

**О МАРКЕРАХ РАЗГОВОРНОЙ РЕЧИ  
(ПРЕДВАРИТЕЛЬНОЕ ИССЛЕДОВАНИЕ ПОДКОРПУСА КИНО  
В НАЦИОНАЛЬНОМ КОРПУСЕ РУССКОГО ЯЗЫКА,  
WWW.RUSCORPORA.RU)<sup>1</sup>**

**THE SPOKEN RUSSIAN MARKERS  
(PRELIMINARY INVESTIGATION OF THE MOVIE SUBCORPUS  
OF THE RUSSIAN NATIONAL CORPUS, WWW.RUSCORPORA.RU)**

*Гришина Е.А. (rudi2007@yandex.ru), Институт русского языка им. В.В.Виноградова РАН*

Статья содержит описание маркеров русской разговорной речи, т.е. тех языковых единиц, определенное количество которых в тексте воспринимается как знак отнесенности данного текста в рамках данной культуры к устной, а не к письменной сфере функционирования. Исследование выполнено на материале транскриптов русских фильмов, включенных в Национальный корпус русского языка. Транскрипты сравнивались с исходными текстами (пьесами и сценариями), а также с субтитрами, которые были составлены для слабослышащих зрителей. Исследование показало, что при трансформации исходный текст → транскрипт → субтитр изменяются, появляются и исчезают практически одни и те же единицы и конструкции. Именно они и должны рассматриваться как маркеры русской разговорной речи. Статья содержит предварительный список таких единиц. Предлагается использовать этот набор в практических целях, в частности, для построения программ, позволяющих анализировать уровень «разговорности» того или иного текста, например, в целях оптимизации поиска разговорных текстов в Интернете.

**1. Постановка задачи**

Как известно, Национальный корпус русского языка включает в себя не только письменные, но и устные тексты (НКРЯ2006: 94-110). Устный подкорпус содержит устную публичную и устную непубличную речь. Кроме того, внутри устного подкорпуса сформирован так называемый Мультимедийный подкорпус русского языка, который, как предполагается, в конечном итоге будет состоять из транскриптов мультимедийной (теле- и радио) рекламы, а также (и в основном) из транскриптов русских фильмов (ср. Grishina 2006). В конце 2006 г. объем подкорпуса стандартной устной речи составил порядка 4 млн словоупотреблений, объем кинематографического подкорпуса – порядка 0,5 млн словоупотреблений<sup>2</sup>. В ближайшие годы мы рассчитываем довести объем стандартного устного подкорпуса до 10 млн словоупотреблений, а кинематографического – до 1,5–2 млн словоупотреблений.

Включение в Корпус такого необычного и нетрадиционного материала, как транскрипты кинофильмов, поставило несколько интересных вопросов, решению которых и посвящено настоящее исследование. Прежде всего, возникает вопрос о правомерности объединения кинематографического и стандартного устного подкорпусов в один подкорпус, или, иначе, вопрос о том, является ли речь кинематографа разговорной? Ведь основными характеристиками разговорной речи, которые обеспечивают ее синтаксическое, семантическое, а также морфологическое своеобразие, являются ее спонтанность, неподготовленность, а также невозпроизводимость. Но в отношении кинематографа об этих характеристиках и речи быть не может: речь кино не только не спонтанна и подготовлена (основана на сценариях разной степени подробности), но и принципиально воспроизводима (сама возможность осуществления нескольких дублей одной сцены в сочетании с последующим озвучиванием фильма снимает идею невозпроизводимости в принципе). С другой стороны, было бы странно считать речь кино письменной.

<sup>1</sup> Исследование осуществляется как составная часть специализированной программы ОИФН Российской академии наук и при поддержке Российского фонда фундаментальных исследований, грант № 06 06-80133а.

<sup>2</sup> К созданию подкорпуса мультимедийной рекламы создатели Корпуса пока не приступали.

Очевидно, решение этой дилеммы таково: речь кино – это имитация устной речи<sup>3</sup>. Но если согласиться с тем, что имитация разговорной речи – базисная характеристика языка кино, то, очевидно, придется признать, что в кинотранскриптах мы должны сталкиваться с некоторым набором лингвистических маркеров, которые в данной культуре воспринимаются как маркеры устной речи<sup>4</sup>. Таким образом, уже на данном этапе, не дожидаясь момента, когда русский кинематографический корпус начнет функционировать в полном объеме, можно поставить вопрос о предварительном определении состава таких маркеров для русского языка. Этому и посвящено предлагаемое исследование.

## 2. Материал

В качестве материала для сопоставительного исследования мы использовали, с одной стороны, транскрипты четырех фильмов, с другой, – их сценарии (или драматические произведения, послужившие их основой)<sup>5</sup>. Кроме того, для исследования были привлечены субтитры к фильмам, созданные специально для слабослышащих зрителей<sup>6</sup> (см. раздел 4.2).

