Translation Quality Evaluation in the Molto Project (II)

Maarit Koponen, Lauri Carlson,
Cristina España-Bonet and Lluís Màrquez

University of Helsinki & Universitat Politècnica de Catalunya

– First year project meeting –

Göteborg, March 9th, 2011

Overview

1. Introduction

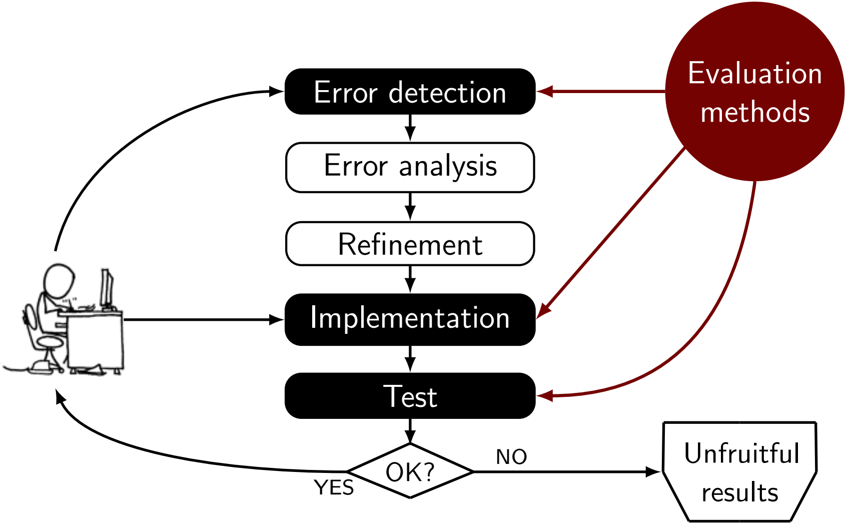
2. Manual Evaluation

3. Automatic Evaluation
   - Motivation
   - The Asiya Software
   - Case of Study: Patents

4. Conclusions

Automatic Evaluation: Motivation

Importance for system development



Error detection

Error analysis

Refinement

Implementation

Test

OK?

YES

NO

Unfruitful results

Evaluation methods

Automatic Evaluation: Motivation

What can be achieved with automatic evaluation?

Automatic metrics notably **accelerate** the development cycle of MT systems:

- **Error analysis**
- **System optimisation**
- **System comparison**

Automatic Evaluation: Motivation

What can be achieved with automatic evaluation?

Automatic metrics notably **accelerate** the development cycle of MT systems:

- **Error analysis**
- **System optimisation**
- **System comparison**

Besides, they are

- **Costless** (vs. costly)
- **Objective** (vs. subjective)
- **Reusable** (vs. non-reusable)

Automatic Evaluation: Motivation

What can be damaged with automatic evaluation?

- **System overtuning** when system parameters are adjusted towards a given metric.
- **Blind system development** when metrics are unable to capture system improvements.
- **Unfair system comparisons** when metrics are unable to reflect difference in quality between MT systems.

Automatic Evaluation: Motivation

Lexical similarity

Metrics based on lexical similarity
(most of the metrics!)

- **Edit Distance**: WER, PER, TER
- **Precision**: BLEU, NIST, WNM
- **Recall**: ROUGE, CDER
- **Precision/Recall**: GTM, METEOR, BLANC, SIA

Automatic Evaluation: Motivation

Lexical similarity

Metrics based on lexical similarity
(most of the metrics!)

- **Edit Distance**: WER, PER, TER
- **Precision**: BLEU, NIST, WNM
- **Recall**: ROUGE, CDER
- **Precision/Recall**: GTM, METEOR, BLANC, SIA

Nowadays, **BLEU** is accepted as **the standard** metric.

Automatic Evaluation: Motivation

Limits of lexical similarity

The reliability of lexical metrics depends very strongly on the **heterogeneity/representativity** of reference translations.

e: `This sentence is going to be difficult to evaluate.`

Ref1: `The evaluation of the translation is complicated.`

Automatic Evaluation: Motivation

Limits of lexical similarity

The reliability of lexical metrics depends very strongly on the **heterogeneity/representativity** of reference translations.

e: This sentence is going to be difficult to evaluate.

Ref1: The evaluation of the translation is complicated.

Ref2: The sentence will be hard to qualify.

Ref3: The translation is going to be hard to evaluate.

Ref4: It will be difficult to punctuate the output.

Automatic Evaluation: Motivation

Limits of lexical similarity

The reliability of lexical metrics depends very strongly on the **heterogeneity/representativity** of reference translations.

e: `This sentence is going to be difficult to evaluate.`

Ref1: `The evaluation of the translation is complicated.`
Ref2: `The sentence will be hard to qualify.`
Ref3: `The translation is going to be hard to evaluate.`
Ref4: `It will be difficult to punctuate the output.`

Lexical similarity is **nor a sufficient neither a necessary condition** so that two sentences convey the same meaning.

