

# METEOR FOR MULTIPLE TARGET LANGUAGES USING DBNARY

---

Zied Elloumi, Hervé Blanchon

Gilles Sérasset & Laurent Besacier

*premon.nom@imag.fr*

LIG – GETALP – Univ. Grenoble Alpes



# Outline

---

## **Situation**

 METEOR, WordNet, **Dbnary**

## **“Dbnary Synsets” Extraction**

## **METEOR Scores on English**

 WordNet vs Dbnary Synsets

## **Correlation with human judgment**

 METEOR without Synset vs METEOR with “Dbnary Synsets”

## **Conclusion and Perspectives**

# SITUATION

# METEOR

---

 Introduced by (*Banerjee and Lavie, 2005*)

- ▶ to overcome several weaknesses of BLEU (*Papineni, 2002*) and NIST (*Doddington, 2002*)
- ▶ to better correlate with human judgment

 A 3-leveled mapping approach  
between a MT Hypothesis and one or several References

 **surface forms overlap** of words

 **stems (lemma) overlap** of surface forms

 tool: a stemmer (lemmatizer) for the language

 **synonymy overlap** through shared **WordNet Synsets**

 resource: a WordNet for the language

# METEOR Recent Extensions

---


## METEOR-NEXT (*Denkowski and Lavie, 2010a*)


➤ to better correlate with HTER (*Snover & al., 2006*)

 a 4<sup>th</sup> mapping level to accommodate **multi-word matches**

 resource: a paraphrase database for the language

## METEOR Universal (*Denkowski and Lavie 2014*)


 tool: automatic extraction of paraphrase tables and function word lists from bitexts

 resources: paraphrase tables for English, Arabic, Czech, French, German, Spanish

 parameter set (learned from human judgments)






## METEOR-WSD (*Apidianaki and Marie, 2015*)

➤ to filter synonyms/paraphrases according to word senses

 English references further disambiguated and annotated using Babelify (*Moro et al., 2014*)


# WordNet

---

-  A large lexical database for English  
(*Fellbaum, 1998*)
-  WordNet links **nouns**, **verbs**, **adjectives** and **adverbs** to sets of cognitive synonyms (Synsets)
-  Different versions of WordNet in other languages  
(Arabic, French, ...)
  -  **pro:** important and a very useful resources
  -  **cons:** not free and/or not available for every language

# METEOR & WordNet

## Pro

-  synonym match increases the chance of the MT output words to match the reference words

## Cons

-  synonym match available only for English

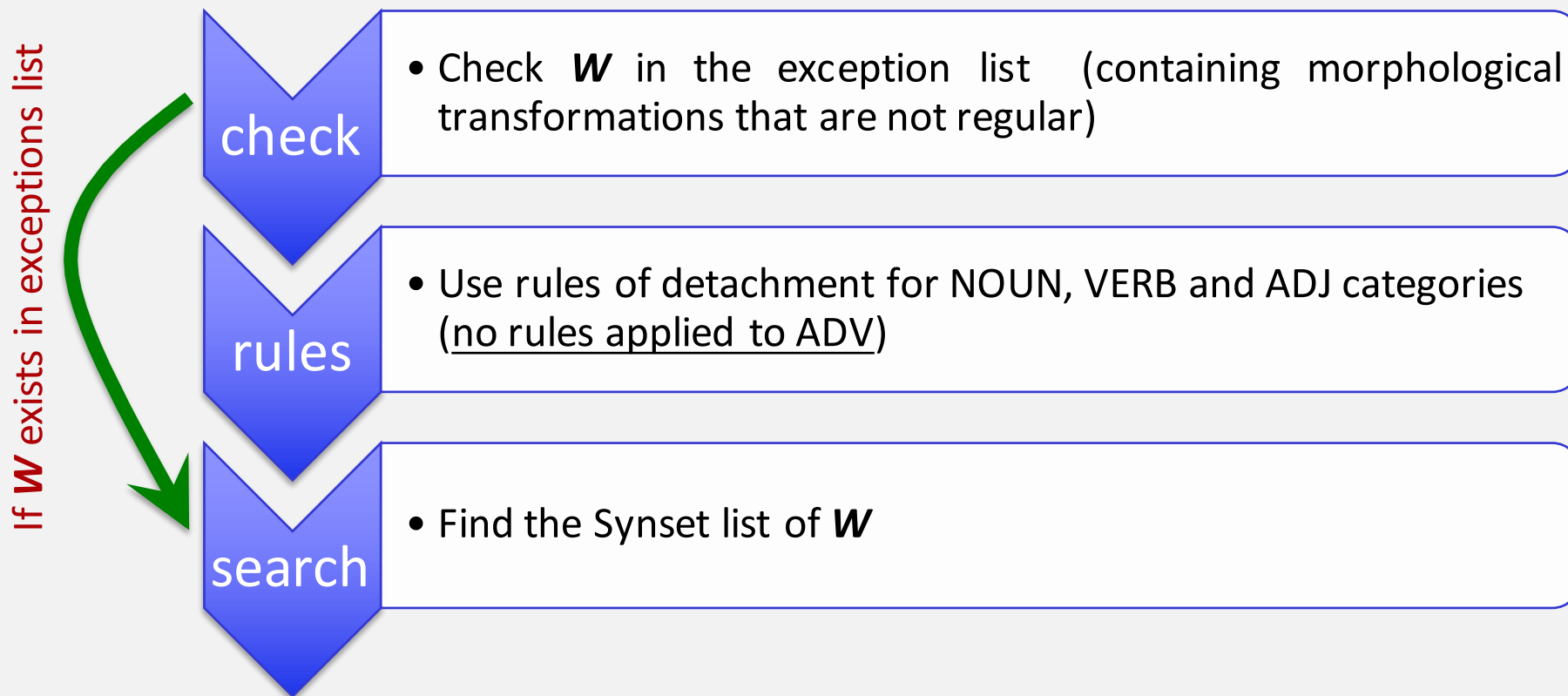
 Latest version of WordNet 3.0 = 117 659 Synsets

