Морфологическая информация, приписываемая произвольному слову в тексте, состоит из четырех «полей», или групп помет:

1. Лексема, которой принадлежит словоформа (указывается «словарная запись» данной лексемы и ее принадлежность к той или иной части речи).
2. Множество грамматических признаков данной лексемы, или словоклассифицирующие характеристики (например, род для существительного, переходность для глагола).
3. Множество грамматических признаков данной словоформы, или словоизменительные характеристики (например, падеж для существительного, число для глагола).
4. Информация о нестандартности грамматической формы, орфографических искажениях и т. п.

Ниже приводим инвентарь всех используемых в корпусе грамматических помет. Для пояснения в скобках даются примеры.

Части речи

S — существительное (*яблоня, лошадь, корпус, вечность*)
A — прилагательное (*коричневый, таинственный, морской*)
NUM — числительное (*четыре, десять, много*)
A-NUM — числительное-прилагательное (*один, седьмой, восьмидесятый*)
V — глагол (*пользоваться, обрабатывать*)
ADV — наречие (*сгоряча, очень*)
PRAEDIC — предикатив (*жаль, хорошо, пора*)
PARENTH — вводное слово (*кстати, по-моему*)
S-PRO — местоимение-существительное (*она, что*)
A-PRO — местоимение-прилагательное (*который, твой*)
ADV-PRO — местоименное наречие (*где, вот*)
PRAEDIC-PRO — местоимение-предикатив (*некого, нечего*)
PR — предлог (*под, напротив*)
CONJ — союз (*и, чтобы*)
PART — частица (*бы, же, пусть*)
INTJ — междометие (*увы, батюшки*)

Значения грамматических категорий

Род:

m — мужской род (*работник, стол*)
f — женский род (*работница, табуретка*)
m-f — «общий род» (*задира, пьяница*)
n — средний род (*животное, озеро*)

Одушевленность:

anim — одушевленность (*человек, ангел, утопленник*)
inan — неодушевленность (*рука, облако, культура*)

Число:

sg — единственное число (*яблоко, гордость*)
pl — множественное число (*яблоки, ножницы, детишки*)

Падеж:

nom — именительный падеж (*голова, сын, степь, сани, который*)
gen — родительный падеж (*головы, сына, степи, саней, которого*)
dat — дательный падеж (*голове, сыну, степи, саням, которому*)

acc — винительный падеж (*голову, сына, степь, сани, который/которого*)
ins — творительный падеж (*головой, сыном, степью, санями, которым*)
loc — предложный падеж (*[о] голове, сыне, степи, санях, котором*)
gen2 — второй родительный падеж (*чашка **чаю***)
acc2 — второй винительный падеж (*постричься в **монахи**; по **два** человека*)
loc2 — второй предложный падеж (*в **лесу**, на **оси́***)
voc — звательная форма (*Господи, Серёж, ребят*)
adnum — счётная форма (*два **часа́**, три **шара́***)

Краткая/полная форма:

brev — краткая форма (*высок, нежна, прочны, рад*)
plen — полная форма (*высокий, нежная, прочные, морской*)

Степень сравнения:

comp — сравнительная степень (*глубже*)
comp2 — форма «*по*+сравнительная степень» (*поглубже*)
supr — превосходная степень (*глубочайший*)

Вид:

pf — совершенный вид (*пошёл, встречу*)
ipf — несовершенный вид (*ходил, встречаю*)

Переходность:

intr — непереходность (*ходить, вариться*)
tran — переходность (*вести, варить*)

Залог:

act — действительный залог (*разрушил, разрушивший*)
pass — страдательный залог (только у причастий: *разрушаемый, разрушенный*)
med — медиальный, или средний залог (глагольные формы на *-ся*: *разрушился* и т.п.)

