

Международная конференция «Диалог 2010»

Круглый стол «Онтологии и лексическая семантика: общие проблемы»

Ведущая Н.В. Лукашевич

Н.В. ЛУКАШЕВИЧ (НИ ВЦ МГУ им. М.В. Ломоносова; Центр информационных технологий): Мы обсуждаем проблемы онтологий, и больше лингвистических, лексических онтологий. Мы можем обсуждать примеры таких онтологий, чем они отличаются или сходны с более формальными онтологиями. С другой стороны, если лингвистические онтологии должны описывать значения реальных лексических единиц, то как формализованные описания в рамках таких онтологий сходны или отличаются от описаний, которые делаются в теории лексической семантики или от словарных описаний в обыкновенных словарях.

Вчера я на дискуссии об этом говорила, повторю — о соотношении формальных и лексических онтологий. В частности интересный вопрос — может ли быть большая, развитая онтология полностью независимой... понятно, что маленькую онтологию можно построить какую угодно... Интересно обсудить, какой размер должен быть у лингвистической онтологии. Должна ли она быть очень подробна и похожа на WordNet, или она должна ограничиваться некоторым количеством концептов (пример такой онтологии — MikroKosmos). Насколько подробно мы можем в онтологических ресурсах описывать различные специфические особенности слов и их значений. И, на мой взгляд, очень важная проблема — что описания в онтологиях очень дискретны, мы должны как бы разложить на систему понятий. Как правильно работать с лексической неоднозначностью и группами близких по смыслу слов. Очень интересный вопрос — соотношение лексических отношений, как они устанавливаются между словами, и более строгими, формализованными онтологическими отношениями и их критериями.

Вот соотношение лингвистических и формальных онтологий. Здесь я в каком-то смысле цитирую работы Йорика Уилкса, который писал, что во многих онтологиях используются слова, чтобы представлять понятия, что имена в концептах часто формируются на естественном языке, и была даже высказана гипотеза, что то, что не может быть описано словами, не может быть представлено в онтологии. С другой стороны, если вы обратите внимание на то, как люди работают в конкретных предметных областях, то вы увидите, что в предметных областях много онтологий очень близких тезаурусам, то есть они базируются на предметной терминологии, и вообще предметное знание очень сильно скрывается в текстах, и возникает вопрос, насколько можно крупные формализованные структуры создать независимо, рассматривая конкретный язык, языковое разбиение предметной области.

Другой пункт из моего доклада, который я бы хотела пояснить, — мне казалось, что один из важных пунктов формального подхода к онтологиям — это различимость понятий. Я привела некоторые положения из онтологических

руководств по тому, как должны быть отличимы понятия. Нужно ли добиваться этого в лингвистических онтологиях или на самом деле это неважно?

Еще один очень сложный момент — насколько подробно нужно и можно представлять значения многозначных слов. Понятно, что это одна из сложнейших проблем автоматической обработки текстов — как автоматически разобраться в разных значениях. Те люди, которые пытались применить WordNet, много писали о том, что значения слов в WordNet разобраны слишком подробно и нет никаких отношений между близкими смыслами. С другой стороны, пример онтологии MikroKosmos, как я обсудила с авторами, является первым экспериментальным опытом теории онтологической семантики, там достаточно много значений было кластеризовано одной концептуальной единицей, то есть провозглашается принцип объединения как можно большего количества значений. Вот тоже интересный вопрос — насколько возможно в этих приложениях объединять похожие, близкие значения, чтобы снизить проблему разрешения многозначности.

Когда мы работаем с онтологиями, вроде получается, что очень дискретно нужно рассматривать значения. С WordNet было очень много усилий, чтобы попытаться разобраться, что, может быть, можно автоматически или другими методами объединять значения, но тем не менее выяснилось, что особенно не очень это получается: разные приложения требуют разных объединений значений. То есть то, что можно объединить с точки зрения информационного поиска, то плохо с точки зрения такого приложения, как, допустим, машинный перевод. С другой стороны, некое объединение похожих значений иногда ведет, с точки зрения формальных подходов, к нарушению онтологической структуры. В частности известна дискуссия, когда известный онтолог Никола Гуарино указывал, что некоторые онтологии толкуют такую сущность, как окно, одновременно как некоторый артефакт, некоторую вещь и как отверстие. То есть для того чтобы избежать такого разрешения многозначности, в частности в такой онтологии, как MikroKosmos, такие ипостаси «окна» как артефакта и отверстия были склеены в единый концепт. Насколько это может повредить или, напротив, улучшить нашу жизнь с точки зрения разрешения многозначности?

Продолжая тему с окном, много обсуждаемую в литературе, — какие последствия того, что мы соединяем эти «окна» в единую концептуальную сущность?

Последствия такие: по жизни, если моделировать мир, мы знаем, что окно появляется в доме как отверстие; с другой стороны, окна как артефакты делаются на заводе, в совершенно другом месте. Если мы четко хотим сказать, что мы имеем в виду отверстие, мы можем сказать «window opening» — и это 351 000 страниц в Гугле; люди говорят так, если хотят сказать точно. А если мы точно хотим сказать, что имеем в виду окно как артефакт, то оно выражается как «оконная рама», «window pane». Если мы для разрешения многозначности делаем неподробный анализ значений, то у нас окно-отверстие получается чуть ли не синонимом оконной рамы.

Мне казалось, что наиболее упоминаемые примеры лингвистических онтологий представляют собой некоторые крайние точки зрения. Конечно, авторы WordNet

не строили онтологий, однако многие называют их ресурс онтологией, поэтому это здесь можно рассмотреть как некоторую модель, то есть очень подробная система концептов, или синсетов, множество разбиений значений многозначных слов, нет отношений между близкими значениями. И, как я сегодня пыталась привести пример, иногда ряды синонимов, квазисинонимов как-то произвольно разбиты на синсеты. С другой стороны, MikroKosmos представляет совершенно другую парадигму, в которой несколько тысяч концептов, онтология и словарь очень жестко разделены, значения многозначных слов обобщаются к некоторому концепту, ряды объединяются в достаточно обобщенные понятия. Но, может быть, это ведет к какому-то переобобщению, нарушению правильных логических отношений?

Если возникнут другие вопросы, то можно будет их обсудить.

YORIK WILKS (University of Oxford): Please don't take anything I say that has been critical of the work of Professor Raskin and his team. We have sparred this for many years. For me, we laid our positions. He has a close associate with whom he has written a book on this.

