



# Word2Vec vs DBnary ou comment (ré)concilier représentations distribuées et réseaux lexico-sémantiques ? Le cas de l'évaluation en traduction automatique

Christophe SERVAN, Zied ELLOUMI, Hervé BLANCHON et Laurent BESACIER

Université Grenoble Alpes - LIG - GETALP

Le 6 Juin 2016

TALN 2016 - INALCO, Paris

# Plan



*C. Servan & al.*

*Word2Vec vs DBnary ou comment (ré)concilier représentations distribuées et réseaux lexico-sémantiques ? Le cas de l'évaluation en traduction automatique*

- ▶ **Introduction** (*Word2Vec vs DBnary*)
- ▶ **Approche** (*Comment réconcilier représentations distribuées et réseaux lexico-sémantiques*)
  - ▶ Augmentation de la métrique METEOR
- ▶ **Experiences** (*Le cas de l'évaluation en traduction automatique*)
  - ▶ Corrélations avec jugements humains
- ▶ **Conclusions & Perspectives**



# Introduction

# Résumé de l'article



- ▶ Ressources lexicales vs Représentations distribuées
- ▶ ... pour mesurer la similarité entre phrases
  - ▶ Synonymie, proximité morphologique, etc.
- ▶ Le cas de l'évaluation en traduction automatique
  - ▶ Problème ouvert
- ▶ Augmentation de la métrique METEOR
  - ▶ Meteor augmenté par module *DBnary* (ressource lexicale)
  - ▶ Meteor augmenté par module *Vecteurs* (représentation distribuée)
  - ▶ Evaluation sur la tâche « *metrics* » de WMT 2014
  - ▶ Analyse des corrélations avec jugements humains
  - ▶ Code mis à disposition sur *github*

# Word2Vec vs Dbnary ...



## Ressources lexicales (lexico-sémantiques)

- ▶ Wordnet
  - ▶ [Miller, 1995 ; Fellbaum, 1998]
- ▶ Babelnet
  - ▶ [Navigli, 2010]
- ▶ DBnary
  - ▶ [Sérasset, 2012]

## Représentations distribuées

- ▶ Word2Vec
  - ▶ [Mikolov et al., 2013; Le & Mikolov, 2014]
- ▶ ... et ses extensions
  - ▶ Bivec
    - ▶ [Luong et al., 2015]
  - ▶ Multivec
    - ▶ [Berard et al., 2016]
- ▶ GloVe
  - ▶ [Pennington & al, 2014]

# Word2Vec vs Dbnary ...



## Ressources lexicales (lexico-sémantiques)

- ▶ « *Beautiful* » data
- ▶ Interposables par un humain
- ▶ Création couteuse
  - ▶ Nouveau domaine, nouvelle langue
- ▶ Granularité fine

## Représentations distribuées

- ▶ « *Big* » data
- ▶ Adaptables
  - ▶ Re-entraînement possible sur nouveaux corpus
- ▶ Plus difficiles à interpréter
- ▶ Distingue usage et contextes des mots ... plutôt que leur sens
- ▶ Gros grain [Pachenko, 2016]

# Associer “*beautiful data*” et “*big data*”

- ▶ Raffiner l'apprentissage de représentations distribuées en utilisant des ressources lexicales
  - ▶ [Faruqui et al., 2014] *Retrofitting word vectors to semantic lexicons*. NAACL 2015
- ▶ Représentations distribuées de sens
  - ▶ [Panchenko, 2016] *Best of both worlds : making word sense embeddings interpretable*. LREC 2016
  - ▶ [Rothe and Schütze, 2015] *Autoextend : Extending word embeddings to embeddings for synsets and lexemes*. ACL 2015
- ▶ Cet article
  - ▶ Comparer et associer ressources lexicales et représentations distribuées
  - ▶ Pour une meilleur métrique d'évaluation en traduction automatique



