

tion of the Synonymy/Antonymy Relationship between Words // Journal of the ACM, 1967, V. 14, -N 1.

30. Murai M. CIRES (Conversational Informational Retrieval System on Documents), its Keywords and its associated Words // Papers presented at the 1973 International Conference on Computational Linguistics. - Pisa, 1973.
31. Rubinstein H., Goodenough Y. Contextual Correlates of Synonymy // Communications of the ACM, 1965, V. 8, -N 10.

А.И. Кузнецова

ОСОБЕННОСТИ МОРФЕМ И МОРФЕМНОЙ СТРУКТУРЫ СЛОВА В ТЕКСТЕ
СРАВНИТЕЛЬНО С МОРФЕМАМИ И МОРФЕМНОЙ ОРГАНИЗАЦИЕЙ СЛОВА
В СЛОВАРЕ

В последнее время в распознающих системах при автоматическом распознавании речи стала использоваться информация о морфологической организации слова, разновидностями которой являются словообразовательная, формообразовательная и морфемная структуры [1]. Использование морфемы (конечной внутрисловной составляющей) "в качестве основной единицы системного представления языковых средств" [1, 11] требует особого внимания к вопросу о соотношении разного рода внутрисловных структур. Сравнение формообразовательной и морфемной структур фактически есть сравнение структур слова и составляющих его морфем в тексте и словаре (в системе). Сопоставление внутрисловных структур можно проводить как в плане фонетическом (изучая многочисленные фонетические характеристики слов и морфем, реализующихся в словаре и тексте), так и в плане чисто морфемном (в этом случае допустимо обращение к письменной форме русской речи).

Формообразовательная структура слова представляет собой формообразующую основу плюс форматив (в понимании этого термина Ю.С. Масловым [2]); формообразующая основа, в свою очередь, имеет сложную морфемную структуру. Исходные формы слов, фиксируемые в словаре и вместе с тем встречающиеся все (за редким исключением¹) в тексте, обладают, естественно, одной и той же морфемной структурой; остальные ("не основные") парадигматические, сугубо "текстовые" словоформы (косвенные падежи существительных, спрягаемые формы глаголов...), не приводимые в словаре, нередко отягчаются по своим морфемным структурам и входящим в состав слова морфем от словарных словоформ.

В настоящей статье, с одной стороны, сравнивается запас (инвентарь) морфем, реализующихся в словаре и тексте, с другой стороны, сопоставляются морфемные структуры словарных словоформ и текстовых форм слова. В тексте могут встретиться такие структуры, такие особенности морфемного состава слов, такие корни и аффиксы, которых нет

¹ Этими исключениями являются, например, рестицируемые в словаре, но реально не сохранившиеся в языке формы им.п. ед. или мн.ч. некоторых существительных, имеющих так называемые дефектные парадигмы. Так, в слове ре обычно дается: уможд, -у: без уможду и т.п.

в словаре, особенно в системе литературного языка. Выявление этих специфических черт текста может пригодиться при машинной обработке текста.

1. Особенности морфемного состава слов в тексте

При изучении морфемного состава слов в тексте можно анализировать материал с разных точек зрения. Релевантными признаками в этом случае будут, например, следующие: количество морфем в слове (глубина слова); неповторяемость морфем в пределах слова; длина морфем в терминах фонем (или букв для письменной разновидности речи). Другим признаком является, в отличие от двух первых морфологических, фонетической характеристикой морфемы) и т.п.

Глубина слова (количество морфем в слове) в исходной словарной форме составляет 4 ± 1 морфема [3 ; 4].

В отличие от словаря глубина слова в тексте в среднем сохраняется такой же в именах существительных и прилагательных (за исключением степеней сравнения), но изменяется (в сторону уменьшения или в сторону увеличения) в глаголах прежде всего за счет появляющихся альтернативно выражаемых форм времени (особенно прошедшего времени), а также за счет форм наклонения (особенно 2 л. мн. ч.). Сравним некоторые глагольные формы в тексте и словаре (цифры в скобках указывают на количество морфем).