Many of you may know, many of you know Sergei Nirenburg. Nirenburg and I wrote an article together some years ago. If you type into Google "Nirenburg, Willks, symbol", you'll get this article. It was a Socratic dialogue between us published like a Greek dialogue: he speaks, I speak, he speaks, I speak. And it was entirely about this very question of what are the objects in an ontology. Are they really words no matter what anybody says, no matter whether they are in lower case or upper case, which is what I believe? Or are they language independent objects, which is what I believe professor Raskin believes and I know Sergei Nirenburg believes.

This is an important difference and it respects on all the things we are talking about. I personally cannot see any difference in the end between linguistic ontologies and formal ontologies. All the attempts I've seen to distinguish them don't work. I mean ontologies in a sense how facts about the world and theories of the world... But dictionaries are full of facts about the world. They tell you horses are mammals. A formal ontology has different senses of words in just like a linguistic ontology. A crane in English can be a big machine for lifting things or can be a bird with a long beak. They both have to be in there. If this is a formal ontology, you probably have to mark it crane1 and crane2 and so on. If the formal ontology is in English and this is an important question. Why should it be in English? Guarino was mentioned by professor Lukashevich. I've struggled with Guarino for years. He is the man who believes he can tell the formal ontology from linguistic ontology. He is the man who thinks word naughty is informal, basically rubbish. But all his arguments I believe are completely false and are based on false analysis. And also he doesn't understand English perfectly, which again is a problem. Because

he starts to make points about “lumps of coal” or things he doesn’t really know the words. No, I’m not criticizing his English because he is Italian. I’m saying this is really difficult because everybody seems to be conducting his arguments and not English words. This is terrible! Why should ontologies always be in English? Why should word be in English? This is absurd! Because people are becoming obsessed by the features of the English language. And since I happened to believe that ontologies are in language, I believe these ontologies always are in English. This is a deep mistake!

People who don’t speak English perfectly often think that they are... I mean I once went on tour in Japan and looked at Japanese machine translation systems, very good ones, like Fujitsu. In the middle of their machine translation system is an Interlingua in English. Some of them were very familiar, some of them were old formal codings I’ve made in the 1970s. They reappeared in the middle of Fujitsu machine translation system. But Japanese loved this. They said: Look, we have Interlingua for doing MT. I said: That’s not Interlingua, that’s English! No-no-no-no! No-no-no-no! This Interlingua! And the trouble was, you see, they don’t understand English perfectly. So they thought of these Interlingua symbols they were using they only have one sense. But anybody who knows English knows they have lots of senses. How do Japanese know which senses it was?

So, actually, I formed up a theorem at that time: maybe it[’s] good idea in ontologies if you use a language as the ontology not to know the language very well. Because if you know the language very well, you won’t believe it can be an Interlingua. And now, that’s said, I think almost any language can be an Interlingua. I think we can recode content. I’m certain of attacking me if it’s professor Raskin which I admire. I think the deep question is: what does it mean to recode content? What does it mean to write out the meanings of words in some other language? What does that mean? It sounds easy! Dictionaries do it. Of course, dictionaries are codings of meanings in other words. But what does it mean to write out the meaning in some other word, nah .. some other language? This discussion was held in linguistics 40 years ago with furthering CATs. Furthering CATs provided the first semantics for (junks...?) grammar. And David Lewis, the American philosopher, famously said the word markerees. He said all furthering CATs’ work is markerees. He said all they are doing is translating English into another piece of English, roughly what I’m saying. You see, here’s the difference: I don’t think that matters. I think You can do that. I think you can translate down to smaller version of English and do computation. Lewis, the philosopher, thought If you translate English down into some English coding like ANIMAT, physical object, that flaws it, that makes it all non-sense. You are just from one language to another, and how do you know it’s a good translation? It’s a machine translation So you can accept that meta-language if you like is the same as the object language and still think that it’s OK, like I do. Or you can be like philosopher Lewis and think it’s terrible. If you know the work of Schank, the same problem rose with him. He had 14 primitive actions to code everything, and him and me, who worked in.. we worked in next door offices, and I’m going to stay with him next month. We’ve been arguing about this for 45 years. He believes those – he doesn’t

care anymore – he believes those 14 primitive objects were not English, but they were English, they obviously were – trans, transfer.

So you see why I'm skeptic about being able to make an important distinction here. Let me end on a positive note. There is a project that some of you know when I spoke about it the other night, but I think very few of you were there, called the Semantic Web. Semantic Web is an attempt, and I'm going on to make the worldwide web in a sense understandable by recoding its content into other forms and therefore having a worldwide web where documents understood by computer. Very much the same project is - the different name, the different form of project – what professor Raskin and his team are talking about. But another worldwide web is going all over the place in a large scale. And if it works - and I think it will, as I think science requires ontologies and a lot of the semantic web will be scientific ontologies. And they are written by hand, and they are in language. They are in, scientific ontologies are in language. They are often in English I'm afraid. If the semantic web works, it proves that you will be able to have worldwide web where the text documents are understood by the system. And that will be proof that you can recode content. And then the only difference comes between me and professor Raskin over: what the status of this methodology is. This is very interesting question, but it's only a philosophical question. It needn't stop us all doing good work.

VIKTOR RASKIN (Purdue University): I'm sorry, I'm going to do something strange, may be stupid, I'm going to respond to some of your arguments. As you know, I'm in a great admire of your work. And I will do it in Russian.

На самом деле, это глупо и с другой точки зрения, потому что мой русский, когда я говорю на бытовые темы, еще туда-сюда, а профессиональный — не очень, потому что я не работаю совсем по-русски, я не знаю по-русски многих терминов, которые сам же и ввел. Так что вы уж меня простите.