Automatic Evaluation: Motivation

Going over lexical similarity

Extension of the reference material:

- Using **lexical variants** such as morphological variations or synonymy lookup or using **paraphrasing** support

Automatic Evaluation: Motivation

Going over lexical similarity

Extension of the reference material:

- Using **lexical variants** such as morphological variations or synonymy lookup or using **paraphrasing** support

Comparing other **linguistic features** than words:

- **Syntactic** similarity: shallow parsing, full parsing (constituents /dependencies).
- **Semantic** similarity: named entities, semantic roles, discourse representations, textual entailment.

Automatic Evaluation: Motivation

Going over lexical similarity

Extension of the reference material:

- Using **lexical variants** such as morphological variations or synonymy lookup or using **paraphrasing** support

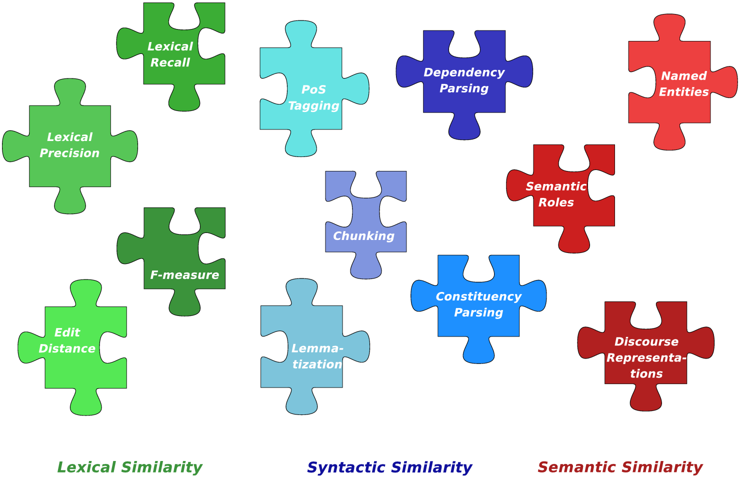
Comparing other **linguistic features** than words:

- **Syntactic** similarity: shallow parsing, full parsing (constituents /dependencies).
- **Semantic** similarity: named entities, semantic roles, discourse representations, textual entailment.

Combination of the existing metrics.

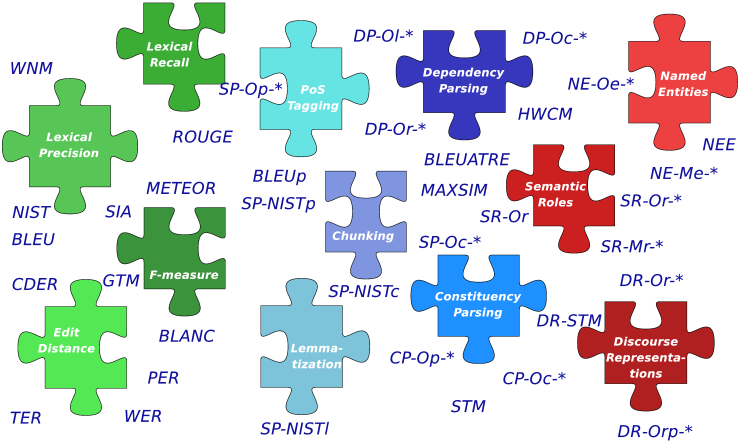
Automatic Evaluation: Motivation

Towards Heterogeneous Automatic MT Evaluation



Automatic Evaluation: Motivation

Towards Heterogeneous Automatic MT Evaluation



Lexical Recall
Lexical Precision
F-measure
Edit Distance
WNM
ROUGE
METEOR
NIST
SIA
BLEU
CDER
GTM
BLANC
PER
TER
WER

PoS Tagging
Dependency Parsing
Chunking
Lemmatization
Constituency Parsing
SP-Op-*
DP-Ol-*
DP-Oc-*
HWCM
DP-Or-*
BLEUp
SP-NISTp
BLEUATRE
MAXSIM
SP-Oc-*
SP-NISTc
CP-Op-*
CP-Oc-*
STM
SP-NISTl

Named Entities
Semantic Roles
Discourse Representations
NE-Oe-*
NEE
NE-Me-*
SR-Or
SR-Or-*
SR-Mr-*
DR-Or-*
DR-STM
DR-Orp-*

Lexical Similarity **Syntactic Similarity** **Semantic Similarity**

Automatic Evaluation: The Asiya Software

Towards Heterogeneous Automatic MT Evaluation

ASIYA

Asiya has been designed to assist both **system** and metric **developers** by offering a rich repository of metrics and meta-metrics.

