
Лексико-функциональная разметка русских текстов

Татьяна Фролова, Ольга Подлесская,
*Лаборатория компьютерной
лингвистики, ИППИ РАН*

Понятие ЛФ-коллоката

- Значение или семантическое отношение, которое может выражаться различными словами, причем выбор конкретного слова зависит как от его собственного лексического значения, так и от того, при каком ключевом слове (аргументе функции) выражается данное общее значение или отношение
-

Типовые смыслы

- MAGN ‘высокая степень’
 - VER ‘правильный’
 - INCEP ‘начало’
 - CAUS ‘каузация’
 - BON ‘хороший’
 - ANTI ‘противоположный’
 - FIN ‘прекращение’
 - MANIF ‘проявление’
 - и др.
-

Пример значений некоторых ЛФ

MAGN 'высокая степень X' – *сильная жара, высокая влажность*

OPER1 'делать X' – *оказывать влияние, вносить вклад*

FUNC0 'X происходит' – *идет экзамен*

REAL1 'использовать X по назначению' – *летать на самолете*

Свойства ЛФ-коллокатов

- Универсальность
предположительно есть во всех языках
- внутриязыковая идиоматичность
палящий зной, но сильная жара, высокая влажность
heavy rain, но extreme heat
- межъязыковая идиоматичность
сильная жара – extreme, intense, great heat
(но не **strong heat, *heavy heat*)

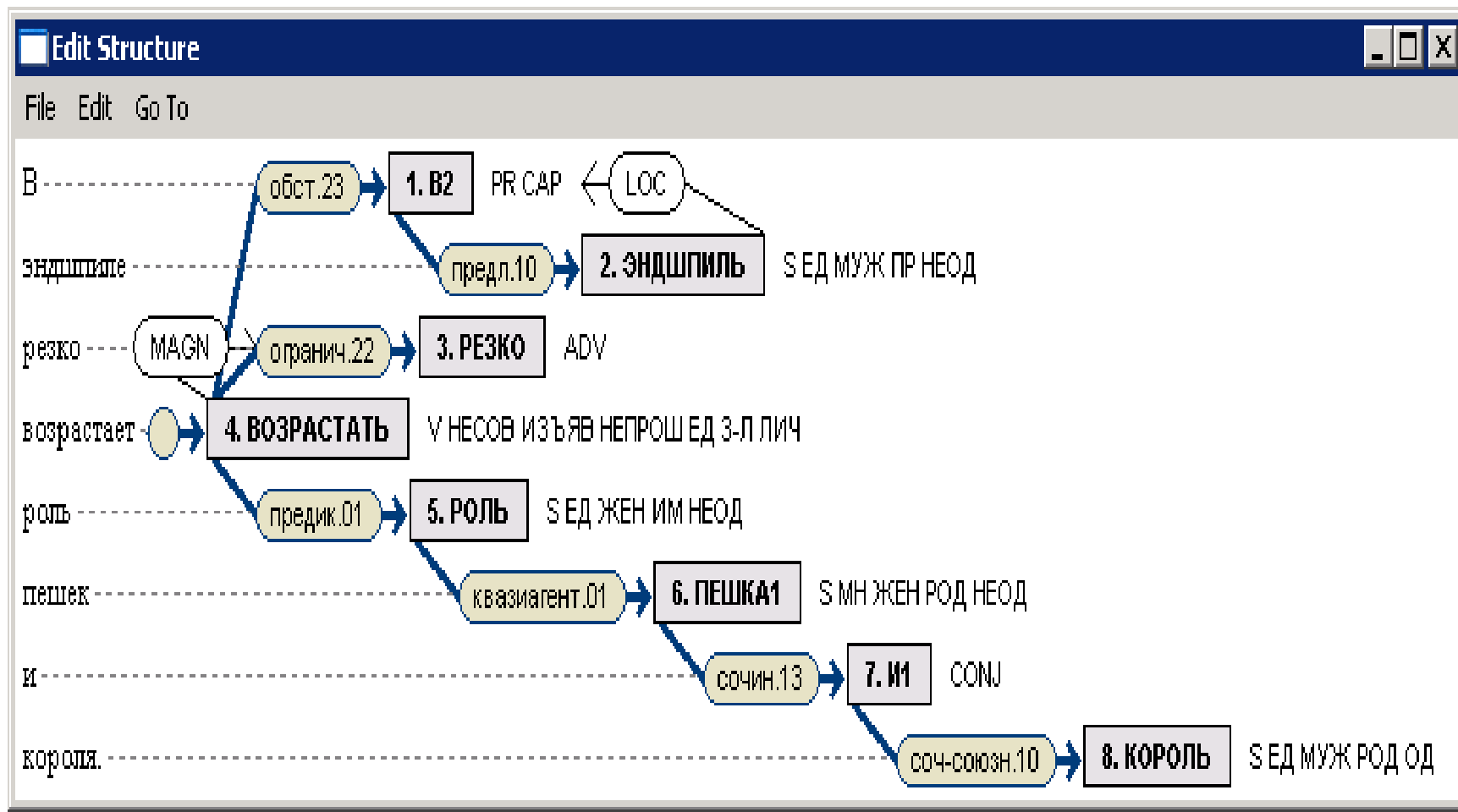
Национальный корпус русского языка (НКРЯ): www.ruscorpora.ru

