

Минимальная длина описания и сравнение лингвистических теорий

М.Б. Ермолаева (МГУ)

Метрика минимальной длины описания (minimum description length, MDL; Rissanen 1978) рассчитывается для грамматики и некоторого корпуса данных и минимизирует сумму двух компонентов, измеряемых в битах: длины грамматики и длины описания данных, порождаемого этой грамматикой. Этот подход одновременно поощряет короткие и простые грамматики, делающие обобщения относительно данных корпуса, и штрафует грамматики, порождающие недопустимые структуры и неспособные эффективно сжимать данные. Применительно к лингвистике, классическая MDL допускает сравнение грамматик только внутри одной формальной теории. А поскольку в литературе отсутствует консенсус по поводу как самой лингвистической теории, так и деталей формализации, использование MDL на практике затруднено.

В докладе описывается трехчленный вариант MDL, разработанный совместно с Дж. Голдсмитом и развивающий концепцию сравнения лингвистических теорий, изначально предложенную в Goldsmith (2015). В дополнение к вышеописанным компонентам эта метрика учитывает длину описания самой лингвистической теории, представленной в виде программы на фиксированном языке программирования, также измеряемую в битах. Такая трехчленная MDL позволяет сравнивать грамматики, созданные в рамках различных теорий.