В книге «Онтологическая семантика» я написал главку, которая содержит серьезные возражения против тезиса Йорика о том, что онтологии — это онтологии на языке. Я недавно перечитал эту главку с обычным для меня отвращением к моей собственной прошлой работе, потому что я не понимаю, чего это я такие глупости писал, как я мог такое думать, потому что всё изменилось. На самом деле я не понимаю аргументов Йорика. Я знаю, что он не может путать ярлычки, labels, концептов со словами, которые используются в качестве этих ярлычков. Я надеюсь, вы сегодня не были запутаны нашим докладом, где в частности мы с Хемпельманном сначала говорили о концепте «eat» и о слове «eat» — это совершенно разные вещи. Эти labels компьютер не читает, они просто маркеры и могут быть написаны на любом языке, они просто должны быть разные, если концепты разные. А потом слова английского языка, русского языка, китайского языка совершенно одинаково описываются в терминах этих концептов. Может быть, и я об этом писал, Йорик имеет в виду, что у носителя русского языка

онтология другая, чем у того, кто говорит на другом языке. Это подозрительно похоже на идею Бенджамина Ли Уорфа, так называемую идею Сепира—Уорфа, что если бы сэр Исаак Ньютон был не англичанин, то сила тяжести имела бы другое числовое значение. Я хотел бы на это посмотреть. Опять же Йорик не может иметь это в виду, так что же тогда остается от языковой идеи онтологии? Не ярлычки и не идея неприменимости онтологии к другому языку. Если мы описываем при помощи одних и тех же концептов лексические значения любого другого языка, как можно говорить, что онтология — это языковое явление?

Я хотел бы также высказать свое изумление по поводу термина «лингвистическая онтология». Никакой лингвистической онтологии нет. Есть только формальные онтологии. Если вы делаете что-то другое, вы занимаетесь другим делом. Наша онтология — это математически формальный объект, который подчиняется всем правилам формальных онтологий и который, в отличие от чисто логических онтологий, использует большее количество свойств. У нас их, скажем, 300, а в онтологиях логических, о которых вы читаете, их меньше.

Но устроены они совершенно так же. Был задан вопрос сегодня о различии наивных онтологий и научных онтологий. Да нет никаких наивных онтологий. Есть научные онтологии. Что такое для меня вообще онтология? Онтология — это просто метаязык лексических описаний. Ничего другого за этим нет. Поэтому WordNet не является никакой онтологией, потому что у них нет никаких описаний, у них есть имплицитные, подразумеваемые сущности, которые позволяют объединять определенные виды слов — кошмарные виды слов, с лингвистической точки зрения, — в одни и те же группы. Если вы употребляете слово «онтология» таким образом, что WordNet оказывается онтологией, я требую, чтобы вы признали, что брикет сливочного масла тоже является онтологией. Если вы хотите, я следующий час буду вам рассказывать, почему это так.

Мы все взрослые люди, и мы знаем, что большинство споров, которые мы ведем, это споры терминологические. И где я согласен с Йориком, это то, что неважно, как это называется, давайте делать хорошую работу.

Последняя вещь, которую я хотел бы сказать, — о нашем коротком споре с Йориком в конце нашего сегодняшнего доклада. Йорик спросил меня, что я имею против того, что называется машинное обучение. Да ничего я против не имею, только я хотел бы, чтобы оно происходило, чтобы результаты были хорошие. Йорик изобрел preference semantics, и я научился от него этому в 70-х годах, когда вообще люди занимались только синтаксисом, и уже о каких-то результатах можно было говорить несемантическими способами. Моя идея состоит в следующем. Для того чтобы заставить машинное обучение работать хорошо, с той точностью, в которой настоящие приложения нуждаются, чтобы люди их использовали, — я не знаю, говорили ли вы когда-нибудь здесь о факторе приятия систем natural language processing — обработки языков носителями языка... Вы знаете, что первые системы машинного перевода были такие жуткие, что ими никто не хотел пользоваться, а последние системы тоже являются такими

жуткими, что ими по-прежнему никто не хочет пользоваться. Это называется прогресс. Конечно, можно способами машинного обучения найти preferences и решить какие-то проблемы. Самые хвастливые из моих друзей, которые занимаются машинным обучением, говорят, что у них 80% успехов по особой метрике, по tagging, которое делается очень примитивно, — пример настоящей проблемы разрешения многозначности. И потом они достигают 80-процентной точности. Я так понимаю, что нам, для того чтобы друг с другом разговаривать и понимать, нужна 95-процентная точность, и то это будет не очень много. Поддерживают такие искусственные игрушки и очень красивые и приятные математические номера — я сам по образованию, как и многие из вас по сути дела математик, и мне это очень нравится, может, поэтому я не могу обманываться этим — всё это очень мило. Почти все исследования этого рода субсидируются американским правительством или Европейским сообществом. Системы, которые мы на эти небольшие деньги можем разработать, — это игрушки. Я не знаю, что может делать с ними Европейское сообщество, а американское правительство кладет эти игрушки на полку, а потом с них часто стирают пыль, потому что пыль — это негигиенично. Поэтому эти системы не подвергаются серьезному испытанию. Поэтому люди, которые занимаются машинным обучением, не знают, что пользователи не будут этим пользоваться. Можно ли добиться результата, которого мы хотим добиться, надеемся добиться, добиваемся в онтологической семантике методами машинного обучения? Можно. Tagging, который делают на test corpus, с выдающейся семантикой, то есть научить, разработать полную семантическую систему, чтобы люди делали маркировку по поводу этой системы. Встает вопрос — если вы разработали такую дивную семантическую систему, зачем догадываться? Просто применяйте ее к описанию, и у вас получится то, что делает онтологическая семантика.

Я закончу одним из своих любимых анекдотов, которые я десять лет пытаюсь рассказывать на западе и не могу найти хотя бы одного американца, который понимает, о чем я говорю. Петька вбегает в избу и говорит: «Василий Иванович, интерконтинентальные баллистические ракеты привезли». Он говорит: «А седла?» — «Уже третий раз привозят». Понимаете, когда мы сделали такую семантику, вам не нужны седла для этого, она всё сделает.

Н.В. ЛУКАШЕВИЧ: Спасибо, но мне кажется, что мы не планировали обсуждать соотношение подходов к машинному обучению, и, мне кажется, что обсуждение осталось на соотношении семантических описаний и онтологий. Потому что, мне кажется, это некая отдельная тема.

В. РАСКИН: Да нет, Наташа, это очень сильно связано.

Л.Л. ИОМДИН (ИППИ РАН): Я нахожусь в очень странном положении, потому что мы никогда онтологиями не занимались. Наша группа занималась глубокой семантикой, словарями, лингвистическими процессорами, машинным переводом и никогда не занимались онтологиями. Потом, как персонаж Мольера, мы узнали, что говорим прозой, и выяснилось, что мы все-таки использовали онтологии. Я был уверен, что это так, но теперь Виктор Раскин говорит, что WordNet — это вовсе никакая не онтология.