Categories	# of Synsets
Verb	13 767
Noun	82 115
Adverb	3 621
Adjective	18 156

**Table 1.** Number of Synsets in WordNet

# METEOR & WordNet




- METEOR uses the **Morphy-7WN** function from WordNet to lemmatize forms
- **Morphy-7WN** uses a two-step process to find lemma of a particular word **W**









# Dbnary (<http://kaiko.getalp.org/about-dbnary/>)

## What is it?

-  a multilingual lexical resource in RDF (*Klyne & Carroll, 2004*) collected at the LIG (*Sérasset, 2015*) and extracted from Wiktionary (currently 21 languages editions)
-  the lexical data is made available as LLOD (Linguistic Linked Open Data)
-  the lexicon structure is defined using the LEMON vocabulary (*McCrae et al., 2011*)

## Availability

-  downloadable files
  -  queried locally using SPARQL
-  Linked Open Data directly accessible to browsers or applications
  -  queried online using SPARQL

### Wiktionary

the dictionary counterpart of Wikipedia

### LEMON

a model for modeling lexicon and machine-readable dictionaries linked to the Semantic Web and the Linked Data cloud

### SPARQL

a standard language for querying linked data

# Dbnary: the dataset

---

## Core data

### **Lexical Entries, Lexical Senses and Translations**

## Additional data

### Semantically enriched Relations


 **Translations:** attached to their source Lexical Sense when possible

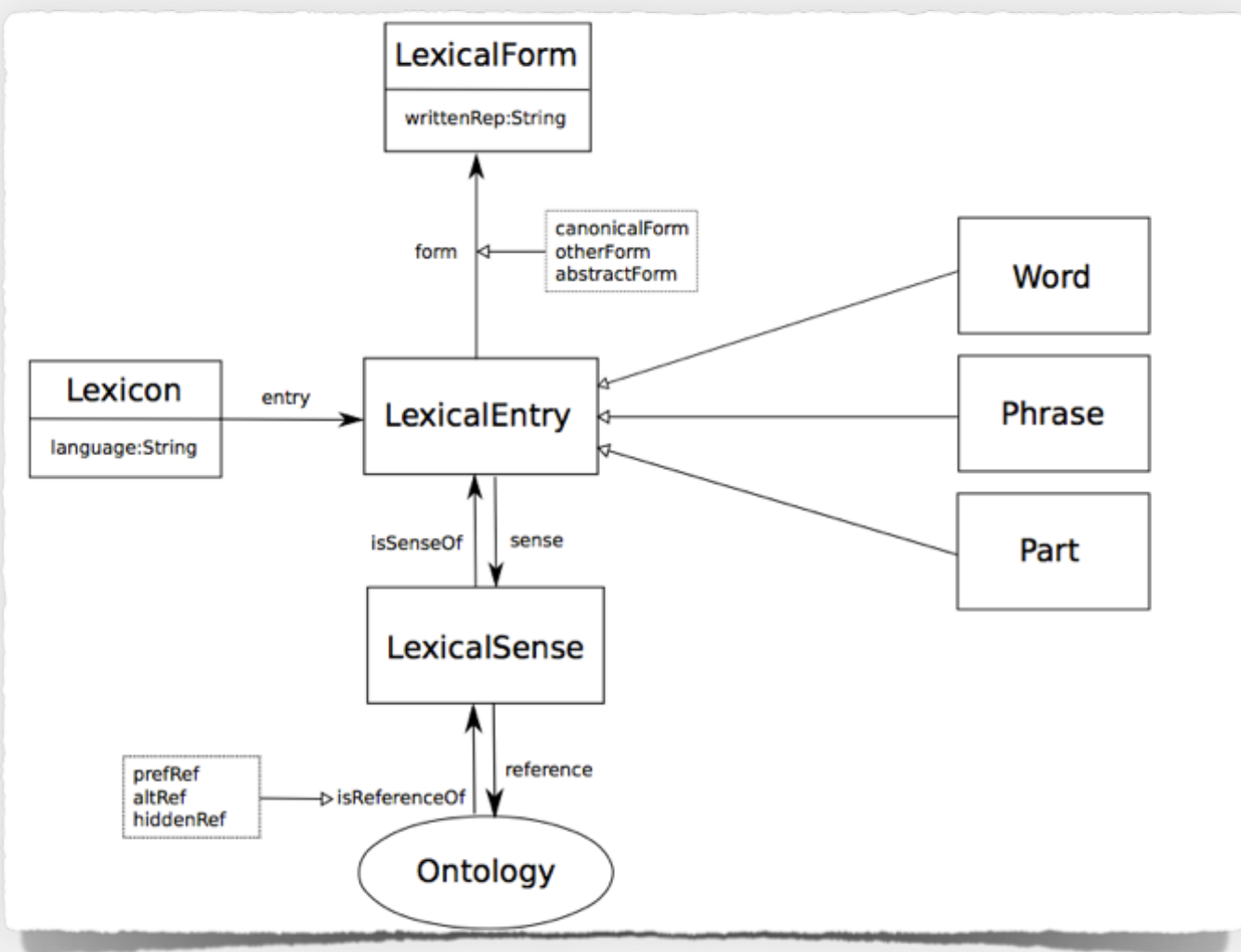
