

A simple Russian-Slovak machine translation system

Radovan Garabík

Jazykovedný ústav Ľ. Štúra

Panská 26

813 64 Bratislava

Slovakia

e-mail: korpus@korpus.juls.savba.sk

www: <http://korpus.juls.savba.sk>

Машинный перевод

- правила
- статистический
- параллельные корпус[ы,а]

текст → морфологическая разметка → снятие

морфологической омонимии → (синтаксическая разметка)

→ *трансфер* → генерирование текста

текст → *трансфер* → снятие омонимии словоформ

Словарь

- **73000 пар слов**
- **уровень омонимии 2.98**
- **общий словарь — попытка включить все возможные переводы слова, без оценки контекста или вероятностей**
- **увеличение омонимии в переводе**

Морфология

- **phr morphology**
- **прямая связь между категориями русской и словацкой морфологии**
- **род, падеж, число, лицо, повелительное наклонение, инфинитив**
- **родительный в русском ⇔ родительный или винительный в словацком (отрицательная форма глаголов)**
- **отличие в падежах после предложений**
- ***после+родительный* ⇔ *ро+местный/предложный***

- настоящее время глагола *быть*
- *он* ⇔ *оп*
- *он* ⇔ *оп је*
- посессивное предложение с *быть*
- согласение рода прилагательных и существительных
- ~50 правил морфологического отношения
- после обработки: ~300 к пар слов, уровень омонимии

11.7

Модель языка

- цепь Маркова 2-го порядка

- $$\textit{translation} = \operatorname{argmax}_{x_1 \dots x_n} \prod_{i=1}^N P(x_i | x_{i-1}, x_{i-2})$$

- вероятность перехода

- $$P(x_n | x_{n-1}, x_{n-2}) = \lambda_3 f(x_n | x_{n-1}, x_{n-2}) + \lambda_2 f(x_n | x_{n-1}) + \lambda_1 f(x_n)$$

- алгоритм Viterbi

Оценка

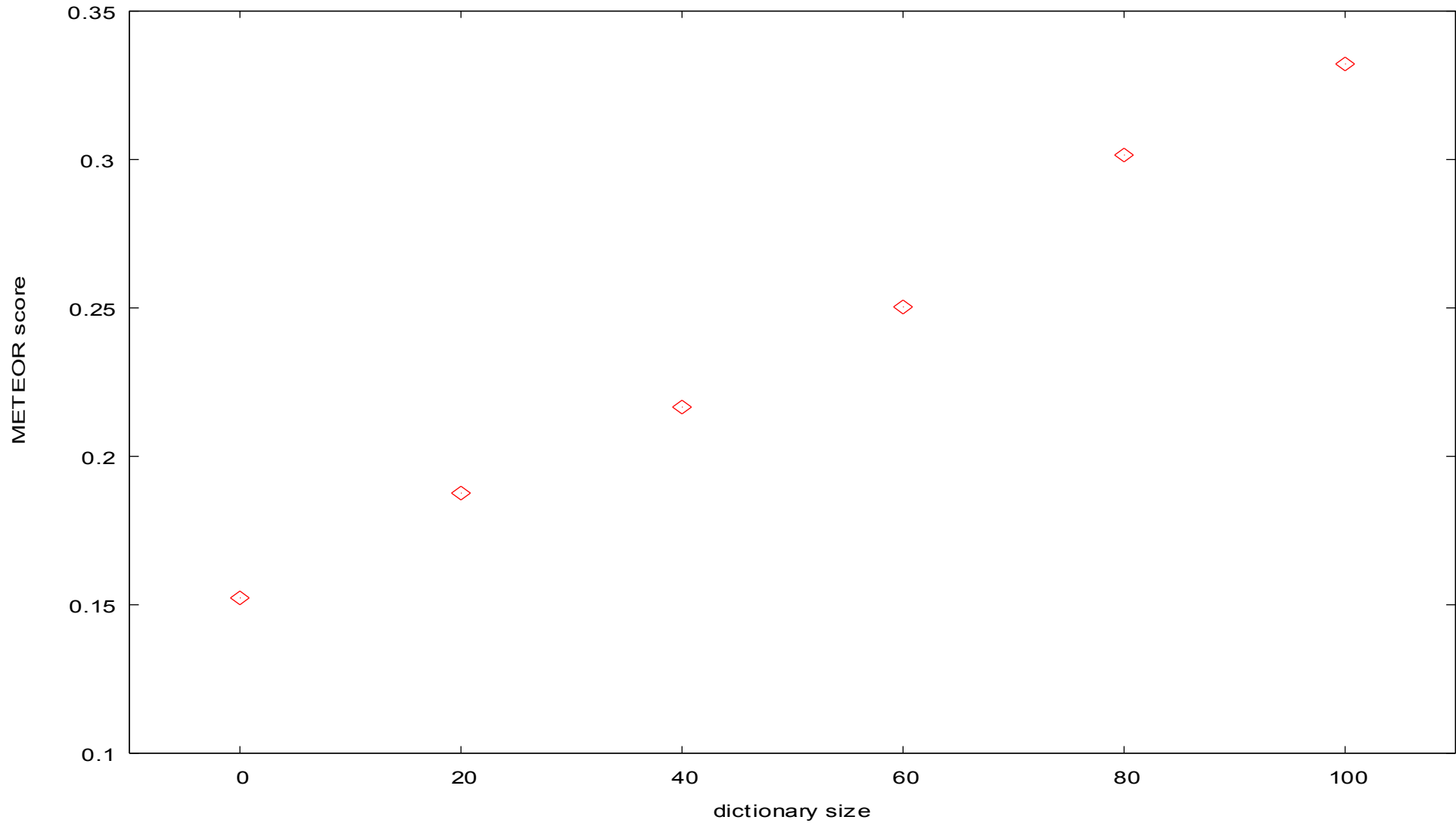
- **METEOR**
- **сравнение с переводчиком-человеком**
- **параметры чешского языка**
- **1000 произвольно избранных идеально сравненных предложений из параллельного русско-словацкого корпуса**