## 3. Технология

Практика показывает, что между реальным текстом фильма, с одной стороны, и его драматургической основой или сценарием, – с другой, обязательно существует то или иное количество расхождений, которые затрудняют пословное сопоставление транскрипта фильма и сценария/пьесы. Поэтому текст транскрипта и сценарий/пьеса были разбиты нами на небольшие блоки (длиной от одного до трех предложений), которые и стали предметом сопоставления. Необходимыми условиями для сравнения двух блоков было следующее: 1) они должны были соответствовать друг другу с точки зрения содержания; 2) они должны были быть одним и тем же речевым актом (сочетанием речевых актов)<sup>7</sup>.

## 4. Результаты сопоставительного анализа

### 4.1. Сценарий/пьеса → транскрипт

Ниже мы перечислим лишь наиболее частые и типичные расхождения между транскриптами фильмов и текстами-первоисточниками (сценариями и драматическими произведениями)<sup>8</sup>.

#### 4.1.1. Межфразовые скрепы

Для транскриптов характерно появление в начале фраз специальных частиц-скреп, а именно:

1. Ну<sub>1</sub>...<sup>9</sup>, вводящее нейтральные констатации (утверждения), которые не содержат противопоставления предыдущему тексту или какому-либо действию или высказыванию собеседника. За скрепой ну<sub>1</sub> в начале фразы может следовать небольшая пауза (такое ну<sub>1</sub> можно называть **ну пугативным**, т.е. следующая за ну<sub>1</sub> фраза подается говорящим как результат определенного размышления):

<sup>3</sup> Вопрос о том, насколько эти маркеры универсальны, т. е. свойственны другим языкам, кроме русского, в данной статье не ставится, хотя, конечно, его исследование представляет немалый интерес.

<sup>4</sup> Несколько ранее сходная идея была сформулирована в работе (Капанадзе 1986), где киноязык рассматривался как член триады киноязык – язык театра – разговорная речь и был проанализирован с точки зрения фонетики. Автор писал о том, что, в отличие от языка театра, который максимально подчеркивает свое (фонетическое) отличие от реальной разговорной речи, чтобы соблюсти дистанцию между сценой и зрительным залом, язык кино принципиально ориентируется на «речь улицы», реальную разговорную речь.

<sup>5</sup> «Старшая сестра» (режиссер Г. Натансон, далее СС) и «Пять вечеров» (реж. Н. Михалков, далее ПВ) А. Володина, «Обыкновенное чудо» (реж. М. Захаров, далее ОЧ) Е. Шварца, «ДМБ» (сценарий И. Охлобыстина, реж. Р. Качанов, далее ДМБ). Общий объем – 32 000 словоупотреблений.

<sup>6</sup> «Афоня» (далее АФ) и «Джентльмены удачи» (далее ДУ; реж. Г. Данелия), «Московские каникулы» (далее МК; реж. А. Сурикова). Общий объем – 20 000 словоупотреблений.

<sup>7</sup> Например, сопоставление вопроса и утверждение не проводилось (следует специально отметить, что достаточно часто одно и то же содержание передается в фильме и сценарии/пьесе разными речевыми актами).

<sup>8</sup> Наверняка при более массивном исследовании могут быть выявлены дополнительные типы расхождений, но лидеры, по-видимому, определены уже на данном этапе.

<sup>9</sup> Подробное и на данный момент, по-видимому, исчерпывающее описание типов функционирования маркера ну в устном тексте можно найти в статье Шмелев 2005.

(1) *Ну / дней / может быть / десять... || Дней, может, десять...*<sup>10</sup> (ПВ).

Но такая пауза необязательна:

(2) *Ну ты же ничего не знаешь / ни пьесу / ни роль / как ты можешь меня учить! || Ты же ничего не знаешь — ни пьесу, ни роль, как ты можешь меня учить!* (СС),

и тогда ну<sub>1</sub> представляет собой наиболее нейтральный, практически десемантизированный, пустой способ соединения констатаций в устном тексте.

**2. Ну<sub>2</sub>...**, вводящее побуждения (может сочетаться как с явным, так и со скрытым императивом):

(3) *Эй вы / приговоренные к смерти! Ну / в путь! || Эй, вы там, приговоренные к смерти! В путь!* (ОЧ).

Частным случаем императивного ну<sub>2</sub> является вопросительное ну<sub>2</sub>, которое подразумевает побуждение к ответу:

(4) *Ну / молодой человек / вы служить собираетесь? Или вы к нам на побывку прибыли? || Вы, молодой человек, собираетесь служить, или вы к нам на побывку?* (ДМБ).

**3. А...**, которое вводит фразы, содержащие то или иное противопоставление предыдущему (тексту или действию):

(5) *А дядька правильно говорит / быть средним инженером ещё куда ни шло... || Дядя правильно говорит: быть средним инженером — куда ни шло...* (СС).

Поскольку ситуация вопроса обычно возникает как результат противопоставления предыдущему (тексту или действию), то скрепа А... часто используется для ввода вопроса:

(6) *А сколько же ей стукнуло...? || Сколько же ей стукнуло?*; (ПВ)

(7) *А вот ответ мне пожалуйста / на такой вопрос || Вот ответ мне на такой вопрос* (ПВ).

