

МОДЕЛЬНОЕ ПРЕДСТАВЛЕНИЕ ПРОЦЕССОВ ПОНИМАНИЯ ТЕКСТА НА ОСНОВЕ КЛЮЧЕВЫХ СЛОВ

Статья посвящена моделированию процессов понимания текста на основе ключевых слов. Выявляется механизм связей однокомпонентных ключевых звеньев в процессе понимания текста и выводятся закономерности процесса осмысления содержания текста информантами.

Ключевые слова: понимание текста, модель, ключевые слова, матрица переходов.

Проблема восприятия речи (текста) является одной из центральных для современной психолингвистики. В многочисленных исследованиях подчеркивается сложность процессов восприятия и понимания текста и указывается их тесная взаимосвязь, поэтому восприятие и понимание принято рассматривать «как две стороны одного явления – сторону процессуальную и сторону результативную» [6, с. 205].

Восприятие текста представляет собой процесс раскрытия опосредованных словами связей. «Установление связи между (как минимум) двумя словами формирует определенное смысловое звено, показателем которого является возможность перевода его на уровень нерасчлененного, монолитного представления – образа» [5, с. 25]. Образ содержания текста соотносим с понятием проекции текста. Под проекцией текста понимается «продукт процесса смыслового восприятия текста реципиентом, в той или иной мере приближающийся к авторскому варианту проекции текста» [4, с. 252]. Однако общеизвестен тот факт, что в интерпретации одного и того же текста разными реципиентами часто наблюдаются значительные расхождения. Прежде всего это относится к художественным текстам, поскольку их содержание зачастую «настолько неоднозначно, что можно говорить о множественности содержаний» [3, с. 29]. Поэтому человек, воспринимающий текст, может создать для себя его собственную проекцию, которая коренным образом может отличаться как от проекций текстов других реципиентов, так и от авторского замысла.

При ориентировке в тексте и в процессах его понимания читатель пользуется разными опорными пунктами. В качестве опорных пунктов могут выступать наиболее важные, существенные для понимания слова текста,

которые принято называть «ключевые слова» (далее КС). Они служат вехами в процессе понимания текста или, другими словами, операторами, с помощью которых происходит членение текста на отдельные смысловые группы и их обобщение.

В процессе прочтения и анализа текста информант переходит от одного предложения к другому, каждое из которых состоит из определенного количества слов. Когда мы говорим о ключевом слове, то думаем об одном слове. Слово, являясь основной психолингвистической единицей восприятия, оказывается «критической единицей сегментации текста, позволяющей проследить взаимодействие между означающими и означаемыми, между данным в тексте и извлекаемым из памяти» [4, с. 243]. Однако на практике информанты выделяют как отдельные слова, так и группы КС, иными словами, звенья КС.

Целью нашего исследования явилось моделирование процессов понимания текста. Мы полагаем, что механизм понимания текста опирается на определенные ритмические закономерности, которые можно вывести путем анализа ритмико-графических параметров функционирования КС в тексте. Ранее ритмико-графическое пространство разрабатывалось при исследовании ритма размера предложений в тексте [1, с. 168; 2, с. 32].

Материалом нашего исследования послужили художественные тексты малой формы прозы, а именно рассказы различных авторов: В.П. Катаева «Лето»; И.С. Тургенева «Череп», «Как хороши, как свежи были розы»; Л.Г. Енгибаряна «Аплудисменты»; И.А. Бунина «Русь», «Волки»; М.М. Зощенко «Сила таланта»; М.М. Пришвина «Расставание и встреча», «Ботик»; А.И. Куприна «Безумие». В качестве методики выявления КС в тексте в наших экспериментах мы используем метод графическо-

го выделения (подчеркивания), позволяющий проследить распределение КС по тексту.