Тем не менее, мне кажется, что все-таки что-то такое должно существовать. И онтология как нечто более высокого порядка, чем языковой словарь, наверное, все-таки имеет место. Когда я сказал про WordNet, который может быть онтологией, а может быть, и нет, — это как раз тот проект, который упоминал профессор Уилкс, когда говорил о японце Учидо, который когда-то работал над системой машинного перевода, а потом сделал там интерлингву на основе английского языка. Вот эта интерлингва называется UNL, universal networking language, то есть он ее назвал UNL через какое-то время после того как он начал это разрабатывать, и мы используем UNL как интерлингву для машинного перевода. Мы ее применяем с разной степенью успеха. А другое, что есть, — мы впервые занялись таким проектом, где есть какие-то ресурсы, которые мы не готовы считать лингвистическими. Этот проект можно было бы назвать чем-то вроде *Interfacing the lexicon in the ontology and semantic analyse*, что бы под этой онтологией не понимать. Я приведу два примера и на этом закончу, потому что это не доклад, а отчет о том, что в этой области делается.

Вот мы хорошо умеем понимать, являются ли какие-то предложения перифразами друг друга, передают ли они одно и то же значение. Вот несколько фраз:

1. Real Madrid and Barcelona will meet in the semi-finals on Thursday.
2. The semi-final match between Real Madrid and Barcelona will take place on Thursday.
3. The adversary of Real Madrid in the semi-finals on Thursday will be Barcelona.

Эти предложения значат одно и то же, это можно доказать или опровергнуть с помощью аппарата лексических функций. Всё это очень хорошо. Представим, однако, что мы еще хотим что-то такое, чтобы можно было синонимию более широкого порядка обнаружить. Вот я прочитаю еще одно предложение:

4. The semi-finals on Thursday will see the champion of the UEFA championship 2008—2009 and the team of Manuel Pelegrini.

Ну, может быть, кто-то и знает, кто такой Мануэль Пелегрини и кто был чемпионом УЕФА в 2008—2009 году, и этот человек сможет эту синонимию определить. Для этого нужна база реальных человеческих знаний, которую, собственно, можно разработать. И мы хотим попробовать эту вещь сделать. Не думаю, что стоит вдаваться в подробности этого проекта, но если нам удастся такие вещи сделать, то это будет хорошо. Это просто реальная картина мира, знания, и такое можно сделать с помощью базы знаний, хотя сделать интерфейс между лингвистической семантикой и такой базой знаний, мне кажется, не очень просто. Но там есть такие challenging задачи, которые кажутся весьма лингвистическими и которые все равно нельзя решить, не выходя за пределы чистой лингвистики. Представьте себе такое предложение:

Столичные «динамовцы» принимают сегодня своих одноклубников из Харькова.

Что это может значить? Наверно, из толкования слова «одноклубники» можно вывести, что одноклубник — это namesake по-английски, который играет в команде, которая так же называется. Можно понять, что «динамовец» — это член команды «Динамо». А что такое «столичные»? Мы знаем, что «столичный» — это относящийся к столице страны, а что за страна? Если мы это прочитали в русской газете, может быть, это Москва. Вряд ли. Мы должны проверкой посмотреть, что динамовцы — из Харькова, Харьков на Украине, и, наверно, эта столица — вовсе не Москва, а Киев. Так вот, для того чтобы это выяснить, нужна эта нетривиальная база знаний. И мы хотим этим воспользоваться. Вот, мне кажется, выход за пределы чистой лингвистики таит в себе перспективы.

В.Б. БОРЩЁВ (ВИНИТИ РАН): На этой конференции больше говорили про лингвистические онтологии и про прикладные задачи, в которых они применяются. Меня больше интересуют онтологии естественного языка, языковые онтологии. Ну онтология — это то, как устроен мир, а онтология естественного языка — это как язык смотрит на мир. То есть это наивная картина мира, что ли, в некотором смысле, и описание объектов в нашей наивной картине мира. Все-таки наивная картина мира в разных языках разная, и такого рода онтологии были бы, наверно, разные.

Ну вообще слово «онтология» в лингвистику, по-моему, впервые принес немецкий логик и философ Герберт Линк, я не помню, когда точно, кажется, это было в 80-е годы. Он работал и работает в области формальной семантики, но, как известно, в формальной семантике Монтегю даже множественное число не описывается. Ну вот Линк предложил описывать множественные объекты, причем он показал достаточно удачно, что множественные объекты, такие счетные объекты, типа «мальчики и столы», и так называемые mass-nouns — субстанции — их можно описывать сходными методами. И он это действительно продемонстрировал и

описал это в своем учебнике по формальной семантике, который вышел вначале по-немецки, а потом по-английски. Для него онтология была не для прикладных работ. Я не хочу ругать прикладные работы, они обязательно нужны, и, может быть, онтологии для прикладных исследований — это немножко другие онтологии, не знаю, но мне интересны именно онтологии естественного языка и вопросы, которые там возникают.

Хотя Линк ввел понятие «онтология» в лингвистику, но по сути дела онтологией в этом смысле всегда занимались лексикографы. Потому что лексикографы классифицировали не только лексику, но и то, как язык смотрит на мир. И на самом деле онтология естественного языка должна быть очень близкой к словарю, это не мешает ей быть формальным объектом (также как и формальная семантика — такая формальная наука), но следовать именно логике естественного языка.

Мы с Барбарой Парти последние лет 10—12 занимались этим немножко на таком понятии как контейнер-субстанция, и для меня не очень ясно до сих пор, не происходит ли тут размножение сущностей между словарем и онтологиями. Но в это я не буду углубляться.

Е.Г. СОКОЛОВА (РГГУ): Ну в отличие от ученых, которые долгое время занимались онтологиями глубоко, я человек свежий в этой области, и я хочу высказать свое наивное представление об онтологиях, с которыми я в какой-то степени сталкивалась. Первое, о чем хотелось бы сказать, — это противопоставление «лингвистические и нелингвистические». Я, честно говоря, не очень хорошо понимаю, что такое формальные онтологии, наверно, я просто этого не знаю.

Нелингвистические онтологии называются также domain model. Это онтологии, которые описывают узкие предметные области и для которых это реально, потому что они обозначают их терминами обычно, термин — это однозначная сущность, как правило, практически однозначная, и поэтому можно построить достаточно строгую сеть или иерархию этих понятий. Кроме того, они обозримы.

Лингвистические онтологии, если так можно говорить, описывают другие понятия — не понятия предметной области, а понятия языка. Какими понятиями пользуется язык? Это значения — лексические и грамматические. Поэтому есть два вида лингвистических онтологий — те, которые описывают лексику языков, и те, которые описывают грамматику языков. Лексика — примеры WordNet и FrameNet, а грамматические ... описывают грамматические значения, которые используются языками, — исчисляемость и неисчисляемость, переходность и непереходность и так далее.