**4. Да...**, которое часто вводит восклицания:

(8) *Это не пассажир / а дрянь какая-то! — Да что вы такое говорите! || Он не пассажир, а дрянь какая-то! — ... Вы что такое говорите!* (ДМБ).

Таким образом, Да... – эмоциональная скрепа, и именно поэтому легко употребляется вместо нейтрального ну<sub>2</sub> при императивах:

(9) *Да молчи ты! || Молчи!* (ОЧ).

#### 4.1.2. Эллипсис союзов

Внутри фразы происходят процессы, обратные процессам на границах между фразами – для внутрифразового синтаксиса характерна редукция союзов (впрочем, это явление гораздо менее частотное, чем появление скреп в начале фразы).

**1. Эллипсис сочинительных союзов.** Это происходит в случаях, когда оба однородных члена устойчиво связаны между собой:

(10) *Ага / в четыре утра... Отличное время. Делу время / потехе час || Четыре утра – отличное время! Делу время, а потехе час;* (ДМБ)

(11) *Внучк'и! Пуля дура / штык молодец || Внучк'и! ... Пуля – дура, а штык – молодец!* (ДМБ);

(12) *...хочешь / вспоминая ее смешные шалости-словечки || ...хочешь, вспоминая ее смешные шалости и словечки* (ОЧ);

(13) *Живете с папой-мамой? || Живете с папой и мамой?* (СС).

**2. Эллипсис подчинительных союзов.** Простые подчинительные союзы целиком подвергаются эллипсу:

(14) *Думаете / не отпустит? || Боитесь, что она вас не отпустит?* (ПВ).

У сложных союзов снимается соотносительное слово в главном предложении («союзная апокопа»):

(15) *...если бы вы меня узнали поближе / вы бы даже примирились с этим недостатком || ...если бы вы узнали меня поближе, то ... вы бы даже примирились с этим недостатком* (СС).

Впрочем, может сниматься и обязательная часть союза в придаточном предложении:

(16) *Случалась неприятность / мы не падали духом || Если у нас случалась какая-нибудь неприятность, мы не падали духом* (СС).

#### 4.1.3. Экспликация перформатива

В ряде случаев расхождение между текстом-источником и транскриптом определяется тем, что в транскрипте на словесном уровне эксплицируется скрытый перформатив последующей или (реже) предшествующей фразы:

<sup>10</sup> Далее в статье цитаты из транскриптов оформляются курсивом, из текстов-первоисточников – подчеркнутым курсивом. Расхождения отмечаются полужирным курсивом.

- (17) **Я не понимаю** / ... зачем ты со мной в кино хотела пойти... || Зачем же тогда со мной в кино пошла? (ПВ);
- (18) **Мне интересно** / ... почему ты от меня скрыл / что получил двойку? || Почему ты от меня скрыл, что получил двойку? (ПВ);
- (19) **Но я считаю** / что человек в жизни должен делать самое большое / что он способен || ...человек должен делать все-таки самое большое, на что он способен (ПВ);
- (20) **Товарищ генерал-лейтенант / я давно хотел спросить** / как с йети быть? || Товарищ генерал, а как же с йети быть? (ДМБ);
- (21) **Я в ужасном горе. Предупреждаю сразу** || Я в ужасном горе (ОЧ);
- (22) **Теперь я вас прошу** / пойдите туда вы и пофантазируйте обратно || ...теперь пойдите туда вы и пофантазируйте обратно (СС).

#### 4.1.4. Введение метатекста

Транскрипты характеризуются довольно частым использованием метатекстовых вставок, которые сами по себе не имеют значения, но выполняют функции текстовых ориентиров, например, сигнализируют о том, что в предыдущей фразе был подведен некий итог или высказана какая-то важная мысль, или что только что было совершенно какое-то важное действие (*Вот. Вот так.*), или предупреждают о том, что нечто важное будет сказано в следующей фразе (*Так вот.*):

- (23) **Предки виноваты. Прадеды / прабабки / внучатые дяди / тети разные / праотцы / ну / и праматери. Вот.** || Предки. Прадеды, прабабки, внучатные дяди, тети разные, праотцы и праматери (ОЧ);
- (24) **Так вот** / Надежда Георгиевна / я прошу вас позвонить моей жене || Я прошу вас позвонить моей жене (СС);
- (25) **Заслуга не в том / чтобы не совершать ошибок / а в том / чтобы вовремя их исправлять. Вот так** || Заслуга не в том, чтобы не делать ошибок, а в том, чтобы вовремя их исправлять (ПВ).

Интересный пример обрамления фразы метатекстом («вводный» метатекст, *так вот*, предложен автором пьесы, а «итоговый» метатекст, *вот так*, появился уже в речи актера):

- (26) **Ну так вот** / товарищ Ильин / эту самую страницу мне читать неохота. **Вот так.** || Ну так вот, эту самую страницу мне читать неохота (ПВ).

Метатекст, как представляется, является производным от скрытого перформатива фразы, что проявляется в тех случаях, когда метатекст и перформатив встречаются одновременно:

- (27) **Дурак ты / вот что**<sup>(метатекст)</sup> **я тебе скажу**<sup>(перформатив)</sup> / дурак... || ... дурак ты такой! (ОЧ).

