

Математические модели в лингвистике

4. Морфо(но)логия: лингвистическое введение

Мати Пентус, Александр Пиперски,
Алексей Сорокин

МГУ, межфакультетский курс,
осенний семестр 2017–2018 учебного года

Понятие слова

- ▶ Что такое слово?
- ▶ Что именно включается в словарь?
- ▶ Проблема не только компьютерной, но и теоретической лингвистики:
морфология

Сколько здесь слов?

варил сидел терпеть
украшать варю терпим терпим
варят варить терплю украшали
сидела варила терпят

**4 слова или 14 слов (в зависимости
от понимания термина «слово»)**

Лексема и словоформа

- ▶ **Словоформа** — конкретная языковая единица, встречающаяся в текстах
- ▶ **Лексема** — множество словоформ, имеющих общее значение и различающихся только грамматическими признаками

Лексема и словоформа

- ▶ **Словоформы** *варил, варю, варить, варила* ∈ лексема **ВАРИТЬ**
- ▶ **Словоформы** *терпеть, терпим, терплю, терпит* ∈ лексема **ТЕРПЕТЬ**

Лексема и словоформа

- ▶ **Лексема** — множество словоформ, имеющих общее значение и различающихся только грамматическими признаками
- ▶ Что такое «общее значение»?
- ▶ Что такое «грамматические признаки»?

грамматический = обязательный

- ▶ Грамматические значения — значения, обязательные к выражению в данном языке у данного класса слов
- ▶ Называя предмет русским существительным, мы обязаны указать его падеж (\approx синтаксическую роль) и число

Число: китайский vs. русский

- ▶ Китайский показатель числа *ten* присоединяется только к обозначениям одушевлённых объектов
- ▶ *wo tai le hua* 'Я купил цветок/цветы'
- ▶ Но ср. рус. *Я купил свёклу для супа*

Падеж: древнекитайский vs. русский

- ▶ *rāu bōk gu* 'враг старого раба'
- ▶ *rāu gu bōk* 'раб старого врага'
- ▶ *gu rāu bōk* 'старый раб врага'

Вид и время: английский vs. русский

- ▶ *I read this book*
- ▶ *I was reading this book*
- ▶ *I have read this book*
- ▶ *I have been reading this book*
- ▶ *Я читал / прочитал эту книгу*

Морфология в разных языках

- ▶ Чем больше в языке грамматических значений, тем больше у каждой лексемы словоформ
- ▶ Количество грамматических значений коррелирует с количеством морфем в слове

Индекс синтетичности

- ▶ Джозеф Гринберг (1915–2001):
индекс синтетичности = среднее количество морфем в слове
- ▶ Количество грамматических значений коррелирует с количеством морфем в слове

Индекс синтетичности: русский язык

- ▶ *Все счастливые семьи похожи друг на друга, каждая несчастливая семья несчастлива по-своему.*
- ▶ *Вс-е счаст-лив-ые семь-и похож-и друг на друг-а, кажд-ая не-счаст-лив-ая семь-я не-счаст-лив-а по-сво-ему.*
- ▶ 28 морфем / 12 слов = 2,33
- ▶ NB: выделение слов и морфем не всегда однозначно

Индекс синтетичности

западногренландский	3,72
санскрит	2,59
суахили	2,55
древнеанглийский	2,12
лезгинский	1,93
немецкий	1,92
английский	1,68
вьетнамский	1,06

Синтетические и аналитические языки

- ▶ **Полисинтетические языки** — высокий ИС
- ▶ **Синтетические языки** — средний ИС
- ▶ **Аналитические языки** — низкий ИС
- ▶ **Изолирующие языки** — очень низкий ИС

NB: любая граница условна

Изолирующие языки: вьетнамский

- ▶ *Chúng tôi muốn đi Việt Nam để học tiếng Việt trong hai tháng nghỉ hè. Tháng bảy, tháng tám là mùa du lịch, vì vậy có rất nhiều người muốn sang Việt Nam.*
- ▶ Много я хотеть идти Вьет Юг поставить учить язык Вьет в два месяц отдыхать лето. Месяц семь, месяц восемь быть месяц гулять расписание, для так иметь очень много человек хотеть ехать Вьет Юг.