 **Lexico-semantic relations:** also attached to their source Lexical Sense

 **syno/anto-nymy**, hypo/hyper-nymy

 mero/holo-nymy, tropo-nymy

### Morphology

 Extensive representation of morphology  
(a set of “lemon:otherForm”)



## LEMON

A quick overview

# Dbnary example: entry *chat* in French

 <http://kaiko.getalp.org/dbnary/fra/chat>

**About: dbnary-fra:chat** [Goto](#) [Sponge](#) [Permalink](#)  
An Entity of Type : dbnary:Vocable, within Data Space : kaiko.getalp.org associated with source dataset(s)  
Type:

---

Attributes	Values	
<u>rdf:type</u>	<u>dbnary:Vocable</u>	
<u>dbnary:refersTo</u>	<u>dbnary-fra:chat_nom_1</u> <u>dbnary-fra:chat_nom_2</u> <u>dbnary-fra:chat_nom_3</u>	sense families
is <u>dbnary:synonym</u> of	<u>dbnary-fra: ws   clavardage_nom_1</u> <u>dbnary-fra: ws   palatine_nom_1</u> <u>dbnary-fra: ws   jeu du loup_nom_1</u>	synonyms
is <u>dbnary:hypernym</u> of	<u>dbnary-fra:chat_sauvage_nom_1</u> <u>dbnary-fra: ws   matou_nom_1</u>	hypernyms
is <u>dbnary:hyponym</u> of	<u>dbnary-fra: ws   animal de compagnie_nom_1</u> <u>dbnary-fra: ws   Félinés_nom_1</u>	hyponyms

fr chat → en cat; fr chat → en chat

**About: dbnary-fra:chat\_nom\_1** [Goto](#) [Sponge](#) [Permalink](#)  
 An Entity of Type : [lemon:Word](#), within Data Space : [kaiko.getalp.org](#) associated with source [dataset\(s\)](#)  
 Type:

