratoire d'Informatique de Grenoble (LIG)
Université Grenoble Alpes



Introduction

Mesures en TAP

RAP

TA

Corpus pour la
traduction de la
parole

Le corpus RAP

Le corpus RAP + TA

Exemple

Systèmes RAP et TA

Mesures de
confiance GETALP

RAP

TA

Performances

Redécodage de
la phrase de
traduction

Travaux antérieurs

Notre approche

Résultats

Exemples de sorties

Conclusion et
perspectives

Pourquoi des mesures de confiance ?

- ▶ Estimer la qualité *a priori* d'une sortie
- ▶ Réécrire ou corriger ?

Dans la pratique

- ▶ Scénarios interactifs
- ▶ Réordonner les N-meilleures hypothèses
- ▶ Redécoder un graphe
- ▶ Si au niveau mot : label *good* ou *bad*
- ▶ RAP et TA traités séparément

L. Besacier, B.
Lecouteux & NQ.
Luong

Introduction

Mesures en TAP

RAP

TA

Corpus pour la
traduction de la
parole

Le corpus RAP

Le corpus RAP + TA

Exemple

Systèmes RAP et TA

Mesures de
confiance GETALP

RAP

TA

Performances

Redécodage de
graphe de
traduction

Travaux antérieurs

Notre approche

Résultats

Exemples de sorties

Conclusion et
perspectives

L. Besacier, B.
Lecouteux & NQ.
Luong

Introduction

Mesures en TAP

RAP

TA

Corpus pour la
traduction de la
parole

Le corpus RAP

Le corpus RAP + TA

Exemple

Systèmes RAP et TA

Mesures de
confiance GETALP

RAP

TA

Performances

Redécodage de
graphe de
traduction

Travaux antérieurs

Notre approche

Résultats

Exemples de sorties

Conclusion et
perspectives

En RAP

- ▶ Détection des MHV (Asadi et al. 1990)
- ▶ Les probabilités *a posteriori* (Young, 1994)
- ▶ À partir du graphe dans un premier temps (Kemp et al. 1997)
- ▶ Étendues avec d'autres traits : graphiques, sémantiques
- ▶ Problème d'apprentissage machine : réseaux Bayésiens naïfs, SVMs, réseaux de neurones ou CRFs.

En TA

- ▶ Syntaxe, topologie du graphe, polysémie (Luong et al. 2013)
- ▶ Similarité sémantique (Bicici. 2013)
- ▶ Les CRFs (Lafferty et al., 2001) semblent bien se prêter aux mesures en TA.

Pas de travaux pour la traduction orale utilisant des paramètres joints RAP+TA.

Corpus traduction

- ▶ Construit par GETALP (Potet et al. 2010)
- ▶ 10881 phrases traduites français->anglais
- ▶ Ensemble des sorties post-éditées

Corpus parole

- ▶ Construit par GETALP (Besacier et al. 2014)
- ▶ 2643 phrases prononcées en Français
- ▶ Détail : 881 phrases prononcées par 3 locuteurs
- ▶ 5h de parole (corpus journalistique)

L. Besacier, B.
Lecouteux & NQ.
Luong

Introduction

Mesures en TAP

RAP

TA

Corpus pour la
traduction de la
parole

Le corpus RAP

Le corpus RAP + TA

Exemple

Systèmes RAP et TA

Mesures de
confiance GETALP

RAP

TA

Performances

Redécodage de
grabe de
traduction

Travaux antérieurs

Notre approche

Résultats

Exemples de sorties

Conclusion et
perspectives

Un quintuplet disponible pour chaque phrase

- ▶ La sortie de Transcription automatique(src-asr)
- ▶ Le verbatim (srf-ref)
- ▶ Traduction automatique du verbatim (tgt-mt)
- ▶ Traduction automatique de la transcription automatique (tgt-slt)
- ▶ La post-édition de la traduction (tgt-pe)

Disponible en ligne :

<https://github.com/besacier/WCE-SLT-LIG>

L. Besacier, B.
Lecouteux & NQ.
Luong

Introduction

Mesures en TAP

RAP

TA

Corpus pour la
traduction de la
parole

Le corpus RAP

Le corpus RAP + TA

Exemple

Systèmes RAP et TA

Mesures de
confiance GETALP

RAP

TA

Performances

Redécodage de
graphe de
traduction

Travaux antérieurs

Notre approche

Résultats

Exemples de sorties

Conclusion et
perspectives

<i>src-ref</i>	quand	notre	cerveau	chauffe
<i>src-asr1</i>	<i>comme</i>	notre	cerveau	chauffe
labels RAP	B	G	G	G
<i>src-asr2</i>	<i>qu'</i>	<i>entre</i>	<i>serbes</i>	<i>au</i>
labels RAP	B	B	B	B G
<i>TA</i>	when	our	brains	<i>chauffe</i>
labels TA	G	G	G	B
<i>tgt-slt1</i>	<i>as</i>	our	brains	<i>chauffe</i>
labels TAP	B	G	G	B
<i>tgt-slt2</i>	<i>between</i>	<i>serbs</i>	<i>in</i>	<i>chauffe</i>
labels TAP	B	B	B	B
<i>tgt-pe</i>	when	our	brain	heats up

