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

DATA AND APPROACH

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-VaC 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

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), Recall (R) and F1 on errors

Baseline: Frequency of occurrence

CONCLUSIONS

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.

EXPERIMENTAL RESULTS

Table 1: System performance (in %) using L2 lexico-semantic features, L1 → L1L2C

Table 2: System performance (in %) using L1 and L2 lexico-semantic features, L1 → L1L1

RESULTS SUGGEST: there may be less semantic variation within a language group than across groups.

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

CROSS-LINGUAL LEXICO-SEMANTIC TRANSFER IN LANGUAGE LEARNING
{ Ekaterina Kochmar, Ekaterina Shutova }  |  Computer Laboratory

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.

Figure 1: The use of L1 data

Table 3: System performance (in %) using L1 and L2 lexico-semantic features, L1 → L1L1

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

L1 → L1L2C: adding L1 lexico-semantic features to L2 features (ft_L2) improves all measures (Table 1)

L1 → REL_L1: adding L1 lexico-semantic features improves Acc and R (Table 2)

L1 → L1L1: a minor effect on Acc and P1 and a more pronounced effect on R (Table 3)

L1 → REL_L1: Acc and P1 improve (Table 4)