Cross-Lingual Lexico-Semantic Transfer in Language Learning

Cambridge ALTA

{ Ekaterina Kochmar, Ekaterina Shutova } Computer Laboratory

University of Cambridge

Introduction

We investigate *whether* and *to what extent* the lexico-semantic models of the **native language** (L1) are transferred to the **second language** (L2).

E.g.: **push the trigger* [Es/Ru] =
pull the trigger [En]

⇒ errors of lexical choice by Ru and Es L1 speakers learning En as L2

Contributions

1. We focus on lexical choice and investigate it in the context of 3 typologically diverse languages: Russian (Ru), Spanish (Es) and English (En).

2. We show that a statistical semantic model learned from L1 data improves automatic error detection in L2 for the speakers of that L1.

3. We investigate whether the semantic model learned from a particular L1 is portable to other, typologically related languages.

Data and Approach

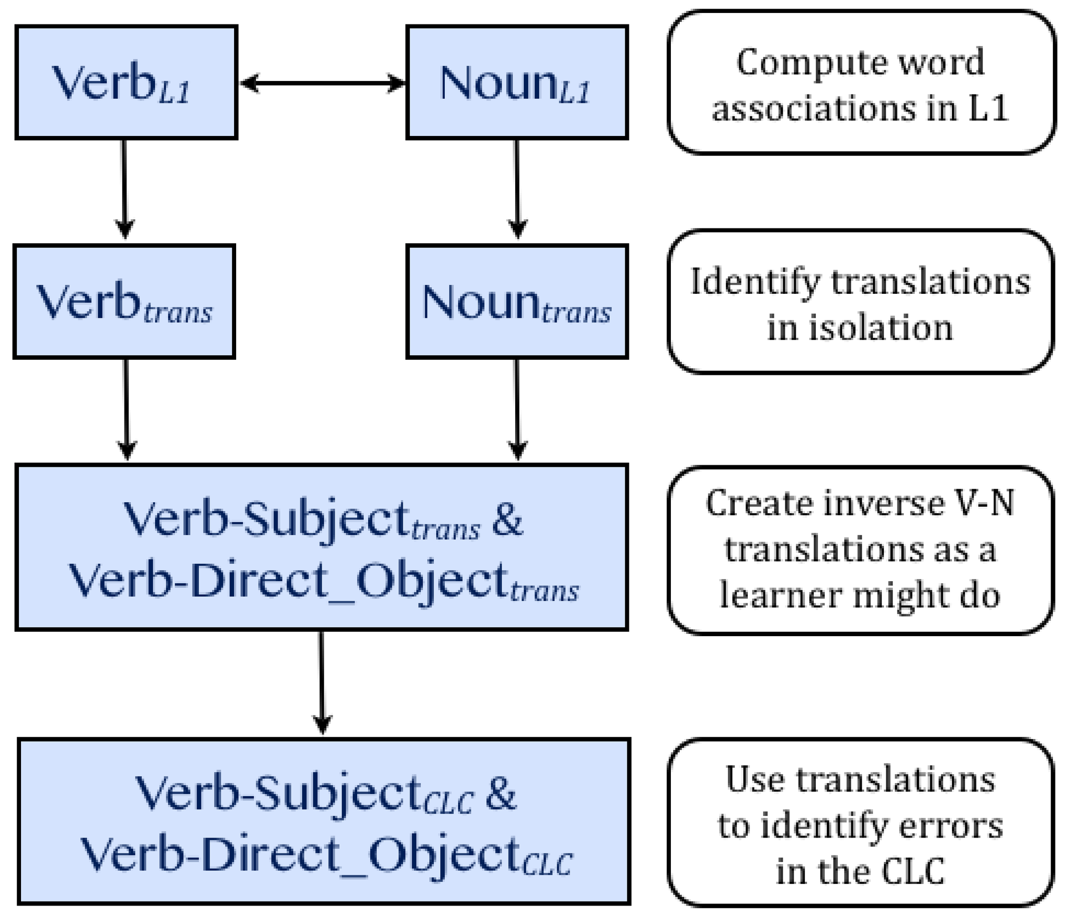

Figure 1: The use of L1 data

♦ **Grammatical relations:** VERB-SUBJECT and VERB-DIRECT_OBJECT

♦ **Spanish data:** extracted from the Spanish Gigaword and parsed using the Spanish Malt parser