Словарь	Текст
свал-и-ть (3)	<u>свал-и-те</u> (3), но: <u>свал-и</u> (2), <u>свал-ит</u> (2), <u>свал-и-л-а</u> (4)
у-держ-а-ть (4)	<u>у-держ-и-те</u> (4), но: <u>у-держ-ит</u> (3), <u>у-держ-а-л-а</u> (5)
пред-у-пред-и-ть (5)	<u>пред-у-пред-и-те</u> (5), но: <u>пред-у-пред-и</u> (4), <u>пред-у-пред-и-л-а</u> (6)

Незначительный сдвиг (обычно на одну морфему) в сторону уменьшения глубины слова в формах императива (особенно в ед. ч.) и сдвиг в сторону увеличения глубины слова в формах прошедшего времени наблюдается не только в словах с глубиной 4 ± 1 морфема, но также в глаголах большой глубины, имеющих в своем составе более семи морфем, и в глаголах малой глубины (двуморфемных глаголах):

нес-ти (2)	<u>нес-ь</u> (2), <u>нес-ет</u> (2), <u>нес-и</u> (2), но: <u>нес-л-а</u> (3), <u>нес-и-те</u> (3)
вос-про-из-вод-и-ть-ся (7)	<u>вос-про-из-вод-и-те-сь</u> (7), но: <u>вос-про-из-вод-ит-ся</u> (6), <u>вос-про-из-вод-и-л-о-сь</u> (8)

пере-о-с-вид-е-тель-ств-ов-а-ть-ся (11) но: пере-о-с-вид-е-тель-ств-уй-ся (9),
пере-о-с-вид-е-тель-ств-у-ет-ся (10),
пере-о-с-вид-е-тель-ств-ов-а-л-о-сь (12)

Изменение глубины слова в тексте возможно за счет вариативности форм времени у некоторых групп глаголов. Так, в русском языке есть более 120 глаголов с безударным -нуть (всего глаголов на -нуть более 1200), образующих прошедшее время двойко: без суффикса -ну (особенно у приставочных глаголов) и с суффиксом -ну [5 ; 6]. За счет этих колебаний в образовании форм прошедшего времени, а также за счет форм императива разброс в глубине слова в текстовых формах особенно заметен:

мерз-ну-ть (3)	<u>мерз-л-а</u> (3), <u>мерз-н-и</u> (3), но: <u>мерз-ь</u> (2), <u>мерз-ну-л-а</u> (4), <u>мерз-н-и-те</u> (4)
мок-ну-ть (3)	<u>мок-л-а</u> (3), <u>мок-н-и</u> (3), но: <u>мок-ь</u> (2), <u>мок-ну-л-а</u> (4), <u>мок-н-и-те</u> (4)

Вариативность форм при глагольном словоизменении. свойственная глаголам типа брызгать, прыгаться (брызгает - прыгает, движется - движется, около 40 глаголов), также ведет к колебаниям в количестве морфем в слове:

<u>двиг-а-ть-ся</u> (4)	<u>двиг-а-ет-ся</u> (4), но: <u>двиг-ет-ся</u> (3)
<u>мяу-ка-ть</u> (3)	<u>мяу-ч-ят</u> (5), но: <u>мяу-ка-ет</u> (4)

В подобных парах одна из форм обычно интерпретируется как разговорная или просторечная, например: клеветает, хныкает, прыгает, лезает (разг.), нянчится, колеблется (прост.) Наряду с литературными формами хвещет, хнычет, щиплет, лазит, нянчит, колеблет [7].

Иногда грамматическая вариативность сопровождается или, наоборот, предшествует расхождению в значениях форм слова. Так, от клеветать возможны формы клеветает (в значении "скрепить металлические части с помощью заклепок") и клеветлет (в значении "клеветать, наговаривать на кого-либо"); от метать - метает (в значении "прошивать или скреплять крупными стежками, намечал линию шва") и мечет (в значении "бросать") и т.п.

Таким образом, в тексте наблюдается осцилляция глубины глаголов - она увеличивается в одних формах глаголов и уменьшается (что случается статистически реже) в других.

Для словаря в тексте характерна неповторяемость в пределах слова одинаковых морфем, хотя незначительный процент рекуррентных аффиксов возможен [8]. Столь же статистически редко явление омографии в рамках одного слова. В тексте омография увеличивается в количественном отношении по сравнению с омографией среди словарных форм. Происходит это за счет совпадений окончаний в длине

выражения с другими морфемами. Иными словами, с морфемных префиксов и суффиксов, корней и префиксов, корней и суффиксов, префиксов и интерфиксов, суффиксов и интерфиксов в тексте столько же, сколько и в словаре. Однако случаев оморфных совпадений окончаний с другими видами морфем в тексте значительно больше, чем в словаре (ср.: [1]). Например:

корень - окончание: объ-ем-л-ем², ом-ом (тв.п.ед.ч. от заимствованного существительного ом "единица сопротивления")

суффикс - окончание: (о) сид-с-лк-а, (о) вес-ом-ом, (о) ис-а-объ-ем-л-щ-ем, цен-я-тел-я

префикс - окончание: у-бег-у, у-пад-у (ср.: с-кук-с-и-л-а-в-[с]³).

