

**НАБОР ОПОРНЫХ СЛОВ КАК ВИД СВЕРТКИ ТЕКСТА  
(В СОПОСТАВЛЕНИИ С НАБОРОМ КЛЮЧЕВЫХ СЛОВ)\*  
SET OF RECOGNIZABLE WORDS AS COMPRESSION TEXTS  
(WITH COMPARISON OF KEY-WORD SET)**

*Ягунова Е.В. (iagounova\_elen@mail.ru)  
Санкт-Петербургский государственный университет*

В докладе рассматриваются основные характеристики набора опорных слов как свертки текста. Под опорными понимаются наиболее распознаваемые слова при восприятии текста в шуме. Наши данные подтверждают гипотезу о том, что существенное значение для смыслового структурирования текста имеет функциональный стиль и степень динамичности текста.

**1. Вводные замечания**

Свертка текста – это упорядоченный набор слов, сопоставленный исходному тексту и в определенных условиях выступающий в качестве его «представителя». Порядок слов в свертке обычно воспроизводит порядок их появления в исходном тексте (хотя возможны и отклонения), сами слова отбираются по принципам, определяемым целью свертывания текста (см. об этом ниже). В настоящей работе анализируются два типа сверток. Первый – это наиболее традиционная разновидность – набор ключевых слов (НКС), в котором задан порядок – порядок введения ключевых слов в текст. Ключевые слова отражают тему текста, а упорядоченность слов в НКС активизирует ассоциативные связи, необходимые для существования любого текста (что проявляется в возможности развертывания НКС в цельный и связный текст) (Мурзин, Штерн 1991; Сахарный и др. 1984).

Наибольший интерес для нас в настоящей работе представляет другой вид сверток – наборы опорных слов (НОС). Под опорными в настоящем докладе понимаются слова, распознающиеся не менее чем 30% испытуемых при восприятии текста в шуме при соотношении сигнал/шум 0 дБ. Распознаваемость этих слов определяется влиянием как фонетических, так и внефонетических признаков. Наиболее разборчивыми могли стать те слова, которые идентифицировались за счет фонетической информации и/или за счет высокой контекстной предсказуемости. Можно предположить, что эти слова несут особую смысловую нагрузку, которой и обусловлена высокая распознаваемость именно этих слов (в частности, соответствующая фонетическая оформленность и/или контекстная предсказуемость опорных слов)<sup>1</sup>.

В качестве исследуемого материала было выбрано 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 сюжетной художественной прозы Нам Као «Ти Фео» с элементами диалога (в дальнейшем «художественный текст»).

Художественный текст является динамическим, т.е. характеризуется сменой ситуаций. Композиционная структура художественного текста соответствует традиционной схеме нарратива: прамбула (ориентация), завязка, развитие сюжета, развязка (по Чейфу). Деловой текст (текст-предписание) является статическим.

В эксперименте по восприятию текста в шуме приняло участие 40 испытуемых. В инструкции испытуемым предлагалось прослушать текст удобными порциями, останавливаясь с помощью клавиши «пауза» и записывая каждый услышанный фрагмент. Текст можно было слушать один раз, не возвращаясь назад.

В настоящем исследовании восприятие текста носителями языка осуществляется – в любых экспериментах – в условиях ограничения на «базу знаний» слушающего, т.к. испытуемыми выступают люди, далекие от предметной области экономики и делопроизводства (деловой текст) и реалий вьетнамской жизни (художественный текст).

При выборе стратегий восприятия особую роль должно играть начало текста как «площадка», в рамках которой происходит подстройка под особенности воспринимаемого текста. Соответственно рассматриваются противопоставления:

\* Работа выполнена при частичном финансировании РГНФ (проект №07-04-00161а)

<sup>1</sup> При восприятии осмысленного текста разделить вклад фонетической и внефонетической информации невозможно. В формат доклад не вошел дополнительный эксперимент по восприятию аналогичного асемантического текста (см., напр., Ягунова 2006), в котором опорными являлись слова, хорошо распознаваемые в силу своих собственно фонетических свойств.