# Augmentation de métrique

Appliqué au METEOR



# Le Score BLEU [Papineni, 2001]

- ▶ Métrique la plus utilisée
- ▶ Précision n-grammes (taille 4)  $P_n$
- ▶ Pondération par une pénalité de Brièveté
- ▶ Inconvénients :
  - ▶ Pas de prise en compte morphologique
  - ▶ Pas de sémantique
  - ▶ Très controversée...

$$BP = \begin{cases} 1 & \text{si } c > r \\ \exp(1 - r/c) & \text{si } c < r \end{cases}$$

$$BLEU = BP \cdot \exp\left(\sum_{n=1}^N w_n \log P_n\right)$$

# Le Score METEOR [Banerjee & Lavie, 2005]



- ▶ Pallier aux inconvénients de BLEU
- ▶ Métrique mieux corrélée avec le jugement humain
- ▶ Intègre des « *modules* » d'adéquation entre :
  - ▶ des formes de surfaces (module « exact »)
  - ▶ des racines des mots (module « stem »)
    - ⇒ Utilisation d'un extracteur de racine de mots (ou stemmer)
  - ▶ des synonymes (module « synonym »)
    - ⇒ Ressource lexico-sémantique : WordNet (**anglais seulement**)

# Extensions de METEOR

- ▶ METEOR-Next [Denkowski & Lavie, 2010] :
  - ▶ Correlation avec le HTER [Snover & al., 2006]
  - ▶ Ajout d'un module de « Paraphrase » (base de donnée de paraphrases)
- ▶ METEOR-Universal [Denkowski & Lavie, 2014] :
  - ▶ Ressource de type « paraphrase » : anglais, tchèque, français, allemand et espagnol
  - ▶ Optimisation des poids de paramètres pour maximiser corrélation avec le jugement humain
- ▶ METEOR WSD [Apidianaki & Marie, 2015] :
  - ▶ Filtrage des synonymes selon leur sens (*BabelFly*)



# Augmentation de METEOR :

## Ajout d'une base lexicale [Elloumi & al., 2015]

- ▶ Utilisation de WordNet compliquée
- ▶ Utilisation de DBnary
  - ▶ Disponibilité dans 17 langues
  - ▶ Extraction de synsets (lemme - synonymes)
  - ▶ Extraction des lemmes grâce à TreeTagger
- ⇒ Existe pour le français, l'allemand, l'espagnol et le russe
- ⇒ N'existe pas pour toutes les langues

# Augmentation de METEOR :

## Ajout d'une base lexicale [Elloumi & al., 2015]

- ▶ Utilisation de WordNet compliquée
- ▶ Utilisation de DBnary
  - ▶ Disponibilité dans 17 langues
  - ▶ Extraction de synsets (lemme - synonymes)
  - ▶ Extraction des lemmes grâce à TreeTagger
- ⇒ Existe pour le français, l'allemand, l'espagnol et le russe
- ⇒ N'existe pas pour toutes les langues
- ⇒ Solution possible : inférer ces données ?

# Augmentation de METEOR :

## Utilisation d'une représentation vectorielle

- ▶ Induction non-supervisée de relations syntaxiques et sémantiques [Blacoe & Lapata, 2012; Mikolov et al., 2013]
  - ▶ Représentation utilisant le contexte du mot
    - ▶ « *Dis moi qui sont tes amis, je te dirais qui tu es* »
  
- ▶ Utilisation du toolkit *MultiVec* [Bérard & al, 2016]

# Augmentation de METEOR :

## Utilisation d'une représentation vectorielle

Création d'un module « vecteurs » :

- ▶ Score issu de la similarité cosinus & normalisé  $S_{cos} \in [0;1]$
- ▶ Seuil de décision  $Th$  tel que :
  - ▶ Si  $S_{cos} \geq Th$ , alors les mots sont alignés
  - ▶ Seuil par défaut 0,80



# Augmentation de METEOR :

## Utilisation d'une représentation vectorielle

Création d'un module « vecteurs » :

- ▶ Score issu de la similarité cosinus & normalisé  $S_{cos} \in [0;1]$
- ▶ Seuil de décision  $Th$  tel que :
  - ▶ Si  $S_{cos} \geq Th$ , alors les mots sont alignés
  - ▶ Seuil par défaut 0,80