В тексте встречаются совпадения иностранного префикса а- и окончания род.п.ед.ч. -а (в примерах выделяются только оморфемные элементы): а-догмизм-а, а-морфизм-а, а-политизм-а, а-театр-а, а-стигматизм-а, а-спиратор-а и др. (в нескольких словах оморфизм этого типа отмечена и в словаре, в существительных жан.р., например: а-театр-а, а-септик-а, а-спират-а). Реже в тексте встречаются оморфными иностранный префикс а- и окончание глагола жан.р. прошедшего времени: а-ранжировал-а, а-трофировал-а-сь, а-нонсировал-а, а-пробировал-а.

В дефектных парадигмах типа без у-держ-у, без у-мож-у, превратившихся в наречия, омонимия префикс - окончание заменилась омонимией префикс - суффикс, типичной для словаря.

В качестве релевантного признака морфем в тексте (в отличие от словаря) можно признать д л и у окончания, которое в тексте постоянно изменяется (от минимальной длины в одну фонему⁴ до длины в три и более). В глаголах настоящего времени и действительных при-

² Все примеры приводятся в орфографии, т.е. для письменной разновидности русской речи; фонетическая транскрипция дается с той степенью точности и тем, где она необходима для изучения морфем и структур слова.

³ Если возвратную частицу -сь относить к окончанию (считать компонентом окончания или одним из двух окончаний) и учитывать нормы старомосковского пронашения, то в тексте возможен (как в приведенном глаголе скуксилась) дополнительный оморфизм префиксов и окончания, отсутствующий в системе. В "Русской грамматике" (1980) -сь, -ся называются постфиксами, т.е. они выделены в особый вид аффикса.

⁴ В морфологии принято говорить о так называемых "нулевых" морфемах (в частности, окончаниях), которые не имеют плана выражения, в связи с чем минимальной длиной окончания в терминах фонем и считается один фонем, а не нуль фонем.

частях на -ся возможно четырех- и пятифонемное окончание, если принять трактовку окончаний причастий типа -ийся (ср.: тв.п. мн.ч. -имися) как двухкомпонентное окончание.

Длиной слогов и длиной слов в слотах как существенными лингвистическими факторами, влияющими на опознание слов в тексте, наряду с другими параметрами, много занимаются в последнее время фонетики. Недавно появились сведения о длине и морфем [1]. Эти сведения важны при морфемном анализе: должен быть изоморфизм длины слова (что можно установить и через длину слогов, и через длину морфем) и его глубины, причем в тексте и словаре изоморфные отношения между длиной и глубиной слова должны быть различными. Как уже отмечалось, глубина слов в тексте испытывает постоянные колебания. Видно, в сторону уменьшения колебаний бывает больше, поскольку средняя длина слова в тексте меньше, чем в словаре (и не только за счет аллегровой редуцированной речи), а именно - 5,4 (по данным Л.Н. Засориной [9]) или - 5,0, по данным Н.Д. Андреева [10]. На ускорение слова в спонтанной речи указывает О.Ф.Кривнова, полагая, что "укорочение... достигается... за счет увеличения числа односложных слов (редких в научной речи) и уменьшения числа слов со слоговой длиной больше трех слогов (в том числе четырех- и пятисложных, достаточно частых в научной речи). Это позволяет считать, что конкретная граница по длине слова в русской речи проходит между тремя и четырьмя слогами" [11, 3]. Уменьшение средней длины слова в тексте по сравнению со средней длиной слова в словаре происходит "за счет первой тысячи самых частых слов, включающей одно-, двух-, трехбуквенные лексемы, которые влияют на среднюю длину слова в тексте" [9, 933]. В словаре "чаще встречаются слова длиной в 8-9 букв, а в тексте - длиной в 1 и 6 букв. Вообще распределение длин слов в словаре и тексте сближаются, начиная с длины в 9 букв. Если снять влияние первой тысячи самых частых слов, то можно наблюдать, что распределение длин слов в словаре и тексте близки друг другу" [9, 933]. Аналогичные результаты представлены в [12, 82], где средняя длина русского словарного слова определяется в 10 букв. Приблизительно такой же результат получен автором настоящей статьи путем переосчета средней глубины слова и средней длины разных морфем в русском языке (средняя длина корня 4+1 буквы, префиксов и суффиксов - 2-3 буквы, окончания - 2).