Процесс анализа выделенных КС проводился следующим образом: из 10 художественных текстов последовательно из каждого предложения в числовой форме выписывалось количество подчеркнутых информантами слов. Под словом мы понимаем графическое слово. Для примера рассмотрим 1 предложение из текста М.М. Зощенко «Сила таланта» информанта Т.В.: «*Он до того был боек в своих движениях, что актриса не могла даже его лица рассмотреть*». Подчеркнутые информантом КС в данном предложении можно представить в виде цепочки следующих друг за другом звеньев КС разного объема: 1–2–3–2. Следующие друг за другом звенья образуют последовательность отдельных слов, фраз, предложений, из которых и строится текст. Таким образом, можно рассмотреть последовательность звеньев разного размера, определить, звено какого размера последует за предыдущим звеном, и установить, какие переходы (последовательности из двух рядом стоящих звеньев) являются наиболее частотными. На основе частотности употребления звеньев они были разделены на три группы: краткие, включающие в себя фрагменты, состоящие из 1 и 2 рядом стоящих слов; средние (3, 4, 5 слов) и длинные (6–18 слов).

Образованные звеньями переходы необходимо рассматривать в системе (матрице переходов). Под матрицей переходов мы понимаем продуктивную модель динамического описания языка, которая помогает проводить анализ сегментации текста, исследовать динамику изменения предложений, слов. Метод моделирования процесса развертки семантического пространства с помощью матрицы переходов был реализован в программном средстве [7].

Для каждого текста была составлена матрица переходов и подсчитано, сколько раз, например, после звена, состоящего из одного слова, стоит звено, состоящее двух из слов, то есть сколько раз звено из одного слова переходит в звено из двух слов и т. д. Рассмотрим для примера матрицу переходов текста «Сила таланта» в таблице 1. По горизонтали и по вертикали отмечается количество звеньев в данном тексте. В ячейке пересечения двух звеньев находится частота встречаемости рассматриваемого перехода.

Из таблицы видно, что наиболее частотными являются краткие переходы 1–1, 1–2 и 2–1. Это означает, что наиболее типичной единицей текста выступает одиночное опорное слово. Испытуемый дробит текст на отрезки, которые воспринимаются им легче. В этих переходах находятся в основном существительные преимущественно в их исходной форме. Информантами было подчеркнуто 233 слова в именительном падеже, например: *успех (55 раз), поклонник (45), слава (25) и т. д.*

Слишком большое отличие в количестве слов в двух рядом стоящих звеньях (например, переходы кратких в длинные: 1–7), в которых наблюдается рост континуальности текста, встречается в единичных случаях в отличие от средних переходов из одного звена в другое с примерно равным количеством слов.

Для определения наиболее важных переходов в рамках всех рассматриваемых текстов была подсчитана общая матрица переходов, включающая в себя сумму всех матриц по каждому тексту. Было выявлено, что частотными выступают следующие переходы: 1–1, 1–2, 1–3, 1–4, 1–5, 2–1, 2–2, 2–3, 2–4, 2–5, 3–1, 3–2, 3–3, 3–4, 4–1, 4–2, 4–3, 4–4, 5–1, 5–2.

Для того чтобы наглядно представить, какие наиболее частотные случаи встречаются во всех текстах, методом случайной выборки из 300 реакций нами были отобраны по три реакции от каждого текста. В реакциях производился подсчет всех значимых переходов, в результате которого было выявлено значительное превалирование переходов 1–1 по сравнению с другими переходами. В этой связи интерес для нас представляет вопрос: какие части речи встречаются наиболее часто в переходах из 1 слова: 1–1.

Таблица 1. Матрица переходов текста «Сила таланта»

	1	2	3	4	5	6	7	8	9	10	11	12
1	262	112	28	11	10	1	1	0	0	0	0	0
2	100	56	20	14	11	1	2	0	0	1	1	0
3	33	10	6	3	6	1	0	1	0	0	0	0
4	10	10	4	3	2	3	0	0	0	0	0	0
5	9	12	4	3	3	2	1	0	0	0	0	0
6	1	0	1	1	2	1	0	1	0	0	0	1
7	1	2	0	1	0	0	0	0	0	0	0	0
8	1	0	0	1	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	1	0	0	0	0	0	0	0	0	0
11	0	1	0	0	0	0	0	0	0	0	0	0
12	0	0	0	1	0	0	0	0	0	0	0	0

Наиболее частотными в переходах 1–1 выступают существительное – существительное. Количество этих переходов составило 39, например: *сцену – цветы; цветы – публикой*. Большую часть их (22 пары существительных) составили существительные в начальной падежной форме: *успех – счастье; счастье – слава; слава – режиссер*. Наличие большего количества переходов существительное – существительное по сравнению с другими переходами свидетельствует о том, что существительные в качестве КС выступают основой понимания текста. Они номинируют субъекты и объекты с помощью конкретной лексики, обозначают реалии мира, благодаря которым информанту становится понятна тематика текста.

