

Ягунова Е.В. Исследование избыточности русского звучащего текста // Избыточность в грамматическом строе языка / Отв. ред. М. Д. Воейкова. СПб.: Наука, 2010. — 462 с. (ACTA LINGUISTICA PETROPOLITANA. Труды Института лингвистических исследований РАН / Отв. редактор Н. Н. Казанский. Т. VI. Ч. 2). с. 90-114
Ягунова Е.В. (СПбГУ)

Исследование избыточности русского звучащего текста

1. Введение

При исследовании процессов восприятия и понимания устной речи неизбежно обращение к вопросам, связанным с информационной избыточностью как неотъемлемым свойством любого текста. Употребляя термин «информационная избыточность» мы подчеркиваем, что для нас подход к исследованию избыточности связан с тем направлением в лингвистике, которое наследует идеи теории информации. Информационная избыточность является тем свойством любого текста, которое обеспечивает возможность успешного восприятия речи (особенно звучащей речи). Подчеркиваем, что с этой точки зрения любой текст на естественном языке характеризуется информационной избыточностью, в противном случае он не может быть воспринят и понят адресатом¹.

В этом принципиальное отличие понятия «информационная избыточность» от более широкого понимания «избыточности» в данном тематическом сборнике². Для того, чтобы исследовать информационную избыточность, необходимо опираться не только на ее качественные, но и на количественные признаки. Они могут быть определены в результате проведения вычислительных экспериментов (ср. многочисленные работы Р.Г. Пиотровского, напр., Пиотровский 1968; Пиотровский 1999; Пиотровский 2006) и экспериментов с информантами (прежде всего, экспериментов по восприятию текста).

В данной статье исследование информационной избыточности основывается на экспериментах, позволяющих оценить предсказуемость тех или иных единиц текста. Для современной теории восприятия речи стало естественным опираться на представление о том, что фонетические характеристики текста не могут содержать того количества информации, которое достаточно для полной фонемной интерпретации всего текста (всех слов текста). Положение о том, что в тексте сосуществуют сегменты полного и неполного типа произнесения, из которых только первые могут распознаваться за счет анализа

¹ Примером текста без информационной избыточности является текст программы, написанный на одном из языков программирования.

² Ср., напр., в (Даль 2009: 26): «...сообщение может поступать в более или менее искаженном виде, но все же может быть понято, если содержит достаточно и з б ы т о ч н о с т и, чтобы оно могло быть восстановлено получателем». Далее – что согласуется с идеями нашей статьи – в этой книге анализируются «информационно автономные» и «информационно неавтономные элементы» текста (в терминологии Э. Даля).

фонетических характеристик, впервые было сформулировано в (Бондарко и др. 1974). Прочие сегменты могут интерпретироваться только в результате контекстной предсказуемости, то есть предсказываться на основании знания контекста. Соотношение сегментов полного и неполного типа произнесения в рамках текста определяется самыми разными характеристиками, прежде всего – функциональным стилем текста. Очевидно, однако, что даже подготовленное дикторское чтение содержит большое количество сегментов неполного типа произнесения (слов, слов, возможно, синтагм и даже фраз), восстанавливаемых на основании присущей тексту избыточности.

Возможно, наиболее иллюстративным примером функционирования избыточности при восприятии текста является роль морфонологических явлений (см. Касевич 2006: 266–267, 280–282). Так, например, большая часть морфонологических характеристик слова может приходиться на безударные сегменты (слоги), тогда – в силу сегментной редукции – собственно морфонологическая информация не может быть извлечена из соответствующего сегмента слова, но лишь на основании более широкого контекста. В этом случае, по-видимому, на первый план могут выступать интегральные характеристики фонетического слова (ФС), которое наряду со знаменательной словоформой может включать и служебные слова (напр., предлоги в предложно-падежных конструкциях). Так, «...если тот или иной морфонологический контекст требует данного фонологического облика, то этот фонологический облик служит, в свою очередь, дополнительным сигналом соответствующей морфонологической характеристики» (Касевич 2006: 266). Именно интегральные признаки, относящиеся к уровню слова (ФС), в целом ряде случаев оказываются доступными восприятию человека. Вероятно, в особенной степени сказанное следует учитывать при исследовании восприятия на материале русского языка, т.к. для него характерны свободный порядок слов, морфонологическая сложность, подвижное разноместное ударение и высокая степень сегментной редукции.

С другой стороны, человек, как известно, не в состоянии проводить пофонемное декодирование слов звучащего текста в силу ограничений своей психофизиологической организации (памяти и быстродействия). В результате этих ограничений и благодаря возможностям контекстной предсказуемости в процедурах восприятия текста человек оперирует сравнительно большими единицами: словами, а возможно – синтагмами и фразами. Как известно, увеличение формата подобных единиц может увеличивать скорость восприятия и понимания (см., например, Грановская 1974). Однако значительное укрупнение единиц (и ускорение восприятия) возможно лишь в тех ситуациях коммуникации, когда максимально «включены» процедуры предсказуемости, в частности,

предполагающие наличие у слушающего соответствующей «базы знаний»: знание предметной области, терминологии и т.д.

Любой естественный текст характеризуется компрессией как результатом эллиптирования некоторого количества информации. Эллиптирование может происходить на самых разных уровнях – от фонетического до смыслового. Уже не раз говорилось о неизбежности присутствия в любом тексте сегментов неполного типа произнесения, что соотносимо с эллиптированием фонетической (а также соответствующей грамматической и/или лексической) информации. Эллиптирование говорящим тех или иных смысловых фрагментов зависит от коммуникативной ситуации, прежде всего – функционального стиля текста и от соответствия «баз знаний» говорящего и слушающего (адресанта и адресата). Например, если слушающий знает предметную область, владеет темой разговора, то говорящий (в силу закона экономии усилий), как правило, опускает ту информацию, которая может быть восстановлена слушающим на основании этого знания. Таким образом, восстановление компрессированного текста слушающим в процессе восприятия оказывается обязательным компонентом, обеспечивающим успешность коммуникации. Возможности такого восстановления лежат в природе избыточности, свойственной не только отдельным естественным текстам, но и шире – речевой и коммуникативной деятельности в целом. Успешность восстановления зависит от типа и степени компрессии текста, что определяется условиями коммуникации (например, уровнем шума при восприятии в шумном месте: в метро, на вокзале, на многолюдной вечеринке), функциональным стилем текста и т.д.

2. Материал и методика

В качестве компрессированных текстов в данной работе рассматривается лакунарный текст. При создании лакунарного текста в исходном тексте вычеркивалось каждое четвертое слово текста (в соответствии с традицией 'cloze tests'³). Эти слова замещались на амплитудно-модулированный белый шум (программа была составлена к.б.н. А.В. Венцовым); таким образом, заменяемый (слово) и замещающий сигнал (белый шум) имели одинаковые энергетические огибающие⁴. При делении текстов на слова (с

³ Как правило, эта методика используется при работе с письменными текстами. К сожалению в последнее время методика 'cloze tests' используется сравнительно редко («вышла из моды»), несмотря на то, что доказала свою убедительность при решении самых разных задач, связанных с контекстной предсказуемостью (ср., например, Aborn et al. 1959; Alderson 1980; Bachman 1985; Berg 1991; Wormuth 1965; Wormuth 1967; Brown 1983b; Brown 1984; Brown 1988; Brown 1989; Brown 1993; Calvo 1999; Chihara et al. 1994; Cziko 1983; Fillenbaum et al. 1963; Oller et al. 1972; Oller et al. 1994; Owens et al. 1997; Nusbaum 1994; Taylor 1953 и др.). В ряде случаев аналогичного рода методика фигурирует в современных работах под названием *syntactic priming* (что, видимо, отражает «моду» на исследования с применением методики прайминга).