Транслитерация

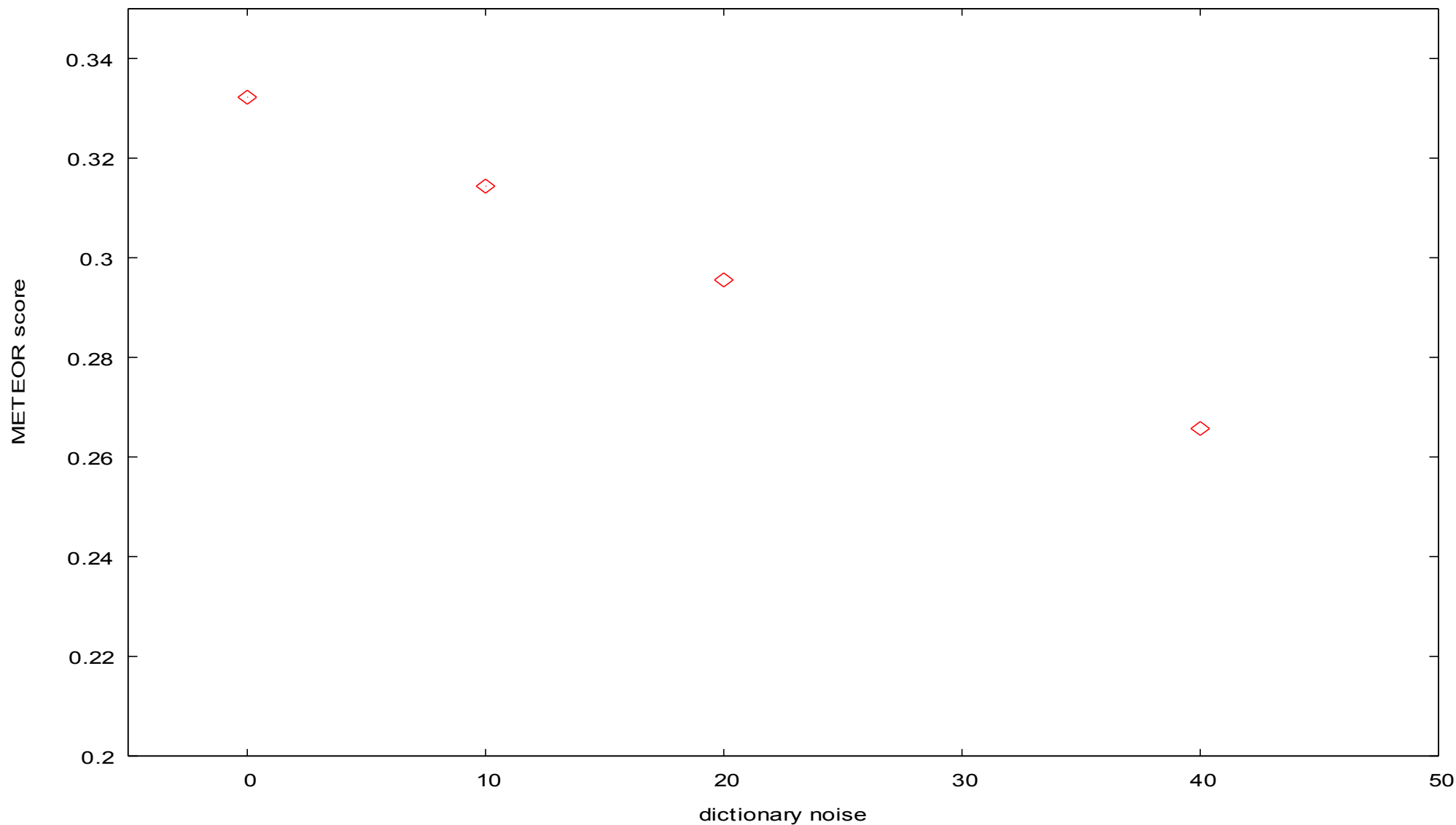
- Правила словацкого правописания — прежде всего собственные имена
- ЭТИМОЛОГИЧЕСКИЕ ИЗМЕНЕНИЯ СЛАВЯНСКИХ ЯЗЫКОВ — слова отсутствующие в словаре ~150 правил

Type	METEOR score
PSP transliteration only	0.1392
Genetic transliteration only	0.1523
Combined transliteration only	0.1490
Translation only	0.2646
Translation + PSP transliteration	0.3243
Translation + genetic transliteration	0.3322
Translation + combined transliteration	0.3279
Human 1	0.5280
Human 2	0.5103

Зависимость качества перевода от размера словаря



Зависимость качества перевода от качества словаря



Уже в 1958 г. составители будущего учебного плана исходили из того, что выпускники отделения должны иметь глубокую подготовку как по языковедению, так и по математике.

Už v 1958 h. zostavovatelia budúceho školského oblasti vychádzali z toho , že absolventi oddelenia náležití byť hlboké prípravu ako z jazykovede , tak práve pre matematike.

Спасибо за внимание