С точки зрения лексических онтологий, по-видимому, есть две традиции — лексикографическая и, скажем так, онтологическая семантика, может быть, Виктора Раскина. Онтологическая семантика — это теория, в которой понятия выстраиваются в соответствии с этой теорией. Лексикографическая традиция —

это традиционное описание языка, в сущности. К этой традиции относится, например, WordNet. Я считаю, что создатели WordNet'a сделали такую интересную вещь — они взяли обычные толкования, толкования для человека неформализованные, самые обычные, и выстроили эти единицы, эти слова, собрали их в группы-синсеты и выстроили между ними вполне формальные отношения — лексико-семантические отношения, типа синонимы, омонимы, квазисинонимы, родо-видовые и так далее. Таким образом получился ресурс, в котором нет формальных толкований, толкования естественные, но зато той стороной, которая представлена отношениями между этими единицами, может пользоваться как человек, так и машина. Их можно использовать прикладными системами, что и делается. Но с другой стороны, я сама не пыталась их использовать как прикладную систему, но читала, что есть мнение людей, которые пытались использовать, скажем, WordNet для информационного поиска, где он теоретически должен был хорошо помогать, потому что эти иерархические синонимические отношения существенно расширяют запрос. Это должно бы давать эффект. А вот вроде бы они отзываются так, что эффекта это дает мало. Это может быть связано с тем, что отношениями связаны не те понятия, которые в тексте взаимозаменяют друг друга. То, как мы осознаем лексику языка, и то, как она используется нами в речи, — это, по-видимому, разные вещи.

Второе, что я хотела сказать, — в докладе Наташи [Лукашевич] говорилось о квазисинонимах. Квазисинонимы и родо-видовые отношения между словами отличаются различительным признаком. Если у нас есть верхние понятия и гипонимы, то гипонимы отличаются от верхнего каким-то признаком. Верхнее — дерево, внизу — сосна, береза и так далее. Сосна — хвойное, береза — лиственное дерево и так далее. Это Наташа подробно рассказывала. С другой стороны, квазисинонимы отличаются тоже различительным признаком. Если башмак и ботинок — это квазисинонимы, они тоже различаются каким-то признаком, стилем или не знаю чем. Мне интересно, какая разница между теми различительными признаками, которые различают по вертикали, и теми различительными признаками, которые различают по горизонтали. Я просто об этом не знаю, может, кто-то из присутствующих знает.

Третье, что я хотела сказать, — такое впечатление, что, говоря об онтологиях и о текстах, в частности прикладывая онтологии к анализу текстов, мы ищем в текстах значения, которые у нас уже описаны, лексические значения. Мне кажется, что это неправильно. Я уж не буду ссылаться на классиков марксизма-ленинизма, которые говорили, что само название — это уже обобщение. На самом деле так оно и есть. В тексте мы обозначаем не значения языковые, текст состоит из понятий, которые обозначены словами естественного языка. И значения — это некоторый инструмент, который позволяет нам это сделать. А вот что мы обозначаем, оно от этого значения отличается. И никто на самом деле не изучал, что это за значение, что это за понятие, которое находится в тексте. Скажем, мы считаем, что озеро — это водоем, у нас в онтологии написано: озеро — это водоем. Водоем — это, скажем,местилище воды. Вот мы должны исходить из этого понятия. Когда мы говорим о водоеме в тексте, мы говорим, что вдалеке

виднелось озеро. Что такое озеро — это точка, мы говорим о ней как о точечном объекте. Никакого там вместилища воды, когда мы об этом говорим, нет. Если мы видим это озеро перед нами и видим в этом озере отражение дома, который стоит рядом с нами на берегу, то это уже другое понятие. Для нас это зеркало, и мы говорим о нем словами как о зеркале. И нельзя искать в тексте то понятие, которое мы описали в онтологии. Этим действительно никто не занимался, и это очень интересно — те условия, тот кусок опыта, в котором у нас это слово зафиксировалось в нашей языковой системе.

Н.В. ЛУКАШЕВИЧ: Поскольку основной набор докладчиков закончился, а были вопросы к моему докладу, я попытаюсь ответить насчет стилистических различий. Мне кажется, мы используем онтологию, чтобы моделировать, описывать внешний мир. На мой взгляд, когда идет различие именно в стиле, то есть где, в какой речи используется то или иное слово, в речи каких людей, но при этом ссылаются на одни и те же объекты — можно сказать «коммунальная квартира» в каком-нибудь законе, «коммуналка» в нашей обыкновенной речи, мы все-таки ссылаемся на один и тот же объект, и кажется, что в данном случае это не ведет, по крайней мере для меня, к возникновению отдельного понятия. То же самое, когда мы «башмак» в каком-то просторечии употребляем именно как ссылку на «ботинки» и знаем, что имеем в виду то же самое, даже, может быть, в таком свободном разговоре, это, на мой взгляд, не должно вести к расширению сущности, потому что здесь нет никакого дополнительного явления мира, и понятно, что есть множество промежуточных, может быть, ситуаций, когда трудно различить, является ли то или иное название только формой, именем одной и той же сущности, или для обозначения этого нужно ввести отдельный концепт в онтологии. В частности, некоторые люди считают, что нужно для уменьшительных названий, если рассматривать серию WordNet как какой-то вид онтологических ресурсов, то разработчики русского WordNet'a считают, что для отражения уменьшительных названий нужно ввести отдельную единицу в сеть синсетов. Тем не менее может быть другой взгляд, что в случае с уменьшительными наименованиями мы все-таки ссылаемся на те же самые объекты внешнего мира, но некоторым образом они могут быть большие, но мы таким образом выражаем свое отношение к ним, тем не менее никаких новых сущностей здесь не производится. Может быть, могут быть другие точки зрения на этот вопрос.