⁴ Непосредственное удаление соответствующих слов в звучащих текстах привело бы к существенному искажению просодической структуры.

последующим вычеркиванием 4-го) слово определялось как орфографическое слово или слово с предлогом *в, с, к* (*во, со, ко*)⁵. Такой принцип деления текста на слова отражает наше внимание к слогу как минимальной оперативной (психолингвистической) единице: единица членения текста не может быть меньше, чем слог (для русского языка такое ограничение заставляет рассматривать названные выше предлоги как элементы соответствующих ФС).

В качестве исходного материала были выбраны два переводных текста на русском языке:

1. Отрывок из официальной публикации «Закон об иностранных инвестициях во Вьетнаме и нормативные акты, изданные на его основе» (в дальнейшем «деловой текст»);
2. Отрывок художественной прозы Нам Као «Ти Фео» с элементами диалога (в дальнейшем «художественный текст»).

Иллюстрация принципа лакунаризации – на месте зашумляемых слов стоит «()».

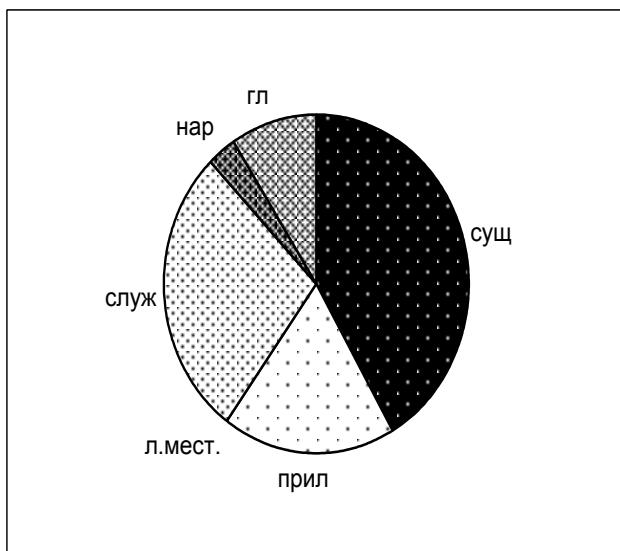
Стороны, участвующие в реализации (1) о деловом сотрудничестве, (2) предприятия с иностранным капиталом (3) право самостоятельно устанавливать (4) и планы своей (5) деятельности.

Удаленные слова: 1 – договора, 2 – и, 3 – имеют, 4 – программы, 5 – производственно-хозяйственной – *Опять притащился! Пора (1) знать своё место, (2) тебе не банк! (3) словами он с размаху (4) монету прямо на (5).*

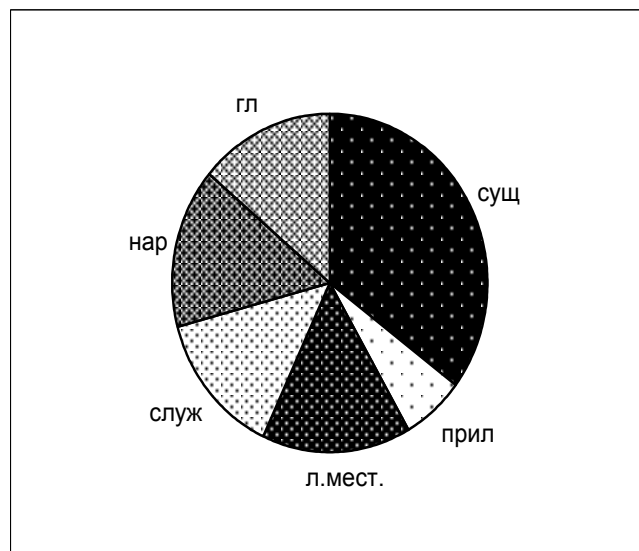
Удаленные слова: 1 – бы, 2 – здесь, 3 – с этими, 4 – швырнул, 5 – землю.

В результате из каждого из текстов было удалено по 65 слов. Таким образом, лакунарные тексты оказываются подвергнутыми незначительной компрессии, принцип удаления слов из текста абсолютно формален, и среди удаленных слов оказываются слова с разными фонетическими и/или внефонетическими характеристиками (служебные и знаменательные, полноударные и клитики, элементы темы и ремы и т.д.). В деловом тексте 28% удаленных слов составляли служебные слова, в художественном тексте доля служебных слов существенно ниже – 14% (см. рис.1). Частеречные классы выделялись следующим образом: существительные, глаголы (в том числе модальные или служебные), прилагательные (в том числе местоимения-прилагательные), личные местоимения (только в художественном тексте), наречия (в том числе предикативы), служебные части речи (предлоги, частицы, союзы). Для удаленных слов обоих текстов примерно равные доли составляют существительные, в отношении остальных частей речи прослеживается гораздо больше различий, чем сходства (см. рис.1).

⁵ В рамках общей работы типологического характера были проведены лакунарные эксперименты с рассматриваемыми текстами (см. ниже) – устными и письменными – на русском и французском языках (см. подробнее Касевич, Ягунова 2006а). Для французского языка при подсчете словоупотреблений текста слово приравнивалось к орфографическому слову или слову с артиклем, притяжательным местоимением и предлогами *à, de, en*.



А



Б

Рис.1. Распределение удаленных слов по частям речи. А – для делового текста, Б – для художественного текста

В ходе лакунарного эксперимента 20 испытуемых должны были восстановить текст, то есть записать его целиком, восстанавливая пропущенные слова.

3. Результаты эксперимента

При идентификации слова в каждой точке текста слушающий, по-видимому, осуществляет выбор между классами и далее – между элементами этих классов. Речь идет, в первую очередь, о таких грамматических классах, как «часть речи» и «член предложения». Грамматический класс удаленного слова может предсказываться на основании синтаксической организации клаузы, фразы или даже всего текста. Наряду с этим можно предположить более близкий путь идентификации для компонентов неоднословной целостности (например, предложно-падежной конструкции, сложной номинации или фразеологизма). В этом случае, вероятно, удаленное слово восстанавливается на основании лексико-синтаксической и/или семантической сочетаемости/несочетаемости. Позиция такого слова, как правило, предсказывается на основании предшествующей (или последующей) позиции. Примерами подобных неоднословных целостностей могут служить сложные номинации *транспортные средства, внутренний рынок* и фразеологизмы *и так далее, во все горло, от всей души*, компоненты которых удалялись из текстов.

Характеристики шума как замещающего сигнала (длительность и амплитудно-частотные характеристики) позволяют извлекать просодическую информацию об удаляемых словах (прежде всего, о длине удаляемого слова в слогах). Очевидно, что эта информация учитывается слушающим при выборе кандидатов в процедурах

восстановления замещенных шумом слов. Результаты рассматриваемых лакунарных экспериментов позволяют предполагать, что в рамках данной методики основная опора на просодические признаки слова происходит при выборе среди элементов некоторого класса, задаваемого синтаксическими, смысловыми или сочетаемостными характеристиками.

В табл. 1 приведены обобщенные результаты восстанавливаемости удаленных слов. Точная восстанавливаемость слов делового и художественного текста различаются незначимо⁶, что резко отличает этот экспериментальный режим от всех прочих. Художественный текст восстанавливается хуже делового по прочим параметрам (грамматические классы «часть речи» и «член предложения», «отказы»).