На втором месте в переходах 1–1 располагается модель глагол – существительное (28), например: *очнулась – поклонника; оправились – прошлое*. Следует отметить, что в этих переходах преобладает предикативная основа (15), например: *запел – соловей; вздрагивает – пес*. Число переходов существительное – глагол почти совпадает и составляет 26, из которых 15 переходов также содержат предикативную основу: *солнце – играет; аплодисменты – вырываются* и т. д.

Модель глагол – глагол является достаточно частотной (20 пар слов). Главным образом были подчеркнуты глаголы изъявительного наклонения (20 пар), единственного числа (12 пар), несовершенного вида (11 пар). Большая часть глаголов (9 пар) употреблена в настоящем постоянном времени: *целует – отодвигает; обнимает – целует* и т. д. У 5 пар глаголов зафиксировано прошедшее время: *умерли – умерли; направился – очнулся* и т. д.

Значимыми явились также модели с прилагательными: модель прилагательное – существительное представлена 26 парами слов, модель существительное – прилагательное – 19 пар, модель прилагательное – прилагательное – 18 пар слов.

В нашем исследовании мы не проводим подробный анализ переходов, количество которых превышает одно слово, поскольку реакции информантов разнятся так, что трудно предложить единую методику их оценки.

В связи с появлением вышеназванных моделей переходов возникает вопрос: какие функции в предложении выполняют выделенные информантами КС, выраженные разными частями речи. В роли КС могут выступать как

главные (подлежащее и сказуемое) (ГЧП), так и второстепенные члены предложения (ВЧП). Причем второстепенные члены предложения чаще выделяются информантами, поскольку переходы со второстепенными членами предложения составляют 125, а ГЧП в сумме равно 102 (см. таблицу 2). Следовательно, количество ВЧП в 1,2 раза превышает количество главных членов предложения. В этой связи интерес для нас представляет рассмотрение явления симметрии и асимметрии между ГЧП и ВЧП. Проявления симметрии/асимметрии между КС представлены в таблице 2.

Как мы видим, при подчеркивании КС явление асимметрии немного превалирует над симметрией: количество переходов, в которых встречается второстепенный член предложения на 10% выше, чем переходов между ГЧП. Симметрия в переходах между ГЧП и ВЧП численно абсолютно равна и составляет 50, что свидетельствует о том, что если информанты в тексте выделяют в качестве КС главные члены предложения с элементами симметрии: когда подлежащее переходит в подлежащее, а сказуемое – в сказуемое, то они подчеркивают также и второстепенные члены в равной пропорции. Незначительное превалирование явления асимметрии над симметрией свидетельствует о стремлении к уравниванию в тексте.

В этой связи становится интересно проанализировать, по какому принципу информанты выделяют КС, что и в какой степени связывают однокомпонентные ключевые звенья между собой: предложения или абзацы. Для этого в каждом переходе отмечалась встречаемость его в одном или в разных предложе-

Таблица 2. Проявление симметрии/асимметрии между КС на синтаксическом уровне

	ГЧП	ВЧП
симметрия	П-П (22) С-С (28)	В-В (50)
асимметрия	П-С (27) С-П (25)	П-В (18) В-П (16) С-В (22) В-С (19)

Примечание: П – подлежащее; С – сказуемое; В – второстепенный член предложения; П-П – подлежащее – подлежащее; П-С – подлежащее – сказуемое; П-В – подлежащее – второстепенный член предложения; С-П – сказуемое – подлежащее; С-С – сказуемое – сказуемое; В-В – второстепенный член предложения – второстепенный член предложения.