В.П. СЕЛЕГЕЙ (АВВУУ): Я никогда не занимался онтологиями профессионально. Зато у меня есть скорбный опыт работы в Редсовете «Диалога» в связи с рецензированием докладов про онтологии. Есть такой фундаментальный вопрос — что такое хорошо, и что такое плохо. Вот у нас вчера было такое замечательное обсуждение вечером, является лингвистика наукой или нет, с участием Максима Кронгауза, и он нас почти убедил, что лингвистика является наукой практически такой же точной, как физика, в ней есть определенные процедуры, верификации которые позволяют сказать, является ли предлагаемое

описание хорошим или плохим. Если предположить, что описание некоторой языковой единицы предсказывает поведение этой единицы в текстах, то, вероятно, можно сравнивать два описания и установить, что одно делает это лучше, а другое -- хуже. И худо-бедно лингвистика при всех своих недостатках, парадигмальных различиях, спорах, тем не менее породила много верифицируемых результатов. И когда вы, например, посылаете на конференцию статью по лексической семантике, у рецензента есть способы доказать, хорошо это или плохо. Примерно такая же ситуация в области формальных онтологий. Мой опыт общения со специалистами по формальным онтологиям подсказывает, что такие же верифицирующие процедуры существуют и здесь. То есть если ваше онтологическое описание предсказывает поведение уже не слова, а моделируемого объекта, значит, ваша онтология хороша. Возвращаясь к началу, мне кажется, что «лингвистическая онтология» — это оксюморон. В общем, это объект, который подлежит какому-то иному, не вполне понятному способу оценки.

Мы на прошлом «Диалоге» немножко говорили о лингвистических ресурсах, об их типизации, о том, что есть ресурсы, так сказать, иллюстрирующие теории, — у них свои способы оценки, а есть ресурсы, ориентированные на практические задачи. Лингвистическая онтология — это, видимо, ресурс, ориентированный не на теорию, а на определенного типа задачи. Единственный способ сказать, хороша она или нет, — это посмотреть на результаты ее применения к конкретной проблеме. В этом смысле мы должны были бы сегодня разойтись и подождать десять лет, пока онтология, сделанная группой Виктора Раскина, будет применена. И тогда мы сможем увидеть, работает она или нет. Я полностью разделяю точку зрения на необходимость разделения лингвистической и онтологической части, как это принято в этом проекте, но, к сожалению, меня не убедила в сегодняшнем докладе ни лингвистическая, ни онтологическая части. Это снова выглядит как ориентированная на какие-то задачи «лингвистическая онтология». Похоже, что и у рецензентов тоже возникли такие проблемы — они не могли ни по чисто лингвистической, ни по чисто онтологической части полностью согласиться с предложенными описаниями, а поскольку они не видели убедительного результата работы системы на конкретной задаче, то осталось очень много сомнений. И это при том, что такое разделение ясно проведено в этом проекте. А когда разделения и вовсе нет, то вообще никаких верифицирующих процедур не остается. То есть нам остается только ждать, будет успешен проект или нет. Если он будет успешен, нам будет интересно узнать, а как у них там всё внутри устроено.

Е.Н. ЗАРЕЦКАЯ (Академия народного хозяйства при Правительстве РФ): Скажи, пожалуйста, а какой результат, которого ты ждешь? Я вот всё слушаю, слушаю — никак не могу понять, что такое практический результат теоретических разработок в области, скажем, формальных онтологий. И чем он отличается от результата семантического описания текста. Вот что это за два разных результата? Если мы разводим семантическое описание и онтологическое описание, да еще с

формальной точки зрения хотим получать результаты, то что это за такой феномен — результат формальной онтологии?

В.П. СЕЛЕГЕЙ: Моя идея состоит в том, что можно разрабатывать лингвистические ресурсы и онтологические ресурсы, каждый с точки зрения определенных научных принципов, в рамках которых они разрабатываются. Вот у нас были доклады, я не помню, приняты были они или нет, — онтологическое описание, соответствующее, допустим, законам классической физики. Если это онтологическое описание позволяет правильно решать определенные типы физических задач, оно корректно, и это можно проверить. То же самое относится к лингвистическим ресурсам, и к связкам лингвистических и онтологических описаний.

Е.Н. ЗАРЕЦКАЯ: А языковое описание этого феномена какую имеет разницу в отношении онтологического описания? Если ты все время говоришь о строгом формальном результате, то как он должен выглядеть в отличие от описания того же самого феномена средствами естественного языка? Это всё равно знаковые системы будут, в любом случае это будет как-то называться. Это знаки явления. Как можно описать явления внешнего мира вне любой знаковой системы? Если это явление континуально, а наша память ограничена, это должна быть дискретная система описания. А если это дискретная система, то есть это семиотическая будет система, то вообще-то дальше это можно называть будет по-разному. Можно картинками, можно средствами естественного языка. Я хочу понять, где такой особый, почти технический феномен реализации онтологического описания. Что мы должны получить на выходе, чтобы ты сказал, что это хорошее описание?

В.П. СЕЛЕГЕЙ: Я не являюсь специалистом по формальным онтологиям, но, те, кто занимается, допустим, темпоральными или пространственными или еще какими-то логиками, имеют методы оценки адекватности тех моделей, которые они создают. Проекция темпоральной или пространственной системы естественного языка на такую формальную экстралингвистическую систему — это сложная задача, которой, безусловно, нужно заниматься. Тут есть правильная постановка вопроса. А мы сплошь и рядом имеем дело с такими ресурсами, в которых все слито в одно. И спорим, какая должна быть структура, сколько должно быть значений, объединять значения или соединять значения?

Е.Н. ЗАРЕЦКАЯ: Это всё частное.

В.П. СЕЛЕГЕЙ: Нет, это конкретные вопросы, возникающие у разработчика.

Е.Н. ЗАРЕЦКАЯ: Я хочу получить ответ на простой обывательский вопрос, сочтите мой вопрос за обывательский, хотя он не очень обывательский. Мы чего хотим получить на выходе, чтобы сказать: это хорошее онтологическое описание и чем оно будет отличаться от хорошего семантического описания?

Н.В. ЛУКАШЕВИЧ: На самом деле все-таки здесь, мне кажется, ответ идет в две стороны. Конечно, один вариант — это всё-таки приложение. То есть как бы мы красиво не создали некую систему, и, может быть, кому-то она нравится, все-таки, если мы имеем в уме, что мы делаем для прикладной области, что мы вкладываем усилия, чтобы это было применено, значит, мы должны доказать и показать приложение.

Е.Н. ЗАРЕЦКАЯ: Пример.