Attributes	Values
<a href="#">rdf:type</a>	<a href="#">lemon:LexicalEntry</a> <a href="#">lemon:Word</a>
<a href="#">dcterms:language</a>	<a href="#">lexvo:fra</a>
<a href="#">lemon:language</a>	fr
<a href="#">dbnary:partOfSpeech</a>	-nom-
<a href="#">dbnary:synonym</a>	<a href="#">dbnary-fra:chat_domestique</a> <a href="#">dbnary-fra:minet</a> <a href="#">dbnary-fra:greffier</a> <a href="#">dbnary-fra:Grippeminaud</a> <a href="#">dbnary-fra:Raminagrobis</a> »more»
<a href="#">lemon:canonicalForm</a>	<a href="#">nodelD://b4173437</a>
<a href="#">lemon:sense</a>	<a href="#">dbnary-fra: ws_6_chat_nom_1</a> <a href="#">dbnary-fra: ws_2_chat_nom_1</a> <a href="#">dbnary-fra: ws_5_chat_nom_1</a> <a href="#">dbnary-fra: ws_3_chat_nom_1</a> <a href="#">dbnary-fra: ws_7_chat_nom_1</a> »more»
<a href="#">lexinfo:partOfSpeech</a>	<a href="#">lexinfo:noun</a>
<a href="#">dbnary:hypernym</a>	<a href="#">dbnary-fra:félicité</a>
<a href="#">dbnary:hyponym</a>	<a href="#">dbnary-fra:chat_domestique</a> <a href="#">dbnary-fra:chat-tigre_du_Bengale</a> <a href="#">dbnary-fra:chat_sauvage</a> <a href="#">dbnary-fra:chat_des_pampas</a> <a href="#">dbnary-fra:chat-tigre</a> »more»
<a href="#">lemon:lexicalVariant</a>	<a href="#">nodelD://b4173438</a>
<a href="#">is dbnary:isTranslationOf of</a>	<a href="#">dbnary-fra: tr_bul_4_chat_nom_1</a> <a href="#">dbnary-fra: tr_ind_1_chat_nom_1</a> <a href="#">dbnary-fra: tr_ces_2_chat_nom_1</a> <a href="#">dbnary-fra: tr_lat_5_chat_nom_1</a> <a href="#">dbnary-fra: tr_ron_5_chat_nom_1</a> »more»

*domestic cat*

senses

**About: dbnary-fra:chat\_nom\_3** [Goto](#) [Sponge](#) [Permalink](#)  
 An Entity of Type : [lemon:Word](#), within Data Space : [kaiko.getalp.org](#) associated with source [dataset\(s\)](#)  
 Type:


Attributes	Values
<a href="#">rdf:type</a>	<a href="#">lemon:LexicalEntry</a> <a href="#">lemon:Word</a>
<a href="#">dcterms:language</a>	<a href="#">lexvo:fra</a>
<a href="#">lemon:language</a>	fr
<a href="#">dbnary:partOfSpeech</a>	-nom-
<a href="#">dbnary:synonym</a>	<a href="#">dbnary-fra:causette</a> <a href="#">dbnary-fra:clavardage</a> <a href="#">dbnary-fra:tchatche</a>
<a href="#">lemon:canonicalForm</a>	<a href="#">nodelD://b4174235</a>
<a href="#">lemon:sense</a>	<a href="#">dbnary-fra: ws_12_chat_nom_3</a> <a href="#">dbnary-fra: ws_13_chat_nom_3</a>
<a href="#">lexinfo:partOfSpeech</a>	<a href="#">lexinfo:noun</a>
<a href="#">lemon:lexicalVariant</a>	<a href="#">nodelD://b4174236</a> <a href="#">nodelD://b4174237</a>
<a href="#">lemon:otherForm</a>	<a href="#">nodelD://b5363181</a>
<a href="#">is dbnary:refersTo of</a>	<a href="#">dbnary-fra:chat</a>

*online conversation*

senses

# Dbnary: a source of Synsets for METEOR?

## The big picture

-  21 languages
-  2.9M lexical entries (pos, canonical form, +{ })
  -  divided into 2.5M senses (def, example)
-  4.9M translations (from 21 languages)

## We will consider the following languages

	English	French	Russian	German	Spanish
# of entries	620,369	322,018	185,910	104,505	86,388
# of senses	498,415	416,323	176,335	116,290	126,411
#of synonyms	35,437	36,019	31,345	33,282	21,024