- Основной, поэтический, устный, обучающий, диалектный, параллельный, акцентологический.
 - Синтаксически размеченный корпус (СинТагРус) > 40 000 предложений со снятой омонимией
-

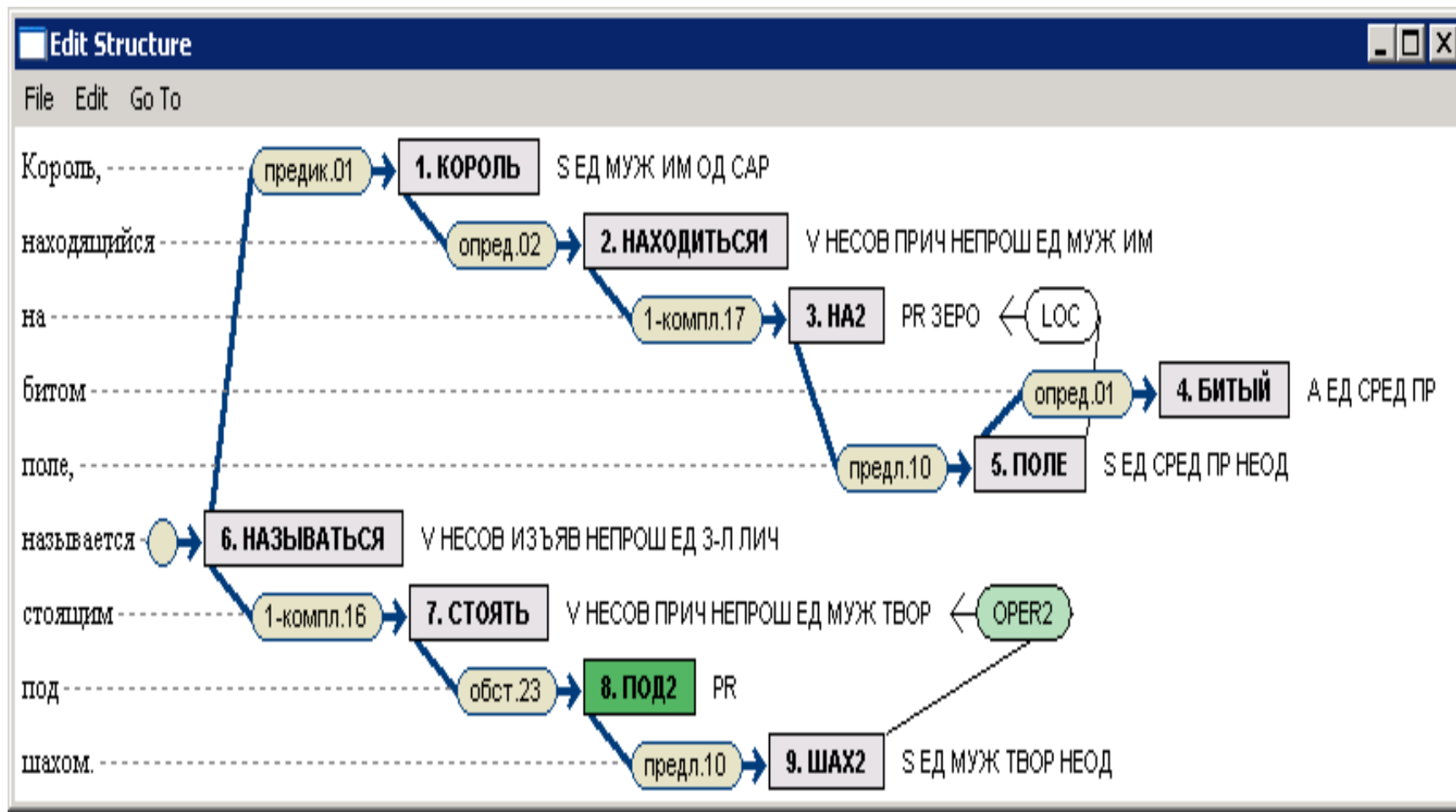
Виды разметки в СинТагРус