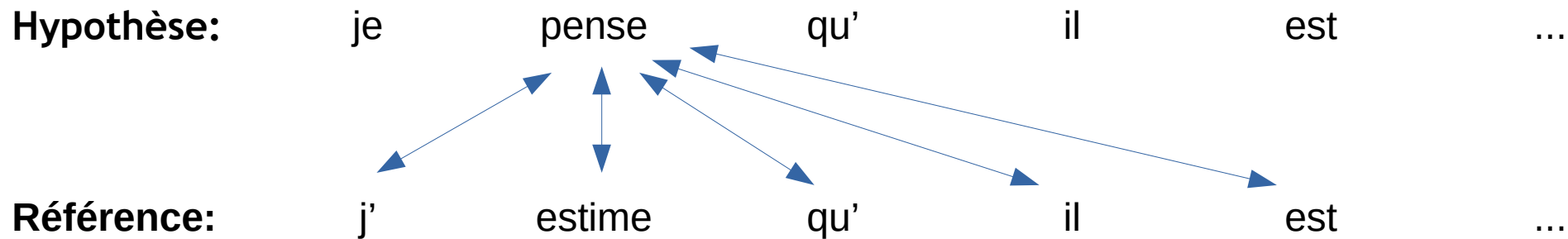
**Hypothèse:**      je            pense            qu'            il            est            ...

**Référence:**      j'            estime            qu'            il            est            ...

# Augmentation de METEOR : Utilisation d'une représentation vectorielle

Création d'un module « vecteurs » :

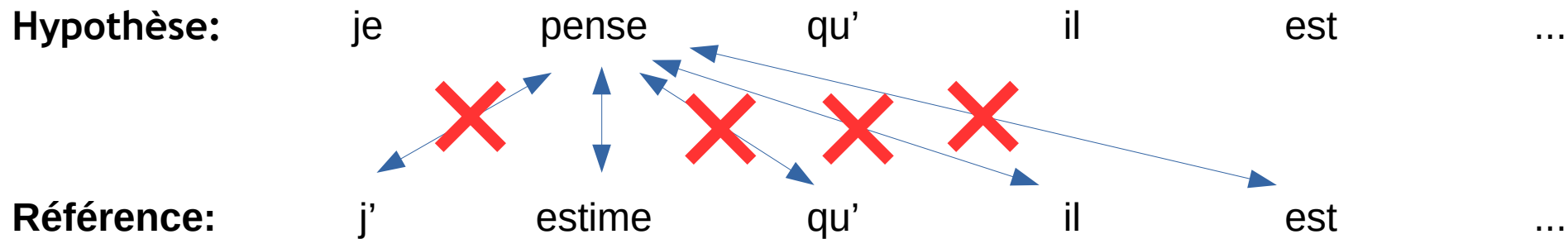
- ▶ Score issu de la similarité cosinus & normalisé  $S_{cos} \in [0;1]$
- ▶ Seuil de décision  $Th$  tel que :
  - ▶ Si  $S_{cos} \geq Th$ , alors les mots sont alignés
  - ▶ Seuil par défaut 0,80



# Augmentation de METEOR : Utilisation d'une représentation vectorielle

Création d'un module « vecteurs » :

- ▶ Score issu de la similarité cosinus & normalisé  $S_{cos} \in [0;1]$
- ▶ Seuil de décision  $Th$  tel que :
  - ▶ Si  $S_{cos} \geq Th$ , alors les mots sont alignés
  - ▶ Seuil par défaut 0,80



# Augmentation de METEOR :

## Utilisation d'une représentation vectorielle

Création d'un module « vecteurs » :

- ▶ Score issu de la similarité cosinus & normalisé  $S_{cos} \in [0;1]$
- ▶ Seuil de décision  $Th$  tel que :
  - ▶ Si  $S_{cos} \geq Th$ , alors les mots sont alignés
  - ▶ Seuil par défaut 0,80

<b>Hypothèse:</b>	je	pense	qu'	il	est	...
<b>Référence:</b>	j'	estime	qu'	il	est	...