2. Особенности инвентаря морфем, реализующихся в тексте, по сравнению с инвентарем морфем, реализующихся в словаре

Запас морфем словаря и текста в основном совпадает. Исключением является лишь набор флексий, которых значительно меньше в словарных формах, чем в словоформах в тексте (~ 1:8 [13; 14]).

Расхождения в альтернирующих морфмах (алломорфмах) корневых и суффиксальных морфем, несомненно, в количественном отношении, заслуживают особого внимания, так как в строении некоторых парадигм (т.е. в тексте) участвуют нередко альтернирующие морфы корней или суффиксов, которые не встречаются в исходной (словарной) форме.

Такие морфы, являясь сигналами разнообразных определенных грамматических функций, выполняемых далеко не всегда и не всеми алломорфами, могут служить особыми приметам для распознавания текста. Кроме того, анализ текстовых алломорфов, отсутствующих в системе литературного языка, позволяет сделать любопытные наблюдения для других подсистем общенародного русского языка.

Корневые и суффиксальные алломорфы в тексте отличаются от алломорфов в словаре. В отличие от словаря морфемы в тексте состоят из большего числа алломорфов, поскольку в тексте реализуются не только все словарные формы, но и формы словоизменения, порою отличные, в силу морфонологических закономерностей, от исходных, словарных форм (т.е. форм им.п. ед.ч. и форм инфинитива).

Повышенное содержание новых по сравнению со словарем алломорфов в тексте наблюдается прежде всего за счет форм спряжения. Обычно новые, отсутствующие в словаре, алломорфы возникают при образовании форм прошедшего времени, типа брёл-, вёл-, вёз-, мёл-, плёл-, цвёл-, щёл- в словах бред-, вед-, вез-, мел-, плел-, а также I и 3 л. ед.ч. настоящего времени, например: обяж- в глаголе обяжет, виш- в глаголе завишу, гляж- в глаголе гляжу, леч- в глаголе лечу (лечать), виж- в глаголе вижу, пояш- в глаголе спояшу. Формы повелительного наклонения также дают не совпадающие со словарными алломорфы, например: беј-, пеј-, шеј- (но не леј-, так как этот корень встречается и в словаре, ср. лейк-а "сосуд для поливки"); ср. императив теш-и ("хоть кол на голове теши!") от глагола тесать.

Дополнительные по сравнению со словарем алломорфы в тексте возникают за счет форм склонения, в частности, особенно много их за счет форм род.п.мн.ч. (душед- при корне в словаре дуща-; кнопк- при словарных алломорфах кнопк-, кнопоч-, банок- и пятен-

при наличии в словаре банк-, баноч- и пятн-) и кратких форм прилагательных (полод- при словарном морфе подн-, пресеи- при преси-, тесеи- при теси- и т.д.). За счет вариативности в склонении, а именно, за счет вариантов род.п. мн.ч. (например, буден - будней), текстовые формы в отдельных случаях могут исчезнуть, уступив место род.п., который образуется от основы "системной" (словарной), как в слове будни - будней, соосуществующей пока с формой буден.

Возможно отдельные текстовые устаревшие формы, сохранившиеся и широко употребляющиеся во фразеологических выражениях и в пословицах: лочид в Бесе (в словарных формах есть только алломорфы бог- и бож-); темна вода во облацех (в словарной системе существуют только корни облак- и облач-); притча во язицех (в словах словаря есть алломорфы язык- и языч-); с холоду ежится, о голоду ичется (в словарных словах отмечается лишь корень ик- - икать, икота) и т.п. (ср. также обороты с энклитической формой местоимения те вместо тебе: Вот те и на! Вот те ре! Вот те крест!).

Наконец, специфически текстовые морфемы (корни) появляются за счет уменьшительных, увеличительных, оценочных образований, не фиксированных словарями (в лучшем случае зарегистрированных среди новых словарных материалов [15]), зачастую лишь синхронных слов: компа-а (при корне компа- в словаре), тон-ят-а (при корне тонн- в словаре; ср. аналогичный случай усечения корня - колонн-а и колонк-а).

При склонении слов нередко меняется и суффикс, но, как правило, возникшие варианты не являются абсолютно новыми по сравнению с имеющимися в словаре: следк-а и следок-а, ср. обратное соотношение - пирожок-а и пирожк-а. Однако при склонении слов с редкими, единичными суффиксами, часто испытываемыми на себе процесс опрощения, в тексте обычно появляются новые алломорфы. Таковы суффиксы в словарных формах -ути/уточ- (прибаутка, прибауточник), -ник/ноч- (косынка, косыночка), -урк/уроч- (дочурка, дочурочка), -удк (танцудка) и т.п. и суффиксы в тексте -уток (прибауток, род.п.мн.ч.), -унок (косынок), -урок (дочурок, пещурок), -улек (танцудек) и т.д.

