*t*Search: Flexible and Fast Search over Automatic Translations for Improved Quality/Error Analysis

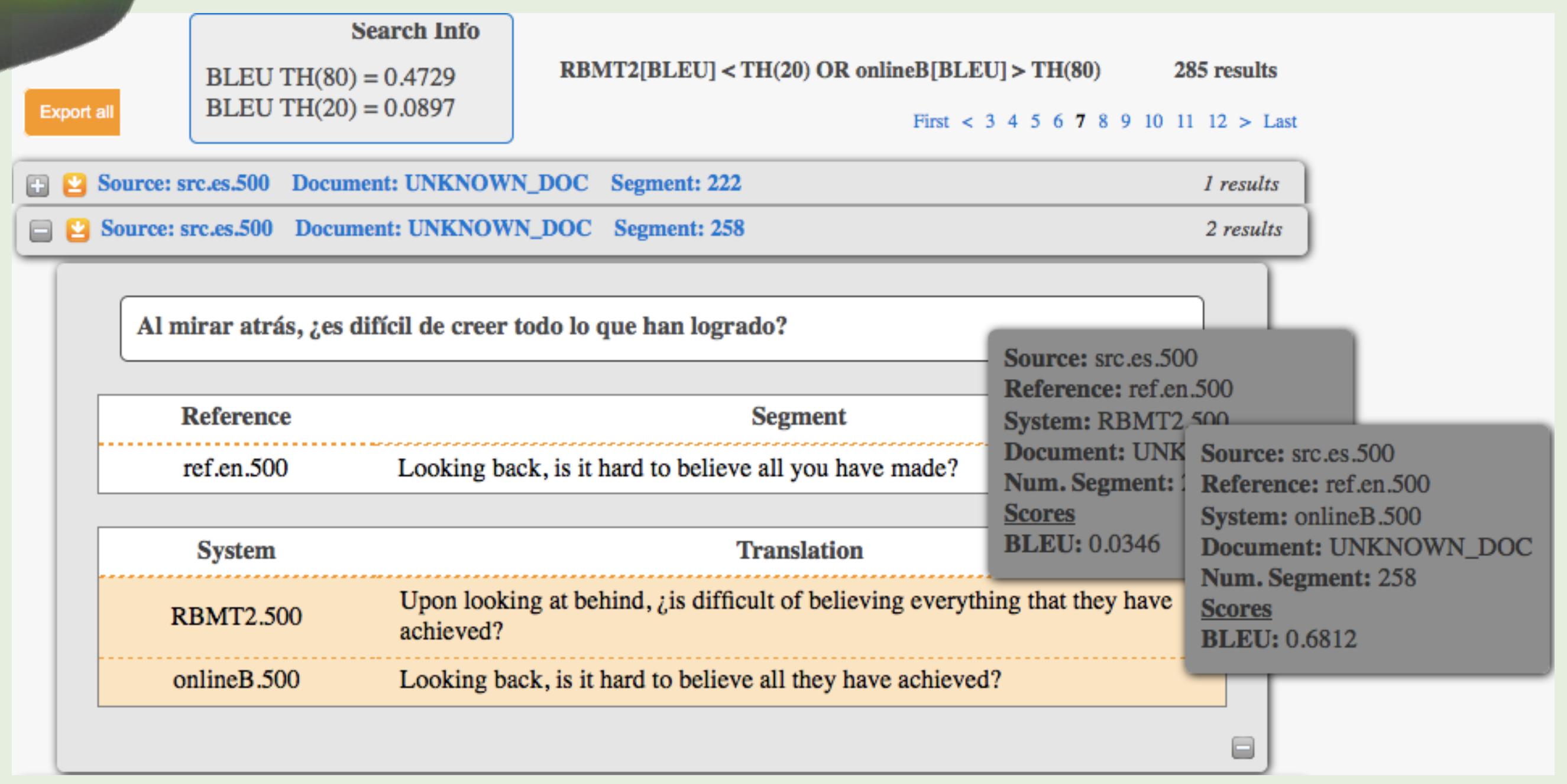
Meritxell Gonzàlez, Laura Mascarell and Lluís Màrquez
TALP Research Center, Univesitat Politècnica de Catalunya

What does *t*Search do?

- ✓ The analysis of MT systems is complex. It becomes very hard when it involves several systems, a large set of diverse measures and a high number of sentences
- ✓ The following is an example taken from WMT2012 test set. It consists of: 12 systems and 3003 sentences

	GTH-UPM	RBMT2	onlineB	uk-dan-moses
BLEU	29.67	22.88	38.90	23.98
METEOR-pa	32.87	32.37	36.60	29.65
SP-Op(*)	46.22	42.17	52.11	40.47
SP-Oc(*)	49.01	43.58	53.67	41.82
CP-Op(*)	45.77	41.43	50.75	29.41
CP-Oc(*)	43.22	37.59	47.33	35.85
DP-Oc(*)	32.80	28.69	37.77	28.03
DP-Or(*)	24.04	20.41	27.75	19.07
NE-Oe(*)	30.10	32.68	38.59	32.36
SR-Or(*)	23.52	18.71	28.10	17.06