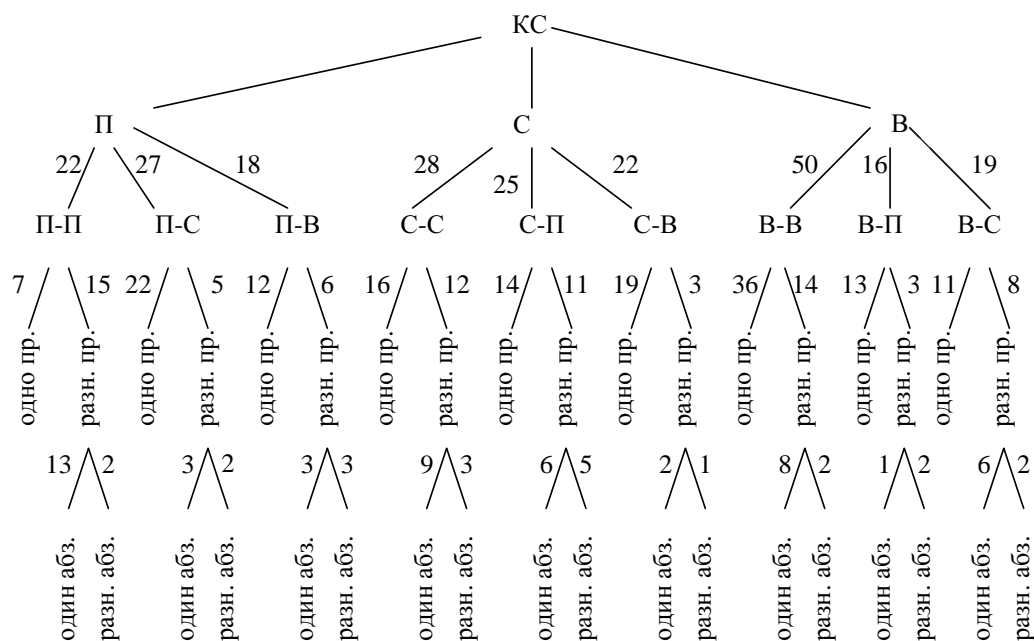
ниях (см. рис. 1). Если же КС были подчеркнуты из разных предложений, то заносились данные об употреблении их в одном или разных абзацах. На рисунке видно, что переход П-П часто встречается в разных предложениях в одном абзаце, что свидетельствует о том, что связь между предложениями в тексте осуществляется через номинацию субъектов или объектов. Например: *Леса там темны, дремучи. Снега выше вековых сосен («Русь»)*.

В рамках одного предложения переход П-П встречается в следующих случаях: в составе сложного предложения с указанием на причину, например: *Вчера вечером на деревне был шум, крик, трусливый лай и визг собак: с удивительной дерзостью, когда по избам уже ужинали, волк зарезал в одном дворе овцу и едва не унес ее – вовремя выскочили на собачий гам мужики с дубинами и отбили ее, уже околевавшую, с разорванным боком («Волки»); Нет, трясовица была на мне, порча на мне была («Русь»); и в качестве однородных подлежащих: *С ужасом глядел я, как двигались и шевелились эти десны и скулы... («Череп»)*.*

Отличие от перехода С-П, встречаемость перехода П-С высока в одном предложении, поскольку информанты, подчеркнув подлежа-

щее, автоматически выделяют и относящееся к нему сказуемое, чтобы охарактеризовать действие, совершаемое субъектом или объектом: *Свеча меркнет и гаснет («Как хороши, как свежи были розы»)*. Переход С-П наблюдается чаще всего в сложных предложениях с инверсированным порядком слов: *Свернувшись в калачик, жметя и вздрагивает у ног моих старый пес, мой единственный товарищ («Как хороши, как свежи были розы»)*. Вероятно, тем самым испытуемые пытаются выделить именно это предложение из фона других.

Интерес представляет переход С-С, в котором КС выявляются информантами почти одинаково как в одном, так и в разных предложениях. Выделение информантами сказуемых подчеркивает значимость действий для субъектного мира. В одном предложении сказуемые отмечались, поскольку выступали в качестве однородных членов предложения или являлись элементами сложного предложения: *Я сижу, забившись в угол; а в голове все звенит да звенит. Я зябну... и все они умерли... умерли («Как хороши, как свежи были розы»)*. Выявление сказуемых в разных предложениях, но в рамках одного абзаца свидетельствуют о том, что действие является связующим



П – подлежащее; С – сказуемое; В – второстепенный член предложения; П-П – подлежащее – подлежащее; П-С – подлежащее – сказуемое; П-В – подлежащее – второстепенный член предложения; С-П – сказуемое – подлежащее; С-С – сказуемое – сказуемое; В-В – второстепенный член предложения – второстепенный член предложения; одно пр. – одно предложение; раз. пр. – разные предложения; один абз. – один абзац; раз. абз. – разные абзацы