#### 4.1.5. Экспликация ролей

Для транскриптов, в отличие от текстов-первоисточников, характерно избыточное, с рациональной точки зрения, появление местоимений первого и второго лица:

- (28) **Ты** плюнь мне в рожу-то / **я тебя** прошу как родного! || Плюй мне в рожу, как родного прошу! (ДМБ);
- (29) **Ты** не сердь меня. **Работай** || Не сердь меня, работай (ПВ);
- (30) ...**другая у нас** просто знаменитость || ...другая просто знаменитость!.. (СС);
- (31) **А вы** его не захваливайте / а то он зазнается || Только не захваливайте, он зазнается (СС)<sup>11</sup>.

Отметим, что ввиду своей избыточности экспликация ролей может также рассматриваться как одна из разновидностей реализации интонационного избытка (см. п. 4.1.9).

#### 4.1.6. Интенсификация фатими

Транскрипты характеризуются повышенным использованием фатических языковых средств.

##### 1. Контактные слова. К таковым можно отнести

1.1. глаголы, привлекающие внимание адресата к высказыванию (обычно в императиве или во 2 л.): *понимаешь/понимаете, знаешь/знаете, видишь/видите, (по)смотри(те), (по)слушай(те), учти(те), понял?* и др.:

- (32) **Слышь** / брат! Ты мне поможешь / я тебе помогу / все люди братья!.. || Брат! ... Ты мне поможешь, я тебе помогу – все люди друзья и братья!.. (ДМБ);
- (33) **Вы знаете** / последнее время некогда читать-то было || Последнее время мне некогда было читать (СС);

<sup>11</sup> Для местоимений третьего лица заметить какой-то закономерности не удалось, может быть, нужно расширение материала, но скорее всего, такое различие между местоимениями, как будет показано ниже, имеет под собой объективные основания.

(34) Новенький заскучал. **Понимаешь** / не ест / не спит / и отвечает как-то невпопад || Новенький мой заскучал, не ест, не пьет, невпопад отвечает (ОЧ);

(35) Здесь же формулы! Адски трудно / **посмотри!** А? || Тут же формулы, адский труд! (ПВ);

1.2. обращения (очень часто):

(36) Ну **братцы** / **братцы** / давайте что-нибудь решать / время же идет || Давайте что-нибудь решать, идет время (СС);

(37) **Ребята** / ну / ну не надо! || Не надо! (ДМБ);

1.3. частицы-обращения (формально они совпадают с частицами-скрепами, см. п. 1, но употребляются обычно (но не обязательно) в конце фразы и произносятся с вопросительной интонацией), с помощью которых говорящий побуждает говорящего занять какую-то позицию по отношению к высказыванию, или совершить какое-то действие:

(38) На что же нам с тобой сыграть? А / боец? || На что же нам с тобой, боец, сыграть? (ДМБ);

(39) Вот народец / а? || Вот народец! (ОЧ);

(40) Рано или поздно я все равно встречу принцессу / поцелую ее и превращусь в медведя / и вот тогда вы... Ну? || Рано или поздно – я разыщу принцессу, поцелую ее и превращусь в медведя... И тут... (ОЧ);

(41) Ну что / я стал бы для тебя хуже? Ну? || Стал бы я для тебя хуже или нет? (ПВ);

(42) Вы кажется / ходили в студию при Дворце культуры / да? || Вы же когда-то ходили в студию при Дворце культуры? (СС).

2. **Выражение оценки.** Оценка только что прозвучавшей реплики (или только что осуществленного действия) выражается

2.1. с помощью полноударных частиц *Да* (согласие, подтверждение) и *Нет* (несогласие, отрицание) в начале ответной реплики:

(43) Ну / в какой-то степени это можно понять. – Нет / нормальный человек это не может понять. || В какой-то степени это можно понять. – Нормальный человек этого не может понять (СС);

(44) Женатые за всеми ухаживают! – Ну да / только не все это поощряют || Женатые за всеми ухаживают. – Но не все их поощряют (СС);

2.2. с помощью некоторых междометий:

(45) Господи **боже мой** / такая удача / над детским домом шефствует театр! || Такое везение: над детским домом шефствует театр! (СС);

(46) **Ой!** У этого / что ли? || У этого, что ли? (ПВ);

(47) Мучается? – Ужасно. – Ах ты! так ему и надо! || Мучается? – Ужасно. – Так ему и надо! (ОЧ).

3. **Имитация спонтанности.** Часть междометий используется для создания иллюзии того, что данная фраза родилась у говорящего только что и является результатом его размышления здесь и сейчас. Для этого используются междометия *э* и *м* (с любым количеством итераций, символизирующим протяженность произношения):

(48) Мясо / э-э / виски с содовой и льда побольше || Мне рулечку жирненькую и винца молодого (ДМБ);

(49) Э-э / что вы будете читать? || Ну, что вы прочитаете? (СС).