## Набор опорных слов как вид свертки текста

- начальный фрагмент делового текста vs. весь деловой текст
- начальный фрагмент художественного текста<sup>2</sup> vs. весь художественный текст

КС определялись – по традиционно используемой методике – в ходе вспомогательного эксперимента, в котором испытуемым предлагалось следовать следующей инструкции: «Прослушайте текст. Подумайте над его содержанием. Выпишите<sup>3</sup> из текста 10-15 слов, наиболее важных с точки зрения его содержания». Испытуемые заполняли анкеты *после* окончания прослушивания текста. Следовательно, НКС, по-видимому, отражает смысловую структуру текста как целостного объекта.

Распределение в тексте наиболее разборчивых слов – НОС – соответствует, вероятно, «следам» восприятия и понимания текста *в текущем режиме времени*<sup>4</sup>, когда смысловая структура извлекается не только из текста целиком, но и из его структурных составляющих.

Каково соотношение двух наборов – НОС и НКС?

Оба вида сверток отражают, тем или иным образом, смысловую структуру текста. Казалось бы, КС как смысловые вехи текста должны формироваться из наиболее частотных в тексте слов и характеризоваться (1) наибольшей разборчивостью (т.е. быть одновременно и опорными словами); (2) наилучшей предсказуемостью на конечном фрагменте текста (в результате подстройки слушающего под особенности текста). Однако реальная ситуация оказывается сложнее. Как показали предыдущие исследования автора (Ягунова 2007), эти характеристики КС (по сравнению с прочими, неключевыми словами (неКС)) обнаруживают значимую зависимость от **коммуникативной ситуации** – отфункционального стиля текста и экспериментального режима прослушивания.

В деловом тексте частотность словоупотреблений (и конструкций) по тексту высокая, порядок следования частотных словоупотреблений соответствует потенциальным путям формирования КС и важен для обеспечения связности текста. В художественном тексте частотность знаменательных словоупотреблений существенно ниже, чем в деловом тексте, т.е. частотных по тексту слов не может быть достаточно для формирования НКС.

В режиме восприятия текста в шуме КС делового текста распознавались существенно лучше, чем неКС, для художественного текста, напротив, неКС распознавались значимо лучше, чем КС. По мере продвижения слушающего по тексту (от начального к конечному фрагменту текста) происходит улучшение распознаваемости КС для обоих текстов.

В парадигме гештальной психологии КС могут рассматриваться как фигура, а неКС – как фон. В случае рассматриваемого художественного текста, характеризующегося сменой ситуаций, соотношение «фигура – фон» может меняться от ситуации к ситуации (т.е. для одних ситуаций фигура плохо выделяется на фоне, для других – хорошо). Можно предположить для этого художественного текста наличие несоответствий между КС, выделяемыми после прослушивания текста и опорными словами, выделяемыми во время прослушивания текста в текущем времени. В случае статического – делового – текста соотношение «фигура – фон» сравнительно стабильно. Соответствие между КС и опорными словами, область их пересечения, таким образом, должны быть существенно выше для делового текста.

Далее будет рассмотрена *гипотеза* о том, что существенное значение для смыслового структурирования и формирования сверток текста имеют признаки «функциональный стиль текста» и «степень динамичности текста».

Функциональный стиль и степень динамичности текста могут рассматриваться как взаимодополняющие друг друга признаки. Проверка гипотезы осуществляется на основании сопоставления двух типов сверток (для КС и опорных слов), показывающего (1) как часто опорные слова являются одновременно и КС; (2) представленность основных частеречных классов в рассматриваемых наборах.

В соответствии со сформулированной гипотезой для статичного делового текста соответствие между НКС и НОС должно быть значительно выше, чем для динамического художественного текста, содержащего несколько последовательных ситуаций. Начальный фрагмент художественного текста, включающий лишь преамбулу и завязку, описывает меньшее число ситуаций и характеризуется меньшей динамичностью, чем весь художественный текст. Следовательно, согласно выдвинутой гипотезе, соответствие между двумя типами сверток для начального фрагмента художественного текста должно быть выше, чем для всего художественного текста.

<sup>2</sup> Начальный фрагмент соответствует примерно трети от каждого текста. Для художественного текста он соотносится с двумя композиционными компонентами: преамбула и завязка.