Рисунок 1. Механизм связей однокомпонентных ключевых звеньев в процессе понимания текста

элементом для создания художественного мира, поскольку сказуемые объединены одной темой. Например, в тексте «Лето» предложения связаны темой смерти: *Лето умирает. Осень умирает.*

Переходы со второстепенными членами встречаются чаще в одном предложении, поскольку они приносят дополнительные характеристики действия или качества названных предметов, например: *Слышится веселый шум семейной деревенской жизни («Как хороши, как свежи были розы»); Она живет бесконечно в недрах вечно изменяющейся материи, только меняет свои формы («Лето»).*

Благодаря введению в исследование инструментария «матрица переходов» нами были выявлены закономерности подчеркивания информантами КС, а следовательно, те закономерности, по которым происходит осмысление содержания текста реципиентом. Допустим, что существует некая матрица переходов МП = [A_{k-m}...A_{n-n}], где A_{k-m} обозначает количество случаев перехода компонента A_k в A_m и A_{k-m} ≠ A_{m-k}. Для данной матрицы переходов соблюдаются 2 условия:

1) для любых индексов k, m из ряда простых чисел от 1 до n, чем меньше $\frac{k+m}{2}$, тем выше A_{k-m} (показатель встречаемости этого перехода). Например, как указывалось выше, переход 1-1 (средний индекс равен единице) встречается чаще, чем любой другой переход;

2) для любых индексов k, m из ряда простых чисел от 1 до n при равенстве средних значений индексов двух пар переходов, чем меньше разница |k - m|, тем выше A_{k-m}. Например, сравним переходы типа 2-2 и 1-3. Средний индекс этих переходов равен 2, однако встречаемость перехода 2-2 будет выше, т. к. |2 - 2| = 0, в то время как |1 - 3| = 2.

Согласно этой закономерности, должны соблюдаться следующие условия (перечислим все возможные варианты): 1-3 < 2-2; 1-4 < 2-3; 1-5 < 2-4; 2-2 > 3-1; 2-4 < 3-3; 2-5 < 3-4; 3-2 > 4-1; 3-3 > 4-2; 4-2 > 5-1; 4-3 > 5-2. Для проверки соблюдения этих закономерностей встает необходимость применить данную методику на:

1) общей матрице, включающей в себя сумму всех переходов;

2) каждом тексте в отдельности.

Мы проанализировали частотные переходы на выполнение/невыполнение условия 1 и

условия 2 и пришли к выводу, что предложенная модель описания переходов КС является продуктивной. Как в первом, так и во втором случае отсутствуют примеры несоблюдения условий. Это свидетельствует о том, что при условии наличия больших чисел, закономерности полностью соблюдаются. Однако, проанализировав переходы с маленькими числовыми данными (переходы по каждому тексту в отдельности), где процент случайного подчеркивания слова высок, мы видим, что закономерности могут выполняться, а могут и не выполняться. Вероятно, этим фактом можно объяснить то, что из 10 проанализированных текстов в семи отмечаются незначительные отклонения. В таблице 3 представлено соблюдение условий по каждому тексту в отдельности и по всем текстам в целом. В первом столбце расположены названия всех текстов, далее – соблюдение, несоблюдение условий и рассчитана доля несоблюдения условия от общего.

Как мы видим, числовые показатели несоблюдения условия 1 достаточно низкие от 0,05 до 0,25. Возможно, такая погрешность как 0,25 связана с размером текста, поскольку «Лето» является самым маленьким текстом. Доля несоблюдения условия 2 варьируется от 0 до 0,52. Наибольшую погрешность имеют такие тексты: «Ботик», «Расставание и встреча» и «Сила таланта». Во всех этих текстах несоблюдение выявляется в следующих переходах: 2-4, 3-3, 4-2, 4-3, 5-2.

