

Incorporating Morphosyntactic Information into Word Alignment and Translation Experiments

(TC-STAR OpenLab 2006 Presentation)



Centre de Tecnologies i Aplicacions del Llenguatge i la Parla

UNIVERSITAT POLITÈCNICA DE CATALUNYA

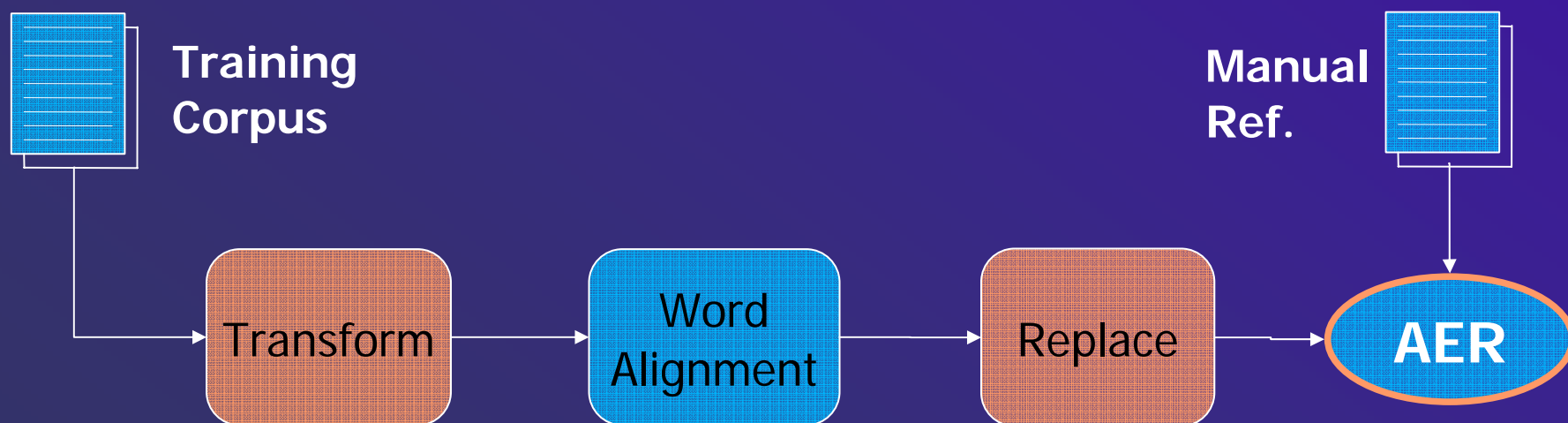
Adrià de Gispert
TALP Research Center
Universitat Politècnica de Catalunya (UPC)

Outline

- **Word Alignment experiments**
 - Experiment description
 - Base forms, Stemming, Full Verb classification, etc.
- Translation experiments
- Next steps

Word Alignment Experiments

- In cooperation with ITC-irst and RWTH
- Evaluation using Alignment Error Rate
- Manual Reference of 400 sentences (~12K words/language)
- Large vs. Small Data Task (1% material)
- Baseline vs **Data Modifications**



Word Alignment Experiments

- Baseline
 - GIZA++ with 50 classes/lang
 - lower case text
- Data Modifications considered:
 - Base forms

Example

asian countries have followed our example too .



asian country have follow our example too .

los países asiáticos han seguido también nuestro ejemplo .



el país asiático haber seguir también nuestro ejemplo .

Word Alignment Experiments

- Baseline
 - GIZA++ with 50 classes/lang
 - lower case text
- Data Modifications considered:
 - Base forms
 - Stemming

Example

asian countries have followed our example too .



asian countri have follow our exampl too .

los países asiáticos han seguido también nuestro ejemplo .



los país asiátic han segu también nuestr ejempl .

