

Математические модели в лингвистике

Вводная лекция, часть 1

Мати Пентус, Александр Пиперски,
Алексей Сорокин

МГУ им. М. В. Ломоносова,
межфакультетский курс,
осенний семестр 2017–2018 учебного года

Что такое лингвистика?

Не только и не столько:

- ▶ изучение иностранных языков
- ▶ знание норм родного языка

На самом деле:

- ▶ наука о том, как устроен человеческий
Язык вообще
- ▶ наука о том, как устроены разные
языки

Норма

- ▶ Норма: осознанный выбор между несколькими существующими в языке вариантами
- ▶ Абсолютное большинство более интересных языковых решений принимается нами неосознанно

Демоверсия ЕГЭ–2018, №4

В одном из приведённых ниже слов допущена ошибка в постановке ударения: НЕВЕРНО выделена буква, обозначающая ударный гласный звук. Выпишите это слово.

нажИвший, Отрочество, жилОсь, вернА,
взЯлась

Демоверсия ЕГЭ–2018, №4

- ▶ *МЫЛ ≠ ВЁЛ ≠ ВЗЯЛ*
- ▶ Более важная задача лингвистики — описать и объяснить такие различия, а не предписать выбор между *о́трочество* и *отро́чество*.

Лингвистические явления: примеры

- ▶ Орфографические правило
касаться ~ коснуться
- ▶ Согласование типа *стоит большой дом,*
*стоят большие дома, но не *стоят*
большая дом
- ▶ Порядок слов: *часа четыре ~ четыре часа*

Венгерский язык

- ▶ *Ez könyv* 'Это книга'
- ▶ *egy könyv* 'какая-то книга'
- ▶ *ez a könyv* 'эта книга'
- ▶ *Egy könyvet lát* 'Он видит (какую-то) книгу'
- ▶ *Ez a könyvet látja* 'Он видит эту книгу'

Глагол согласуется с дополнением по определённости!

Чередования

Язык палау [Schane 1974]:

наст. (для себя)	прич. буд. (устар.)	прич. буд. (нов.)	
[mə-dáŋəb]	[dəŋób-l]	[dəŋəb-áll]	накрывать
[mə-téʔəb]	[təʔíb-l]	[təʔəb-áll]	вытягивать

- ▶ Как лучше всего описать такую ситуацию?

Фонология

- ▶ **Глубинное представление:**
[mə-dáŋob], [daŋób-l], [daŋob-áll]
- ▶ Правила перехода:
редукция безударных гласных в ə
- ▶ **Поверхностное представление:**
[mə-dáŋəb], [dəŋób-l], [dəŋəb-áll]

Правил могло бы быть и больше и их было бы сложнее организовывать!

Русское склонение

- ▶ Сколько в русском языке типов склонения?
- ▶ Достаточно ли наших обычных представлений для правильного образования всех форм?
- ▶ *стена = жена?*
- ▶ *яблоко = облако = окно = число?*
- ▶ [Зализняк 1967, 302]: около 700 типов склонения

«Грамматический словарь русского языка»

секунда	ж	1а	слобода	ж	1f
мунда	п	0 (языки мунда)	вода	ж, 1d' // устар. 1f' ◇	
ерунда	ж	1b—		на воду // на воду;	
¹ рында	мо	<жо 1а> (истор.: оруженосец)		спуститься (уйти и т. п.) под воду [// под	
² рында	ж	1а (судовой колокол)		воду]; ходить (пойти)	
дурьнда	мо-жо	1а		по воду; вымыть	
гирлянда	ж	1а		в двух (трех и т. д.)	
ода	ж	1а		водах	
свобода	ж	1а	подвода	ж	1а
несвобода	ж	1а	воевода	мо	<жо 1а>
			пагода	ж	1а

Перестановка слов

- ▶ *Peter is reading — Is Peter reading?*
- ▶ Достаточно ли для построения вопроса на английском языке правила «Поменяй местами первое и второе слово»?
- ▶ *The boy is reading — *Boy the is reading?*
- ▶ *When Jim was singing, Peter was reading — *Jim when was singing, Peter was reading?*

Сербский язык

- ▶ *Петар ми га даје* 'Пётр даёт его мне'
- ▶ *Дајете му је* 'Вы даёте её ему'
- ▶ *Даје нам их* 'Он даёт их нам'

Другие порядки слов невозможны

- ▶ ? 'Вы даёте нам его'

Применение математики в лингвистике

- ▶ 1950-е–1970-е гг.: стремление сделать лингвистику точной наукой, интерес в первую очередь к математической логике
- ▶ 1950-е гг. и далее: компьютерная лингвистика
- ▶ XXI век: использование статистических методов

Порядок русских дополнений

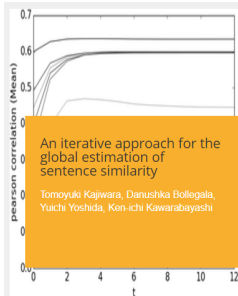
Ира дала книжку Васе vs.