Таблица 1. Правильность восстановления удаленных слов (%)

эксп. вариант стиль текста	+	ч.р.	ч.п.	О.
художественный	33	50	52	39
деловой	39	66	66	24

Условные обозначения: «+» – точное восстановление исходного слова, «ч.р.» и «ч.п.» – восстановление грамматического класса «часть речи» и «член предложения», «О.» – отказы (здесь и далее)

4. Обсуждение результатов

4.1. Грамматические классы. Обобщенные данные

Лакунарный эксперимент на материале делового текста явился сложным заданием для испытуемых, далеких от этой предметной области. Протоколы испытуемых содержали ошибки даже в интактных словах. На месте удаленных слов оказалось 24% отказов (пропуски этих слов). В табл. 2 представлены результаты восстановления удаленных слов в зависимости от частеречных классов.

Таблица 2 Правильность восстановления удаленных слов делового текста в зависимости от частеречного класса (%)

частеречный класс	+	ч.р.	ч.п.	О.
существительное	37	65	63	28
прилагательное	13	51	55	31
служебные части речи	53	65	65	31
наречие	43	58	58	40
глагол	26	58	59	32

⁶ На данной стадии исследования статистическая обработка данных (здесь и далее по всей работе) представляла собой проверку по критерию χ^2 гипотезы об однотипности сравниваемых распределений, а затем по t -критерию гипотезы о равенстве средних при уровне значимости 0,05 (если гипотеза о равенстве средних по t -критерию отвергалась при уровне значимости 0,1, утверждалась значимость различий на уровне тенденции).

Наилучшей восстанавливаемостью обладает класс служебных частей речи (количество отказов примерно такое же, как и у остальных частеречных классов) (см. табл.2). Максимальное число отказов дают союзы; вероятно, это является следствием того, что позиции союзов обладают меньшей силой предсказуемости, чем позиции предлогов в составе предложно-падежной конструкции.

Лакунарный эксперимент на материале художественного текста явился менее сложным заданием для испытуемых (художественный сюжетный текст гораздо более привычен наивному носителю языка), что следовало из замечаний самих испытуемых, как правило, работавших с двумя текстами. Однако на месте удаленных слов оказалось 39% отказов (пропуски этих слов). В табл. 3 представлены результаты восстановления удаленных слов в зависимости от частеречных классов, в качестве параметров наряду с теми, что использовались при анализе результатов делового лакунарного текста (точное восстановление, грамматические классы и отказ), использовались и параметры «точно для текста» и «точно для высказывания»⁷. Последние два параметра являются результатом оценки уместности восстановления единицы, прежде всего, с точки зрения смысловой структуры текста или высказывания⁸.

Таблица 3. Правильность восстановления удаленных слов в зависимости от частеречного класса (%)⁹

частеречный класс	+	т.т.	т.в.	ч.р.	ч.п.	О.
существительное	30	48	52	53	54	40
прилагательное	30	45	49	49	49	50
личное местоимение	43	55	61	48	59	33
служебные части речи	42	56	57	52	55	32
наречие	30	47	51	38	38	41
глагол	28	53	55	55	55	42

4.2. Служебные слова

Предлог в составе предложно-падежной конструкции рассматривается нами как компонент неоднословной целостности (ср. предположение о том, что предложно-

⁷ Например, при восстановлении в позиции «вопли» – *Однако соседи давно уже не обращали внимания на его вопли* – слов *крик, крики, желания* фиксировалось «точное для высказывания» восстановление, из них варианты *крик* и *крики* соответствуют «точному для текста» восстановлению.

⁸ В силу специфики делового текста анализ по этим параметрам оказался малоинформативным. Информативность такого рода точности восстановления присутствовала лишь для замен *сбывать* на *продавать, реализовывать* (что отдельно рассматривалось выше) и немногочисленных замен *импорт*, например, на *покупку, ввоз*. В ряде случаев автор затруднялся оценить точность восстановления по этим параметрам: например, упоминавшаяся выше замена *заинтересованными сторонами* на *заинтересованными лицами*, или замена *стороны* на *страны*.

⁹ Для служебных слов мы также рассматриваем такой параметр как «член предложения», хотя, конечно, в этом случае он носит несколько условный характер: мы считаем, что этот параметр может принимать дополнительное значение «не являться членом предложения»; таким образом учитывается синтаксический статус восстановленного слова как для знаменательных, так и служебных слов.

падежная конструкция может рассматриваться как наиболее вероятный кандидат на роль единицы перцептивного словаря Ягунова 2008). В табл. 4 приведены результаты восстановления пропущенных первообразных предлогов. Ошибки в восстановлении предлога чаще всего связаны с неправильной интерпретацией падежной информации (например, *на выдачу* вместо *о выдаче*, *с заявками* вместо *по заявкам*, *с просьбой* вместо *по просьбе*), но встречаются и замены наподобие *в основе* вместо *на основе* (см. табл. 4). Восстановление *при равных* как *в равных* встретилось чаще, чем правильное (см. табл. 4), что соответствует значениям общеязыковой частоты встречаемости этих единиц: в Национальном корпусе русского языка (<http://www.ruscorpora.ru>) *при равных* встретилось 33 раза (в 30 документах), а *в равных* 136 раз (в 117 документах).

Таблица 4. Результаты восстановления предлогов делового текста (%)

фрагмент текста с удаленным предлогом	+	ч.р.
Заявка с просьбой () выдаче разрешения...	80	85
...в соответствии с договором () деловом сотрудничестве,...	90	90
...устанавливаются один раз на весь срок действия договора или ежегодно () заявкам сотрудничающих сторон,	60	85
...импортные квоты () просьбе предприятия могут быть изменены	55	65
() равных коммерческих условиях преимущество имеют закупки товаров...	30	75
расчеты () эти товары производятся в иностранной или вьетнамской валютах...	60	70
...участвующие в реализации договора () деловом сотрудничестве...	90	90
и предприятия с иностранным капиталом могут продавать самостоятельно или через вьетнамские хозяйственные организации () основе хозяйственных договоров...	70	75

Условные обозначения: рассматриваемые предлоги заменены скобками – (), другие подлежащие удалению слова выделены шрифтом с подчеркиванием; «+» – точное восстановление исходного предлога, «ч.р.» – восстановление грамматического класса «часть речи»

Наилучшей восстанавливаемостью обладают классы служебных частей речи и личных местоимений, они же представлены минимальным количеством отказов.

Класс удаленных служебных частей речи для художественного текста более неоднороден, чем для делового текста (см. табл. 5). Максимальная предсказуемость характеризует отрицательную частицу и предлог в предложении *Я пришел не затем, чтобы кланчить у тебя пять хао!* Это предложение восстанавливается максимально четко, лакунарizations в нем подверглись лишь эти служебные части речи. Семантическая и синтаксическая структура высказывания восстанавливается как за счет его самого, так и за счет предшествующего – *Возьми их себе!* – исключаящего иную интерпретацию (например, без отрицания).

Таблица 5. Результаты восстановления служебных частей речи художественного текста (%)

п.п.	фрагмент текста с удаленными служебными словами	+	т.т.	т.в.	ч.р	ч.п	О.
11	Пора <u>бы</u> знать свое место	5	25	35	10	10	45
16	Забирай и катись, <u>чтобы</u> духу твоего не было, живо!	60	85	85	85	85	15
20	Глаза Ти Фео засверкали <u>от</u> ярости...	25	70	70	65	65	30
24	Я пришел <u>не</u> затем	80	85	85	85	85	5
25	чтобы клянчить <u>у</u> тебя пять хао!	75	75	75	75	75	25
38	<u>не</u> выдержав, расхохотался советник	60	70	70	70	70	30
39	что <u>же</u> , в добрый час	15	20	20	20	45	45
40	хотя <u>и</u> поздновато	20	20	20	20	20	50
51	<u>И</u> , выхватив нож, Ти Фео бросился на советника	40	50	50	40	40	45

Условные обозначения: рассматриваемые служебные слова выделены ирифтом с подчеркиванием

Другой предлог (*от ярости*) восстанавливается сравнительно плохо, т.к. конструкция *от ярости*, как правило, заменялось испытуемыми на близкую по смыслу конструкцию *в ярости*. Возможно, одной из причин такой замены является большая частотность сочетания лексем «в» + «ярость» (372) по сравнению с «от» + «ярость» 156; несмотря на то что частотность предложно-падежных конструкций сопоставима¹⁰.