# Protocole



- ▶ Comparaison du score de METEOR avec le jugement humain :
  - ▶ Metric shared task [Machacek & Bojar, 2014]
  - ▶ Couples de langues choisis
    - ▶ anglais-français, anglais-allemand, anglais-russe (dans les 2 sens)
- ▶ Mesure utilisée :
  - ▶ Corrélation entre le score de METEOR avec le jugement humain
  - ▶ Coefficient Tau ( $\tau$ ) de Kendall au niveau de la phrase

# Apprentissage des représentations distribuées

- ▶ Apprentissage sur corpus monolingue de la tâche WMT14

Langue	corpus	# de lignes	# de mots
anglais	Europarl V7 + news commentary V10	2,2 M	60 M
français	Europarl V7 + news commentary V10	2,2 M	67 M
allemand	Europarl V7 + news commentary V10	2,1 M	57 M
russe	Common Crawl + news commentary V10 + News Crawl 2013 (10%)	2,4 M	50 M

- ▶ Taille des vecteurs : 50
- ▶ Fenêtre considérée :  $\pm 5$
- ▶ Temps d'apprentissage  $\approx$  15 min

# Tests contrastifs



- ▶ **METEOR Baseline** : modules *Exact*, *Stem*, *Synonymes* et *Paraphrase* pour l'anglais et utilisation des modules *Exact*, *Stem* et *Paraphrase* pour les autres langues ;
- ▶ **METEOR DBnary** : utilisation du module *Synonymes* « *DBnary* » plutôt que celles issues de WordNet [Elloumi & al.,2015]
- ▶ **METEOR Vecteurs** : les modules *Stem* et *Synonymes* sont remplacés par le module *Vecteurs* utilisant les représentations distribuées ;
- ▶ **METEOR Baseline + Vecteurs** : le *METEOR Baseline* est enrichi avec le module *Vecteurs* ;
- ▶ **METEOR DBnary + Vecteurs** : le *METEOR DBnary* est enrichi avec le module *Vecteurs*.



# Résultats



## ► Scores de METEOR obtenus sur une tâche fr-en

Métriques	Systèmes				
	online A	online B	online C	rbmt 1	rbmt 4
<i>METEOR Baseline</i>	36,33	36,71	31,19	33,00	31,65
<i>METEOR DBnary</i>	36,93	37,33	32,01	33,69	32,42
<i>METEOR Vecteurs (th=0,80)</i>	37,00	37,34	31,87	33,67	32,34
<i>METEOR Baseline + Vecteurs</i>	37,08	37,40	31,96	33,75	32,45
<i>METEOR DBnary + Vecteurs</i>	37,53	37,88	32,60	34,32	33,05

# Corrélation avec le jugement humain

## ► Scores de corrélation Tau de Kendall

Langues	fr,de,ru-en		en-fr		en-de		en-ru	
	Seuil	Tau	Seuil	Tau	Seuil	Tau	Seuil	Tau
<i>METEOR Baseline</i>	–	0,356	–	0,280	–	0,238	–	0,427
<i>METEOR DBnary</i>	–	0,357	–	0,284	–	<b>0,239</b>	–	0,435
<i>METEOR Vecteurs</i>	0,80	0,355	0,80	<b>0,282</b>	0,80	0,235	0,80	0,420
<i>METEOR Baseline + Vecteurs</i>	0,80	0,359	0,80	0,280	0,80	0,238	0,80	0,428
<i>METEOR DBnary + Vecteurs</i>	0,80	<b>0,360</b>	0,80	0,281	0,80	0,238	0,80	<b>0,436</b>
<i>METEOR Vecteurs</i>	0,79	0,357	0,72	<b>0,287</b>	0,70	0,236	0,90	0,421
<i>METEOR Baseline + Vecteurs</i>	0,82	0,360	0,85	<b>0,287</b>	0,88	<b>0,240</b>	0,83	0,429
<i>METEOR DBnary + Vecteurs</i>	0,78	<b>0,362</b>	0,91	0,285	0,77	<b>0,241</b>	0,84	<b>0,437</b>

# Analyse des appariements trouvés

## ▶ Exemples d'appariement avec *METEOR Exact*:

Hypothèse : je pense qu' il est concevable que ces données soient employées pour le bénéfice mutuel .