Аналитические языки: английский

- ▶ *The Dursleys had everything they wanted, but they also had a secret, and their greatest fear was that somebody would discover it.*

Синтетические языки: русский

- ▶ *Все счастливые семьи похожи друг на друга, каждая несчастливая семья несчастлива по-своему.*

Синтетические языки: турецкий

- ▶ *ev-ler-im-iz-de*
дом-PL-1-PL-LOC
'в наших домах'
- ▶ Много морфем, каждая из которых выражает ровно одно значение — **агглютинативный** язык
- ▶ В этом примере ИС = 5, но это близкий к максимуму случай для турецкого языка

Полисинтетические языки: адыгейский

- ▶ *Мышъэм мэзым укъыщысигъэлъэгъугъ*
'Медведь показал мне тебя в лесу'
- ▶ *Мышъэм мэзым*
медведь лес
у-къы-щы-с-и-гъэ-лъэгъу-гъ
тебя-DIR-LOC-мне-он-CAUS-смотреть-PAST

Лексемы и словоформы

- ▶ В изолирующих языках число словоформ примерно равно числу лексем
- ▶ В синтетических и полисинтетических языках число словоформ заметно больше числа лексем
- ▶ Хранятся ли все они в голове?

Представление слов

Две задачи:

- ▶ Обеспечить экономное описание
- ▶ Описать (психологическую) реальность

Возможные механизмы

- ▶ Computation (вычисление) \leftrightarrow
грамматика
- ▶ Storage (хранение) \leftrightarrow лексикон

[Noteboom et al. (eds.) 2002]

Лексикон

- ▶ "Virtually all items of the lexicon belong to the *substantive categories*, which we will take to be noun, verb, adjective, and particle, putting aside many serious questions about their nature and interrelations"[Chomsky 1995, 5]
⇒ единица хранения — слово
- ▶ [Jackendoff 2002]: единица хранения — от морфемы до текста («Илиада»)

Модель двойного механизма

[Marcus et al. 1995], Стивен Пинкер

- ▶ Часть форм берётся из памяти (1-й шаг), часть строится по правилам (2-й шаг)
- ▶ Самый популярный пример: прошедшее время в германских языках
- ▶ *come* – *came* — из памяти, *exaggerate* – *exaggerated* — по правилам

Модель двойного механизма

- ▶ Что берётся из памяти, а что строится по правилам?
- ▶ Возможно, часть регулярных форм тоже берётся из памяти

Др.-англ.: *sēman* – *sēmde* ‘казаться’

cēpan – *cēpte* ‘хранить’

sēmen – *sēmde* ‘казаться’

Ср.-англ.: *kēpen* – *kepte* ‘хранить’

Модели морфологии

[Hockett 1954]:

- ▶ Item and Process
- ▶ Item and Arrangement

Единство морфемы

- ▶ Одна из ключевых идей в Item and Process — единство морфемы на глубинном уровне
- ▶ До какой степени это достижимо?

Русский язык

- ▶ /стóл/ — /сталу́/ — /стал'э́/
- ▶ /су́т/ — /судóф/ — /суд'йт' / — /сужу́/

Единство морфемы и проблема абстрактности

[Jensen 2004]

- ▶ Нулевая гипотеза: говорящий запоминает поверхностные формы
- ▶ Если нулевая гипотеза задаёт слишком громоздкую грамматику, от неё приходится отказаться

Необходимость абстрактности

Простота: глубинные представления, отличные от поверхностных, нужно вводить, если это сокращает грамматику

Английский язык

Окончание 3 ед. наст. имеет вид /s/, /z/ или /ɪz/ в таких-то случаях + то же про мн. ч. и притяжательную форму существительных и краткие формы глаголов *is* и *has*

Допустимые уровни абстрактности

1. Конкретное глубинное представление (ГП): совпадает с единственным поверхностным представлением (ПП) (*back* /bæk/)
2. Более абстрактное ГП: совпадает хотя бы с одним ПП (/стол/, /z/)

Допустимые уровни абстрактности

3. Ещё более абстрактное ГП: каждый сегмент совпадает хотя бы с одним своим ПП

Язык палау

[Schane 1974]:

наст.

(для себя)

прич. буд.

(устар.)

прич. буд.

(нов.)