- морфосинтаксическая (on-line):
 - грамматические характеристики слов
 - синтаксические отношения между словами
 - лексико-семантическая
 - **лексико-функциональная**
-

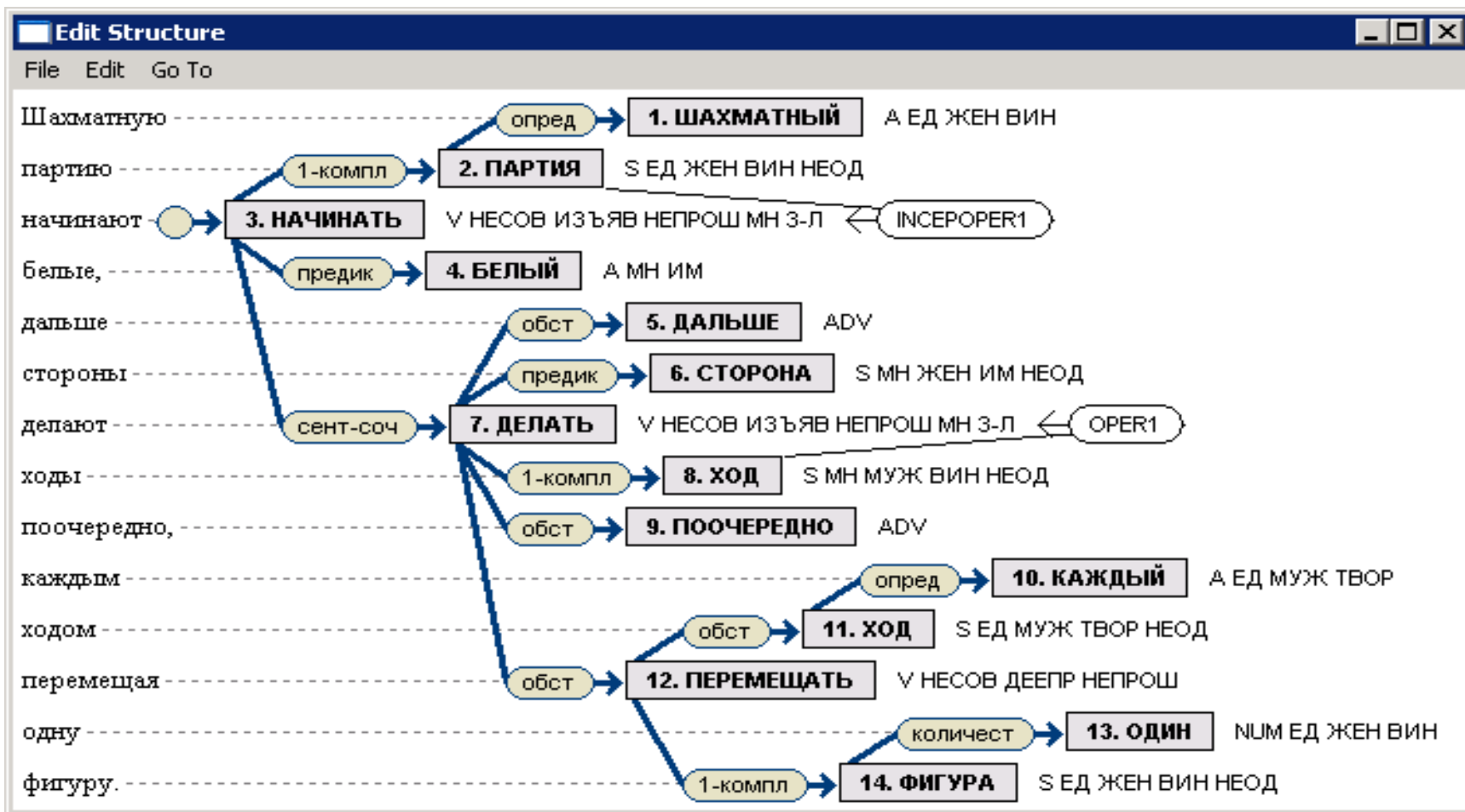
«В эндшпиле резко возрастает роль пешек и короля»