`http://www.lsi.upc.edu/∼nlp/Asiya/`

Automatic Evaluation: The Asiya Software

Language-dependent evaluation

The number of available metrics depends on the available **linguistic procesors**. Currently implemented:

English: Lexical, Syntactic and Semantic similarity

Spanish: Lexical and Syntactic similarity

German, French and others: Lexical similarity

Automatic Evaluation: The Asiya Software

Language-dependent evaluation

The number of available metrics depends on the available **linguistic procesors**. Currently implemented:

English: Lexical, Syntactic and Semantic similarity

Spanish: Lexical and Syntactic similarity

German, French and others: Lexical similarity

Soon! Widening Spanish, German, French, Czech & Catalan

(FAUST project, FP7-ICT-2009-4-247762)

Web interface (OPENMT2 project, TIN2009-14675-C03)

Automatic Evaluation: Case of Study, Patents

Evaluation of the SMT systems

System 1 (MOLTO SMT baseline)

- **Corpus**. Chemical domain, A61P patents
- **Translation Engine**. Moses-based translator

System 2

- **Google**

System 3

- **Bing**

Automatic Evaluation: Case of Study, Patents

English, German & French: Lexical similarity

Lexical metrics available in **Asiya**:

```
metrics_BLEU =  BLEU, BLEU-1, BLEU-2, BLEU-3, BLEU-4,
                BLEUi-2, BLEUi-3, BLEUi-4
metrics_GTM =  GTM-1, GTM-2, GTM-3
metrics_METEOR =  METEOR-ex, METEOR-pa, METEOR-st, METEOR-sy
metrics_NIST =  NIST, NIST-1, NIST-2, NIST-3, NIST-4, NIST-5,
                NISTi-2, NISTi-3, NISTi-4, NISTi-5
metrics_O =  Ol
metrics_PER =  -PER
metrics_ROUGE =  ROUGE-1, ROUGE-2, ROUGE-3, ROUGE-4,
                 ROUGE-L, ROUGE-S*, ROUGE-SU*, ROUGE-W
metrics_TER =  -TER, -TERbase, -TERp, -TERp-A
metrics_WER =  -WER
```

Automatic Evaluation: Case of Study, Patents

English, German & French: Lexical similarity

Lexical metrics available in **Asiya**:

```
metrics_BLEU =  BLEU, BLEU-1, BLEU-2, BLEU-3, BLEU-4,
                BLEUi-2, BLEUi-3, BLEUi-4
metrics_GTM =  GTM-1, GTM-2, GTM-3
metrics_METEOR =  METEOR-ex, METEOR-pa, METEOR-st, METEOR-sy
metrics_NIST =  NIST, NIST-1, NIST-2, NIST-3, NIST-4, NIST-5,
                NISTi-2, NISTi-3, NISTi-4, NISTi-5
metrics_O =  Ol
metrics_PER =  -PER
metrics_ROUGE =  ROUGE-1, ROUGE-2, ROUGE-3, ROUGE-4,
                 ROUGE-L, ROUGE-S*, ROUGE-SU*, ROUGE-W
metrics_TER =  -TER, -TERbase, -TERp, -TERp-A
metrics_WER =  -WER
```

{**-WER,-PER,-TER,BLEU,NIST,ROUGE-W,GTM-2,METEOR-pa**}

Automatic Evaluation: Case of Study, Patents

English-German Translations, scores

	DE2EN			EN2DE		
METRIC	**Bing**	**Google**	**Domain**	**Bing**	**Google**	**Domain**
1-WER	0.52	0.64	**0.72**	0.42	0.51	**0.69**
1-PER	0.66	0.76	**0.82**	0.56	0.64	**0.77**
1-TER	0.59	0.67	**0.76**	0.45	0.53	**0.71**
BLEU	0.43	0.58	**0.65**	0.33	0.45	**0.58**
NIST	8.25	9.67	**10.12**	6.53	8.05	**9.40**
ROUGE-W	0.40	0.48	**0.52**	0.34	0.41	**0.48**
GTM-2	0.30	0.40	**0.47**	0.25	0.32	**0.43**
METEOR-pa	0.60	0.69	**0.74**	0.36	0.45	**0.57**
ULC	0.09	0.29	**0.41**	0.03	0.19	**0.43**

Automatic Evaluation: Case of Study, Patents

English-German Translations, examples

Why such good scores?

DE	Verwendung nach Anspruch 23 , worin das molare Verhältnis von Arginin zu Ibuprofen 0,60 : 1 beträgt .
EN	The use of claim 23 , wherein the molar ratio of arginine to ibuprofen is 0.60 : 1 .

Automatic Evaluation: Case of Study, Patents

English-German Translations, examples

Why such good scores?

DE	Verwendung nach Anspruch 23 , worin das molare Verhältnis von Arginin zu Ibuprofen 0,60 : 1 beträgt .
EN	**The use** of claim 23 , wherein the molar ratio of arginine to ibuprofen **is** 0.60 : 1 .
Domain	The use of claim 23 , wherein the molar ratio of arginine to ibuprofen is 0.60 : 1 .
Google	The **method** of claim 23 , wherein the molar ratio of arginine to ibuprofen 0.60 : 1 **is** .
Bing	The Use of claim 23 , wherein the molar ratio of arginine to ibuprofen is 0.60 : 1 .

Automatic Evaluation: Case of Study, Patents

English-German Translations, examples

What's wrong?