Word Alignment Experiments

- Baseline
 - GIZA++ with 50 classes/lang
 - lower case text
- Data Modifications considered:
 - Base forms *Example*
 - Stemming asian countries have followed our example too .
 - Spanish Noun-Adjective swapping

los países asiáticos han seguido también nuestro ejemplo .



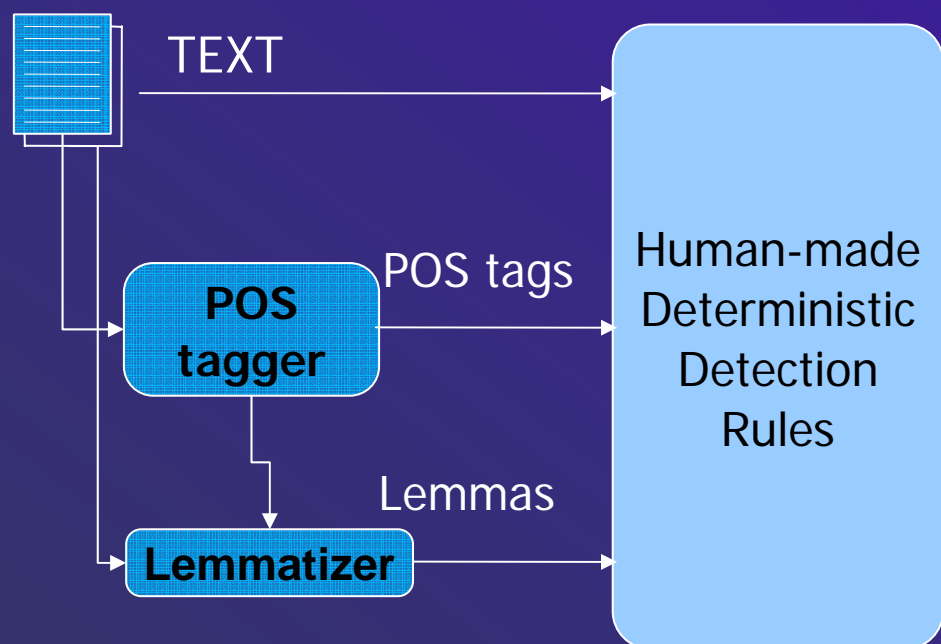
los asiáticos países han seguido también nuestro ejemplo .

Word Alignment Experiments

- Baseline
 - GIZA++ with 50 classes/lang
 - lower case text
- Data Modifications considered:
 - Base forms
 - Stemming
 - Spanish Noun-Adjective swapping
 - Full verb form classification
 - other ...

Full verb form Classification

- Human-made deterministic detection rules



Example

asian countries V[follow] our example too .
(have followed)

los países asiáticos V[seguir] también nuestro
ejemplo .
(han seguido)

- Details: English 'gapped' forms
- Alternative: Statistical chunkers, but formats, pronouns, etc...

AER Results

- AER Results

	1 % CORPUS			FULL CORPUS		
	Recall	Precision	AER	Recall	Precision	AER
baseline	73.37	69.43	28.77	78.42	86.43	17.56
base forms	73.93	75.01	25.51	76.73	87.90	17.82
stems	74.66	75.65	24.82	77.81	88.94	16.74
Spa N-A	73.43	69.59	28.65	78.10	85.93	17.97
verbs	73.96	71.36	27.45	78.60	87.37	16.97
verbs + stems	75.47	75.17	24.69	78.36	88.82	16.42

AER Results

- AER Results

	1 % CORPUS			FULL CORPUS		
	Recall	Precision	AER	Recall	Precision	AER
baseline	73.37	69.43	28.77	78.42	86.43	17.56
base forms	73.93	75.01	25.51	76.73	87.90	17.82
stems	74.66	75.65	24.82	77.81	88.94	16.74
Spa N-A	73.43	69.59	28.65	78.10	85.93	17.97
verbs	73.96	71.36	27.45	78.60	87.37	16.97
verbs + stems	75.47	75.17	24.69	78.36	88.82	16.42