<sup>3</sup> Использование глагола «выписать» в инструкции продиктовано желанием (1) акцентировать испытуемых на задачу представления в своих списках слов, действительно присутствующих в тексте и (2) следовать традиции для возможности сопоставления, напр., НКС, полученных на материале устных и письменных текстов. Ключевые слова записывались испытуемыми *после* прослушивания. Эксперимент был проведен в двух вариантах: при предъявлении всего текста (КС, НКС) и его начального фрагмента (КСнач, НКСнач). В разных экспериментах (или сериях) участвовали разные испытуемые.

<sup>4</sup> Текст испытуемые могли слушать только один раз, удобными порциями, не возвращаясь назад.

## 2. Сопоставление наборов ключевых слов и наборов опорных слов

В настоящей работе в качестве фрагмента, на котором предположительно происходит подстройка слушающего под смысловые особенности текста, был выбран начальный фрагмент. Этот фрагмент соответствует трети от всего текста. Следовательно, при прослушивании начального фрагмента в наименьшей степени задействовано возможное противоречие между собственно смысловой структурой всего текста и его структурных компонентов.

В табл. 1а и 1б приведены свертки трех видов: НОС, НКС и НКСнач. Слова в этих наборах соответствуют словоупотреблениям из текста, а НКСнач обозначает набор ключевых слов, выделенных по традиционной методике на основании эксперимента, в котором предъявлялся только начальный фрагмент делового и художественного текстов.

НОС	НКСнач	НКС
<b>стороны</b>	<b>стороны</b>	<b>стороны</b>
участвующие		
	<b>договора</b>	<b>договора</b>
	о деловом	о деловом
	<b>сотрудничестве</b>	<b>сотрудничестве</b>
	и предприятия	и предприятия
<b>с иностранным капиталом</b>	<b>с иностранным капиталом</b>	<b>с иностранным капиталом</b>
имеют		
право		
самостоятельно		
устанавливать		
<b>программы</b>	<b>программы</b>	
	хозяйственной	
	показатели	
		вьетнамскими
	государственными органами	государственными
	экономическом	
		разрешения
	на вложение инвестиций	
	на деловое	на деловое
<b>стороны договора</b>	<b>сотрудничество стороны договора</b>	<b>сотрудничество стороны договора</b>
	о деловом	о деловом
<b>сотрудничестве</b>	<b>сотрудничестве</b>	<b>сотрудничестве</b>
	предприятий	предприятий
<b>с иностранным капиталом</b>	<b>с иностранным капиталом</b>	<b>с иностранным капиталом</b>
	показатели	

Таблица 1а. Три вида наборов для начального фрагмента делового текста

НОС	НКСнач	НКС
когда	стороны	стороны
подобные		
<b>мысли</b>	<b>мысли</b>	
трудно		

### Набор опорных слов как вид свертки текста

<b>настроение</b>	<b>настроение</b>	
	и выдержку	
		Ти Фео
		клянчить
	на выпивку	
	советник	советник
	бросить	
	хао	хао
пора бы		
знать		
свое		
<b>место</b>	<b>место</b>	
	не банк	
	швырнул	
	монету	
		и крикнул
		притащился
	забирай	
	и катись	
	духу	
	научись	
	жить	
	честно	

Таблица 1б. Три вида наборов для начального фрагмента художественного текста

Примечания к табл. 1а и 1б: словоупотребления упорядочены в соответствии с порядком введения в исходный текст<sup>5</sup>

#### Деловой текст

Слова из НОС в 86% случаев принадлежат НКСнач и в 79% случаев НКС<sup>6</sup>. Таким образом, опорные слова в подавляющем числе случаев являются и ключевыми словами.

Ключевые словоупотребления – НКС – в 63% случаев являются опорными словами, НКСнач лишь в 48% случаев являются опорными.

#### Художественный текст

Для НОС художественного текста лишь три слова (*мысли, настроение, место*) находятся на пересечении с НКСнач (33% опорных слов являются ключевыми и 18% ключевых являются опорными); пересечения НОС и НКС отсутствуют. Как и следовало ожидать, наибольший объем имеет НКСнач (см. табл. 1б).