**Table 2.** Number of entries, senses, and synonyms in Dbnary for the target languages considered in this study.

# SYNSET EXTRACTION FROM DBNARY

# Querying Dbnary

- SPARQL queries to extract every synonym (**?s**) in the Dbnary database for each word (**?w**) in a specific **language**

```
SELECT distinct ?w ?s
WHERE { ?s dbnary:synonym ?w.
        ?w dbnary:refersTo ?le.
        ?le lemon:language 'en'.}
```

## Example

**?w = "cut"**

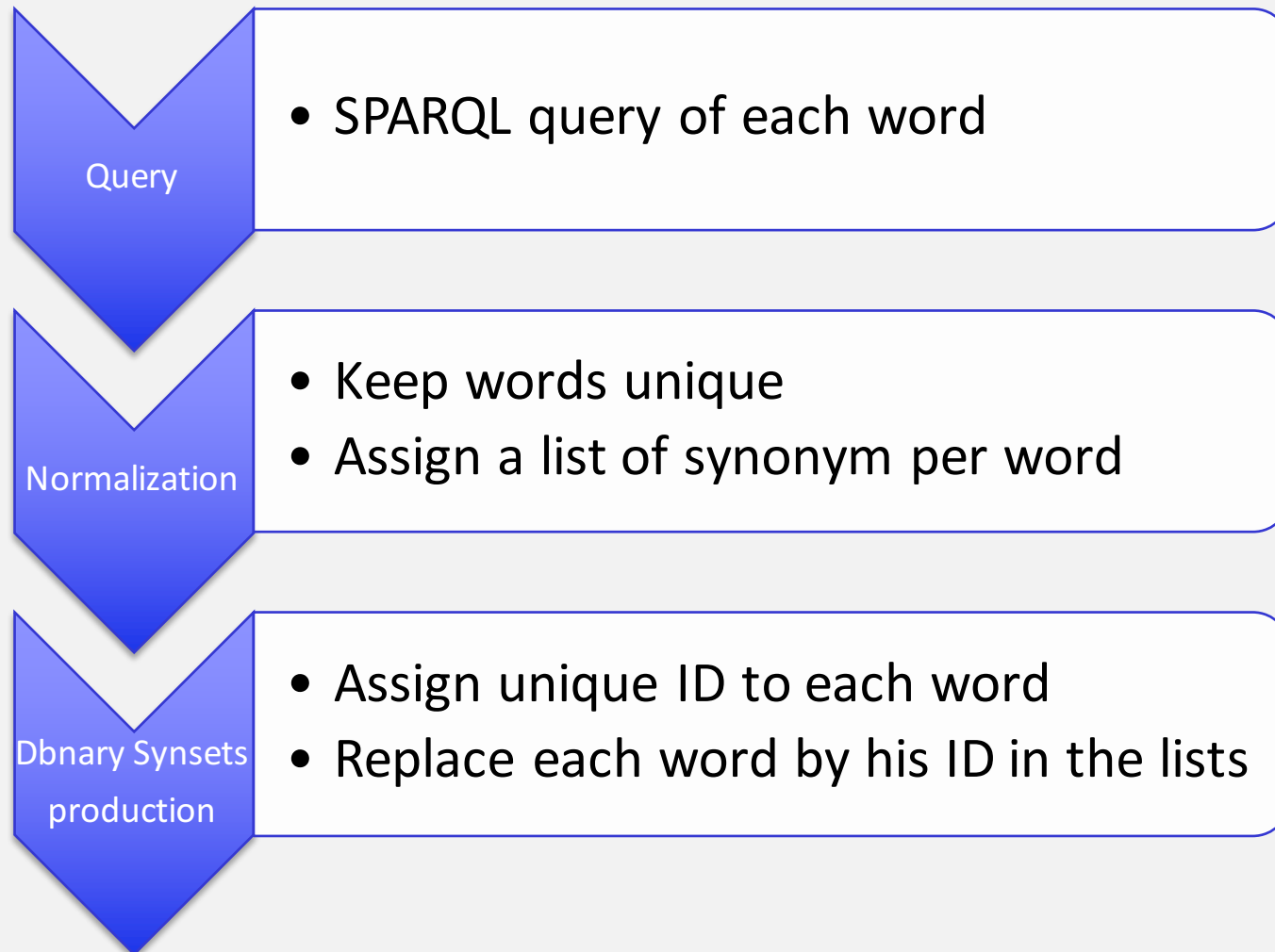
lower	reduce	juice	decrease
vigorish	decrease	ripped	cutting