- Combining 3 best (FULL CORPUS):

Recall	Precision	AER
78.50	90.04	15.79

Baseline system

- Log-linear combination of multiple Statistical Models
 - Bilingual N-gram translation model (vs Phrase-Based)
 - Target Language Model, Word Penalty
 - IBM 1 lexicon model (src -> trg , trg -> src)
- Derived from maximum entropy approach

$$\hat{t}_1^I = \operatorname{argmax}_{t_1^I} \sum_{m=1}^M \lambda_m h_m(t_1^I, s_1^J)$$

- Weight optimization on development set (simplex alg.)

Bilingual N-gram model

- Standard N-gram of bilingual units (tuples)

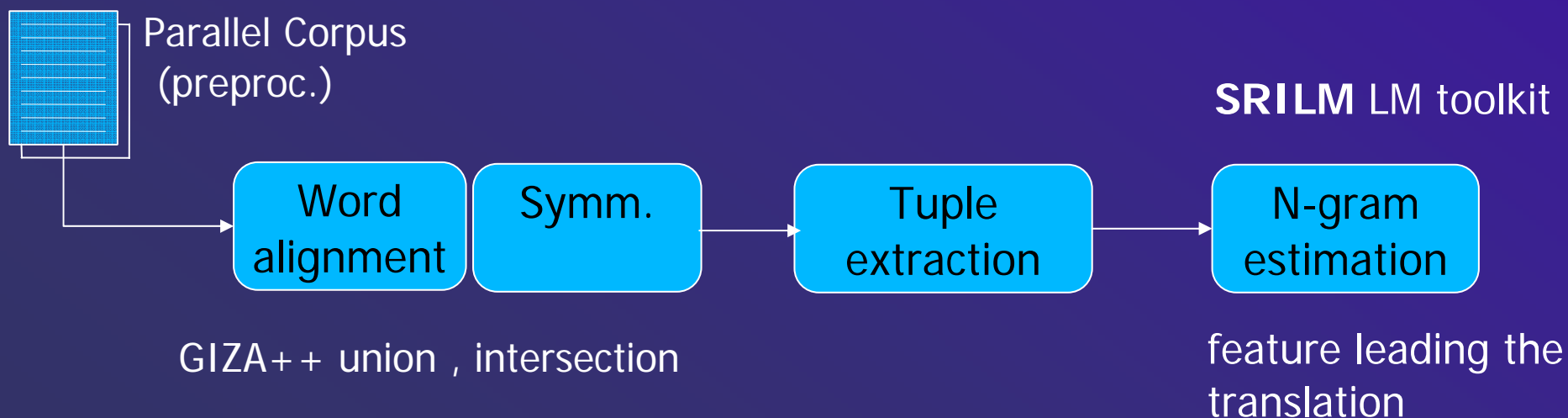
3-gram

$$h_{TM}(t, s) = \log \prod_{n=1}^N p((t, s)_n | (t, s)_{n-1}, (t, s)_{n-2})$$

tuple

$$(t, s)_n = (t_{i(n)} \dots t_{i(n)+I(n)}, s_{j(n)} \dots s_{j(n)+J(n)})$$

- Training



Additional feature models (1)

- N-gram Target Language Model

3-gram

$$h_{LM}(t,s) = h_{LM}(t) = \log \prod_{n=1}^I p(t_i | t_{i-1}, t_{i-2})$$

SRILM LM toolkit

- Word Penalty

*Compensates the LM
preference for short translations*

$$h_{WP}(t,s) = I$$

Additional feature models (2)

- IBM 1 lexicon model

*from src-to-trg
alignment*

$$h_{IBM1}(t,s)_n = \log \frac{1}{(I'+1)^{J'}} \prod_{j=1}^{J'} \sum_{i=0}^{I'} p(t_i | s_j)$$