Данных о частеречных характеристиках словоупотреблений из рассматриваемых наборов приведены в таблицах 1в и 1г).

текст	часть речи	НОС	НКСнач	НКС
деловой	существительные	57	68	63
	глаголы	21	0	0
	прилагательные	14	32	37
художественный	существительные	33	59	50
	глаголы	11	35	50
	прилагательные <sup>7</sup>	56	0	0

Таблица 1в. Представленность основных частеречных классов в рассматриваемых свертках (начальный фрагмент) (%)

<sup>5</sup> Пересечения между НОС и НКС (или НКСнач) в рамках таблицы выделены полужирным шрифтом; наличие такого пересечения фиксировалось, если одна и та же лексема присутствовала в этих наборах вне зависимости от порядка следования (ср. лексему «договор»)

<sup>6</sup> О функциональном подобии структур «КС vs. неКС» и «КС vs. неКС» для начального фрагмента делового текста см. (Ягунова 2007).

<sup>7</sup> Для НОС художественного текста этот частеречный класс включает не только прилагательные, но и наречия, и предикативы

текст	часть речи	НОС	НКС
деловой	существительные	43	65
	глаголы	13	4
	прилагательные	14	37
художественный	существительные	19	70
	глаголы	19	25
	прилагательные	62	5

Таблица 1г. Представленность основных частеречных классов в рассматриваемых наборах (весь текст) (%)

Результаты анализа НОС (в сопоставлении с НКС) для делового и художественного текстов подтверждают выдвинутую гипотезу.

✓ Наборы ключевых слов (НКС и НКСнач) существенно различны для делового и художественного текстов: они содержат больше глаголов для художественного текста, что отражает бóльшую степень его динамичности по сравнению с деловым текстом.

✓ Пересечение НОС и НКС существенно выше для делового текста по сравнению с художественным. Для делового текста подавляющее число опорных слов являются ключевыми (86% vs. 33%) для начального фрагмента, более половины опорных слов являются ключевыми при рассмотрении целого текста (52% vs. 13%).

✓ Предположение о том, что смысловая структура, заложенная в НОС, является более динамичной, чем смысловая структура, представленная в виде НКС, подтверждается только для целых текстов. Для них это проявляется в увеличении доли глаголов для делового текста и глаголов, «дополненных» предикатной лексикой для художественного текста.

✓ Для начального фрагмента текста на основании анализа НОС сделать заключение о степени динамичности исходного текста, по-видимому, невозможно. Для начального фрагмента в случае НОС – в отличие от наборов ключевых слов (НКС и НКСнач) – нивелируются различия между степенью динамичности исходного текста.

### 3. Восстановление текста на основе наборов опорных слов

Признаки «функциональный стиль» и «степень динамичности», заложенные в НОС, необходимо проследить также через восстановление внутритекстовых ассоциативных связей между словами (в том числе связей в разных составляющих: синтагмах, фразах, ситуациях и т.д.), т.е. через эксперимент по восстановлению текстов на основе НОС.

В основу описываемого далее эксперимента легли следующие *гипотезы*:

- опора на НОС позволяет осуществить построение целостного связного текста;
- НОС задают функциональный стиль развертываемого текста;
- НОС отражают предметную (тематическую) область развертываемого текста;
- развертывание НОС позволяет определить позиции (контексты) слов и/или конструкций, обладающих максимальной предсказуемостью.

Проверка трех последних гипотез возможна лишь в том случае, если первая гипотеза верна, т.к. исследованию подлежат именно восстановленные испытуемыми тексты.

Эксперимент проводился в письменно-письменной форме. Опорные слова представляли собой фонетические слова, т.е. знаменательные словоформы с клитиками (предлогами, союзами, частицами). Анкета для испытуемых включала два НОС, записанных в столбцы под названием «текст 1» и «текст 2», и письменной инструкции. Столбцы соответствовали наборам для делового («текст 1») и художественного («текст 2») текстов одного экспериментального режима. В докладе рассматриваются некоторые результаты двух серий эксперимента: для начального фрагмента текста и всего текста. В письменной инструкции было указано: «*Перед Вами последовательность слов, извлеченных из текста. Попробуйте на их основе восстановить текст*», письменная инструкция дополнялась устным требованием сохранять грамматическую форму слов, вводить слова в текст в указанной последовательности и пожеланием минимизирования литературных и «философских» фантазий. Время на выполнение задания не ограничивалось.