♦ **Russian data:** extracted from the RU-WaC corpus and parsed using the Russian Malt parser

♦ **Dictionaries and translation:** English-Spanish and English-Russian editions of Wiktionary

♦ **L2 data:** BNC and UKWaC parsed with RASP

♦ **Learner data:** CLC preprocessed with RASP; CLC error annotation used to split the data into correct combinations and errors

Method

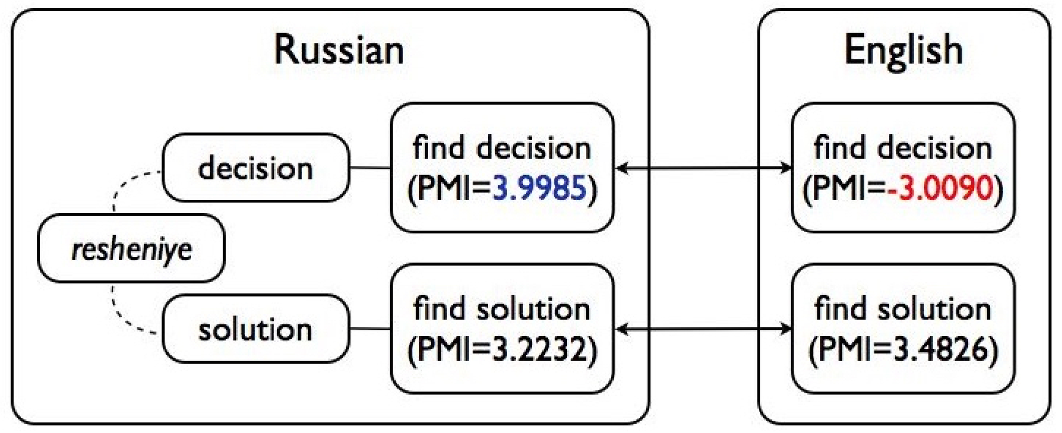

Figure 2: Lexico-semantic transfer Ru → En

Method: Binary classification, linear SVM

Features:

♦ **L2 lexico-semantic features:**
- pmi in L2
- verb and noun identity
- semantic vector space features

♦ **L1 lexico-semantic features:**
- pmi in L1
- difference between the L1 and L2 PMI

Evaluation: Accuracy (Acc) & Precision (P_e), Recall (R_e) and F_{1_e} on *errors*

Baseline: Frequency of occurrence

Experimental Results

L1	Features	Acc	P_e	R_e	F_{1_e}
Ru_{dobj}	baseline	55.68	47.77	61.44	53.55
	ft_{En}	64.79	59.87	47.56	53.01
	$+\text{pmi}_{L1}$	**66.05**	58.74	**62.72**	**60.67**
Ru_{subj}	baseline	54.48	46.30	63.96	53.17
	ft_{En}	67.64	59.88	62.17	60.98
	$+\text{pmi}_{L1}$	**68.68**	**62.10**	**69.61**	**64.38**
Es_{dobj}	baseline	56.74	52.25	74.44	61.36
	ft_{En}	64.34	61.80	59.67	60.71
	$+\text{pmi}_{L1}$	**66.89**	**63.01**	**68.61**	**65.68**
Es_{subj}	baseline	54.45	46.71	70.31	56.00
	ft_{En}	69.51	61.79	68.58	65.00
	$+\text{pmi}_{L1}$	**71.19**	**62.10**	**77.66**	**69.00**

Table 1: System performance (in %) using L2 lexico-semantic features, L1 → L1_{CLC}.