- ✓ *t*Search speeds up the qualitative analysis of a testbed
- ✓ It helps to discover the translation errors and identify the system's weaknesses
- ✓ **Example**: *segments from RBMT2 having the lowest BLEU scores and from onlineB having the highest BLEU scores*



Search Info
BLEU TH(80) = 0.4729
BLEU TH(20) = 0.0897

RBMT2[BLEU] < TH(20) OR onlineB[BLEU] > TH(80)   285 results

Source: src.es.500 Document: UNKNOWN_DOC Segment: 222 — *1 results*
Source: src.es.500 Document: UNKNOWN_DOC Segment: 258 — *2 results*

Al mirar atrás, ¿es difícil de creer todo lo que han logrado?

Reference	Segment
ref.en.500	Looking back, is it hard to believe all you have made?

System	Translation
RBMT2.500	Upon looking at behind, ¿is difficult of believing everything that they have achieved?
onlineB.500	Looking back, is it hard to believe all they have achieved?

Source: src.es.500
Reference: ref.en.500
System: RBMT2.500
Document: UNK
Num. Segment:
Scores
BLEU: 0.0346

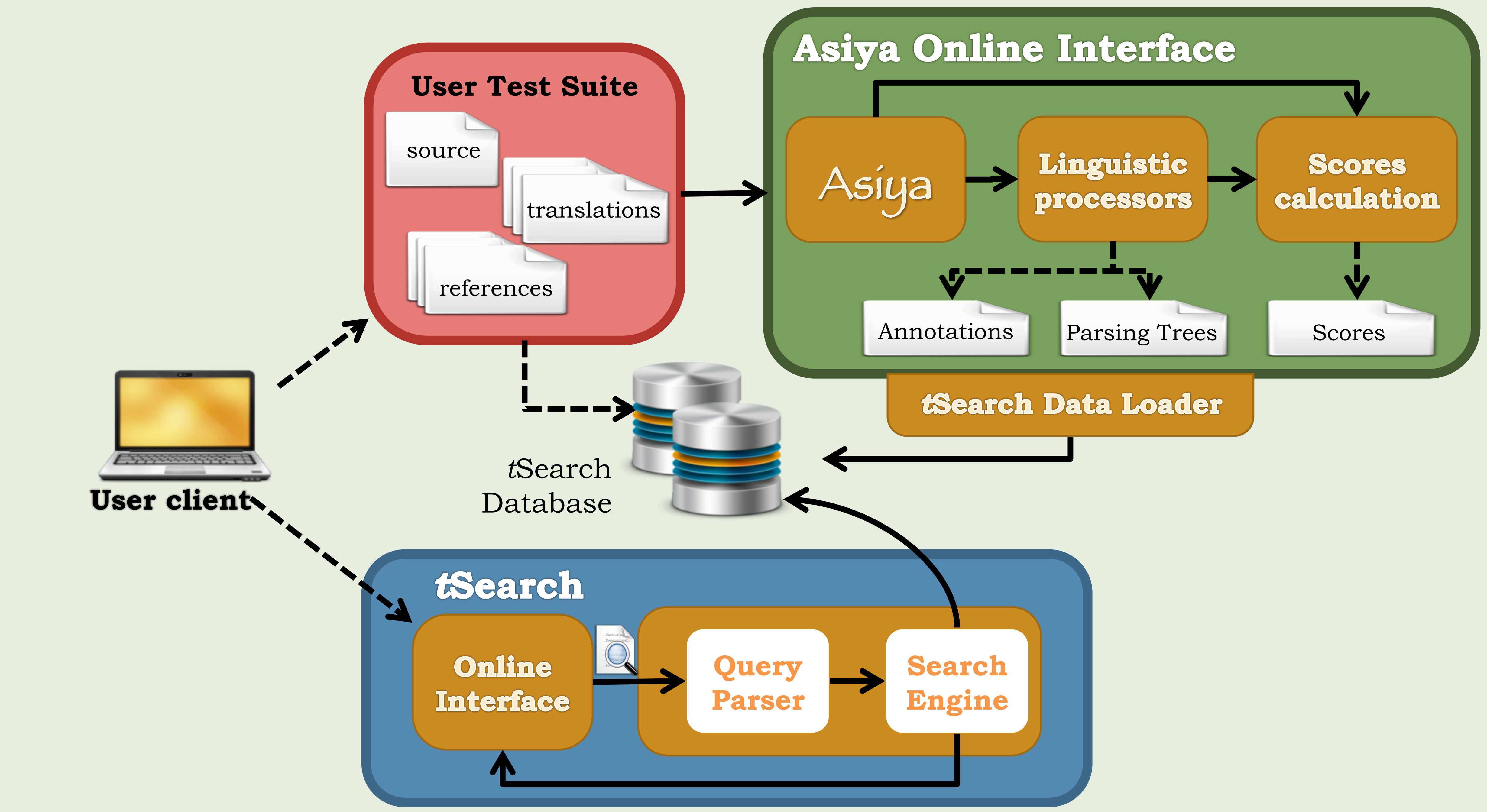
Source: src.es.500
Reference: ref.en.500
System: onlineB.500
Document: UNKNOWN_DOC
Num. Segment: 258
Scores
BLEU: 0.6812