#### 4.1.7. Апокопа

Для имитации устной речи в транскриптах используется стандартный набор усеченных слов. Апокопа употребляется в тех случаях, когда усечение последней гласной слова не приводит к коммуникативному провалу, т.е. фраза понимается слушающим правильно. Список стандартных апокоп (примеры приводить не будем, они многочисленны): хотя → хоть, чтобы → чтоб, бы → б, же → ж, прямо → прям, это → эт (последняя не отражается на письме, хотя реально встречается достаточно часто). Как стандартную апокопу следует рассматривать также т.н. новый русский вокатив (Маша → Маш, т.е. формальное совпадение вокатива имен а-склонения с GenPl). Лексической апокопой является также довольно частая трансформация вводного выражения может быть → может, а также усечение сложных союзов (см. п. 4.1.2-2).

#### 4.1.8. Рематизация и эмфатизация текста

Для транскриптов по сравнению с текстами-первоисточниками характерно использование разного рода средств рематизации, т.е. интенсификации ремы для облегчения ее восприятия, а также использование средств эмфатизации, т.е. лексической поддержки интонационной эмфазы:

1. использование частицы *-то* в качестве стандартного рематизатора и эмфатизатора, который облегчает слушающему выделение ремы, новой или особо важной информации в реплике говорящего, а также является стандартным средством лексикализации эмфазы:

(50) *Как же я без указательного пальца /... жениться-то буду?* || *Как же я без указательного пальца ... жениться буду?* (ДМБ);

(51) *У нас-то есть / о чем поговорить* || *У нас есть о чем поговорить* (ОЧ);

(52) *Ну что я пою-то?* || *Ой, что я пою!* (ПВ);

(53) *А ты-то счастлива?* – Я / нет || *А ты счастлива? – Я? Нет* (СС);

2. Для аналогичных целей используется также конструкция *что (чего) ..., что ли?*, которая обрамляет рему (впрочем, может использоваться только первая или только вторая часть этой конструкции) и тем самым интенсифицирует ее:

(54) *Не / вы чего / издеваетесь / что ли?* || *Что вы, издеваетесь!?* (ДМБ);

(55) *Этот / что ли? – Еще чего!* || *Этот? – Еще чего!* (ОЧ);

(56) *У тебя что / маникюр / что ли?* || *Маникюр у тебя, что ли?* (ПВ);

(57) *Вы что / не знаете / что ли?* || *Неужели вы не знаете?* (ПВ);

3. изменение порядка слов в целях интенсификации ремы<sup>12</sup>; здесь возможны два варианта:

3.1. перемещение ремы к началу фразы, с тем, чтобы самая важная информация поступала к слушающему в первую очередь<sup>13</sup>:

(58) *...молодые люди / скоро кладбище открывается / люди плакать<sup>(2)</sup> придут<sup>(1)</sup> / им вас видеть никакого удовольствия* || *...молодые люди, скоро кладбище открывается, люди придут<sup>(1)</sup> плакать<sup>(2)</sup>, им вас видеть никакого удовольствия* (ДМБ);

(59) *Никто никогда не подходил так близко<sup>(2)</sup> ко мне<sup>(1)</sup>* || *Никто никогда не подходил ко мне<sup>(1)</sup> так близко<sup>(2)</sup>* (ОЧ);

(60) *Он влюбился<sup>(2)</sup> в нее<sup>(1)</sup> до беспамьятства* || *Он в нее<sup>(1)</sup> влюбился<sup>(2)</sup> до беспамьятства* (ПВ);

(61) *Лида взбалмошная<sup>(2)</sup> немножко<sup>(1)</sup>* || *Лида немного<sup>(1)</sup> взбалмошная<sup>(2)</sup>* (СС);

3.2. интенсификация ремы посредством ее удвоения; последнее достигается отрывом от ремы зависимых от нее слов:

(62) *Только у меня условие / никого<sup>(1)</sup> сюда не водить<sup>(2)</sup>* || *Только у меня условие: сюда никого<sup>(0)</sup> не водить<sup>(1)</sup>* (ПВ);

(63) *...поссорился с ней Слава / она лекции-то<sup>(1)</sup> ему и не дает<sup>(2)</sup>!* || *...поссорился с ней Слава, она ему конспекты<sup>(0)</sup> не дает<sup>(1)</sup>* (ПВ);

(64) *... хорошо / это<sup>(1)</sup> мы учтем<sup>(2)</sup>* || *Хорошо, мы это<sup>(0)</sup> учтем<sup>(1)</sup>* (СС);

(65) *Вот поэтому Лида так<sup>(1)</sup> тебе и ответила<sup>(2)</sup>* || *Поэтому Лида тебе так<sup>(0)</sup> и ответила<sup>(1)</sup>* (СС).

#### 4.1.9. Реализация интонационного избытка

Транскрипты отражают такую особенность устной речи, как использование тех или иных лексических средств только для того, чтобы на них могла осуществиться определенная интонация, нужная говорящему (ср. «слово как носитель нужной мне интонации», М.М.Бахтин).