Ира дала Васе книжку

Возможные факторы:

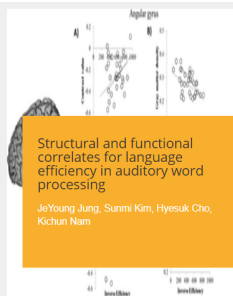
- ▶ Определённость объекта и адресата
- ▶ Прайминг (влияние ранее услышанных конструкций того же типа)
[Фёдорова 2009]
- ▶ ...

Лингвистика в PLOS ONE (1)



An iterative approach for the global estimation of sentence similarity

Tomoyuki Kajiwara, Danushka Bollegala, Yuichi Yoshida, Ken-ichi Kawarabayashi



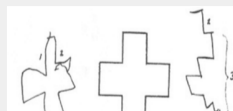
Structural and functional correlates for language efficiency in auditory word processing

JeYoung Jung, Sunmi Kim, Hyesuk Cho, Kichun Nam

	Lineup Performance		Confidence	
	False ID	No ID	False ID	No ID
Wake	68% (26)	34% (18)	3.69 (1.11)	4.58 (1.48)
Sleep	42% (18)	58% (25)	3.67 (1.48)	4.60 (1.91)

Sleep and eyewitness memory: Fewer false identifications after sleep when the target is absent...

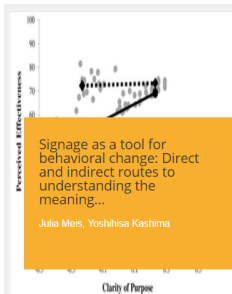
Michelle E. Stepan, Taylor M. Dehnke, Kimberly M. Fenn



Are baboons learning "orthographic" representations? Probably not

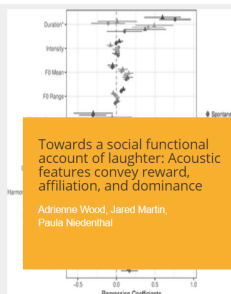
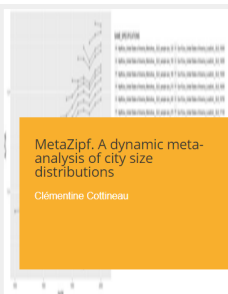
Maja Linke, Franziska Broker, Michael Ramscar, Harald Baayen

Лингвистика в PLOS ONE (2)



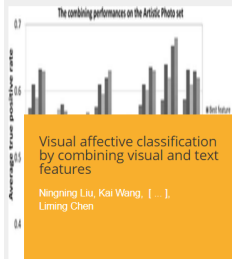
Signage as a tool for behavioral change: Direct and indirect routes to understanding the meaning...

Julia Meis, Yoshihisa Kashima



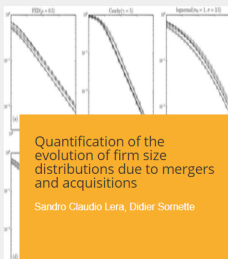
Towards a social functional account of laughter: Acoustic features convey reward, affiliation, and dominance

Adrienne Wood, Jared Martin, Paula Niedenthal



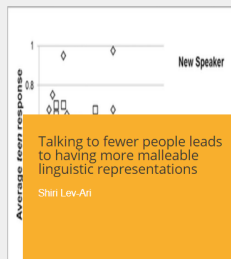
Visual affective classification by combining visual and text features

Ningning Liu, Kai Wang, [...], Liming Chen






Quantification of the evolution of firm size distributions due to mergers and acquisitions

Sandro Claudio Lera, Didier Sornette



Talking to fewer people leads to having more malleable linguistic representations

Shiri Lev-Ari

-  Зализняк А. А. 1967. *Русское именное словоизменение*. Москва: Наука.
-  Фёдорова О. В. 2009. *Основы экспериментальной психолингвистики. Синтаксический прайминг*. Москва: Спутник+.
-  Schane, Sanford. 1974. How abstract is abstract? In Anthony Bruck, Robert A. Fox & Michael W. LaGaly (eds.), *Papers from the parasession on Natural Phonology*. Chicago, Chicago Linguistic Society. 297–317.