GIZA++ model 1

- IBM 1 inverse lexicon model

*from trg-to-src
alignment*

$$h_{IBM-1}(t,s)_n = \log \frac{1}{(J'+1)^{I'}} \prod_{i=1}^{I'} \sum_{j=0}^{J'} p(s_j | t_i)$$

Error Analysis

- Eng→Spa examples:

: Mr | President | , | with | the | special sitting | on | 21 | December | , |
 : Señor | Presidente | , | con | la | sesión especial | NULL | del 21 | de diciembre | , |

you | have | given | us | something | of | an | early | Christmas present | . |
 NULL | NULL | NULL | NULL | NULL | de | una | pronta | regalo de Navidad | . |

: Will | we | get | more | information | from | this | sitting | than | we | will |
 : ¿ Se | NULL | obtener | más | información | de | esta | sesión | que | NULL | NULL |

be able | to | get | from | the | media | ? |
 pueda | NULL | obtener | de | los | medios de comunicación | ? |

: I call | for | greater | transparency | , | immediate information | and | full publication |
 : Pido | NULL | una mayor | transparencia | , | información inmediata | y | publicación completa

of | the | analyses | relating | to | both | cases | . |
 de | los | análisis | relativas | a los | dos | casos | . |

Target POS Language Model

- Addresses agreement problems
- Incorporated as additional feature during decoding
- Does not require POS-tagging during translation (only in training)

TRAIN tuples	trgPOS
I declare # Declaro	VMIP1S0
resumed # reanudado	VMP00SM
the # el	DA0MS0
session # periodo de sesiones	NCMS000 SPS00 NCFP000
of # del	SPCMS
the # NULL	
European Parliament # Parlamento Europeo	NCMS000 AQ0MS0
adjourned on Thursday , # , interrumpido el	Fc VMP00SM DA0MS0
28 # 28	Z
March # de marzo	SPS00 NCMS000
1996 # de 1996	SPS00 Z
. # .	Fc
....

Target POS LM: Results

- EPPS task (2004 data)

		BLEU	mWER
Spa-to-Eng	baseline	0.5412	34.98
	+POSLM	0.5461	34.47
Eng-to-Spa	baseline	0.4714	40.22
	+POSLM	0.4750	40.42

- Not a big impact in BLEU
- Qualitative improvement on manual inspection

: I call | for | greater | transparency | , | immediate information | and | full publication |
 : Pido | NULL | una mayor | transparencia | , | información inmediata | y | publicación completa

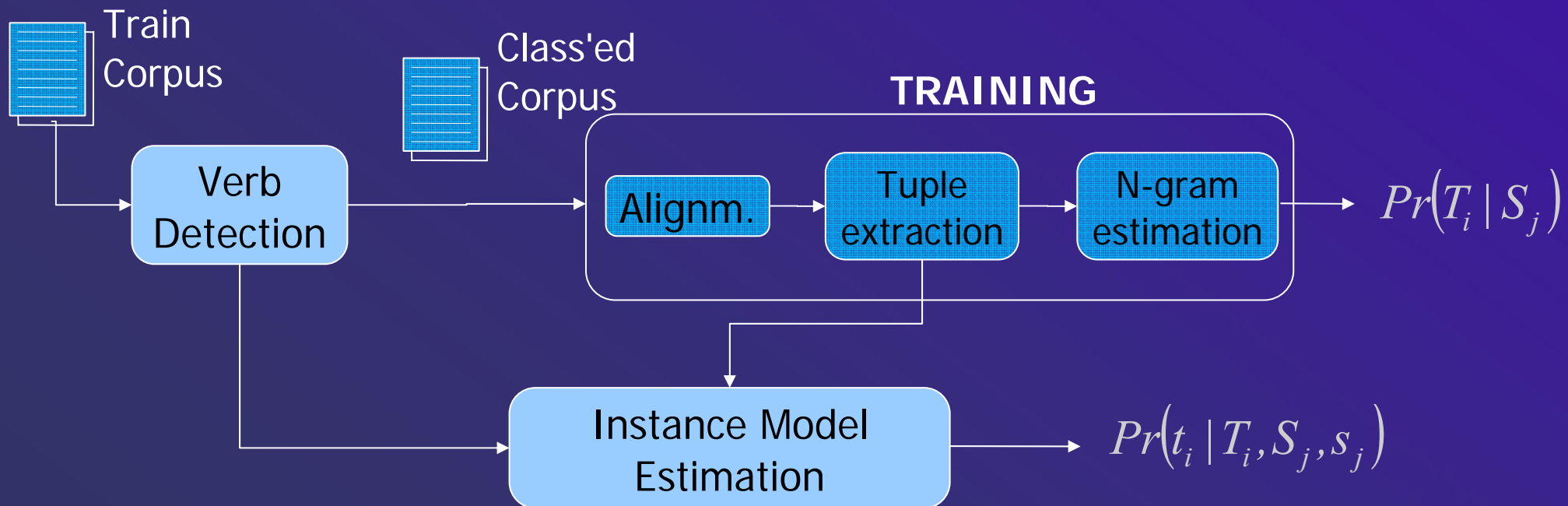
of | the | analyses | **relating** | to | both | cases | . |
 de | los | análisis | **relativos** | a los | dos | casos | . |