Дефектность парадигмы (особенно глагольной) создает иногда дум-смысленные варианты: неизвестно, можно или нельзя образовать формы I л.ед.ч. от глаголов пидесосить, гудеть и т.п. и какие они будут.

В общей сложности среди рассмотренных более чем 4000 корневых морфов русского языка, от которых образовано 52 тыс. русских слов, лишь 150 корней (3,61% от 4150) имеют в тексте такие алломорфы, которые не встречаются в морфемных структурах словарного слова, т.е.

они появляются только при склонении и спряжении, выполняя особые узкограмматические функции.

Совершенно иная картина предстает при обращении к текстам других подсистем русского языка, особенно к диалектам, просторечиям. В этих сферах текстовые алломорфы корня, обнаруженные в подсистеме литературного языка, оказываются уже словарными формами, т.е. формами, присущими словарной системе того или иного диалекта, например: в литературном языке корень друз- (друзья) встречается только в парадигме множественного числа, а по головам зафиксировано слово друз-як-я "уживчивый человек"; иди: диал. корень эдл- (эджий, эолить) при литературном словарном корне эдл- (элой, эмить; ср. краткая форма прилагательного эдлй др.

Практически почти все "текстовые" корни литературного языка оказываются "системными", словарными формами в диалектах и просторечии.

Рассмотрение инвентаря морфем и структуры слов в тексте русского языка позволило установить в этой области следующие особенности текста:

- при сохранении средней глубины слова, свойственной словоформам словаря (4±1 морфема), глубина различных текстовых форм постоянно испытывает колебания то в сторону ее уменьшения, то в сторону увеличения;
- количество омоформных префиксов, флексий и некоторых других морфем в составе слова увеличивается в тексте по сравнению со словарными формами;
- в тексте литературного языка отмечаются такие алломорфы корней, которые не встречаются в словаре, в то время как в диалектах корневые алломорфы текста и словаря, как правило, не отличаются друг от друга. Это обстоятельство лишней раз подтверждает необходимость и разумность при морфемном анализе привлекать весь доступный последователю материал русского языка, взятый из всех его подсистем.

Л и т е р а т у р а

1. Уровни языка в речевой деятельности. К проблеме лингвистического обеспечения автоматического распознавания речи. - Л., 1986.
2. Маслов В.С. Понятие основы и формата в дериватологии и парадигматике // Актуальные проблемы русского словообразования. Т. I. - Самарканд, 1972.
3. Кузнецова А.И. Глубина слова как фактор, регулирующий процесс образования новых слов // Деривация в семантике: слово-предложение-текст. - Пермь, 1986.

4. Кузнецова А.И. Морфемная глубина слова в русском языке (слова малой и большой глубины как периферийное явление русского языка) // Проблемы структурной лингвистики. - М., 1988.
5. Виноградов В.В. Русский язык. - М., 1947.
6. Mazon A. Morphologie des aspect du verbe russe. - Paris, 1908.
7. Граудина Л.К., Ицкович В.А., Катлинская Л.П. Грамматическая правильность русской речи. Опыт частотно-стилистического словаря вариантов. - М., 1976.
8. Кузнецова А.И. Словообразовательный круг в русском языке // Публикация отделения структурной и прикладной лингвистики. - М., 1973. - Вып. 6.
9. Частотный словарь русского языка /Под ред. Л.Н.Засориной. - М., 1977.
10. Андреев П.Д. Статистико-комбинаторные методы в теоретическом и прикладном языкознании. - Л., 1967.
11. Крильова О.Ф. Ритмика фонетического слова в разных видах русской прозаической речи // Тез. докл. и сообщений XII Всесоюзного семинара "Автоматическое распознавание слуховых образов". - Киев, 1982.
12. Корсакова Т.В., Меньшиков И.И., Мордань В.И., Ткачук В.И. К вопросу об определении средней длины слова // Вопросы прикладной лингвистики. - Днепропетровск, 1976. - Вып. 6.
13. Волоцкая З.М., Моложная Т.Н., Николаева Т.М. Опыт описания русского языка в его письменной форме. - М., 1964.
14. Апресян Ю.Д., Богуславский И.М., Исидия Л.Л., Крысан Л.П., Лазурский А.В., Перцов Н.В., Санников Р.З. Лингвистическое обеспечение в системе автоматического перевода третьего поколения. - М., 1978.
15. Новое в русской лексики. Словарные материалы-77. - М., 1980-1987.