DE	(±)-N-(3-Aminopropyl)-N,N-dimethyl-2,3-bis(syn-9-tetradecenyloxy)-1-propanaminiumbromid
EN	(±)-N-(3-**a**minopropyl)-N,N-dimethyl-2,3-bis(syn-9-tetradeceneyloxy)-1-propanaminium **bromide**

Automatic Evaluation: Case of Study, Patents

English-German Translations, examples

What's wrong?

DE	(±)-N-(3-Aminopropyl)-N,N-dimethyl-2,3-bis(syn-9-tetradecenyloxy)-1-propanaminiumbromid
EN	(±)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(syn-9-tetradeceneyloxy)-1-propanaminium **bromide**
Domain	(±)-N-(3-Aminopropyl)-N,N-dimethyl-2,3-bis(syn-9-tetradecenyloxy)-1-propanaminiumbromid
Google	(±)-N-(3-aminopropyl)-N , N-dimethyl-2 , 3-bis (syn-9-tetradecenyloxy) is 1-propanaminiumbromid
Bing	(±)-N-(3-Aminopropyl)-N,N-dimethyl-2,3-bis(syn-9-tetradecenyloxy)-1-propanaminiumbromid

Automatic Evaluation: Case of Study, Patents

English-French Translations, scores

	FR2EN			EN2FR		
METRIC	**Bing**	**Google**	**Domain**	**Bing**	**Google**	**Domain**
1-WER	0.54	0.66	**0.78**	0.57	0.63	**0.73**
1-PER	0.71	0.78	**0.86**	0.68	0.75	**0.82**
1-TER	0.59	0.70	**0.80**	0.60	0.66	**0.74**
BLEU	0.45	0.62	**0.70**	0.43	0.53	**0.62**
NIST	8.52	10.01	**10.86**	8.39	9.21	**9.96**
ROUGE-W	0.41	0.50	**0.54**	0.39	0.45	**0.49**
GTM-2	0.32	0.43	**0.53**	0.31	0.36	**0.45**
METEOR-pa	0.61	0.72	**0.77**	0.57	0.65	**0.71**
ULC	0.07	0.28	**0.44**	0.10	0.23	**0.39**

Automatic Evaluation: Case of Study, Patents

German-French Translations, scores

	DE2FR			FR2DE		
METRIC	**Bing**	**Google**	**Domain**	**Bing**	**Google**	**Domain**
1-WER	0.42	0.52	**0.76**	0.30	0.43	**0.65**
1-PER	0.58	0.68	**0.77**	0.46	0.59	**0.74**
1-TER	0.47	0.56	**0.68**	0.32	0.46	**0.66**
BLEU	0.29	0.43	**0.56**	0.24	0.39	**0.53**
NIST	6.72	8.21	**9.10**	5.35	7.30	**8.88**
ROUGE-W	0.31	0.38	**0.45**	0.29	0.37	**0.44**
GTM-2	0.24	0.30	**0.41**	0.21	0.28	**0.41**
METEOR-pa	0.45	0.56	**0.64**	0.26	0.39	**0.51**
ULC	0.03	0.22	**0.41**	-0.03	0.19	**0.44**

Conclusions

Evaluation within MOLTO

- MOLTO uses both manual and automatic evaluation.
- For a fast development process automatic metrics are very useful.
- But, for the automatic evaluation one needs reference translations.
- Manual evaluation assures a high quality final evaluation.

Translation Quality Evaluation in the Molto Project (II)

Maarit Koponen, Lauri Carlson,
Cristina España-Bonet and Lluís Màrquez

University of Helsinki & Universitat Politècnica de Catalunya

– First year project meeting –

Göteborg, March 9th, 2011

Conclusions

System evaluation with Asiya

```
Asiya.pl -eval single,ulc -g sys Asiya.config
```

Conclusions

System evaluation with Asiya

```
Asiya.pl -eval single,ulc -g sys Asiya.config
```

```
input=raw

SRCLANG=de
TRGLANG=en
SRCCASE=cs
TRGCASE=cs

#SRC ==================================================
src=./data/patsA61P.test.de
#REF ==================================================
ref=./data/patsA61P.test.en
#OUT ===================================================
sys=./data/patsA61P.test.trans.de2en
sys=./data/patsA61P.test.trad.google.de2en
sys=./data/patsA61P.test.trad.bing.de2en
#-------------------------------------------------------
```

Conclusions

System evaluation with Asiya

```
Asiya.pl -eval single,ulc -m metrSet Asiya.config
```