Вторая из рассматриваемых отрицательных частиц (*не выдержав, расхохотался советник*), восстанавливается сравнительно хорошо (60%). Несколько более низкие значения восстанавливаемости этой позиции можно объяснить лишь тем, что вся конструкция (*не выдержав*) имеет низкую синтаксическую и семантическую значимость; при восстановлении текста испытуемыми она могла пропускаться целиком.

Позиция союза в *чтобы духу твоего не было* обладает большой силой предсказуемости: 0% точного восстановления союза *чтобы* и 85% – союза *чтобы* или его варианта *чтоб* (см. табл. 5). Таким образом, среди наиболее предсказуемых служебных частей речи есть и предлог, и частицы, и союз.

Служебные части, представленные частицами *бы, же, и*, легко опускаются без существенного ущерба для понимания текста. Именно для этих позиций служебных слов характерно максимальное число отказов и минимальное число восстановленных единиц (по всем параметрам: точное восстановление, точно для текста, точно для высказывания, восстановление грамматических классов). Вынесенный в начало предложения союз «и» (*И, выхватив нож,*) восстанавливается не более, чем в половине случаев, то есть демонстрирует сравнительно большое число отказов (хотя и меньше, чем для частиц *бы, же, и*).

¹⁰ Значения частоты встречаемости приведены по сервису С.А. Шарова <http://corpus1.leeds.ac.uk/ruscorpora.html>, позволяющему подсчитать частоты встречаемости сочетаний (биграмм) на базе НКРЯ по состоянию на 2008 год.

4.3. Личные местоимения

Для художественного текста рассматривался «расширенный» класс личных местоимений, в который были также включены местоимения *кто* и *тот*. Местоимения можно рассматривать как переходный класс от служебных к полнозначительным словам. Как и в случае класса служебных слов, причина неоднородности класса личных местоимений кроется, прежде всего, в степени обязательности этого элемента (с точки зрения слушающего).

Таблица 6. Результаты восстановления личных местоимений художественного текста (%)

п.п.	фрагмент текста с удаленным местоимением	антецедент	+	т.т.	т.в.	ч.р.	ч.п.	О.
7	Советник решил бросить ему пять хао	Ти Фео	30	40	45	30	35	55
23	– Возьми их себе!	пять хао	40	40	60	55	70	30
30	– Сказано тебе: не за деньгами пришел	Советник	55	55	55	55	55	30
34	Впервые слышу от тебя такое!	Ти Фео	75	75	80	80	80	20
35	Что же тебе надо?	Ти Фео	70	70	75	75	80	20
45	Как мне стать теперь порядочным?	Ти Фео	10	40	45	10	10	55
46	Кто сотрет с моего лица эти шрамы?		30	40	40	40	55	45
50	Сейчас ты узнаешь, что именно!!!	Советник	70	70	70	75	75	10
53	И, выхватив нож, Ти Фео бросился на советника. Тот в страхе привстал	Советник	20	45	50	20	45	50
62	Однако соседи давно уже не обращали внимания на его вопли. Когда же наконец они явились, то увидели Ти Фео распростертым на земле в луже крови рядом с трупом советника	Соседи	25	75	85	40	85	15

Около половины испытуемых пропустило местоимения в тех случаях, когда эллипсирование этих элементов вполне допустимо с точки зрения русского языка, а соответствующий актант легко восстанавливается из контекста (*бросить ему, как мне стать порядочным*) (см. табл. 6). Однако если высказывание маркировано особой логической и/или эмоциональной выделенностью, то позиция соответствующих местоимений восстанавливается в большинстве случаев (*Впервые слышу от тебя такое!*, *Что же тебе надо?*, *Сейчас ты узнаешь, что именно!!!*). Промежуточное положение занимает позиция местоимения в «*Сказано тебе: не за деньгами пришел*», что, возможно, связано со степенью маркированности в контексте общего смысла текста (противостояния основных действующих лиц).

Позиция местоимения в «*Возьми их себе*» оказывается сравнительно сложной для точного восстановления, т.к. возможны неоднозначные прочтения референциальных связей. Во-первых, цепочка референциальных связей включает разные наименования: «*пять хао*» в третьем предложении, «*монета*» в шестом предложении, «*их*» – в десятом предложении (и – в качестве подтверждения – «*пять хао*»). Кроме того, в качестве

потенциального антецедента может рассматриваться обобщающее наименование «деньги». Во-вторых – «база знаний» испытуемых не содержит соответствующие реалии вьетнамской жизни, возможные характеристики референта восстанавливаются слушающим с трудом (например, *хào* – денежная единица, она может быть представлена в виде одной монеты).

Позиция местоимения в «*Когда же наконец они явились, ...*» обладает хорошей предсказуемостью с точки зрения восстановления смысла и структуры высказывания и текста; однако восстанавливается чаще всего единица, отличная от исходной (например, *люди, все*¹¹).

Восстановимость позиций в «*Кто сотрет с моего лица эти шрамы?*» и «*Тот в страхе привстал,...*» – низкая (как за счет большого количества отказов, так и за счет замен) (см. табл. 6).

4.4. Прилагательные

Минимальной правильной восстановимостью характеризуются прилагательные, количество пропусков (отказов) для этих слов – максимальное. Отметим, что сравнительно многочисленный частеречный класс «прилагательные» (12 слов) содержит лишь одно местоимение-прилагательное (*в соответствии со своими ежегодными экспортными планами*). Первоначальное слово для прилагательных восстанавливается значимо хуже, чем для существительных. Восстановление части речи и члена предложения для прилагательных и для существительных различаются незначимо.

Вероятно, сложность восстановления пропущенного прилагательного может возрасти при наличии между удаленным прилагательным и определяемым существительным других прилагательных и/или сложности восприятия синтаксической конструкции. Наихудшие результаты восстановления демонстрирует словоформа *вьетнамскими* из фразы *Плановые показатели / вьетнамскими государственными органами / не устанавливаются*¹²: исходное слово не было восстановлено, классы «часть речи» и «член предложения» восстановлены лишь в 10% случаев. Так, например, если испытуемые нажимали клавишу «пауза» до глагола, то варианты восстановления пропуска могут содержать глаголы: *Плановые показатели определяются государственными органами, Плановые показатели задаются государственными органами*. Слабой предсказуемостью обладает позиция удаляемого слова

¹¹ Именно за счет появления существительных в этой позиции грамматический класс часть речи восстанавливается сравнительно плохо.

¹² Знак «/» обозначает синтагматическую границу в звучащем тексте.

заключаемых¹³ из ...могут продавать самостоятельно или через вьетнамские хозяйственные организации на основе хозяйственных договоров, заключаемых в соответствии с действующим во Вьетнаме законодательством. 0% точного восстановления, 40% восстановления грамматического класса «часть речи» и 40% отказов. Наилучшие результаты показал компонент «транспортные» сложной номинации «транспортные средства» во фрагменте *Импортные квоты на оборудование, машины, запасные части, транспортные средства,...*: в 50% оно восстановлено правильно, в 5% правильно восстановлены грамматические классы (замены *оборотные, необходимые, другие* и т.д.).

