**Частеречная разметка: спорные случаи**

1. **«То есть» (0:10:29; 0:11:18)**

Ни в Грамматическом словаре Зализняка, ни в НКРЯ «*то есть*» не рассматривается как неделимая единица, что обуславливает необходимость поделить данное сочетание на два токена. В НКРЯ для «*то*» предлагаются варианты местоимение-существительное и союз. Вариант CONJ оказывается предпочтительнее, так как в данной конструкции первична именно союзная функция. Бóльшие проблемы вызывает слово «*есть*» - НКРЯ довольно закономерно предлагает вариант глагол, так как в примерах с составным союзом «*то есть*» авторы разметки учитывают историческое развитие данного сочетания (которое включало в себя глагол «*быть*», впоследствии застывший в данном сочетании в форме наст. времени). Грамматический словарь Зализняка предлагает помимо прочего вариант «междометие», однако такая трактовка в данном контексте недопустима. Кажется, что глагольная семантика в «*то есть*» довольно сильно стёрта, поэтому можно было бы отнести «есть» к разряду частиц, однако тогда не совсем ясна его функция. Можно было бы также отнести «*есть*» к союзам, что указало бы на совместное употребление «*то*» и «*есть*» в данном контексте, однако при отдельном поиске по корпусу такое решение выглядело бы странно. В итоге было решено отнести «есть» к частицам со значением уточнения, так как мне кажется важным отделять нормальное употребление глагола «*быть*» в настоящем времени (а он в данной форме почти не употребляется, и разметив сочетания так, мы бы повысили его показатели в нашем корпусе) от «*есть*» в составе союза, а по семантике «*есть*» здесь похоже на уточняющие частицы типа «*именно*», «*как раз*» и т.д. Хотя отнесение его к глаголам, возможно, даже более оправданно и вероятно.

1. **«В целом» и «в общем» (0:10:29; 0:10:51)**

«*В целом*» и «*в общем*» также не распознаются НКРЯ и Грамматическим словарём как вводные конструкции, что ставит перед разметчиками вопрос о том, к каким частям речи отнести токены «*целом*» и «*общем*». По своим изначальным морфологическим характеристикам они относятся к именам прилагательным, однако в данном случае существительное при них отсутствует, поэтому гораздо более оправдано описывать их как субстантивы. Поэтому в разметке «*целом*» и «*общем*» отнесены к существительным.

1. **«Такое (происходит)» (0:10:46)**

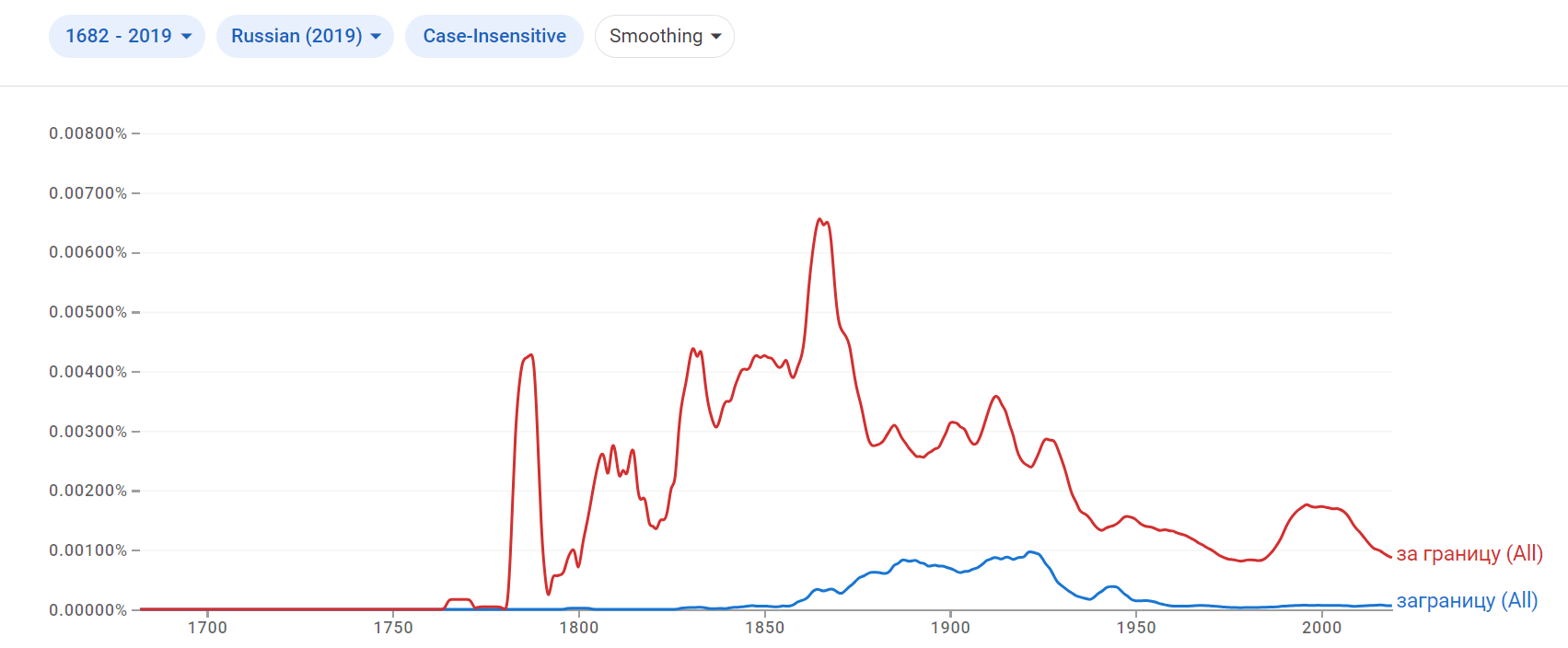
В данном случае основным является конфликт между морфологическим и синтаксическим подходами к определению частей речи. С точки зрения классической морфологии «такое» является указательным местоимением, однако в данном случае оно выступает в функции подлежащего, отчего оно сближается с субстантивами. Тем не менее, видимо, в качестве доминирующего подхода следует выбрать морфологический, так как при разметке слова «такое» как существительного возникнут проблемы с семантикой из-за неполнознаменательности слова. Поэтому выбрана разметка в качестве местоимения-прилагательного.