# Producing the Synsets

---

## Produce *a la* WordNet Synsets from Dbnary



# 2 dictionaries of synonyms

## DB-4-catg

 with the 4 WordNet categories: Verb, Noun, Adverb, Adjective

## DB-all-catg

 with all the existing categories in Dbnary

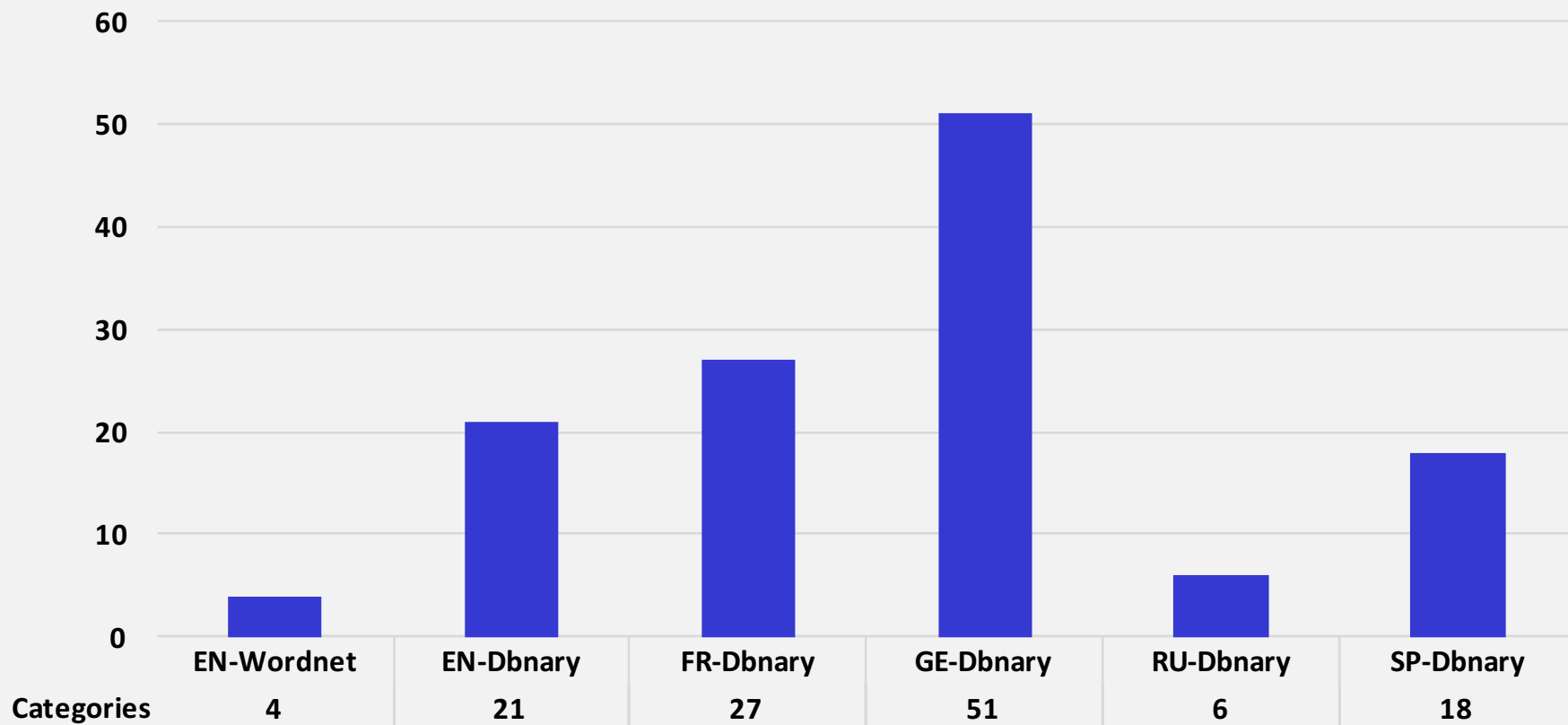
category		
Noun	Phrase	Proverb
Adjective	Suffix	Numeral
Verb	Pronoun	Determiner
Adverb	Prep_phr	Symbol
Proper_noun	Conj	Card_num
Interjection	Prefix	Infix
Preposition	Particle	Idiom

