М. Рожанец

МОРФОЛОГИЧЕСКАЯ РАЗМЕТКА КОРПУСА ТЕКСТОВ ПРОГРАММОЙ LANCSBOX

В современной лингвистике при проведении многих исследований необходимы большие объемы текстов в виде корпусов, обрабатываемых с применением компьютерной техники. Корпусы дали возможность сравнивать и анализировать не единицы текстов, а их совокупности, достигающие миллиардов словоупотреблений. Релевантные данным задачам предоставляет специализированное программное обеспечение *LancsBox*.

С помощью корпусного менеджера *LancsBox* был создан и автоматически размечен репрезентативный одноязычный корпус текстов современных американских авторов, опубликованных в 2021, включающий в себя 5 основных стилей (художественный, публицистический, научный, официально-деловой, разговорно-бытовой) и насчитывающий 57 тысяч словоупотреблений.

Наибольшее количество ошибок в автоматической разметке было допущено при тегировании прилагательных 8,6 %.

Половину всех собранных контекстов включали наречия, принятые за прилагательные (**50,3 %**): *They're at a well-resourced university, which you're paying a tremendous amount for them to attend, quite* ***likely***.

Различные существительные были размечены как прилагательные (**9,5 %**): *As the youngest students begin to get vaccinated, Maryland education officials are rethinking the* ***state's*** *mask mandate for schools and how long it needs to stay in place.* В данном контексте существительное *state* находится в форме притяжения с помощью *'s*.

Самым распространённой частью речью, которая была отмечена как существительное, было местоимение с глаголом в сокращённой форме (**40 %**): *"After all that,* ***I'm*** *done with festivals," he said*.

Существительное также было принято за глагол (**27,1 %**): *Two belonged to* ***people*** *he later learned had died at the festival*. Существительное ***people*** стоит перед предлогом *to*, что во многих случаях указывает на глагол в форме инфинитива.

Yeah, because humans are ***amazing***. Вспомогательный глагол *are* с прилагательным ***amazing*** были приняты за конструкцию с глаголом в форме настоящего продолженного времени (*Present Continious*) по формуле *to be+Ving*.

Некоторые предлоги были приняты за глаголы (**12,5 %**): *Pleasing also said that the line was inspired* ***by*** *pearls, which Styles frequently wears, calling it an "ode to the beauty found in a simple shell".*