The *t*Query Language

- ✓ Queries can be applied at segment-, document- and/or system-level
- ✓ Creation of group of systems or metrics limit the search to certain types of systems (e.g., rule-based vs. statistical) or specific metrics (e.g., lexical vs. syntactic)

Query Type	Examples	Description
Arithmetic Comparison	➢ BLEU > 0.4	Operators: >, <, >=, <=, =
Statistical Functions	➢ BLEU > AVG ➢ BLEU > TH(40)	Precalculated statistical variables: average, median, min, max, percentiles [1..100], thresholds.
Range	➢ BLEU IN [0.2,0.3) ➢ BLEU IN Q(4) ➢ BLEU IN [TH(20),TH(30)]	In a range of values
Linguistic Elements	➢ LE[SP(NN), DP(conj), CP(PP)] ➢ LE[SP(Fz, VC, Vai)]	Llinguistic processor: shallow (SP), constiency (CP), dependency (DP), semantic (SR). Linguistic elements (LE): PoS, lemma, Nes, categories, relationships, roles and even N-grams
Logical Composition	➢ BLEU > 0.5 AND –PER < 0.7	Logical operators to concatenate several conditions
System- and document- level queries	➢ s_1:BLEU > 0.3 ➢ news:BLEU > 0.3 ➢ s_1:news:BLEU > 0.3	Segments from system s_1 translations (1) , from the news document (2), and both (3), having a BLEU score above 0.3
Groups of Systems and/or Metrics	➢ s_{rb}:LEX > AVG AND s_{rb}:SYN < AVG Where, s_{rb}={s_1, s_2}; LEX={BLEU, NIST} and SYN={CP-Op(*), SP-Oc(*)}	Segments from s_{rb} having good scores for lexical measures and bad scores for syntactic measures, and same segments for s_3 having bad and good scores for lexical and syntactic measures, respectively.

*t*Search Architecture



The *t*Database

- ✓ **High volume** of data per testbed
- ✓ **High speed** response for complex queries
- ✓ NoSQL (Cassandra Apache)
- ✓ Data model based on Column Families
  CF = set of rows uniquely identified
- ✓ Each row have **a set of columns** as values

Scores CF

CF keys	CF values				
BLEU	**0.2**	**…**	**0.6**	**1.0**	
	s_1{seg3,seg5}, s_2{seg4}	…	*s_1{seg8}*	*s_2{seg8}*	
MTR-ex	**0.0**	**…**	**0.8**	**0.85**	**1.0**
	s_1{seg2}	…	*s_1{seg6}, s_2{seg7}*	*s_2{seg5}*	*s_2{seg8}*

Stats CF

CF keys	CF values							
	MIN	**MAX**	**AVG**	**MEDIAN**	**PERC(1)**	**..**	**PERC(50)**	**PERC(100)**
BLEU	0.0	1.0	0.34	0.27	0.0-0.1	…	0.34-0.36	0.99-1.0
MTR-ex	0.1	1.0	0.83	0.87	0.1-0.2	…	0.83-0.83	1.0-1.0

Linguistic Elements CF

CF keys	CF values					
SP	**DT**	**NN**	**VBZ**	**NNP**	**JJ**	
	s_1{seg1,seg2}, s_2{seg1,seg2}	*s_2{seg1}*	*s_1{seg1,seg2}*	…	*s_2{seg1}*	
CP	**ADJP**	**ADVP**	**CONJP**	**NP**	**PP**	**WHPP**
	s_1{seg3}	…	*s_1{seg1}, s_2{seg2,seg5}*	*s_1{seg1,seg2,seg3}*	*s_2{seg1}*	…
DP	**N_nsubj_V**	**D_nsubj_V**	**C_cc_V**	**I_prep_N**	**N_pobj_I**	**M_aux_V**
	s_1{seg1,seg2,seg3}	*s_2{seg4}*	*s_1{seg1,seg2}*	…	*s_2{seg1}*	*s_2{seg1}*
SR	**A0**	**A1**	**AM-TMP**	**AM-ADV**	**AM-LOC**	**R-AM-LOC**
	s_1{seg1}, s_2{seg2,seg5}	…	*s_1{seg2}*	*s_1{seg1,seg2}, s_2{seg1,seg2}*	*s_2{seg1}*	…

This research has been partially funded by the Spanish Ministry of Education and Science (OpenMT-2, TIN2009-14675-C03), the European Community's Seventh Framework Programme under grant agreement number 247762 (FAUST, FP7-ICT-2009-4-247762) and the EAMT Sponsorhip of Activities: Small research and development project, 2012.