**Table 3.** All category extracted from Dbnary for English

# # of Dbnary categories/language

 # of categories for the languages considered

 English, French, German, Russian, and Spanish



Scores comparison with reported results of WMT14 on French-**English**

- ✓ WordNet original Synsets (4 categories)
- ✓ “Dbnary 4 cats Synsets” (**DB-4-catg**)
- ✓ “Dbnary 21 cats Synsets” (**DB-all-catg**)




## **METEOR SCORES ON ENGLISH WORDNET VS DBNARY**

# Impact of the “Synsets”

METEOR	Baseline (WordNet)	DB-4-catg	DB-all-catg
online A	36.97 %	36.91 %	37.13 %
rbmt-1	33.74 %	33.60 %	33.89 %

**Table 4 .** METEOR-Baseline vs METEOR-Dbnary for 2 randomly picked up systems from WMT14 data (French-English MT)

## Comments

-  similar scores for the **Baseline & DB-4-catg**
  -  the size of the WordNet dictionary is 2,5 times larger than the size of Dbnary (4-catg).
-  small increase (>0.2, >0.6%) using all 21 Dbnary categories with **DB-all-Catg**


# The second hidden parameter

---

 METEOR uses the **Morphy-7WN** function to find the lemma of a given English word

 what would we do for the other languages?


 Idea

 Use Treetagger (*Schmid, 1994*) to lemmatize forms for any language

 Cons

 Using Treetagger while computing METEOR score will slow down the execution time

 Solution


 preprocess the data (hypo, ref) to get lists of pairs (word, lemma)

# Impact of the lemmatizer


	METEOR-Morphy	METEOR-TTG
online A	36.97 %	37.00 %
rbmt-1	33.74 %	33.76 %


**Table 5.** Impact of lemmatization; METEOR-Morphy vs METEOR-TTG for 2 randomly picked up systems from WMT14 data (French-English MT)

## Comment

 A slight increase between the scores of METEOR-Morphy and METEOR-TTG

 Possible explanation

 TreeTagger lemmatizes all categories

 Morphy-7WN lemmatizes only three categories (Noun, Verb and Adjective)

Correlation comparison with previously reported results

- ✓ English–Spanish (WMT13)
- ✓ French–English, English–French, English–Russian, English–German (WMT14)

## **CORRELATION WITH HUMAN JUDGMENT**

### **METEOR WORDNET VS DBNARY**



# Goal


---

 Compare correlation of METEOR and METEOR-Dbnary with human judgments of MT hypotheses


 WMT13 Metrics Shared Task (*Machacek and Bojar, 2013*)


 English–Spanish

 WMT14 Metrics Shared Task (*Machacek and Bojar, 2014*)

 French–English, English–French, English–German and English–Russian

 Evaluation measures

 System-level: Pearson correlation coefficient between system rankings based on human judgments vs automatic score


 Segment-level: Kendall's  $\tau$  rank correlation coefficient between system rankings based on human judgments vs automatic score


# Setup

---

 “Dbnary Synsets” for all the target languages:


FR, SP, RU, GE

 weight of 0.8 for the synonyms for each language

 same weight as the English synonym module in the METEOR default setting

 Two configurations of METEOR

 **METEOR-Baseline:** METEOR Universal (v1.5) with the synonym module activated for English only with WordNet





 **METEOR-Dbnary:** METEOR Universal with the synonym module activated for EN, FR, SP, RU, GE, using “Dbnary Synsets”

# Results for Pearson Correlation Coeff.

	WMT14				WMT13
	FR-EN	EN-FR	EN-RU	EN-GE	EN-ES
Meteor-Baseline	<b>.975</b>	.941	.923	.263	.886
Meteor-Dbnary	.973	<b>.943</b>	<b>.928</b>	<b>.320</b>	<b>.895</b>

**Table 6.** System-level correlations (Pearson Correlation Coefficient) between METEOR- Baseline (or METEOR-Dbnary) and the WMT13/WMT14 human rankings

## Comments



-  when WordNet Synsets are available (**FR-EN**)
  -  slight degradation ( $\text{size}(\text{Dbnary}) \ll \text{size}(\text{WordNet})$ )
-  when WordNet Synsets are not available (**EN-XX**)
  -  use of “Dbnary Synsets” slightly improves system-level correlations of METEOR score with human judgment