[mə-'daŋəb]

[də'ŋob-l]

[dəŋə'b-all]

накры-
вать

[mə-'teʔəb]

[tə'ʔib-l]

[təʔə'b-all]

вытяги-
вать

Допустимые уровни абстрактности

4. Максимально абстрактное ГП:
сегменты не обязаны совпадать ни с
одним из своих ПП

Язык йоулумне (< америндские; США)

Дубитатив	Аорист	Перевод
xatal	xathin	есть
giy'al	giy'hin	касаться
me:k'al	mek'hin	глотать
ʂa:pał	ʂaphin	жечь

$V: \rightarrow V / _ C C$

(V — гласный, C — согласный, : — долгота)

$ʂa:p + hin = ʂa:phin > saphin$

Генеративная фонология: правила

- ▶ Контекстные замены вида $A \rightarrow B / X _ Y$
- ▶ «А переходит в В в положении после X и перед Y»
- ▶ X и Y могут опускаться

Язык йоулумне (< америндские; США)

Дубитатив	Аорист	Перевод
paʔtal	paʔiṭhin	драться
ʔilkaɫ	ʔilikhin	петь
lihmaɫ	lihimhin	бежать
ʔaʔuʔaɫ	ʔaʔiʔuhin	толкать лодку шестом

∅ > i / C _ C C

Язык йоулумне (< америндские; США)

Дубитатив	Аорист	Перевод
?amlal	?a:milhin	помогать
ʂalk'al	ʂa:lik'hin	будить
ʂeŋtal	ʂe:niṯhin	нюхать

- ▶ Глубинные представления?
- ▶ Порядок правил?

Английский язык

- ▶ Окончание множественного числа существительных **-z** + правила
- ▶ Порядок правил?

Неконкатенативность

- ▶ Неконкатенативная морфология — не соединение отдельных морфем друг с другом, а модификация морфем (обычно корневых)
- ▶ Английский язык: *foot* ~ *feet*
- ▶ **Семитские языки**

Арабский язык

- ▶ трёхсогласный корень
- ▶ КТВ ‘писать’, DRS ‘учить’, QTL ‘убивать’
- ▶ Заполнители из гласных:
 - а-а- ‘прошедшее время глагола’;
 - и-і- ‘то же + пассивный залог’
 - ā-ī- ‘активное причастие наст. вр.’
 - ...
- + комбинации с приставками и суффиксами

Арабский язык

- ▶ KaTaBa – он написал
- ▶ KāTaBa – он переписывался
- ▶ KuTiBa – он был написан
- ▶ maKTūB – написанный
- ▶ KāTiB – пишущий, писатель
- ▶ KuTTāB – писатели
- ▶ KiTāB – книга
- ▶ KuTuB – книги

Арабский язык






- ▶ 'aKТаВа – он продиктовал
- ▶ 'iKtiTāВ – подписка
- ▶ yaKTuВи – он пишет
- ▶ KiTāВа – писание, письмо
- ▶ maKТаВ – письменный стол
- ▶ maKТаВа – библиотека
- ▶ KiTВа – почерк
- ▶ miKāTiВ – корреспондент

Выводы (1)

- ▶ Разные языки выражают разное количество морфологических значений
- ▶ Разное среднее число морфем в слове: от изолирующих до полисинтетических языков
- ▶ Аддитивная модель морфологии иногда требует применения фонологических правил

Выводы (2)

- ▶ Бывает неконкатенативная морфология
- ▶ На следующих лекциях: как всё это моделировать?

-  Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, Massachusetts: MIT Press
-  Hockett, Charles F. 1954. Two models of grammatical description. *Word* 10: 210–234.
-  Jackendoff, Ray. 2002. What's in the lexicon? In [Noteboom et al. (eds.) 2002]. 23–58.
-  Jensen, John T. 2004. *Principles of Generative Phonology: An Introduction*. Amsterdam/Philadelphia: John Benjamins.
-  Markus, Gary F., Ursula Brinkmann, Harald Clahsen, Richard Wiese & Steven Pinker. 1995.

German inflection: the exception that proves the rule. *Cognitive Psychology* 29.3: 189–256.



Nooteboom, S. G., Fred Weerman and Frank Wijnen. 2002. *Storage and computation in the language faculty*. Dordrecht: Kluwer Academic.



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.