Référence : j' estime qu' il est concevable que ces données soient utilisées dans leur intérêt mutuel .

Hypothèse : le créateur de SAS disait il faisait un genre du feuilleton géopolitique .

Référence : le père de SAS disait faire un genre de feuilleton géopolitique .

# Analyse des appariements trouvés

▶ Exemples d'appariement avec *METEOR Exact* + **DBnary**:

Hypothèse : je pense qu' il est concevable que ces données soient **employées** **pour** le bénéfice mutuel .

Référence : j' estime qu' il est concevable que ces données soient **utilisées** **dans** leur intérêt mutuel .

Hypothèse : le **créateur** de SAS disait il **faisait** un genre du feuilleton géopolitique .

Référence : le **père** de SAS disait **faire** un genre de feuilleton géopolitique .

# Analyse des appariements trouvés

► Exemples d'appariement avec *METEOR Exact* + **DBnary** + **Vecteurs** :

Hypothèse : je pense qu' il est concevable que ces données soient employées pour le bénéfice mutuel .  
 Référence : j' estime qu' il est concevable que ces données soient utilisées dans leur intérêt mutuel .

Hypothèse : le créateur de SAS disait il faisait un genre du feuilleton géopolitique .  
 Référence : le père de SAS disait faire un genre de feuilleton géopolitique .

# Analyse des appariements trouvés

- ▶ Exemples d'appariement avec *METEOR Exact* + *DBnary* + *Vecteurs* :

Hypothèse : je pense qu' il est concevable que ces données soient employées pour le bénéfice mutuel .  
 Référence : j' estime qu' il est concevable que ces données soient utilisées dans leur intérêt mutuel .

Hypothèse : le créateur de SAS disait il faisait un genre du feuilleton géopolitique .  
 Référence : le père de SAS disait faire un genre de feuilleton géopolitique .

- ▶ Appariements non trouvés

# Analyse des appariements trouvés

▶ Exemples d'appariement avec *METEOR Exact* + Vecteurs :

Hypothèse : je pense qu' il est concevable que ces données soient employées pour le bénéfice mutuel .  
 Référence : j' estime qu' il est concevable que ces données soient utilisées dans leur intérêt mutuel .

Hypothèse : le créateur de SAS disait il faisait un genre du feuilleton géopolitique .  
 Référence : le père de SAS disait faire un genre de feuilleton géopolitique .

▶ Appariements non trouvés

# Analyse des appariements trouvés

▶ Exemples d'appariement avec *METEOR Exact* + Vecteurs :

Hypothèse : je pense qu' il est concevable que ces données soient employées pour le bénéfice mutuel .  
 Référence : j' estime qu' il est concevable que ces données soient utilisées dans leur intérêt mutuel .

Hypothèse : le créateur de SAS disait il faisait un genre du feuilleton géopolitique .  
 Référence : le père de SAS disait faire un genre de feuilleton géopolitique .

▶ Appariements trouvés en abaissant le seuil de décision





# Conclusions & Perspectives

# Conclusions & Perspectives



## Conclusions

- ▶ Nouvelle métrique augmentée pour l'évaluation de la TAS
- ▶ Représentations distribuées peuvent pallier au manque de données
- ▶ Léger intérêt de la combinaison *Dbnary* et *Vecteurs* (Word Embeddings + Données Lexico-sémantiques)
- ▶ Outil disponible en ligne : <https://github.com/cservan/METEOR-E>

## Perspectives

- ▶ Extension à d'autres métriques [Le & al., 2016]
- ▶ Utilisation dans les mesures de confiance [Le & al., 2016]
- ▶ Corrélation entre l'utilisabilité de la traduction et le temps de post-édition
- ▶ Création de représentations adaptées à la tâche



# Merci de votre attention

Questions ?

<https://github.com/cservan/METEOR-E>