# Results for Kendall's $\tau$ rank corr. coeff.

	WMT14				WMT13
	FR-EN	EN-FR	EN-RU	EN-GE	EN-ES
Meteor-Baseline	.406	.280	.238	.427	.184
Meteor-Dbnary	.406	<b>.284</b>	<b>.240</b>	<b>.435</b>	<b>.187</b>

**Table 7.** Segment-level correlations (Kendall's  $\tau$ ) between METEOR-Baseline (or METEOR-Dbnary) and the the WMT13/WMT14 human rankings

## Comments

-  Same trend that before for segment-level correlations
-  Dbnary can be a useful resource for MT evaluation to bring synonyms as an added feature

# Changes in the METEOR score


	WMT14			WMT13
	EN-FR	EN-RU	EN-GE	EN-ES
Meteor-Baseline	50.94	36.21	38.06	49.88
Meteor-Dbnary	<b>52.34</b>	<b>37.60</b>	<b>41.51</b>	<b>51.04</b>

**Table 8** : Comparison of METEOR-Baseline without synonyms vs METEOR-Dbnary (for *rbmt-1* system)



## Comments

 METEOR-Dbnary scores are better

## Explanation

 Using Dbnary as a lexical resource for synonymy, the metric maps more words with the same meaning

# Example 1

-  **Reference:** [...] alors les **dirigeants** d'entreprise sont sûrement **aussi** des cibles potentielles.
-  **Hypothesis:** [...] alors sûrement les **chefs** de file des affaires sont **également** les cibles potentielles.

## Synonym match



Word	Lemma	Synonym list
<b>dirigeants</b>	dirigeant	[ <b>chef</b> , maître, leader, directeur]
<b>chefs</b>	chef	[tête, maître, cuisinier, leader, maître_queux, patron]
<b>aussi</b>	<u>aussi</u>	[ainsi, <u>également</u> , itou]
<b>également</b>	<u>également</u>	[ <u>aussi</u> , pareillement, de_même, par_ailleurs]

➤ **Segment score:**

**METEOR-Baseline: 0.6762**

**METEOR-Dbnary : 0.7290**

# Example 2

-  **Reference:** J'estime qu'il est concevable que ces données soient **utilisées** dans leur intérêt mutuel.
-  **Hypothesis:** Je pense qu'il est concevable que ces données soient **employées** pour le bénéfice mutuel.



## Synonym match

Word	Lemma	Synonym list
<b>utilisées</b>	<u>utiliser</u>	[user]
<b>employés</b>	employer	[occuper, <u>utiliser</u> ]

➤ **Segment score :**

**METEOR-Baseline : 0.6609**  
**METEOR-Dbnary : 0.7133**

# Example 3

-  **Reference:** Il me parlait, m'encourageait constamment, il **habitait** mon corps.
-  **Hypothesis:** Il me parlerait, m'encouragent constamment, il a **vécu** dans mon corps.

## Synonym match

Word	Lemma	Synonym list
habitait	<u>habiter</u>	[occuper]
vécu	vivre	[ <u>habiter</u> , nourriture ]

➤ **Segment score :**

**METEOR-Baseline : 0.6743**






**METEOR-Dbnary : 0.7688**



# CONCLUSION AND PERSPECTIVES




# Contribution

---

-  Dbnary as a good source of Synsets for the languages without a free WordNet for METEOR with synonymy overlap
  -  better scores & better correlation for the tested conditions (WMT13, WMT14) for EN-to-(FR, RU, GE, SP)
-  Dbnary as reasonable alternative to WordNet for English
  -  may even get better as the English Wiktionary grows
-  METEOR-Dbnary is available for download at:  
<http://kaiko.getalp.org/about-dbnary/meteor-with-dbnary>

# Perspectives

---

-  Apply the same adaptation of synonym matches to TER-Plus (*Snover et al., 2009b*)
-  Use WSD (*Apidianaki and Marie, 2015*) in conjunction with “Dbnary Sysnets”
  -  to better improve correlation between automatic evaluation metrics and human judgments?
-  The METEOR framework with **synonymy overlap** can be used for more target languages (21 as of today) using “Dbnary Synsets” extracted from Dbnary
  -  Bulgarian, Dutch, English, Finnish, French, German, Indonesian, Italian, Japanese, Latin, Lithuanian, Malagasy, Greek (modern), Norwegian, Polish, Portuguese, Russian, Serbo-Croatian, Spanish, Swedish, Turkish