TABLE: Exemple de quintuplet avec étiquettes associées

Les mots sont étiquetés en utilisant TERp-A

Système RAP

- ▶ Basé sur KALDI (Povey et al., 2011)
- ▶ Modèle de langue 3-gramme (ESTER + Gigaword)
- ▶ Lexique 55K
- ▶ Modèles acoustiques SGMM
- ▶ Post-traitement pour la TA

Système de TA

- ▶ Basé sur MOSES (Koehn et al., 2007), phrase-based
- ▶ Modèle de langue 3-gramme
- ▶ Post-éditions disponibles
- ▶ Hypothèse forte : post-éditions initiales valables pour la sortie de Traduction de Parole

L. Besacier, B.
Lecouteux & NQ.
Luong

Introduction

Mesures en TAP

RAP

TA

Corpus pour la
traduction de la
parole

Le corpus RAP

Le corpus RAP + TA

Exemple

Systèmes RAP et TA

Mesures de
confiance GETALP

RAP

TA

Performances

Redécodage de
traçage de
traduction

Travaux antérieurs

Notre approche

Résultats

Exemples de sorties

Conclusion et
perspectives

Mesures de confiance RAP

- ▶ Traits acoustiques : durée
- ▶ Graphiques : nb de chemins, probabilités *a posteriori*
- ▶ Linguistiques : unigramme, 3-gramme, repli utilisé
- ▶ Étiquettes morpho-syntaxiques
- ▶ Classifieur utilisé : Bonzaiboost (Laurent et al. 2014)
- ▶ Estimées sur BREF 120

L. Besacier, B.
Lecouteux & NQ.
Luong

Introduction

Mesures en TAP

RAP

TA

Corpus pour la
traduction de la
parole

Le corpus RAP

Le corpus RAP + TA

Exemple

Systèmes RAP et TA

Mesures de
confiance GETALP

RAP

TA

Performances

Redécodage de
graphie de
traduction

Travaux antérieurs

Notre approche

Résultats

Exemples de sorties

Conclusion et
perspectives

Mesures de confiance TA

- ▶ Mot cible, séquence 2/3 gramme précédant le mot
- ▶ Mots source alignés avec la cible
- ▶ Contexte d'alignement
- ▶ Probabilité a posteriori du mot cible
- ▶ Topologie du graphe
- ▶ Comportement du repli (source et cible)
- ▶ Étiquettes morphosyntaxiques
- ▶ Analyse syntaxique
- ▶ Traits sémantiques
- ▶ Classifieur basé sur des CRFs : WAPITI (Lavergne et al., 2010)

L. Besacier, B.
Lecouteux & NQ.
Luong

Introduction

Mesures en TAP

RAP

TA

Corpus pour la
traduction de la
parole

Le corpus RAP

Le corpus RAP + TA

Exemple

Systèmes RAP et TA

Mesures de
confiance GETALP

RAP

TA

Performances

Redécodage de
graphe de
traduction

Travaux antérieurs

Notre approche

Résultats

Exemples de sorties

Conclusion et
perspectives

Tâche	RAP (WER)	TA (BLEU)	% G (correct)	% B (in- correct)
TA	0%	52.8%	82.5%	17.5%
TAP	26.6%	30.6%	65.5%	34.5%

TABLE: Performances de traduction de texte (TA) et de parole (TAP) sur notre corpus (2643 phrases)

Tâche	Est. conf. RAP	Est. conf. TA	Est. conf. TAP	Est. conf. TAP	Est. conf. TAP
Traits issus de	RAP	TA	TA	RAP	RAP+TA
G	87.85%	87.65%	77.17%	76.41%	77.54%
B	37.28%	42.29%	39.34%	38.00%	43.96%

TABLE: Résumé des performances de nos estimateurs de confiance (différentes tâches, différents traits utilisés)

Introduction

Mesures en TAP

RAP

TA

Corpus pour la
traduction de la
parole

Le corpus RAP

Le corpus RAP + TA

Exemple

Systèmes RAP et TA

Mesures de
confiance GETALP

RAP

TA

Performances

Redécodage de
graphe de
traduction

Travaux antérieurs

Notre approche

Résultats

Exemples de sorties

Conclusion et
perspectives

Travaux antérieurs

- ▶ Seconde passe de post-édition (Parton et al. 2012)
- ▶ Re-ordonnancement des N-meilleures hypothèses
- ▶ Redécodage de graphe basé sur :
 - ▶ Minimum Bayes Risk
 - ▶ Probabilités *a posteriori*
 - ▶ Autre modèle de langue