Результаты предсказуемости классов знаменательных частей речи художественного текста – существительных, прилагательных, наречий и глаголов – приблизительно одинаковы.

Класс прилагательных художественного текста насчитывает всего 4 единицы (см. табл. 7). Они демонстрируют приблизительно одинаковую предсказуемость по большинству рассмотренных параметров (точное восстановление, точно для текста, точно для высказывания, восстановление грамматических классов), кроме точного восстановления. Наименьшее число точного восстановления у позиции определения в конструкции «*хочу стать порядочным человеком*», но предлагаемые замены *наподобие стать хорошим человеком, стать настоящим человеком* полностью отражают смысл высказывания и текста¹⁴ и стилистически не менее уместны. Наибольшее число точных восстановлений приходится на потенциальную неоднословную целостность «*от всей души*», имеющую два варианта *от всей души* или *от души*.

Таблица 7. Результаты восстановления прилагательных художественного текста (%)

п.п.	фрагмент текста с удаленным прилагательным	+	т.т.	т.в.	ч.р.	ч.п.	О.
13	С этими словами	40	40	40	40	40	55
36	порядочным человеком	5	40	55	55	55	45
41	от всей души	55	55	55	55	55	45
47	эти шрамы	20	45	45	45	45	55

4.5. Существительные

Предсказуемость позиции существительных зависит от интерпретации синтаксической организации структурных составляющих текста, от лексико-синтаксической сочетаемости контактно расположенных слов, а также от понимания

¹³ Удаленное слово *заключаемых* в рассматриваемом делении на части речи было отнесено к прилагательным, т.к. в ином случае пришлось бы вводить класс с одним элементом.

¹⁴ Инструкция позволяла лишь однократное прослушивание, поэтому последующее повторение конструкции «*стать порядочным человеком*» не приводило к исправлениям в протоколах испытуемых.

текста (и степени подстройки под текст). Из 27 удаленных существительных рассмотрим 13 единиц, обладающих максимальной (не менее 75%) и минимальной (не более 5%) предсказуемостью (см. табл. 8)¹⁵.

Таблица 8. Результаты восстановления существительных делового текста (%)

п.п.	удаленное слово	+	ч.р.	ч.п.	О.
4	программы	0	50	50	40
8	к_заявке	5	55	55	45
12	сотрудничестве	75	80	75	20
16	(на_)импорт	0	50	50	25
17	средств	85	90	90	10
21	раз	85	90	90	10
24	топливо	5	30	30	60
26	(на_)цели	0	45	45	30
30	раз	85	90	90	10
38	преимущество	0	50	10	30
39	во_Вьетнаме	5	25	40	45
47	стороны	75	85	85	10
58	рынке	90	90	90	10

Условные обозначения: п.п. – порядковый номер среди удаляемых слов, «+» – точное восстановление исходного существительного, «ч.р.» и «ч.п.» – восстановление грамматического класса «часть речи» и «член предложения», «О.» – отказы, предлоги в скобках удалены не были согласно правилам лакунаризации

Среди наиболее предсказуемых позиций могут быть названы компоненты сложных номинаций (на внутреннем рынке, транспортных средств). Высокой силой семантической и синтаксической предсказуемости обладает позиция *раз* в

¹⁵ Контексты для рассматриваемых существительных:

Стороны, участвующие в реализации договора о деловом сотрудничестве, и предприятия с иностранным капиталом имеют право самостоятельно устанавливать программы и планы своей производственно-хозяйственной деятельности.

В технико-экономическом обосновании, прилагаемом к заявке с просьбой о выдаче разрешения на вложение инвестиций, или на деловое сотрудничество, стороны договора о деловом сотрудничестве и учредители предприятий с иностранным капиталом должны включать соответствующие показатели.

Заявка с просьбой о выдаче разрешения на импорт оборудования, машин, транспортных средств, сырья, материалов, которые иностранная сторона ввозит во Вьетнам в качестве своего вклада в уставной фонд, подаётся только один раз.

Импортные квоты на оборудование, машины, запасные части, транспортные средства, природное сырьё, топливо, материалы и так далее, ввозимые во Вьетнам на цели производства в соответствии с договором о деловом сотрудничестве, утверждённым Государственным комитетом по сотрудничеству и инвестициям, устанавливаются один раз на весь срок действия договора или ежегодно по заявкам сотрудничающих сторон.

При равных коммерческих условиях преимущество имеют закупки товаров во Вьетнаме по сравнению с их импортом.

Стороны, участвующие в реализации договора о деловом сотрудничестве, и предприятия с иностранным капиталом имеют право самостоятельно экспортировать продукцию в соответствии со своими ежегодными экспортными планами, или поручить экспорт этой продукции правительствам иностранного участника или вьетнамским внешнеторговым организациям в соответствии с договорённостью между заинтересованными сторонами.

Продукцию, которую разрешено сбывать на вьетнамском внутреннем рынке, стороны, участвующие в реализации договора о деловом сотрудничестве, и предприятия с иностранным капиталом могут продавать самостоятельно или через вьетнамские хозяйственные организации на основе хозяйственных договоров, заключаемых в соответствии с действующим во Вьетнаме законодательством.

устанавливаются (подается) один раз. Вероятно, таким образом отражается целостность конструкции *один раз*: воспроизводимость этой конструкции обеспечивается высокой общеязыковой частотностью¹⁶. Предсказуемость позиции *о деловом сотрудничестве* определяется высокой частотой встречаемости данной конструкции в тексте: она является третьей по счету в тексте. Более того, даже в данной фразе в разных падежных вариациях она появляется два раза (...или на деловое сотрудничество, стороны договора о деловом сотрудничестве...). В процессе подстройки слушающего под особенности текста такого рода частота встречаемости, по-видимому, может переводить конструкцию в разряд воспроизводимых единиц.

В общей сложности лексема *сторона* в тексте встречается 7 раз; лакунаризации подверглись второе, пятое и шестое вхождения (см. табл.5). Эта лексема в тексте встречалась, как правило, в однотипном семантическом (тематическом) контексте: первый раз – *Стороны, участвующие в реализации договора о деловом сотрудничестве*, третий – *иностранная сторона*, четвертый – *сотрудничающих сторон*. Второе вхождение данной лексемы соответствует типовому – для данного текста – контексту, но имеет низкую восстанавливаемость. Вероятно, интерпретация этого результата состоит в том, что подстройка под особенности текста еще не произошла. Эти данные соответствуют выдвинутому ранее предположению о том, что на начальном фрагменте текста, на который и приходится данное вхождение, происходит подстройка слушающего под структурные особенности текста (Ягунова 2007; Ягунова 2008). Пятое вхождение также соответствует типовому контексту и восстанавливается хорошо; оно входит в конечный (неначальный) фрагмент текста, на котором уже произошла подстройка слушающего¹⁷. Шестое вхождение приходится на словосочетание «*заинтересованные стороны*», однако это словосочетание обладает меньшей общеязыковой частотой встречаемости, чем, например, словосочетание «*заинтересованные лица*». Данные по частоте контактной встречаемости заданной последовательности лексем «*заинтересованный*» + «*лицо / сторона*» (BigramsFreq) и объему взаимной информации (MI) (см. табл. 9), предоставленные А. Авериным¹⁸, подтверждают роль частотности в выборе слушающим

¹⁶ В Национальном корпусе русского языка (<http://www.ruscorpora.ru>) *один раз* встретилось 5209 раз (в 2275 документах). Данные по частоте контактной встречаемости заданной последовательности лексем (16451) и по статистическим мерам, позволяющим оценить неслучайность появления «раз» сразу после «один» (логарифм правдоподобия и t-score) по сервису С.А. Шарова <http://corpus1.leeds.ac.uk/ruscorpora.html> – также подтверждают высокую степень неслучайности данной конструкции (значения этих мер выше лишь для сочетания «один из»).