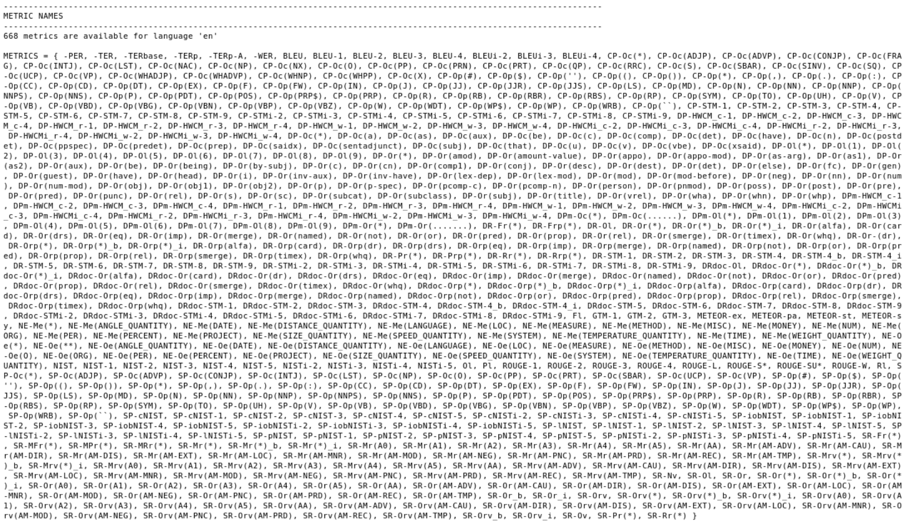
```
SRCLANG=de
TRGLANG=en

#SRC =================================================
src=./data/patsA61P.test.de
#REF =================================================
ref=./data/patsA61P.test.en
#OUT ==================================================
sys=./data/patsA61P.test.trans.de2en
#-----------------------------------------------------

metrSet=1-PER 1-TER 1-WER BLEU-4 CP-Oc-* CP-Op-* CP-STM-9 DP-HWC-c-4
DP-HWC-r-4 DP-HWC-w-4 DP-Oc-* DP-Ol-* DP-Or-* DR-Or-* DR-Orp-* DR-STM-9
GTM-1 GTM-2 GTM-3 MTR-exact MTR-stem MTR-wnstm MTR-wnsyn NE-Me-* NE-Oe-*
NE-Oe-** NIST-5 RG-L RG-S* RG-SU* RG-W-1.2 SP-Oc-* SP-Op-* SP-cNIST-5
SP-iobNIST-5 SP-lNIST-5 SP-pNIST-5 SR-Mr-* SR-Mrv-* SR-Or SR-Or-* SR-Orv
```