Н.В. ЛУКАШЕВИЧ: Мы, например, работаем с задачами информационного поиска. Например, кластеризация новостных текстов. И вот мы пробежались по разным методам кластеризации и поняли, что ряд методов работают, если мы сравниваем их по словам, приблизительно одинаково. И замерили некоторым образом полноту, точность, насколько хорошо, понравится, как наша, допустим, автоматическая система кластеризует поток новостей. Пока нет онтологии, пока мы работаем со словами. При этом мы понимаем, что когда идет поток текстов, очень много синонимов, очень много видовых, родовых понятий. И, естественно, если мы хотим, чтобы обработка была более глубокой, чтобы мы узнавали, как-то могли это использовать, могли увидеть синонимы и понять, что это одно и то же, поэтому эти тексты надо объединять в единый кластер, они на одну тему. И мы специально, например, исследовали фактор того, что поможет ли нам лучше кластеризовать эти новостные тексты, если мы будем применять не только слова, но и специальный индекс, который специально делается для текста, что называется, концептуальный индекс, когда мы синонимы собрали, а различные значения слов по мере нашей возможности... Это реальный эксперимент по современному потоку новостных текстов. Выяснилось, что с помощью добавления дополнительной информации — синонимов, разбора значений — получаем дополнительные 7% качества. В таких информационно-поисковых приложениях очень много завязано на слова. Но если мы добавим еще информацию — синоним, значение, то на такой задаче как кластеризация мы получаем дополнительное качество. Много это или мало, но если мы никаким другим способом, пробежав, допустим, пять разных методов кластеризации, не смогли получить лучшего результата, как добавив синонимы и значения, то в какой-то

мере наши усилия дали нам какой-то дополнительный бонус. Таким образом, речь идет о том, как, вкладывая дополнительное знание, упорядочивая его, интегрируя с поверхностной обработкой, — потому что что бы мы не описали, мы всегда отстаем от мира, и мы многое должны делать на поверхностном, скажем, уровне, — но создание дополнительных ресурсов такого онтологического плана дает нам дополнительные возможности по тому, чтобы углублять как бы обработку. Мы много усилий тратим на то, чтобы мерить, делать внутренний evaluation, чтобы понять, в какой мы точке находимся и чтобы видеть, насколько мы сумели в этом сложном современном потоке текстов новостей получать дополнительное преимущество.

Е.Н. ЗАРЕЦКАЯ: Какая разница между тем, что 40 лет назад высказал Мельчук, сказав, что значение — это инвариант синонимических преобразований, назвав это вообще семантикой, вот какая разница между тем, что вы сейчас сказали, и тем, что он сказал 40 лет назад? И при чем здесь онтология? Это, может быть, совсем не к вам вопрос. Потому что как вы ответили, в рамках реализации синонимических рядов и так далее, — это не онтология, это семантика. Это, простите, языковое выражение некоторого знания разными способами. Почему онтология?

VIKTOR RASKIN: Я хотел бы коротко ответить Володе Селегею и Лене [Зарецкой] тоже. То, что мы делаем, мы не хотели, чтобы ждали десять лет результат, мы хотели, чтобы вы нам помогали, боролись против нас, делая, что нужно делать. Научное сообщение имеет такую цель. Володя признает, что формальные онтологии имеют критерий успеха. У нас есть формальная онтология, мы полностью подчиняемся этим критериям успеха. Володя признает, что лексикографическое описание тоже можно судить, хорошее оно или плохое, — я думаю, что это грубо, умея это делать лучше, чем многие другие. Мы делаем лексикографическое описание — не только значения слов, но и значения предложений. Мы делаем это путем привлечения того, о чем Леня [Иомдин] говорил, что является точной онтологией — он занимается онтологией, он делает ровно то, что нужно делать, потому что, как он совершенно справедливо, грамотно и красиво сказал, онтологии покрываются знаниями, которые лингвистические, а мы их используем для понимания каждого предложения, потому что каждое предложение значит гораздо больше, из него выводится гораздо больше, чем слова, которые в нем употреблены. Помимо этого мы еще и применяем это к приложению. Таким образом мы подчиняемся всем критериям оценки.

В.П. СЕЛЕГЕЙ: Я могу пояснить. Было сказано в докладе, что вас не интересует сегодня evaluation. Избегаете, например, соревнований методов разрешения лексической неоднозначности.

VIKTOR RASKIN: Потому что они к нам неприменимы. Потому что мы совершенно не этим занимаемся. У нас результат — вот они делают какую-то частичную вещь и могут эту часть измерять. Нам это совершенно неинтересно.

В.П. СЕЛЕГЕЙ: Но тогда с чем вы оставляете рецензента? Вы делаете лингвистическую часть, описание которой, я уверен, у каждого сидящего в зале лингвиста будут возражения против каждого ее...

VIKTOR RASKIN: А я думаю, что нет. Потому что мы вам даем пример disambiguation, который сделан при помощи наших ресурсов хорошо, и вы умеете судить, что это было сделано хорошо. Или вы будете с нами несогласны.

В.П. СЕЛЕГЕЙ: Нет, это сделано и на бумаге...

VIKTOR RASKIN: Нет, это сделано компьютером.

В.П. СЕЛЕГЕЙ: Ну тогда нужно показать, как проводился evaluation, который вы отрицаете, это и есть попытка получить все-таки объективный результат. Как делалась тестовая выборка. Как считалось. Там не видно методики.

VIKTOR RASKIN: Я могу поделиться с вами этим результатом, Володя. Понимаете, может быть, мы сделали это плохо. Может быть, у вас не было рецензентов, которых это интересует. Это совершенно внешний факт по отношению к «Диалогу», который его нисколько не компрометирует, внешний факт по отношению к нам. Я часто думаю о том, что natural language processing без лингвистики и без семантики — это не наша область, нам надо перестать прикидываться, что мы этим занимаемся. Я организовываю новый журнал, который будет заниматься вычислительной семантикой, Йорик [Уилкс] и все другие великие люди в computational semantics, а on the board могут быть 13-14 групп, которые этим будут заниматься... И, может быть, это хорошо, и, собственно, и конференции, где будут нас понимать. Если б вы меня спросили, кому послать на рецензию наш доклад, мы были бы очень впечатлены людьми, которые это бы сделали, они нас ругали. Джон Сова, например, обругал нашу с Сережей [Ниренбургом] книгу «Computational linguistics» — дай Бог каждому такую рецензию. Он говорил: «Это важная книга, с которой я не согласен». Бога ради! Так что я не принимаю ваших аргументов по поводу невозможности оценивать

нашу работу. Мы дали вам все для этого возможности — у нас не получилось, у вас не получилось — не имеет никакого значения. Мы вас любим, пожалуйста, любите нас тоже.

Теперь относительно вопроса Лены. В некотором смысле то, что я уже сказал, отвечает на ваш вопрос.