По условиям эксперимента испытуемые не владели предметной областью экономики, делопроизводства (деловой текст), были незнакомы с реалиями жизни во вьетнамской деревне (художественный текст). Это требование позволило рассматривать процедуры понимания текста в условиях ограничений на «базу знаний» адресата.

## *Набор опорных слов как вид свертки текста*

Анкеты, не удовлетворяющие требованиям инструкции, далее не рассматривались. В результате для серии «начальный фрагмент текста» было получено более 30 анкет, для серии «весь текст» – 20 анкет. Каждый испытуемый участвовал только в одной из серий эксперимента.

Определение того, принадлежат ли восстановленные тексты тому же функциональному стилю, производилось на основании двух критериев:

- экспертная оценка:
  - заключение эксперта о принадлежности восстановленного текста к данному функциональному стилю (деловому или художественному),
  - степень статичности vs. динамичности смены описываемых ситуаций (как дополнительный признак);
- количественные (формальные) критерии:
  - коэффициент лексического разнообразия текстов (КЛР)<sup>8</sup>, что отражает степень разнообразия лексических средств при построении текста и соотносится с функциональным стилем текста,
  - длина текстов в словах (как дополнительный признак).

Предполагается, что разнообразие лексики (высокий КЛР) характеризует динамический (художественный) текст, а клишированность (низкий КЛР) – статический деловой текст.

### **4. Результаты эксперимента. Выводы**

На основании данных, полученных в результате эксперимента по восстановлению текста, можно сформулировать следующие выводы.

	деловой		художественный	
	НОС для всего текста	НОС для начального фрагмента	НОС для всего текста	НОС для начального фрагмента
длина текстов (в с/у)	124	51	118	64
КЛР для словоформ	0,62	0,81	0,80	0,86
КЛР для лексем	0,51	0,74	0,72	0,79

*Таблица 2. Средние значения КЛР и длины восстановленных текстов*

#### *Деловой текст*

##### **• НОС для всего текста**

1. Все восстановленные на основе НОС тексты – как и исходный текст – относятся к деловому функциональному стилю. Большинство развернутых текстов можно отнести к жанру нормативных актов. Все развернутые тексты – как и исходный текст – относятся к статическому варианту: как правило, предписание, регулирующее положение дел, или – в ряде случаев – описание некоторого положения дел.

2. Высокая степень повторяемости словоформ (коэффициент лексического разнообразия КЛР 0,46) в предъявляемом НОС провоцирует высокую повторяемость слов в восстанавливаемых испытуемыми текстах (см. табл. 2).

3. При развертывании текста испытуемыми восстанавливались КС (как присутствующие, так и отсутствующие в НОС), т.к. они соответствовали предметной области данного текста. Подавляющее число используемых слов относятся к деловому функциональному стилю<sup>9</sup>.

##### **• НОС для начального фрагмента**

1. Все развернутые тексты – как и исходный текст – относятся к деловому функциональному стилю. Большинство развернутых текстов можно отнести к жанру нормативного акта.

2. Значения КЛР для начального фрагмента исходного делового текста показывает сравнительно высокую степень повторяемости лексических единиц, однако степень клишированности исходного делового текста на всем тексте существенно выше. Клишированность текста лишь отчасти задается через НОС (КЛР 0,82). КЛР для лексем восстановленных текстов в среднем близок к соответствующему показателю исходного текста. КЛР для словоформ восстановленных текстов в среднем ниже, чем для исходного текста (для этого фрагмента исходного тек-

<sup>8</sup> КЛР подсчитывается следующим образом: в числителе – число разных единиц, в знаменателе – общее число рассматриваемых единиц в тексте.