Conclusions

Metrics in Asiya (English)



```
------------------------------------------------------------------------------------------------------------------------
METRIC NAMES
------------------------------------------------------------------------------------------------------------------------
668 metrics are available for language 'en'

METRICS = { -PER, -TER, -TERbase, -TERp, -TERp-A, -WER, BLEU, BLEU-1, BLEU-2, BLEU-3, BLEU-4, BLEUi-2, BLEUi-3, BLEUi-4, CP-Oc(*), CP-Oc(ADJP), CP-Oc(ADVP), CP-Oc(CONJP), CP-Oc(FRA
G), CP-Oc(INTJ), CP-Oc(LST), CP-Oc(NAC), CP-Oc(NP), CP-Oc(NX), CP-Oc(O), CP-Oc(PP), CP-Oc(PRN), CP-Oc(PRT), CP-Oc(QP), CP-Oc(RRC), CP-Oc(S), CP-Oc(SBAR), CP-Oc(SINV), CP-Oc(SQ), CP
-Oc(UCP), CP-Oc(VP), CP-Oc(WHADJP), CP-Oc(WHADVP), CP-Oc(WHNP), CP-Oc(WHPP), CP-Oc(X), CP-Op(#), CP-Op($), CP-Op(''), CP-Op((), CP-Op()), CP-Op(*), CP-Op(,), CP-Op(.), CP-Op(:), CP
-Op(CC), CP-Op(CD), CP-Op(DT), CP-Op(EX), CP-Op(F), CP-Op(FW), CP-Op(IN), CP-Op(J), CP-Op(JJ), CP-Op(JJR), CP-Op(JJS), CP-Op(LS), CP-Op(MD), CP-Op(N), CP-Op(NN), CP-Op(NNP), CP-Op(
NNPS), CP-Op(NNS), CP-Op(P), CP-Op(PDT), CP-Op(POS), CP-Op(PRP$), CP-Op(PRP), CP-Op(R), CP-Op(RB), CP-Op(RBR), CP-Op(RBS), CP-Op(RP), CP-Op(SYM), CP-Op(TO), CP-Op(UH), CP-Op(V), CP
-Op(VB), CP-Op(VBD), CP-Op(VBG), CP-Op(VBN), CP-Op(VBP), CP-Op(VBZ), CP-Op(W), CP-Op(WDT), CP-Op(WP$), CP-Op(WP), CP-Op(WRB), CP-Op(``), CP-STM-1, CP-STM-2, CP-STM-3, CP-STM-4, CP-
STM-5, CP-STM-6, CP-STM-7, CP-STM-8, CP-STM-9, CP-STMi-2, CP-STMi-3, CP-STMi-4, CP-STMi-5, CP-STMi-6, CP-STMi-7, CP-STMi-8, CP-STMi-9, DP-HWCM_c-1, DP-HWCM_c-2, DP-HWCM_c-3, DP-HWC
M_c-4, DP-HWCM_r-1, DP-HWCM_r-2, DP-HWCM_r-3, DP-HWCM_r-4, DP-HWCM_w-1, DP-HWCM_w-2, DP-HWCM_w-3, DP-HWCM_w-4, DP-HWCMi_c-2, DP-HWCMi_c-3, DP-HWCMi_c-4, DP-HWCMi_r-2, DP-HWCMi_r-3,
 DP-HWCMi_r-4, DP-HWCMi_w-2, DP-HWCMi_w-3, DP-HWCMi_w-4, DP-Oc(*), DP-Oc(a), DP-Oc(as), DP-Oc(aux), DP-Oc(be), DP-Oc(c), DP-Oc(comp), DP-Oc(det), DP-Oc(have), DP-Oc(n), DP-Oc(postd
et), DP-Oc(ppspec), DP-Oc(predet), DP-Oc(prep), DP-Oc(saidx), DP-Oc(sentadjunct), DP-Oc(subj), DP-Oc(that), DP-Oc(u), DP-Oc(v), DP-Oc(vbe), DP-Oc(xsaid), DP-Ol(*), DP-Ol(1), DP-Ol(
2), DP-Ol(3), DP-Ol(4), DP-Ol(5), DP-Ol(6), DP-Ol(7), DP-Ol(8), DP-Ol(9), DP-Or(*), DP-Or(amod), DP-Or(amount-value), DP-Or(appo), DP-Or(appo-mod), DP-Or(as-arg), DP-Or(as1), DP-Or
(as2), DP-Or(aux), DP-Or(be), DP-Or(being), DP-Or(by-subj), DP-Or(c), DP-Or(cn), DP-Or(comp1), DP-Or(conj), DP-Or(desc), DP-Or(dest), DP-Or(det), DP-Or(else), DP-Or(fc), DP-Or(gen)
, DP-Or(guest), DP-Or(have), DP-Or(head), DP-Or(i), DP-Or(inv-aux), DP-Or(inv-have), DP-Or(lex-dep), DP-Or(lex-mod), DP-Or(mod), DP-Or(mod-before), DP-Or(neg), DP-Or(nn), DP-Or(num
), DP-Or(num-mod), DP-Or(obj), DP-Or(obj1), DP-Or(obj2), DP-Or(p), DP-Or(p-spec), DP-Or(pcomp-c), DP-Or(pcomp-n), DP-Or(person), DP-Or(pnmod), DP-Or(poss), DP-Or(post), DP-Or(pre),