1. Прежде всего, для этой цели используются повторы разной степени длины (в принципе, такой повтор не ограничен ничем, кроме а) интонационного «размаха» говорящего и б) продолжительности экранного действия, которое сопровождается данным высказыванием и данным повтором). В текстах-первоисточниках такие повторы либо отсутствуют, либо только обозначаются – двойным, максимум, тройным повтором той или иной лексической единицы:

(66) *Ты проверь / проверь* || *Ты проверь* (ДМБ); *Вот / вот-вот! Я хочу / чтоб ты почитал. Я хочу / чтобы ты почитал* || *Вот, хочу, чтоб ты прочитал* (ПВ);

(67) *Постойте-постойте-постойте / как ваша фамилия?* || *Постойте, как ваша фамилия?* (СС);

(68) *Это он! Он! Он! Он!* || *Это он, это он!* (ОЧ);

(69) *Ах ты наглец / а? Ну ты наглец! Ну ты наглец! Ну наглец!* || *Ах ты наглец!* (ОЧ).

Повтор может быть не чисто лексический, а синонимический:

(70) *Все сидели уже в машинах. А кругом все плачут / режут...* || *Они уже сидят на машинах, все кругом плачут* (ПВ).

<sup>12</sup> Разумеется, в текстах транскриптов встречаются перестановки, никак не затрагивающие рему высказывания, и в их употреблении не удалось заметить никакой закономерности.

<sup>13</sup> См. об этом РРР 1973, с. 380-393).

2. Кроме того, избыток интонации реализуется на словах-интенсификаторах, типа *очень*, *ничего* и *никакой* (при отрицании), *ровно* (при обозначении количеств, сроков), *чрезвычайно* и некоторых других:

(71) *Это то место / где вы интересно и **чрезвычайно** увлекательно проведете ближайшие два года || ...это не склад, это то место, где вы интересно проведете два года* (ДМБ);

(72) *В стройбат у меня нет **никакого** настроения || В стройбат у меня нет настроения* (ДМБ);

(73) *Но в то же время моя неуверенность **вовсе** не означает / что я ни на что не способен || Но в то же время моя неуверенность не означает, что я действительно ни на что не способен* (СС);

(74) *Вам здесь будет **очень** удобно || Вам здесь будет удобно* (ПВ);

(75) *В детстве я **всегда** завидовала девочкам / у которых были братья || В детстве я завидовала девочкам, у которых есть братья;*

(76) *Вы привлекательны / я / **чертовски** привлекателен / чего зря время терять? || Вы привлекательны, я привлекателен – чего же тут время терять?* (ОЧ).

#### 4.2. Транскрипт → субтитры

Субтитры, как известно, перемещаются по экрану, и скорость их перемещения определяется двумя довольно жесткими факторами: 1) субтитры должны быть привязаны к изображению, 2) зритель должен успевать не только читать субтитры, но и следить за изображением. Тонкое соотношение этих двух разнонаправленных факторов определяет скорость движения слов по экрану. Но какова бы эта скорость ни была, она так или иначе ведет к сокращению реального «фильмотекста». Соответственно, интересно проверить, по каким параметрам проходит это сокращение, т.е., проще говоря, интересно определить, чем именно из содержания «фильмотекста» готовы пожертвовать авторы субтитров, – при том, что фильм должен остаться понятным зрителю.

Версия, которая приходит в голову первой, заключается в том, что «под нож» идут все элементы фразы, которые не несут «предметного» содержания, т.е. выполняют не коммуникативную (в якобсоновском смысле), а фатическую, прагматическую и поэтическую функции. Следовательно, по этой версии, состав элиминируемых частей высказывания должен в общем и целом совпадать с составом наращений на исходных драматический текст, перечисленных в разделе 4.1.

Анализ субтитров к трем фильмам показал, что эта версия, похоже, правильная: порядка 84% расхождений между транскриптами и субтитрами прямо укладываются в приведенную выше классификацию.

##### 4.2.1. Межфразовые скрепы

###### 4.2.1-1. Ну<sub>1</sub>:

(77) <Ну><sup>14</sup> у тебя получается, а у меня нет (АФ)

###### 4.2.1-2. Ну<sub>2</sub>:

(78) <Ну> ныряй! Быстро! (ДУ)

###### 4.2.1-3. А...:

(79) <А> не хотите ли чего-нибудь выпить? (МК)

###### 4.2.1-4. Да...:

(80) <Да> нету больше, кончились (АФ)

...

##### 4.2.3. Перформативы:

(81) Ясно? <Я спрашиваю,> ясно? (ДУ)

##### 4.2.4. Метатекст:

(82) Ну, <так>! У вас свои законы, у нас свои! (МК)

##### 4.2.5. Свертывание ролевой структуры:

(83) <Я тебе> верю (МК);

(84) Пойми <ты>, люди тебе помочь хотят (АФ);

(85) Здесь <у нас> будет спортзал (ДУ);

(86) Я <вам> говорю, его нет! (АФ)

##### 4.2.6. Редукция фатических средств

###### 4.2.6-1.1. Глаголы:

(87) <Слышь, Доцент>? Сразу расколется! (ДУ)

###### 4.2.6-1.2. Обращения:

(88) Беда нашего искусства, <Сань,> не в том, что ему плохо, а в том, что ему недостаточно плохо (МК)

###### 4.2.6-2. Скрытие оценки

<sup>14</sup> Усечения в субтитрах по сравнению с транскриптами берутся в угловые скобки.