¹⁷ Ср. данные эксперимента по восприятию текста в шуме о том, что с увеличением степени знакомства с текстом – продвижения от начального к срединному и конечному фрагментам текста – улучшается словесная разборчивость (Ягунова 2008).

¹⁸ К сожалению, в настоящее время поиск биграмм А. Аверина (<http://www.aot.ru/cgi-bin/BigramsCGI>) работает только для лексем, но не для словоформ. Объем корпуса – 448 703 525 tokens. Подробнее см.

замещающей единицы. Использование меры MI (mutual information, коэффициента взаимной информации) позволяет нам сравнивать зависимые контекстно-связанные частоты с независимыми, как если бы слова появлялись в тексте совершенно случайно. В результате мы имеем основания говорить, что оба сочетания характеризуются высоким значением этой меры (степени неслучайности), при этом вероятность выбора лексемы «лицо» выше, чем вероятность выбора лексемы «сторона».

Человечно добавить два предложения, обобщающих результаты Аверина, чтобы не заставлять всех лезть туда и разбираться: что значит «объем взаимной информации»?

Таблица 9. Частота встречаемости и объем взаимной информации последовательности лексем «заинтересованный» + «лицо / сторона»

Word1	Word2	WordFreq1	WordFreq2	BigramsFreq	MI
заинтересованные	лицо	3227	487953	495	7.140111
заинтересованные	сторона	3227	386839	255	6.518189

Условные обозначения: WordFreq1 – частота встречаемости в корпусе первого слова, WordFreq2 – частота встречаемости в корпусе второго слова, BigramsFreq – частота контактной встречаемости заданной последовательности лексем, MI – объем взаимной информации

Таблица 10. Результаты восстановления вхождений лексемы «стороны»(%)

фрагмент текста с удаленным словом	+	ч.р.	ч.п.	О.
с т о р о н ы договора о деловом сотрудничестве	10	45	30	45
С т о р о н ы, участвующие в реализации договора о деловом сотрудничестве	75	85	85	10
в соответствии с договоренностью между заинтересованными с т о р о н а м и	50	90	95	5

Обобщенные результаты восстановимости большинства существительных в замещенных шумом позициях художественного текста приведены в табл. 11. Интерпретация синтаксической организации структурных составляющих художественного текста, как правило, не составляла труда для испытуемых. Различие в степени предсказуемости существительных художественного текста в большей степени зависит от понимания текста, умения извлечь из него характеристики описываемых реалий, исходно не входящих в «базу знаний» испытуемых, нежели от синтаксической организации высказываний.

(Аверин 2006). Выбор сервиса – А.Аверина или С.Шарова – определяется особенностью лемматизации, Выбор статистической меры зависит от типа рассматриваемого сочетания (MI используется преимущественно для анализа неоднословных номинаций, терминов и т.д.).

Таблица 11. Результаты восстановления существительных художественного текста (%)

п.п.	фрагмент текста с удаленным существительным	+	т.т.	т.в.	ч.р.	ч.п.	О.
1	Когда одолевают подобные мысли	40	60	60	60	60	40
2	трудно сохранить хорошее настроение	45	55	55	55	55	45
4	а тут еще принесла нелегкая Ти Фео	5	20	25	5	5	75
6	явился клянчить на выпивку	10	35	40	45	45	55
15	швырнул монету прямо на землю	10	40	50	55	50	45
19	глаза Ти Фео засверкали от ярости	60	65	65	65	65	30
21	и, подняв руку, он наставил палец	15	25	50	55	55	40
26	советник решил	45	45	45	45	50	45
33	О небо!	0	25	25	35	30	55
42	желаю тебе успеха на радость всей деревне	5	50	55	55	60	35
43	желаю тебе успеха на радость всей деревне	5	40	55	45	60	40
44	Ти Фео	50	50	50	50	50	45
52	Ти Фео	50	60	65	55	60	35
55	успел крикнуть только один раз	50	50	50	65	50	35
56	Ти Фео	45	55	55	45	50	45
57	удар за ударом	65	70	70	70	75	25
58	орал во все горло	80	85	85	85	85	15
59	однако соседи давно уже не обращали внимания на его вопли	0	60	60	70	80	20
61	однако соседи давно уже не обращали внимания на его вопли	10	55	65	70	70	30
63	Ти Фео	40	45	45	50	50	50
64	увидали Ти Фео распростертым на земле	0	35	45	55	50	45
65	рядом с трупом	10	35	35	50	55	20

Как и следовало ожидать, наибольшей предсказуемостью обладает компонент фразеологизма (неоднословной целостности) «и орал во все горло» (см. табл. 11). Хорошо предсказуемая позиция существительного в составе «наносил удар за ударом» также рассматривается нами как позиция компонента устойчивого словосочетания (фразеологизма), который в ряде случаев испытуемыми заменялся на одно существительное «наносил удары». Сравнительно хорошей восстановимостью исходного слова характеризуется существительное «глаза»: на эту позицию сильное воздействие оказывает последующая позиция глагола «засверкали». Эти позиции не являются контактными, но глагол входит в ту же клаузу и ту же синтагму¹⁹. Наихудшей восстановимостью обладает вьетнамское имя впервые упоминающегося действующего лица (Ти Фео); естественно, что большинство испытуемых пропускают данную позицию (75%). Плохо восстанавливается позиция восклицания (О_небо!), по-видимому, за счет того, что его опущение не влечет изменения смысловой и семантико-синтаксической

¹⁹ Можно лишь предполагать, что контактность увеличила бы силу предсказуемости, а наличие синтагматической границы между ними, напротив, уменьшило бы ее.

структуры²⁰. Сравнительно низкая точная восстановимость многих существительных связана со сложностью восприятия собственно смысловых особенностей текста: имен действующих лиц, места действия (например, деревня – *на радость всей деревне*, нахождение в доме или на улице – *швырнул на землю, распростертым на земле*) (см. табл. 11). Ряд пропущенных существительных достаточно часто заменялся на синонимы, например, *желаю удачи, ... счастья, ... радости* вместо *желаю успеха*; *крики, крик* вместо *вопли* (параметры «точно для текста» и «точно для высказывания»; см. табл. 11). Существительное *соседи* часто восстанавливается как *люди, никто*²¹; эти замены можно считать контекстными эквивалентами. Уточнение *на выпивку* (*явился кланчить на выпивку*), вероятно, многими испытуемыми считалось или избыточным (55% отказов), или вариативным (например, *кланчить на игру*). Предсказуемость позиции «руку» (*и, подняв руку, он наставил палец*) мог бы обеспечивать последующий контекст (*наставил палец*), однако этот последующий контекст входит в другую клаузу и другую синтагму²². В результате сила влияния последующего контекста оказывается незначительной; испытуемыми восстанавливаются другие существительные (см. табл. 11), например, *подняв голову, подняв монету*.

4.6. Наречия и глаголы

Среди удаляемых слов классы «наречие» и «глагол» малочисленны и неоднородны²³. В класс «наречие» попало две единицы: *далее* (как компонент неоднословной целостности *и так далее*) и *самостоятельно*. *Далее* обладает высокой предсказуемостью. *Самостоятельно* может опускаться, что и нашло отражение в том, что более половины испытуемых не восстановило никакой единицы на месте данной позиции (см. табл. 12).