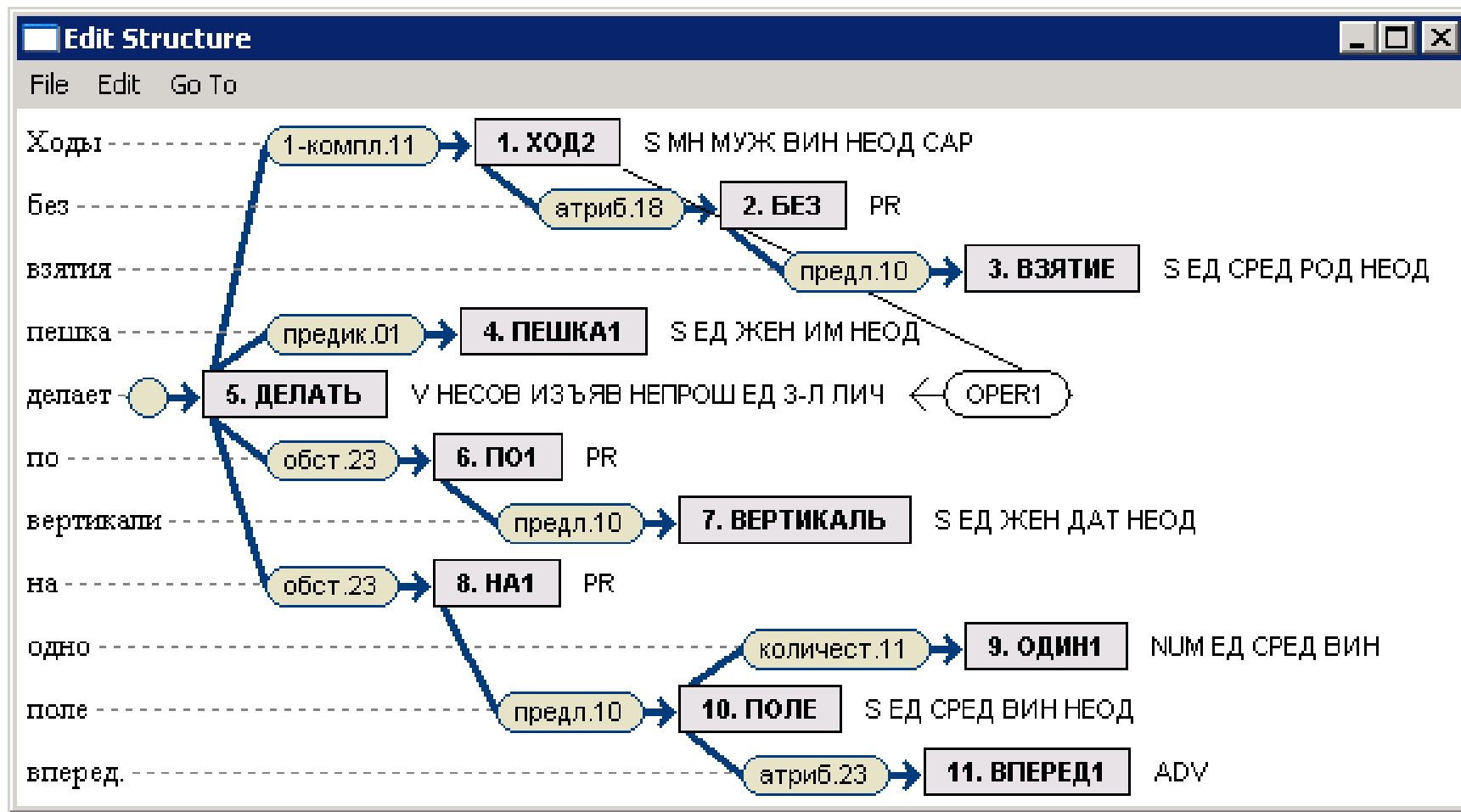
«Король, находящийся на битом поле,
называется СТОЯЩИМ ПОД ШАХОМ»