При отнесении лексики к деловому функциональному стилю использовался сделанный С.А. Шаровым на основе НКРЯ прототип частотных списков по жанрам (<http://corpus.leeds.ac.uk/serge/frqlist/>), в котором представлен «Частотный словарь административных текстов, значимая лексика»

ста КЛР 0,40 и 0,61, соответственно).

3. Различия между КЛР между восстановленными деловыми текстами на основании «НОС для всего текста vs. НОС для начального фрагмента» существенны как для словоформ, так и для лексем (см. табл. 2). По-видимому, это различие соотносится с важностью текстового «окна сверки» для восприятия делового текста (Ягунова 2007).

#### *Художественный текст*

##### **• НОС для всего текста**

1. Все развернутые тексты – как и исходный текст – относятся к художественному функциональному стилю. Они характеризуются динамичностью. Обычно восстанавливается сюжет с двумя действующими лицами, в тексте присутствуют диалоговые фрагменты; характеристики действующих лиц (как и место действия) могут быть самыми разными. Общее содержание развернутых текстов может быть охарактеризовано как реализации сценария «конфликт» (сопровождающийся каким-либо требованием).

2. Художественный текст с динамично развивающимся сюжетом характеризуется разнообразием знаменательной лексики. Повторяемыми являются, главным образом, местоименная лексика, в какой-то степени – наименования действующих лиц.

3. КЛР для словоформ и лексем восстановленных текстов несколько выше, чем для исходного текста. Это незначительное расхождение, вероятно, связано с индивидуальными стратегиями порождения текста (различаются богатством сюжетной линии, использованием синонимических ресурсов и местоименной лексики).

##### **• НОС для начального фрагмента**

1. Все развернутые тексты – как и исходный текст – относятся к художественному функциональному стилю. Большинство развернутых художественных текстов характеризуется статичностью (отсутствием смены ситуаций). Как правило, при разворачивании НОС для начального фрагмента текста не восстанавливается сюжет с двумя действующими лицами, в тексте отсутствуют диалоговые фрагменты.

2. Степень разнообразия лексем может быть связана со степенью статичности текста: восстановленные тексты являются более статичными, чем исходные, и КЛР для лексем восстановленных текстов ниже, чем для исходного текста. Возможным подтверждением статичности восстанавливаемого фрагмента художественного текста служит то, что КЛР для лексем восстановленных художественных и деловых текстов (начальные фрагменты) различаются лишь на уровне тенденции (см. табл. 2).

Главным результатом является подтверждение гипотезы о том, что НОС задают функциональный стиль текстов, восстанавливаемых испытуемыми в эксперименте. Значимость различий между КЛР (для словоформ и лексем) восстановленных (по НОС для всего текста) деловых и художественных текстов является количественным показателем принадлежности текстов к разным функциональным стилям.

Полученные данные не противоречат гипотезе о том, что НОС (для начального фрагмента текст) задают функциональный стиль тех текстов, что восстанавливаются испытуемыми в эксперименте. То, что восстановленные тексты принадлежат тому же функциональному стилю, что и исходные, подтверждается качественными и – отчасти – количественными критериями.

Результаты эксперимента по разворачиванию НОС подтверждают выдвинутую гипотезу о том, что существенное значение для смыслового структурирования текста имеет функциональный стиль и степень динамичности текста. Наиболее ярко взаимодействие факторов «*функциональный стиль*» и «*степень динамичности текста*» проявляется именно в случае наличия противоречия этих факторов.

#### *Список литературы*

1. Мурзин Л.Н., Штерн А.С. Текст и его восприятие. Свердловск, 1991
2. Сахарный Л.В., Сибирский С.А., Штерн А.С. Набор ключевых слов как текст // Психолого-педагогические и лингвистические проблемы исследования текста – Пермь, 1984
3. Ягунова Е.В. Мелодические признаки и опорные элементы при восприятии текста // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог-2006» (Бекасово, 31 мая – 4 июня 2006 г.) / Под ред. И.М. Кобозевой, А.С. Нариньяни, В.П. Селегея.– М.: Наука, 2005
4. Ягунова Е.В. Коммуникативная и смысловая структуры текста и его восприятие // Вопросы языкознания. 2007, №6