Таблица 12. Результаты восстановления наречий делового текста (%)

фрагмент текста с удаленным словом	+	ч.р.	О.
и так далее	75	75	25
Стороны ... имеют право самостоятельно экспортировать Продукцию	10	40	55

Класс удаленных глаголов насчитывает 6 единиц, из них один глагол модальный (*могут продавать*). Позиция этого модального глагола обладает максимальной для глаголов предсказуемостью (см. табл. 13), немногочисленные замены представляют

²⁰ Кроме того, это восклицание, на наш взгляд, может казаться стилистически неожиданным для данного текста.

²¹ Необходимость согласования по числу с глаголом не оказывала влияния на принятие решения испытуемых, т.к. глагол также был удален.

²² Членение высказывания на синтагмы, в скобках указана длительность межсинтагменных пауз: *Глаза Ти Фео засверкали от ярости* (424 мс) / *и, подняв руку* (383 мс) / *он наставил палец* (0 мс) / *прямо в лицо Советнику*.

²³ Обобщенные количественные показатели для этих классов (см. табл. 2) до некоторой степени условны.

синонимическое выражение *имеют право*, опущение модальных глаголов может рассматриваться как типичное.

Таблица 13. Результаты восстановления глаголов делового текста(%)

фрагмент текста с удаленным словом	+	ч.р.	О.
...и предприятия с иностранным капиталом могут продавать самостоятельно...	45	55	40
Плановые показатели вьетнамскими государственными органами не устанавливаются.	5	45	55
Продукцию, которую разрешено сбывать на вьетнамском внутреннем рынке, ...	5	90	10

Позиции глаголов (*не*) *устанавливается* и *сбывать* обладают наименьшей силой предсказуемости точной единицы (5%, см. табл. 13), однако причины и механизмы этого явления существенно различны. Несколько выше уже рассматривалась возможность неправильной интерпретации синтаксической структуры фразы *Плановые показатели вьетнамскими государственными органами не устанавливаются*: позиция удаленного прилагательного *вьетнамскими* восстанавливается как позиция глагола в том случае, если эта интерпретация имела место до появления позиции глагола *не устанавливаются*. В результате в протоколе испытуемого остаются варианты наподобие *Плановые показатели определяются государственными органами*, в дальнейшем не исправляемые. За глаголом следует большая пауза (1с 405мс), которая не позволяет слушающему присоединить глагол к следующей фразе, в результате чего более половины испытуемых не заполняет данную позицию. В случае с глаголом *сбывать* семантико-синтаксическая и смысловая структура – и, соответственно, частеречный класс – как правило, восстанавливаются (90%, см. табл. 13). Правильный выбор исходного элемента этого класса, напротив, маловероятен. Согласно последней версии частотного словаря С.А. Шарова (на 28 февраля 2010 года), у *сбывать* – общеязыковая частота встречаемости 1,95 ipm²⁴, в тексте же этот глагол встретился один раз в лакунаризируемой позиции. Чаще всего испытуемые заменяют этот глагол на более частотный глагол *продавать* (его частота – 50,04).

²⁴ ipm (instances per million words) – частота встречаемости оценивается в числе употреблений на миллион слов корпуса

Таблица 14. Результаты восстановления глаголов художественного текста (%)

п.п.	фрагмент текста с удаленным глаголом	+	т.т.	т.в.	ч.р.	ч.п.	О.
10	но при этом не выдержал и крикнул	5	35	45	45	45	55
14	он с размаху швырнул монету прямо на землю	15	50	50	50	50	45
17	чтобы духу твоего не было	85	85	85	85	85	15
18	сколько раз говорил	20	45	45	55	55	45
27	ладно, ладно, бери	5	35	45	30	30	55
31	не за деньгами пришел	45	45	45	45	45	45
32	ответил Ти Фео	5	55	55	55	55	45
54	но успел крикнуть только один раз	55	60	60	60	60	40
60	соседи давно уже не обращали внимания	20	65	65	70	70	30

В художественном лакунарном тексте было пропущено 9 глаголов. Несколько неожиданным результатом анализа восстановимости этих позиций является сравнительно большое число пропусков, по-видимому, свидетельствующих о возможности эллиптирования этих единиц. Вероятно, это связано, прежде всего, с глаголами говорения, вводящими прямую речь (*крикнул, говорил, ответил*). Глагол «*бери*» также может легко эллиптироваться как легко восстанавливающийся из контекста; на материале данного эксперимента он восстанавливался с точностью до синонимов (см. табл. 14). Наилучшая точная восстановимость характеризует глагол как компонент фразеологизма «*чтобы духу твоего не было*». Высокая предсказуемость характеризует глагол как компонент устойчивого словосочетания «*обращали внимание*»; как упоминалось выше, как правило, глагол восстанавливался в единственном числе (*не обращал внимания*): в 20% случаев *обращали* и в 45% случаев *обращал* (*никто не обращал*) при 30% отказов. Позиция глагола *успел*, по-видимому, предсказывается последующим контекстом «*только один раз*»; вероятно, сила влияния обеспечивается и тем, что этот последующий контекст не оторван синтагматической границей от глагола.

Группа удаленных наречий насчитывает 10 единиц; результаты их восстановления приведены в табл. 15.

Таблица 15. Результаты восстановления наречий художественного текста (% , кроме ФВС)

п.п.	фрагмент текста с удаленным наречием	ФВС	+	т.т.	т.в.	ч.р.	ч.п.	О.
3	а тут еще принесла нелегкая Ти Фео	2	50	60	60	60	60	40
5	не иначе как явился клянчить на выпивку	0	40	75	75	65	65	20
8	чтобы поскорее убрался	0	5	50	50	10	10	50
12	здесь тебе не банк!	0	15	50	50	25	25	50
22	швырнул монету прямо на землю	3	35	40	40	40	40	50
28	больше у меня нет	3	30	30	30	30	30	55
29	добавил он уже более миролюбивым тоном	2	25	40	40	40	40	55
37	что же тебе надо?	2	45	45	80	55	55	10
48	понятно?...	3	0	0	0	0	0	65
49	Теперь мне осталось одно!	3	50	80	80	50	55	10

Условные обозначения: ФВС – степень фонетической выделенности слова (в абсолютных числах)²⁵

Как и следовало ожидать, однословное высказывание *Понятно?...* не восстанавливается испытуемыми. В целом же – результаты восстановления или невосстановления наречий²⁶ зависят от стратегии испытуемого: стремления записывать текст, ориентируясь прежде всего на его содержание или в максимальной степени пытаться восстанавливать все лакунаризованные позиции (слова, замененные шумом). По-видимому, для восстановления этих позиций в наибольшей степени (по сравнению с другими частями речи) испытуемыми использовалась следующая просодическая информация, частично сохраненная в характеристиках амплитудно-модулированного замещающего шума: длина слова (чем короче слово, тем скорее оно будет пропущено) и логическая и/или эмфатическая выделенность (чем больше выделенность, тем скорее позиция будет заполнена).

5. Заключение

Как уже говорилось, информационная избыточность является неотъемлемым свойством любого текста на естественном языке. Благодаря этой избыточности обеспечивается возможность успешного восприятия и понимания речи; особенно ярко эта закономерность проявляется при восприятии звучащей речи. Именно контекст позволяет предсказывать и интерпретировать большую часть слов звучащего текста (ср. эксперименты по восприятию текста в условиях разных типов искажения).

Результаты вычислительных экспериментов (данные о частоте встречаемости рассматриваемых единиц и/или конструкций) и экспериментов с информантами

²⁵ Фонетическая выделенность слова определялась тремя экспертами, которые должны были при многократном прослушивании подчеркнуть в орфографической записи текста то, что они считают фонетически выделенными словами (без указания критериев выделенности). Под степенью фонетической выделенности слова понимается число экспертов, выделивших это слово.

²⁶ Предикативы условно отнесены нами к классу наречий.