Verb instance model

- **Goal:** Developing a model incorporating linguistic classes of translation units
 - Verb phrases

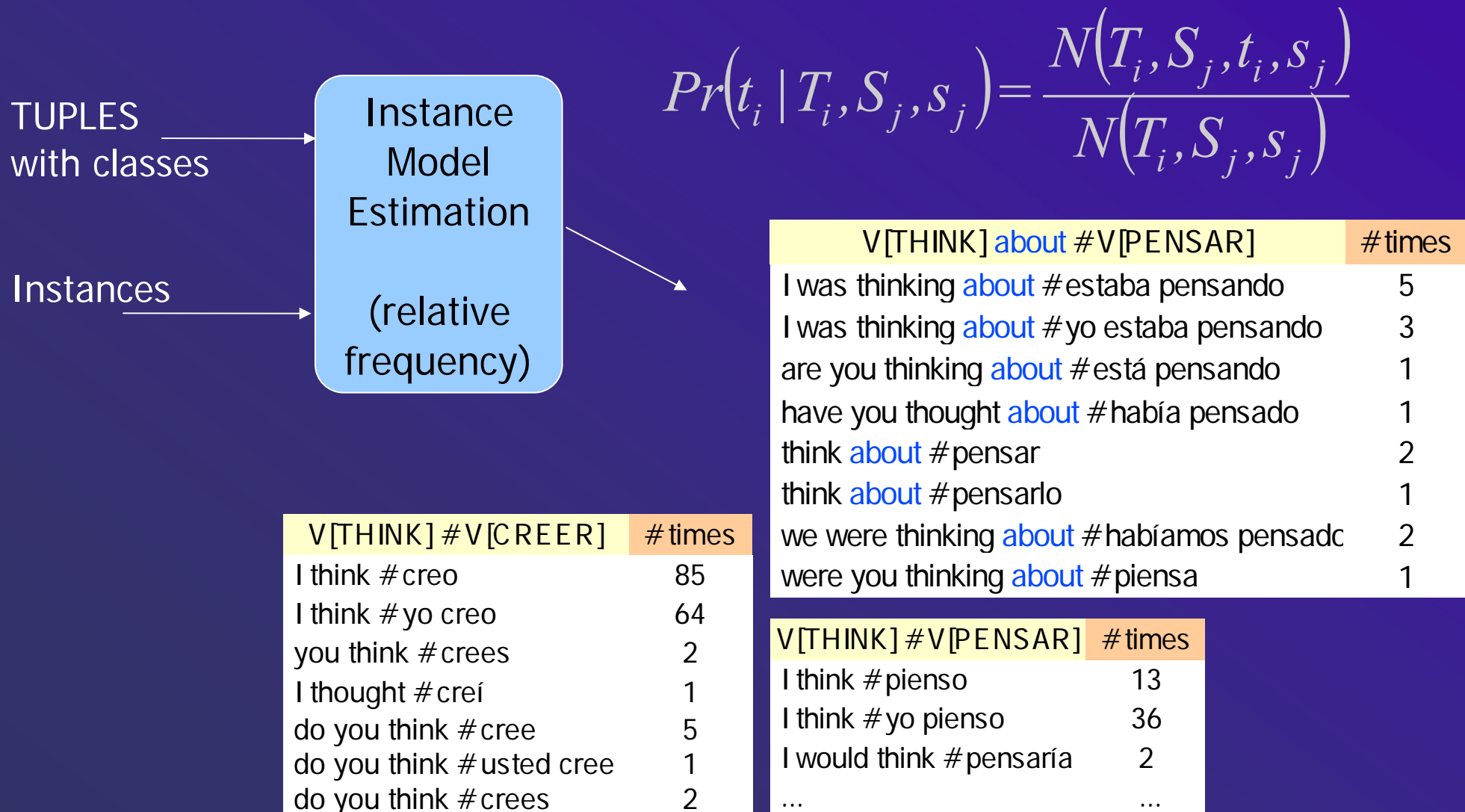
$$h_{TM}(t_i | s_j) \approx Pr(T_i | S_j) Pr(t_i | T_i, S_j, s_j)$$