1. **«Вообще» - вводное слово или наречие (0:10:53)**

Со словом «*вообще*» возникает довольно большой вопрос. С одной стороны, в предложении «*мы с семьёй вообще не очень часто путешествовали за границу*» слово «*вообще*» не очень выпадает из синтаксической структуры и может рассматриваться как обстоятельство меры и степени. С другой стороны, вписывание его по семантике в данную группу обстоятельств кажется довольно натянутым: «вообще» гораздо ближе к оценке меры, чем к описанию самой меры, что сближает его с вводными словами. Так как в НКРЯ, в отличие от, например, Грамматического словаря А. А. Зализняка и от Нового частотного словаря русской лексики, «*вообще*» всё же классифицируется как вводное слово, мы поступим аналогичным образом разметим его так же.

1. **«Заграницу» (0:10:54)**

В данном случае вопрос касается не только непосредственно частеречной разметки, но и количества токенов. По данным Google Ngram Viewer предложно-падежное сочетание «*за границу*» встречается сильно чаще, чем наречие «заграницу» со слитным написанием.



При этом вхождение «*заграницу*» в основном подкорпусе НКРЯ встречается 932 раза (правда против 10 547 «*за границу*»), хотя по современным нормам русского языка нормативно только сочетание «за границу», а написание «заграницу» является ошибочным с точки зрения нормы. Однако в узусе довольно очевидно, что такое сочетание тяготеет к устойчивому наречному употреблению (аналогично наречиям «*домой*», «*навстречу*», «*наружу*»): оно в основном следует за глаголами движения, выполняя при них роль обстоятельства места. Тем не менее в данном случае следует пойти на поводу у современной нормы (и показателям частотности употребления на данный момент), и остановится на выделении предложно-падежного сочетания.

1. **Союзные слова на примере «*что*» (0:11:33)**

В данном случае мы вновь сталкиваемся с проблемой выбора того, что оказывается семантически и синтаксически важнее – союзная функция слова или его роль внутри клаузы. В нашей разметке такая проблема возникла в предложении *«… в трёх часах езды от Риги,* ***что*** *по меркам Латвии очень далеко*». В данном случае «*что*» служит для присоединения придаточного, однако внутри него выполняет роль подлежащего. Следовательно, при частеречной разметке приходится выбирать, какой грамматический тэг лучше присвоить в данном случае: тэг CONJ или PRON. С одной стороны, союзная функция «*что*» безусловно важна, но, с другой, в такой разметке не будут различаться «*что*» в качестве полноценного союза и в качестве союзного слова. Для избежания синтаксической омонимии слова «что» в разметке было отнесено к местоимениям.