позволяют сравнить информационную избыточность единиц текста (прежде всего, слов) как в качественном, так и в количественном отношении (Касевич, Ягунова 2006б; Ягунова 2006). Исследование восстановления пропущенных слов в лакунарном эксперименте – как показателя информационной избыточности текста – позволяет выделить основные типы контекстной предсказуемости, определяющиеся взаимодействием разных характеристик, например, таких, как:

- грамматический класс слова (например, часть речи, синтаксическая позиция),
- позиция с точки зрения синтагматического, коммуникативного и смыслового структурирования текста,
- величина контекста (эта характеристика может принимать значения от минимального (предшествующее слово) до максимального (весь текст от начала до рассматриваемой единицы)),
- частотность слова (словоформы и лексемы) и конструкции, в которую входит рассматриваемая единица в языке и данном тексте,
- тип текста (прежде всего, функциональный стиль текста).

Именно взаимодействие этих характеристик обеспечивает успешность восприятия и понимания текста, а количественный и качественный анализ экспериментальных данных позволяет оценить информационную избыточность для разных единиц текста. Более того, для текстов разного типа – художественного и делового – роль этих характеристик оказывается различной. В данной статье основное внимание было сосредоточено на роли признака «грамматический класс слова», однако невозможно рассматривать один признак, игнорируя роль других. Например, на материале восстановления существительных можно было увидеть, что характеристика «общезыковая частотность словоформ» играет существенную роль; что на конечном фрагменте делового текста значительно улучшается предсказуемость слов, частых в тексте, но редких в языке, а в художественном тексте существенного улучшения не происходит.

Можно ли сделать выводы об избыточности текстов разных стилей и их компонентов (не только о влиянии факторов)?

Литература

- Аверин А. Н. Разработка сервиса поиска биграмм слов // Труды международной конференции «Корпусная лингвистика–2006»: 10-14 октября 2006 г., Санкт-Петербург. – СПб., 2006. – С. 5-15.
- Бондарко Л.В., Вербицкая Л.А., Гордина М.В., Зиндер Л.Р., Касевич В.Б. Стили произношения и типы произнесения // Вопросы языкознания, 1974, №2
- Грановская Р.М. Восприятие и модели памяти. Л., 1974.

- Даль Э. Возникновение и сохранение языковой сложности. М., 2009.
- Касевич В. Б. Морфонология // Касевич В. Б. Труды по языкознанию. – Т.1. – СПб, 2006. – С. 241-371.
- Касевич В.Б., Ягунова Е.В. Контекстная предсказуемость слов в тексте (на материале русского и французского языков) // Вестник Пермского ун-та, вып. 3. – 2006а. – С. 61-70.
- Касевич В. Б., Ягунова Е.В. Корпуса письменных текстов и моделирование восприятия речи // Вестник СПбГУ. Серия 2. 2006, Вып.3, 2006б.
- Пиотровский Р.Г. Информационные измерения языка. Л., 1968
- Пиотровский Р.Г. Лингвистический автомат (в исследовании и непрерывном обучении) СПб., 1999
- Пиотровский Р.Г. Лингвистическая синергетика: исходные положения, первые результаты, перспективы СПб., 2006
- Ягунова, Е. В. Неоднословные целостности в словаре и корпусе // Труды международной конференции «Корпусная лингвистика – 2006» : 10-14 октября 2006 г., Санкт-Петербург. – СПб., 2006. – С. 395-412.
- Ягунова Е.В. Коммуникативная и смысловая структуры текста и его восприятие // Вопросы языкознания, 6, 2007
- Ягунова Е.В. Вариативность стратегий восприятия звучащего текста (экспериментальное исследование на материале русскоязычных текстов разных функциональных стилей) / Пермь, 2008а
- Ягунова Е.В. Исследование контекстной предсказуемости единиц текста с помощью корпусных ресурсов // Труды международной конференции "Корпусная лингвистика–2008". – СПб. : СПбГУ, 2008б. – С. 396-403.
- Aborn M., Rubenstein H. Sources of contextual constraint upon words in sentences / M. Aborn, , Th. D. Sterling // J. of Exp. Psychology, 1959, vol. 57. – P. 171-180.
- Alderson J. C. Native and non-native speaker performance on cloze tests // Language Learning, 1980, vol. 30. – P. 59-76.
- Auer E.,T. (Jr.), Luce P.A. Probabilistic phonotactics in spoken word recognition / E. T. Auer (Jr.), // Handbook of Speech Perception / D. B. Pisoni and R. E. Remez (Eds). – Blackwell, 2005. – P. 620-630.
- Bormuth J. R. Validities of grammatical and semantic classifications of cloze test scores // Reading and inquiry : International Reading Association Conference Proceedings, 10 / J. A. Figurel (Ed.). – Newark, 1965. – P. 283-286.
- Bormuth J. R. Comparable cloze- and multiple-choice comprehension test scores // Journal of Reading, 1967, №10. – P. 193-98.
- Brown J. D. A closer look at cloze: Validity and reliability // Issues in Language Testing Research / J. W. Oller, Jr. (Ed.). – Rowley, MA, 1983. – P. 237-250.
- Brown J. D. A cloze is a cloze is a cloze? // TESOL '83 / J. Hand-scombe, R. A. Orem, B. P. Taylor (Eds). – Washington, DC, 1984. – P. 109-119.
- Brown J. D. Tailored cloze: Improved with classical item analysis techniques / J. D. Brown // Language Testing, 1988, vol. 5. – P. 19-31.
- Brown J. D. Cloze item difficulty // JALT Journal, 1989, vol. 11. – P. 46-67.
- Brown J. D. What are the characteristics of natural cloze tests? // Language Testing, 1993, vol. 10. – P. 93-116.
- Calvo M. G. et al. On-line predictive inferences in reading: Processing time during versus after the priming context // Memory & Cognition, 1999, vol. 27(5). – P. 834-43.
- Chihara T., Oller John W., Jr., Weaver K., Chavez-Oller M. A. Are cloze items sensitive to constraints across sentences? // Cloze and coherence / J. W. Oller, Jr., J. Jonz (Eds). – Cranbury, NJ, 1994. – P. 63-73.
- Cziko G. Another response to Shanahan, Kamil, and Tobin: Further reasons to keep the cloze case open // Reading Research Quarterly, 1983, vol.18. – P. 361-365.

- Fillenbaum S., Jones L., Rapoport L.A. The predictability of words and their grammatical classes as a function of rate of deletion from a speech transcript // *J. of Verbal Learning and Verbal Behavior*, 1963, vol. 2. – P. 186-194.
- Nusbaum H. C. et al. Why cloze procedure? // *Cloze and coherence* / J.W. Oller, Jr., J. Jonz (Eds) – Cranbury, NJ, 1994. – P. 1-20.
- Oller J. W., Jr., Bowen J. D., Dien T. T., Mason V. Cloze tests in English, Thai, and Vietnamese. Native and non-native performance // *Language Learning*, 1972, vol. 22, № 1. – P. 1-15.
- Oller J. W., Jr., Grover Kh Yii, Greenberg L.A., Hurtado R. The learning effect from textual coherence measured with cloze // *Cloze and coherence* / J. W. Oller, Jr., J. Jonz (Eds). – Cranbury, NJ, 1994. – P. 247-268.
- Owens M., O'Boyle P., McMahon J., Ming J., Smith Fj. A comparison of human and statistical language model performance using missing-word tests // *Language and speech*, 1997, vol. 40, №4. – P. 377-389.
- Taylor W. L. Cloze procedure: a new tool for measuring readability // *Journalism Quarterly*, 1953, vol. 30. – P. 415-433.