- Proposed scheme:



Verb instance model definition

- Guides the decoder through the target instances given the source instance and the classes defined by the tuple



Verb forms: Generalization

- Unseen verb phrases

V[THINK]
(we think)

V[THINK] # V[CREER]	# times
I think # creo	85
I think # yo creo	64
you think # crees	2
I thought # creí	1
do you think # cree	5
do you think # usted cree	1
do you think # crees	2

V[THINK] # V[PENSAR]	# times
I think # pienso	13
I think # yo pienso	36
I would think # pensaría	2
...	...

V[THINK] # V[TENER] V[PENSAR]	# times
we thought # teníamos pensado	2
were you thinking # tenía pensado	3
...	...

...

Verb forms: Generalization

- Unseen verb phrases

V[THINK]
(we think)

V[THINK] # V[CREER]	# times
I think # creo	85
I think # yo creo	64
you think # crees	2
I thought # creí	1
do you think # cree	5
do you think # usted cree	1
do you think # crees	2

we think # **creemos** (87/200)

we think # **nosotros creemos**
(64/200)

V[THINK] # V[PENSAR]	# times
I think # pienso	13
I think # yo pienso	36
I would think # pensaría	2
...	...

we think # **pensamos** (13/200)

we think # **nosotros pensamos**
(36/200)

V[THINK] # V[TENER] V[PENSAR]	# times
we thought # teníamos pensado	2
were you thinking # tenía pensado	3
...	...

...

Verb Phrases: Results Eng→Spa

- LC-STAR task (30K sentences)

	mWER	BLEU
baseline	23.16	0.671
class	22.22	0.686
class + gen	21.65	0.692

- EPPS task (2005 data, ~1.2M sentences)

	mWER	BLEU
baseline	40.42	0.479
class	41.37	0.465
class + gen	41.29	0.466

Next steps

- Verb instance model Error Analysis
 - Why is harmful in large-vocabulary task?
 - How can we address the problem better?
- Use of additional linguistical error measures
- Other more-structured models
 - Defined on POS, base forms and words
 - To capture more complex relationships
 - Syntax-based approaches

Thanks for your attention

Grazie mille!



*Centre de Tecnologies i Aplicacions del Llenguatge i la Parla
TALP Research Center
Universitat Politècnica de Catalunya (UPC)
Barcelona*