Форма (репрезентация) глагола:

inf — инфинитив (*украшать*)
partcp — причастие (*украшенный*)
ger — деепричастие (*украшая*)

Наклонение:

indic — изъявительное наклонение (*украшаю, украшал, украшу*)
imper — повелительное наклонение (*украшай*)
imper2 — форма повелительного наклонения 1 л. мн. ч. на *-те* (*идемте*)

Время:

praet — прошедшее время (*украшали, украшавший, украсив*)
praes — настоящее время (*украшаем, украшающий, украшая*)
fut — будущее время (*украсим*)

Лицо:

1p — первое лицо (*украшаю*)
2p — второе лицо (*украшаешь*)
3p — третье лицо (*украшает*)

Прочие признаки:

persn — личное имя (*Иван, Дарья, Леопольд, Эстер, Гомер, Маугли*)
patrn — отчество (*Иванович, Павловна*)

famn — фамилия (*Николаев, Волконская, Гумбольдт*)
zoon — кличка животного (*Шарик, Дочка*)
0 — несклоняемое (*шоссе, Седых*)

Часть указанных помет (а именно, второй винительный падеж, звательная форма, счётная форма, форма по+сравнительная степень, общий род, переходность, несклоняемость) присутствуют только в корпусе со снятой грамматической омонимией.

Множественные разборы

В отдельных случаях в морфологической разметке допускается указание у одной и той же словоформы нескольких разборов, а именно:

- Для прилагательных, совпадающих с причастиями (*открытый*), в неоднозначных случаях в качестве исходной дается как лексема-прилагательное (ОТКРЫТЫЙ), так и глагол (ОТКРЫТЬ).
- Ставится множественная помета в случаях, когда однозначный выбор лексемы или грамматического значения в данном контексте невозможен (*не видел родного отца* — **gen/acc**; *манекену* — **anim/inan**;*спазмами* — исходная форма СПАЗМ/СПАЗМА и т. п.)

Информация о нестандартности и особенностях записи

В корпусе со снятой грамматической омонимией предусмотрен ряд помет, указывающих на нестандартность и/или особенности записи входящей в Корпус словоформы. Отстутствие таких особенностей обозначается пометой **normal**.

anom («Аномальная форма») — различного рода морфологические аномалии, возможные у устаревших или просторечных нелитературных форм (*три дни* при нормативном *три дня*, *ляжь* при нормативном *ляг*)
distort («Искаженная форма») — орфографическое и/или фонетическое искажение слова, часто передающее различные особенности произношения (*дэвушка*, *това'ищи*, *про-хо-ди*, *низнаю*).
ciph («Цифровая запись») — запись числительного, числительного-прилагательного или прилагательного (полностью или частично) при помощи цифр (*73*, *LXXIII*, *73-й*, *22-летний*). Для этих словоформ в поле «Лексема» также употребляется цифровая запись; число и падеж указываются только в тех случаях, когда выписано окончание (типа *14-му*).
INIT («Инициал») — запись вида «заглавная буква с точкой» (*М., Р.*). В поле «Лексема» инициал не раскрывается; грамматические признаки не указываются.
abbr («Сокращение») — сокращенная запись (*тов., гг., ч.*). В поле «Лексема» сокращение (кроме инициалов) раскрывается, указывается грамматическая форма, соответствующая контексту. Специально отметим, что акронимы вроде *ООН, вуз* и усеченные слова вроде *зав*, *зам*, записываемые без точки и не раскрываемые при чтении, не получают пометы abbr и трактуются как обычные слова (склоняемые или несклоняемые).

Абзацы выделяются XML-тегами , предложения — <se></se>, слова <w></w>.

Грамматические пометы находятся в теге <ana></ana>. В атрибуте lex приводится начальная форма (лемма, лексема), в атрибуте gr — грамматические признаки в указанном выше порядке. Атрибут joined указывает на нераздельное написание словоформ (together — слитное, hyphen — дефисное). Большинство словоформ нарицательных имён и наиболее частотные словоформы имён собственных акцентуированы (знак ` перед ударной гласной).