 DP-Or(pred), DP-Or(punc), DP-Or(rel), DP-Or(s), DP-Or(sc), DP-Or(subcat), DP-Or(subclass), DP-Or(subj), DP-Or(title), DP-Or(vrel), DP-Or(wha), DP-Or(whn), DP-Or(whp), DPm-HWCM_c-1
, DPm-HWCM_c-2, DPm-HWCM_c-3, DPm-HWCM_c-4, DPm-HWCM_r-1, DPm-HWCM_r-2, DPm-HWCM_r-3, DPm-HWCM_r-4, DPm-HWCM_w-1, DPm-HWCM_w-2, DPm-HWCM_w-3, DPm-HWCM_w-4, DPm-HWCMi_c-2, DPm-HWCMi
_c-3, DPm-HWCMi_c-4, DPm-HWCMi_r-2, DPm-HWCMi_r-3, DPm-HWCMi_r-4, DPm-HWCMi_w-2, DPm-HWCMi_w-3, DPm-HWCMi_w-4, DPm-Oc(*), DPm-Oc(......), DPm-Ol(*), DPm-Ol(1), DPm-Ol(2), DPm-Ol(3)
, DPm-Ol(4), DPm-Ol(5), DPm-Ol(6), DPm-Ol(7), DPm-Ol(8), DPm-Ol(9), DPm-Or(*), DPm-Or(.......), DR-Fr(*), DR-Frp(*), DR-Ol, DR-Or(*), DR-Or(*)_b, DR-Or(*)_i, DR-Or(alfa), DR-Or(car
d), DR-Or(drs), DR-Or(eq), DR-Or(imp), DR-Or(merge), DR-Or(named), DR-Or(not), DR-Or(or), DR-Or(pred), DR-Or(prop), DR-Or(rel), DR-Or(smerge), DR-Or(timex), DR-Or(whq), DR-Or-(dr),
 DR-Orp(*), DR-Orp(*)_b, DR-Orp(*)_i, DR-Orp(alfa), DR-Orp(card), DR-Orp(dr), DR-Orp(drs), DR-Orp(eq), DR-Orp(imp), DR-Orp(merge), DR-Orp(named), DR-Orp(not), DR-Orp(or), DR-Orp(pr
ed), DR-Orp(prop), DR-Orp(rel), DR-Orp(smerge), DR-Orp(timex), DR-Orp(whq), DR-Pr(*), DR-Prp(*), DR-Rr(*), DR-Rrp(*), DR-STM-1, DR-STM-2, DR-STM-3, DR-STM-4, DR-STM-4_b, DR-STM-4_i
, DR-STM-5, DR-STM-6, DR-STM-7, DR-STM-8, DR-STM-9, DR-STMi-2, DR-STMi-3, DR-STMi-4, DR-STMi-5, DR-STMi-6, DR-STMi-7, DR-STMi-8, DR-STMi-9, DRdoc-Ol, DRdoc-Or(*), DRdoc-Or(*)_b, DR
doc-Or(*)_i, DRdoc-Or(alfa), DRdoc-Or(card), DRdoc-Or(dr), DRdoc-Or(drs), DRdoc-Or(eq), DRdoc-Or(imp), DRdoc-Or(merge), DRdoc-Or(named), DRdoc-Or(not), DRdoc-Or(or), DRdoc-Or(pred)
, DRdoc-Or(prop), DRdoc-Or(rel), DRdoc-Or(smerge), DRdoc-Or(timex), DRdoc-Or(whq), DRdoc-Orp(*), DRdoc-Orp(*)_b, DRdoc-Orp(*)_i, DRdoc-Orp(alfa), DRdoc-Orp(card), DRdoc-Orp(dr), DR
doc-Orp(drs), DRdoc-Orp(eq), DRdoc-Orp(imp), DRdoc-Orp(merge), DRdoc-Orp(named), DRdoc-Orp(not), DRdoc-Orp(or), DRdoc-Orp(pred), DRdoc-Orp(prop), DRdoc-Orp(rel), DRdoc-Orp(smerge),
 DRdoc-Orp(timex), DRdoc-Orp(whq), DRdoc-STM-1, DRdoc-STM-2, DRdoc-STM-3, DRdoc-STM-4, DRdoc-STM-4_b, DRdoc-STM-4_i, DRdoc-STM-5, DRdoc-STM-6, DRdoc-STM-7, DRdoc-STM-8, DRdoc-STM-9
, DRdoc-STMi-2, DRdoc-STMi-3, DRdoc-STMi-4, DRdoc-STMi-5, DRdoc-STMi-6, DRdoc-STMi-7, DRdoc-STMi-8, DRdoc-STMi-9, Fl, GTM-1, GTM-2, GTM-3, METEOR-ex, METEOR-pa, METEOR-st, METEOR-s
y, NE-Me(*), NE-Me(ANGLE_QUANTITY), NE-Me(DATE), NE-Me(DISTANCE_QUANTITY), NE-Me(LANGUAGE), NE-Me(LOC), NE-Me(MEASURE), NE-Me(METHOD), NE-Me(MISC), NE-Me(MONEY), NE-Me(NUM), NE-Me(
ORG), NE-Me(PER), NE-Me(PERCENT), NE-Me(PROJECT), NE-Me(SIZE_QUANTITY), NE-Me(SPEED_QUANTITY), NE-Me(SYSTEM), NE-Me(TEMPERATURE_QUANTITY), NE-Me(TIME), NE-Me(WEIGHT_QUANTITY), NE-O
e(*), NE-Oe(**), NE-Oe(ANGLE_QUANTITY), NE-Oe(DATE), NE-Oe(DISTANCE_QUANTITY), NE-Oe(LANGUAGE), NE-Oe(LOC), NE-Oe(MEASURE), NE-Oe(METHOD), NE-Oe(MISC), NE-Oe(MONEY), NE-Oe(NUM), NE
-Oe(O), NE-Oe(ORG), NE-Oe(PER), NE-Oe(PERCENT), NE-Oe(PROJECT), NE-Oe(SIZE_QUANTITY), NE-Oe(SPEED_QUANTITY), NE-Oe(SYSTEM), NE-Oe(TEMPERATURE_QUANTITY), NE-Oe(TIME), NE-Oe(WEIGHT_Q