Notre approche

- ▶ Intégrer des mesures externes
- ▶ Redécodage par pénalité/récompense

Notre approche

- ▶ Récompense des chemins “corrects”
- ▶ Pénalité pour les chemins “incorrects”
- ▶ Estimateurs de confiance calculés sur les N-meilleures hypothèses

Peut être assimilé à une seconde passe

- ▶ Première passe : N-meilleures hypothèses
- ▶ Étiquetage des mots des hypothèses
- ▶ Réestimation des chemins du graphe

Introduction

Mesures en TAP

RAP

TA

Corpus pour la
traduction de la
parole

Le corpus RAP

Le corpus RAP + TA

Exemple

Systèmes RAP et TA

Mesures de
confiance GETALP

RAP

TA

Performances

Redécodage de
graphe de
traduction

Travaux antérieurs

Notre approche

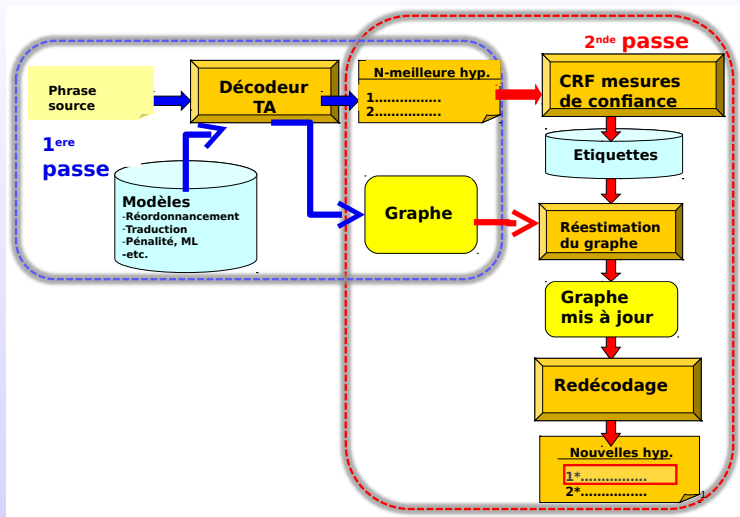
Résultats

Exemples de sorties

Conclusion et
perspectives

Introduction

Mesures en TAP

RAP
TACorpus pour la
traduction de la
paroleLe corpus RAP
Le corpus RAP + TA
Exemple
Systèmes RAP et TAMesures de
confiance GETALPRAP
TA
PerformancesRedécodage de
graphe de
traductionTravaux antérieurs
Notre approche
Résultats
Exemples de sortiesConclusion et
perspectives

H_k : hypothèses du graphe de recherche

c_{ij} : étiquette de qualité du j^{eme} mot de l'hypothèse T_i

$\#mots(H^*)$ est le nombre de mots cible dans H^*

α (>0) pondère l'impact de la pénalité ou de la récompense

$$transition'(H_k) = transition(H_k) + \begin{cases} recompense(t_{ij}) & \text{si } c_{ij} = G \\ penalite(t_{ij}) & \text{sinon} \end{cases} \quad (1)$$

La mise à jour des scores étant faite de la façon suivante :

$$penalite(t_{ij}) = -recompense(t_{ij}) = \alpha * \frac{score(H^*)}{\#mots(H^*)} \quad (2)$$

L. Besacier, B. Lecouteux & NQ. Luong

Introduction

Mesures en TAP

RAP

TA

Corpus pour la traduction de la parole

Le corpus RAP

Le corpus RAP + TA

Exemple

Systèmes RAP et TA

Mesures de confiance GETALP

RAP

TA

Performances

Redécodage de graphe de traduction

Travaux antérieurs

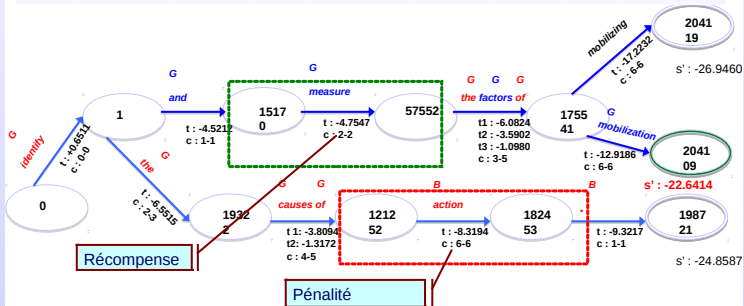
Notre approche

Résultats

Exemples de sorties

Conclusion et perspectives

Rang	Score	Hypotheses + étiquettes						
1	-29.9061	identify	the	cause	of	action	.	
		G	G	G	G	B	B	
2	-40.0868	identify	and	measure	the	factors	of	mobilization
		G	G	G	G	G	G	G



système	TAP base- line (BLEU)	TAP re- décodage (BLEU)	TAP re- décodage (BLEU)	TAP re- décodage (BLEU)
estimateurs conf.	aucun	RAP	TA	RAP+TA
<i>Perf.</i>	30.60%	31.12%	31.89%	32.82%