L1	Features	Acc	P_e	R_e	F_{1_e}
Ru_{dobj}	baseline	55.13	50.17	72.14	58.99
	ft_{En}	63.58	59.73	57.98	58.85
	$+\text{pmi}_{L1}$	**64.60**	58.81	**70.69**	**64.20**
Ru_{subj}	baseline	54.56	47.95	71.10	56.71
	ft_{En}	64.42	57.27	62.64	59.83
	$+\text{pmi}_{L1}$	**64.99**	57.24	**68.17**	**62.21**
Es_{dobj}	baseline	59.35	55.38	71.87	62.51
	ft_{En}	64.32	61.89	63.47	62.67
	$+\text{pmi}_{L1}$	**65.75**	**61.90**	**71.37**	**66.30**
Es_{subj}	baseline	58.34	50.90	66.97	57.48
	ft_{En}	65.57	58.32	64.09	61.06
	$+\text{pmi}_{L1}$	**66.54**	**58.80**	**68.72**	**63.36**

Table 2: System performance (in %) using L1 and L2 lexico-semantic features, L1 → ALL L1s.

Effect on Related L1s

L1	Features	Acc	P_e	R_e	F_{1_e}
Ru_{dobj}	baseline	57.08	51.80	71.58	59.78
	ft_{En}	64.20	60.99	55.36	58.04
	$+\text{pmi}_{L1}$	**65.77**	**61.06**	**64.78**	**62.86**
Ru_{subj}	baseline	56.43	49.52	62.04	54.24
	ft_{En}	62.26	55.84	50.02	52.76
	$+\text{pmi}_{L1}$	**62.78**	**56.02**	**54.48**	**55.21**
Es_{dobj}	baseline	59.18	51.44	72.31	59.97
	ft_{En}	65.14	59.82	53.83	56.66
	$+\text{pmi}_{L1}$	**66.24**	58.92	**67.00**	**62.70**
Es_{subj}	baseline	58.10	52.95	77.43	62.45
	ft_{En}	66.29	61.24	68.45	64.64
	$+\text{pmi}_{L1}$	**67.00**	**61.68**	**70.50**	**65.78**

Table 3: System performance (in %) using L1 and L2 lexico-semantic features, L1 → L1_GROUP.

L1	Features	Acc	P_e	R_e	F_{1_e}
Ru_{dobj}	baseline	55.04	47.68	63.87	53.81
	ft_{En}	64.73	59.76	46.05	52.01
	$+\text{pmi}_{L1}$	**65.15**	**60.63**	45.77	**52.16**
Ru_{subj}	baseline	53.30	44.77	61.09	51.29
	ft_{En}	61.84	54.63	35.81	43.22
	$+\text{pmi}_{L1}$	**62.53**	**57.24**	35.11	43.18
Es_{dobj}	baseline	55.25	51.67	76.79	61.21
	ft_{En}	64.06	62.30	56.01	58.98
	$+\text{pmi}_{L1}$	**65.21**	**63.44**	**58.13**	**60.66**
Es_{subj}	baseline	54.34	47.76	68.73	56.23
	ft_{En}	62.71	58.80	43.09	49.69
	$+\text{pmi}_{L1}$	62.44	58.46	41.71	48.60

Table 4: System performance (in %) using L1 and L2 lexico-semantic features, L1 → REL_L1.

♦ **L1 → L1_{CLC}**: adding L1 lexico-semantic features to L2 features (ft_{En}) improves all measures (Table 1)

♦ **L1 → ALL L1s**: adding L1 lexico-semantic features improves **Acc** and $\mathbf{R_e}$ (Table 2)

♦ **L1 → L1_GROUP**: a minor effect on **Acc** and $\mathbf{P_e}$ and a more pronounced effect on $\mathbf{R_e}$ (Table 3)

♦ **L1 → REL_L1**: **Acc** and $\mathbf{P_e}$ improve (Table 4)

♦ **Results suggest**: there may be less semantic variation *within* a language group than *across* groups.

Conclusions

Statistical semantic models learned from L1s significantly improve error detection in L2 data produced by the speakers of the respective L1s. Moreover, L1 models improve the coverage of the error detection system on a range of other L1s.

Most reliably identified errors include lexical choice errors (e.g., **offer plan* vs. *suggest plan*, **say idea* vs. *tell idea*). Many of the errors missed by the classifier are context-dependent and do not result from an L1 lexico-semantic transfer.

Contact Information

Email Ekaterina.Kochmar@cl.cam.ac.uk
Email Ekaterina.Shutova@cl.cam.ac.uk

Data www.cl.cam.ac.uk/~ek358/cross-ling-data.html