UANTITY), NIST, NIST-1, NIST-2, NIST-3, NIST-4, NIST-5, NISTi-2, NISTi-3, NISTi-4, NISTi-5, Ol, Pl, ROUGE-1, ROUGE-2, ROUGE-3, ROUGE-4, ROUGE-L, ROUGE-S*, ROUGE-SU*, ROUGE-W, Rl, S
P-Oc(*), SP-Oc(ADJP), SP-Oc(ADVP), SP-Oc(CONJP), SP-Oc(INTJ), SP-Oc(LST), SP-Oc(NP), SP-Oc(O), SP-Oc(PP), SP-Oc(PRT), SP-Oc(SBAR), SP-Oc(UCP), SP-Oc(VP), SP-Op(#), SP-Op($), SP-Op(
''), SP-Op((), SP-Op()), SP-Op(*), SP-Op(,), SP-Op(.), SP-Op(:), SP-Op(CC), SP-Op(CD), SP-Op(DT), SP-Op(EX), SP-Op(F), SP-Op(FW), SP-Op(IN), SP-Op(J), SP-Op(JJ), SP-Op(JJR), SP-Op(
JJS), SP-Op(LS), SP-Op(MD), SP-Op(N), SP-Op(NN), SP-Op(NNP), SP-Op(NNPS), SP-Op(NNS), SP-Op(P), SP-Op(PDT), SP-Op(POS), SP-Op(PRP$), SP-Op(PRP), SP-Op(R), SP-Op(RB), SP-Op(RBR), SP
-Op(RBS), SP-Op(RP), SP-Op(SYM), SP-Op(TO), SP-Op(UH), SP-Op(V), SP-Op(VB), SP-Op(VBD), SP-Op(VBG), SP-Op(VBN), SP-Op(VBP), SP-Op(VBZ), SP-Op(W), SP-Op(WDT), SP-Op(WP$), SP-Op(WP),
 SP-Op(WRB), SP-Op(``), SP-cNIST, SP-cNIST-1, SP-cNIST-2, SP-cNIST-3, SP-cNIST-4, SP-cNIST-5, SP-cNISTi-2, SP-cNISTi-3, SP-cNISTi-4, SP-cNISTi-5, SP-iobNIST, SP-iobNIST-1, SP-iobNI
ST-2, SP-iobNIST-3, SP-iobNIST-4, SP-iobNIST-5, SP-iobNISTi-2, SP-iobNISTi-3, SP-iobNISTi-4, SP-iobNISTi-5, SP-lNIST, SP-lNIST-1, SP-lNIST-2, SP-lNIST-3, SP-lNIST-4, SP-lNIST-5, SP
-lNISTi-2, SP-lNISTi-3, SP-lNISTi-4, SP-lNISTi-5, SP-pNIST, SP-pNIST-1, SP-pNIST-2, SP-pNIST-3, SP-pNIST-4, SP-pNIST-5, SP-pNISTi-2, SP-pNISTi-3, SP-pNISTi-4, SP-pNISTi-5, SR-Fr(*)
, SR-MFr(*), SR-MPr(*), SR-MRr(*), SR-Mr(*), SR-Mr(*)_b, SR-Mr(*)_i, SR-Mr(A0), SR-Mr(A1), SR-Mr(A2), SR-Mr(A3), SR-Mr(A4), SR-Mr(A5), SR-Mr(AA), SR-Mr(AM-ADV), SR-Mr(AM-CAU), SR-M
r(AM-DIR), SR-Mr(AM-DIS), SR-Mr(AM-EXT), SR-Mr(AM-LOC), SR-Mr(AM-MNR), SR-Mr(AM-MOD), SR-Mr(AM-NEG), SR-Mr(AM-PNC), SR-Mr(AM-PRD), SR-Mr(AM-REC), SR-Mr(AM-TMP), SR-Mrv(*), SR-Mrv(*
)_b, SR-Mrv(*)_i, SR-Mrv(A0), SR-Mrv(A1), SR-Mrv(A2), SR-Mrv(A3), SR-Mrv(A4), SR-Mrv(A5), SR-Mrv(AA), SR-Mrv(AM-ADV), SR-Mrv(AM-CAU), SR-Mrv(AM-DIR), SR-Mrv(AM-DIS), SR-Mrv(AM-EXT)
, SR-Mrv(AM-LOC), SR-Mrv(AM-MNR), SR-Mrv(AM-MOD), SR-Mrv(AM-NEG), SR-Mrv(AM-PNC), SR-Mrv(AM-PRD), SR-Mrv(AM-REC), SR-Mrv(AM-TMP), SR-Nv, SR-Ol, SR-Or, SR-Or(*), SR-Or(*)_b, SR-Or(*
)_i, SR-Or(A0), SR-Or(A1), SR-Or(A2), SR-Or(A3), SR-Or(A4), SR-Or(A5), SR-Or(AA), SR-Or(AM-ADV), SR-Or(AM-CAU), SR-Or(AM-DIR), SR-Or(AM-DIS), SR-Or(AM-EXT), SR-Or(AM-LOC), SR-Or(AM
-MNR), SR-Or(AM-MOD), SR-Or(AM-NEG), SR-Or(AM-PNC), SR-Or(AM-PRD), SR-Or(AM-REC), SR-Or(AM-TMP), SR-Or_b, SR-Or_i, SR-Orv, SR-Orv(*), SR-Orv(*)_b, SR-Orv(*)_i, SR-Orv(A0), SR-Orv(A
1), SR-Orv(A2), SR-Orv(A3), SR-Orv(A4), SR-Orv(A5), SR-Orv(AA), SR-Orv(AM-ADV), SR-Orv(AM-CAU), SR-Orv(AM-DIR), SR-Orv(AM-DIS), SR-Orv(AM-EXT), SR-Orv(AM-LOC), SR-Orv(AM-MNR), SR-O
rv(AM-MOD), SR-Orv(AM-NEG), SR-Orv(AM-PNC), SR-Orv(AM-PRD), SR-Orv(AM-REC), SR-Orv(AM-TMP), SR-Orv_b, SR-Orv_i, SR-Ov, SR-Pr(*), SR-Rr(*) }
```