4.2.6-2.1. Да, нет:

(89) <Да **нет**>, у нас совсем другой телефон (АФ);

(90) <Да,> за шлем бы нам срок сбавили бы (ДУ)

4.2.6-2.2. Междометия:

(91) <**О-ой!**> Не надо. У меня есть итальянское средство (МК)

4.2.6-3. Снятая спонтанность:

(92) <**Э-э**>... А мы <**э-э**> быстро ее зарежем, и вы спокойно сможете ее похоронить (МК)

...

...

...

4.2.8. Дерематизация текста

4.2.8-1. -го:

(93) Ты где работаешь<**-то**>, Архимед? (АФ)

4.2.8-2. что ...что ли:

(94) Ты <**что**>, глухой, что ли? (ДУ);

(95) <Ну,> а я чего, отдыхаю, <**что ли**>;

(96) <**Что ж это**> она? В Кремле, <**что ли**,> скучает? (МК)

...

...

...

4.2.9. Интонационный избыток

4.2.9-1. Повторы:

(97) Да завязал я, <**завязал**> (АФ)

4.2.9-2. Слова-интенсификаторы:

(98) Примус <**очень**> худой, пожар может быть (ДУ)

Обратим внимание на то, что пропущенные пункты **4.1.7. Апокопа**, **4.1.8-3. Рематизация посредством перестановки слов** сохраняют исходное количество слов, а пункт **4.1.2. Эллипсис союзов** и вовсе ведет к сокращению количества слов в транскрипте по сравнению с исходным текстом (пьесой или сценарием). Таким образом, по этим пунктам в субтитрах, основная задача которых – сокращение текста транскрипта, транскрипт не должен претерпевать никаких изменений. Это, однако, не так: хотя и редко, но происходит:

4.2.2. Восстановление союзов

4.2.2-1. сочинительных:

(99) Лестница ... **отвратительная** – **грязная**, **ужасная** {**и**}<sup>15</sup> **кривобокая!** (МК)

4.2.2-2. подчинительных:

(100) Вася, ... я ему давеча сказал, {**что**} у меня насморк (ДУ)

4.2.7. Использование полных форм вместо усеченных:

(101) А если {**бы**} я у тебя там микроскопы стал крутить? (АФ)

– реально ...если **б**...

4.2.8. Восстановление порядка слов

4.2.8-3.1. Перемещение ремы слева направо:

(102) Нам <по пол-> **молочного**, **Афанасий Николаевич**, и **картофельные**<sup>(2)</sup> **котлеты**<sup>(1)</sup> (АФ)

– в транскрипте **котлеты картофельные**;

(103) У тебя там в машине есть такая **маленькая**<sup>(2)</sup> **кнопочка**<sup>(1)</sup>, **черненькая**, **непонятная** (МК)

– в транскрипте **кнопочка маленькая**

Эти факты, по-видимому, показывают, что помимо фактора сокращения исходного транскрипта фильма при создании субтитров, связанного с необходимостью соблюсти определенную скорость перемещения текста по экрану, действует, хотя и в гораздо более слабой степени, также фактор перевода устного текста в письменный, что и сказывается отчетливо в тех случаях, когда количество слов в субтитрах не только не уменьшается, но иногда и возрастает.

В заключение этого раздела отметим, что из единиц, которые подвергаются в субтитрах систематическому сокращению, но при этом пока не замечены в систематическом добавлении при переходе от исходного текста к транскрипту, нужно назвать этикетные формулы (*пожалуйста, будьте добры, извините, здравствуйте* и т.д.), а также вводные слова и выражения (*кстати, в общем, между прочим* и т.д.). Впрочем, обе эти группы попадают в нашу классификацию (пп. 4.1.6-1 и 4.1.6-2).

<sup>15</sup> В фигурных скобках мы показываем союз, появившийся в субтитрах и отсутствующий в транскрипте фильма.



### 5. Предварительные результаты проведенного анализа

Как видим, при трансформации письменный текст (пьеса/сценарий) → устный текст (транскрипт) → письменный текст (субтитры) появляются и исчезают, а также претерпевают те или иные изменения одни и те же группы единиц. Практически все описанные выше процессы хорошо укладываются в давно известный принцип, лежащий в основе организации устной речи: устная речь в нормальном случае является компромиссом между интересами слушающего (получить как можно больше информации, прилагая наименьшее количество усилий) и интересами говорящего (передать как можно больше информации, прилагая наименьшее количество усилий), при безусловном приоритете интересов слушающего<sup>16</sup>: см. таблицу 1<sup>17</sup>.

|   | Интересы слушающего | Интересы говорящего |
|---|---------------------|---------------------|
| 1. Межфразовые скрепы<br>3. Экспликация перформатива<br>4. Введение метатекста<br>5. Экспликация ролей<br>6. Интенсификация фатики: контактные слова<br>8. Рематизация текста: использование рематизаторов (частиц и конструкций) | +                   | –                   |
| 8. Рематизация текста: перестановка слов  | +                   | 0                   |
| 2. Эллипсис союзов<br>7. Апокопа  | 0                   | +                   |
| 6. Интенсификация фатики: выражение оценки, имитация спонтанности<br>9. Реализация интонационного избытка (повторы, слова-интенсификаторы)  |                     |                     |

Таблица 1

### 6. Возможности практического использования

Изложенные результаты, как представляется, имеют некоторый теоретический интерес – по крайней мере, становится достаточно очевидным, что транскрипты кинофильмов можно включать в устные подкорпуса. Однако возможные и некоторые практические следствия.