Таблица 3. Соблюдение условий по каждому тексту

Название текста	Условие 1			Условие 2		
	соблюдение	несоблюдение	доля несоблюдения условия	соблюдение	несоблюдение	доля несоблюдения условия
Лето	9	3	0,25	8	2	0,2
Череп	19	1	0,05	16	0	0
Аплудисменты	16	4	0,2	15	1	0,06
Русь	18	2	0,1	12	5	0,29
Сила таланта	18	2	0,1	8	9	0,52
Расставание и встреча	17	3	0,15	9	7	0,43
Как хороши, как свежи были розы	18	2	0,1	15	3	0,16
Ботик	18	2	0,1	10	7	0,4
Безумие	17	3	0,15	13	4	0,23
Волки	19	1	0,05	16	0	0
Все тексты	20	0	0	16	0	0

Обобщая все вышесказанное, приходим к выводу, что для понимания смысла текста читателем из общего фона выделяются некоторые его элементы, которые осознаются как ключевые. Графически выделенные КС в тексте способны объединяться в звенья разного объема, измеряемого в количестве слов. По этим звеньям можно проследить процесс обработки реципиентом текстовой информации. Звенья КС образуют переходы, которые

необходимо рассматривать в системе (матрице переходов). Построение матриц переходов способствовало выявлению наиболее типичных единиц понимания текста, в качестве которых выступают в основном существительные преимущественно в их исходной форме, а также помогло вывести определенные ритмические закономерности процесса понимания текста информантами.

16.09.2012

**Исследование выполнялось при финансовой поддержке
Российского гуманитарного научного фонда (проект № 12-34-01087– а1)**

Список литературы:

1. Белоусов, К.И. Деятельностно-онтологическая концепция формообразования текста [Текст]: дис. ... докт. филол. наук. – Оренбург, 2005. – 374 с.
2. Белоусов, К.И. Физические и синтаксические текстовые структуры в аспекте пространственной модели текста // Вестник Оренбургского государственного университета. – 2008. – №11. – С. 31–37.
3. Белянин, В.П. Психолингвистические аспекты художественного текста. – М., 1988.
4. Залевская, А.А. Понимание текста: психолингвистический подход. – Калинин, 1988.
5. Зимняя, И.А. Смысловое восприятие речевого сообщения // Смысловое восприятие речевого сообщения (в условиях массовой коммуникации). – М., 1976.
6. Зорькина, О.С. О психолингвистическом подходе к изучению текста // Язык и культура. – Новосибирск, 2003. – С. 205–210.
7. Информационные потоки в тексте: программное средство [электр. издание] / И.В. Влацкая, К.И. Белоусов, И.А. Билалова, К.Г. Журенко. – М.: ВНТИЦ, 2009. – Инвентарный номер ВНТИЦ 50200300243 от 03.02.2009.

Сведения об авторе:

Осипова Ирина Анатольевна, ассистент кафедры немецкой филологии и методики преподавания немецкого языка, аспирант кафедры электронных СМИ Оренбургского государственного университета 460018, г. Оренбург, пр-т Победы, 13, тел. (3532) 372434, e-mail: irinastone@rambler.ru

UDC 81:1**Osipova I.A.**

Orenburg state university, e-mail: irinastone@rambler.ru

MODELING REPRESENTATION OF PROCESSES OF TEXT UNDERSTANDING ON THE BASIS OF KEY-WORDS

Article is devoted to modeling of processes of text understanding on the basis of keywords. Mechanism of links of one-component key parts by text understanding are revealed and regularities of understanding of text contents by informants are deduced.

Key words: text understanding, model, keywords, matrix of transitions.

Bibliography

1. Belousov, K.I. Activity-ontological concept of text forming [Text]: dis. ... doc. of Philology. – Orenburg, 2005. – 374 p.
2. Belousov, K.I. Physical and syntactic text structures in aspect of spatial model of the text // Bulletin of the Orenburg state university. – 2008. – №11. – P. 31–37.
3. Belyanin, V.P. Psycholinguistic aspects of art text. – M., 1988.
4. Zalevskaya, A.A. Understanding of the text: psycholinguistic approach. – Kalinin, 1988.
5. Zimnyaya, I.A. Semantic perception of speech message // Semantic perception of speech message (in conditions of mass communication). – M., 1976.
6. Zorkina, O.S. About psycholinguistic approach to text study // Language und culture. – Novosibirsk, 2003. – P. 205–210.
7. Information streams in the text: software [electronic edition] / I.V. Vlackaya, K.I. Belousov, I.A. Bilyalova, K.G. Zhurenko. – М.: VNTIZ, 2009. – Accession number VNTIZ 50200300243 from 03.02.2009.