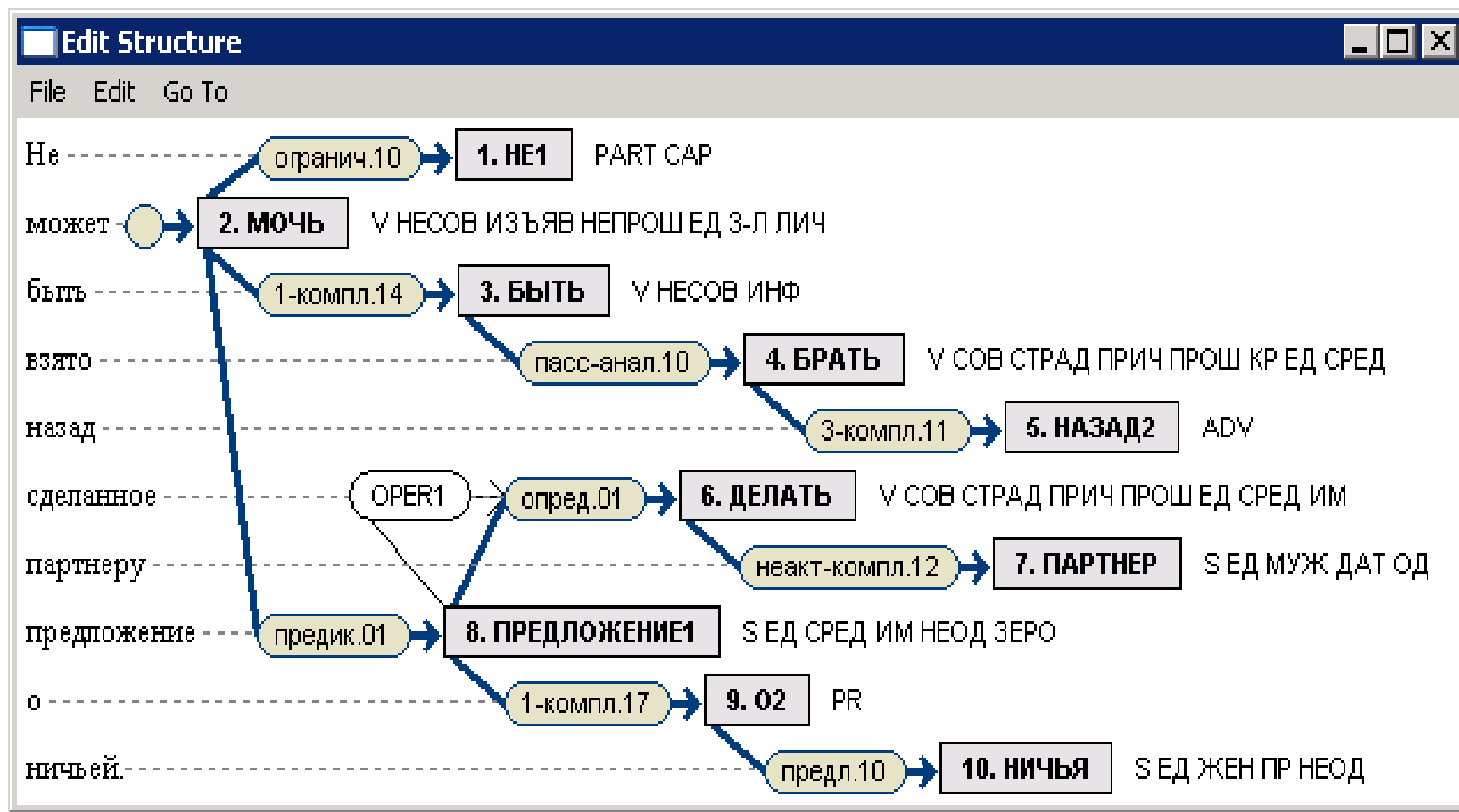
«Шахматную партию начинают белые, дальше стороны делают ходы поочередно, каждым ходом перемещая одну фигуру»



«Ходы без взятия пешка делает по вертикали на одно поле вперед»



«Не может быть взято назад сделанное
партнеру предложение о ничьей»



Полученный корпус

- **Корпус из 4500 предложений с размеченной ЛФ-структурой (>5700 ЛФ-сочетаний)**

Спасибо за внимание