В частности, на более формализованный уровень можно перевести оценку того или иного текста как «устно-го» или «письменного». Эта проблема не возникает, когда исследователь знает источник происхождения данного текста, и уж тем более эта проблема неактуальна, если исследователь имеет дело с устным первоисточником, т.е., например, сам расшифровывает некоторую диктофонную запись.

Однако достаточно часто мы сталкиваемся с ситуацией, когда 1) не известен первоисточник текста, 2) не известна степень его аутентичности (т.е. степень его редакторской обработки при расшифровке). И тогда приходится полагаться на свою интуицию – носителя языка и лингвиста, – с тем, чтобы определить степень «устности» того или иного текста.

Более того, возможна постановка задачи, когда с помощью тех или иных программных фильтров из большого количества текстов (например, из Интернета) автоматически должны быть отобраны тексты, которые мо-

<sup>16</sup> «Видите ли, если мне хочется одного, а миссис Мак-Вильямс другого и мы решаем поступить так, как хочется миссис Мак-Вильямс, – а мы всегда решаем только так, – это называется у нас "пойти на компромисс"» (М. Твен. Мак-Вильямсы и автоматическая сигнализация от воров).

<sup>17</sup> Не покрываются этим объяснением – без натяжек – только случаи 6.2 и 6.3 – выражения оценки и имитация спонтанности.

гут – в предварительном порядке – претендовать на статус устных, живых и т.д.

В таких случаях учет маркеров, описанных в данной статье, становится не теоретической прихотью, а жесткой практической необходимостью. Представляется, что на некотором наборе отчетливо письменных текстов нужно определить некий числовой стандарт употребления следующих единиц:

- ну (пп. 4.1.1-1, 4.1.1-2, 4.1.6-1.3)
- а (пп. 4.1.1-3, 4.1.6-1.3)
- да (пп. 4.1.1-4, 4.1.6-1.3, 4.1.6-2.1)
- глаголы речи, путативные глаголы в 1Sg (4.1.3)
- вот, так (4.1.4)
- местоимения 1-2Sg-Pl (4.1.3, 4.1.5)
- глаголы знания и восприятия *знать, понимать, слышать, слушать, видеть* в 2 л. и в императиве + специальные формы типа слышь (п. 4.1.6-1.1)
- именительные падеж одушевленных существительных – в устной речи таких форм больше из-за их формального совпадения с вокативами (п. 4.1.6-1.2)
- нет 4.1.6-2.1
- междометия (4.1.6-2.2, 4.1.6-3)
- апокопированные формы (4.1.7)
- личные имена а-склонения в GenPl, формально в большинстве случаев совпадающем с новым русским вокативом (4.1.7)
- то (4.1.8-1)
- что ли (4.1.8-2)
- тройные и более повторы (4.1.9-1)
- слова-интенсификаторы, прежде всего *очень, никогда, ничего, никакой, всегда* и нек. др.

Соответственно, текст, в котором хотя бы один из этих показателей будет значимо выше «письменной» нормы, автоматически попадает под подозрение в том, что он на самом деле устный. Значимое превышение по всем перечисленным пунктам гарантирует аутентичность устного текста практически на 100%<sup>18</sup>.

#### Список литературы

1. Крейдлин Г.Е. Невербальная семиотика // М.: Новое литературное обозрение, 2002.
2. Якобсон Р.О. О лингвистических аспектах перевода // Вопросы теории перевода в зарубежной лингвистике. М.: 1978. С.16-24.
3. Капанадзе Л.А. Разговорная речь и киноязык // Л. А. Капанадзе. Голоса и смыслы. Избранные работы по русскому языку. М.: 2005. С. 228-231.
4. Русская разговорная речь // М.: 1973 (PPP 1973).
5. Шмелев А.Д. «Заполнители пауз» как коммуникативные маркеры. // Язык. Личность. Текст. М.: 2005.
6. Grishina Elena. Spoken Russian in the Russian National Corpus (RNC) // LREC2006 5th International Conference on Language Resources and Evaluation. 22-28 May 2006. Genoa Italy. Proceedings. P. 121-124.

<sup>18</sup> Впрочем, требует проверки роль маркеров разговорной речи в «квази-разговорных» текстах, характерных для электронной коммуникации (чаты, форумы и под.). Не исключено, что числовые параметры этих текстовых жанров будет близки к характеристикам разговорных текстов.