TABLE: Performance de TAP (BLEU) *avec* ou *sans* re-décodage de graphes - 2643 phrases

- ▶ Gains observés significatifs (p-valeur dans l'intervalle [0.00 ;0.01] pour le meilleur système 2 passes)
- ▶ Traits RAP et TA complémentaires
- ▶ Chacun des estimateurs apporte des gains

Introduction

Mesures en TAP

RAP
TACorpus pour la
traduction de la
paroleLe corpus RAP
Le corpus RAP + TA
Exemple
Systèmes RAP et TAMesures de
confiance GETALPRAP
TA
PerformancesRedécodage de
la phrase de
traductionTravaux antérieurs
Notre approche
Résultats
Exemples de sortiesConclusion et
perspectives

<i>src-ref1</i>	une démobilisation des employés peut déboucher sur une démoralisation mortifère
<i>src-asr1</i>	une démobilisation des employés peut déboucher sur une démoralisation mort y faire
<i>tgt-slt1</i> base-line	a demobilisation employees can lead to a penalty demoralisation
<i>tgt-slt1</i> avec redécodage	a demobilisation of employees can lead to a demoralization death
<i>tgt-pe1</i>	demobilization of employees can lead to a deadly demoralization

- ▶ Légère amélioration
- ▶ 'a' et 'penalty' étiquetés comme incorrects
- ▶ Émergence d'une nouvelle hypothèse
- ▶ Malheureusement : pas de correction de l'ASR (1 hypothèse)

Introduction

Mesures en TAP

RAP
TACorpus pour la
traduction de la
paroleLe corpus RAP
Le corpus RAP + TA
Exemple
Systèmes RAP et TAMesures de
confiance GETALPRAP
TA
PerformancesRedécodage de
graphie de
traductionTravaux antérieurs
Notre approche
Résultats
Exemples de sortiesConclusion et
perspectives

<i>src-ref2</i>	celui-ci a indiqué que l'intervention s' était parfaitement bien déroulée et que les examens post- opératoires étaient normaux
<i>src- asr2</i>	celui-ci a indiqué que l' intervention c' était parfaitement bien déroulés , et que les examens post opérateur étaient normaux .
<i>tgt-slt2</i> base- line	it has indicated that the speech that was well conducted , and that the tests were normal post route
<i>tgt-slt2</i> avec redéco- dage	he indicated that the intervention is very well done , and that the tests after operating were normal
<i>tgt-pe2</i>	he indicated that the operation went perfectly well and the post-operative tests were normal

- ▶ Légère amélioration
- ▶ 'it has', 'speech that was' et 'post route' étiquetés comme incorrects
- ▶ Émergence d'une nouvelle hypothèse : pronom correct et meilleure qualité en fin de phrase

Introduction

Mesures en TAP

RAP

TA

Corpus pour la
traduction de la
parole

Le corpus RAP

Le corpus RAP + TA

Exemple

Systèmes RAP et TA

Mesures de
confiance GETALP

RAP

TA

Performances

Redécodage de
grafe de
traduction

Travaux antérieurs

Notre approche

Résultats

Exemples de sorties

Conclusion et
perspectives

<i>src-ref3</i>	general motors repousse jusqu'en janvier le plan pour opel
<i>src-asr3</i>	general motors repousse jusqu' en janvier le plan pour open
<i>tgt-slt3</i> base- line	general motors postponed until january the plan to open
<i>tgt-slt3</i> avec redéco- dage	general motors puts until january terms to open
<i>tgt-pe3</i>	general motors postponed until january the plan for opel

- ▶ Légère dégradation
- ▶ 'to open' bien étiqueté comme incorrect
- ▶ Émergence d'une nouvelle hypothèse encore plus mauvaise !

Conclusion

- ▶ Les mesures de confiance peuvent améliorer le décodage
- ▶ Gains significatifs
- ▶ Estimateurs joints plus performants
- ▶ Constitution d'un corpus mis à disposition pour la communauté

Perspectives

- ▶ Proposer une tâche d'estimation de confiance pour la TAP (IWSLT)
- ▶ Entraîner un classifieur à partir de traits joints
- ▶ Utiliser les mesures dans un scénario interactif
- ▶ Extension du corpus