Е.Н. ЗАРЕЦКАЯ: У меня как раз никаких претензий к докладу.

VIKTOR RASKIN: Я хочу сказать, что Леня [Иомдин] на самом деле дал вам ответ на вопрос. Я уже об этом говорил, коротко повторю. Он говорил, что для того чтобы понимать значение каждого предложения, и человеку, и машине, которая должна имитировать эту деятельность человека, нужно пользоваться информацией, которой нет в языке. И для этого вам нужна какая-то база данных, knowledge base, база знаний, которую надо описать при помощи какого-то инструмента. Это и есть онтология. Помимо этого мы хотим, и делаем это, на этом же языке, knowledge base, описывать значения слов и значения предложений, таким образом ставя их на четкую, математически формальную основу.

Е.Н. ЗАРЕЦКАЯ: У меня последняя реплика. Если онтология — это возможность через анализ текстов на естественном языке выйти за затекстовое знание, то можем ли мы утверждать, что этот вывод, который мы сделали относительно явлений неязыковых, будет гипотетическим, неточным, альтернативным?

VIKTOR RASKIN: Ну понимаете, давайте вспомним сэра Карла Поппера, который был последний великий философ, Поппер говорил — нету теории, есть только гипотезы, которые еще не опровергнуты. А другая проблема, о которой вы говорите, и Володя в частности этого касался, я максимально серьезно отношусь к тому, что он сказал, — у нас есть проблема убеждать людей, которые должны нас понимать, что мы делаем что-то, что имеет достаточное качество, и не очень хорошо это делаем. Вы знаете, что есть философия науки, одна из проблем этой науки — верификация теории, и некоторые теории удачны, их можно верифицировать путем опыта: берешь яблоко, сажаешь Ньютона, оно ему падает на голову с той силой, о которой он говорит. Кстати, это не так, оно никогда не совпадает. И есть области, в частности гуманитарные, в которых всё остается гипотезой. Есть теории, когда мы соглашаемся, никаких споров нет, а оказывается — ерунда. Так что это серьезно. Есть современный философ Сюзен Хаген, которая разработала гибридный метод верификации, где можно обосновать это на эмпирике, а соединить это всё так, чтобы это было coherent, имело смысл. Мы, наверно, не сумеем никогда думать лучше и иметь теории. Спасибо.

Н.В. ЛУКАШЕВИЧ: Я хотела как раз вспомнить такую историю из конференции по искусственному интеллекту в России, когда, если сейчас удалось получить вопрос различия информации, хранимой в онтологии и в семантических теориях, то как раз людям, занимающимся искусственным интеллектом, было сложнее, потому что они считали, что достаточно много терминов было до появления слова «онтология», то есть были «модели», «предметные области», «концептуальные модели». И где-то, скажем, в 2000 году в российской конференции по искусственному интеллекту была достаточно острая дискуссия, нужно ли пускать в российский искусственный интеллект такое слово, как «онтология», потому что считалось, что все слова, которые были нужны, они и так уже были, типа понятийные модели, предметные области. Тем не менее «онтология» прижилась и живет, и, мне кажется, с этой точки зрения, действительно, отличить сложнее, потому что «модели», «понятийные модели», «предметные области» — мы все-таки описываем знание о мире, и в онтологиях тоже описываем знание о мире, но все-таки когда появилась тема shared ontologies, разделяемых онтологий, когда люди стали их сливать, разливать, как некоторый такой объект, — наверное, это некое новое качество таких моделей описаний мира, которое, на мой взгляд, привело к появлению нового слова для описания этого. То есть онтология — уже как тот объект, которым можно делиться, брать, и это немножко другое, чем просто модель, предметная область, которая твоя личная. Вот такой тоже был исторический момент.

А.А. БОНЧ-ОСМОЛОВСКАЯ (МГУ им. М.В. Ломоносова): Я хотела бы продолжить мысль Елены Соколовой, что слова в тексте оказываются каким-то образом функционально нагружены, и оказывается, что это не совсем имеет отношение к их значению, как бы они ни были представлены — в виде словаря или в виде некой семантической сети, более они ориентированы на лексику и словари, или более формализованы. На самом деле даже из всех этих ваших примеров про зеркало и про точку — понятно, что это не ниоткуда возникает, а это все-таки какие-то фреймы, которые каким-то образом навешиваются. Наверно, из глагола, потому что «отражается», наверно, из ситуации. И, может быть, у нас есть такой тип ресурса, как FrameNet, где каким-то образом делается попытка как-то это всё формализовать и выстроить, хотя, на мой взгляд, не очень внятная. Суть в том, что онтология — это, оказывается, вещь статическая по отношению к разным регистрам функционирования текстов, поэтому не очень понятно, как к этому может подключаться язык, как это может учитываться. Я приведу еще один пример — при анализе лексики в негативных новостях, которые оцениваются как негативные, самое значимое слово — это слово «происходить». По-человечески это даже понятно. Нигде — ни в онтологиях, ни в словарях про это записать нельзя, на самом деле. Это понятно, когда мы это узнали, а заранее мы это не предугадаем. Наверно, когда-нибудь мы придем к чему-то третьему, «статическая—динамическая» — это не очень понятная оппозиция, но какая-то такая.

Е.Г. СОКОЛОВА: Я так отвечу. В онтологиях мы пытаемся записать такое философское значение, мы пытаемся толковать, классифицировать мир. А в текстах возникают понятия, с которыми мы соприкасаемся в каждодневной жизни. Эти понятия мы воспринимаем с каким-то куском опыта. Например, как пройти в библиотеку? А вот походи прямо, поверни направо. Мы же не объясняем, что там улица какая-то пересекается с какой-то... Как надо действовать. То же самое у нас происходит с очень многими понятиями. С тем же самым озером — мы его воспринимаем совершенно не так, как оно описано в энциклопедии, а так, как мы его чувствуем. Видим вдали маленький точечный объект. То есть мы должны каким-то образом, чтобы находить эти понятия в тексте, научиться записывать свои ракурсы или куски опыта, которые связаны у нас с этими словами, с этими понятиями.

А.А. БОНЧ-ОСМОЛОВСКАЯ: Вот я считала, что эти ракурсы как раз...

Е.Г. СОКОЛОВА: Это фреймы. Там ничего не навязывается глаголом.

А.А. БОНЧ-ОСМОЛОВСКАЯ: Я не говорю, что это навязывается глаголом, глагол используется для обозначения ситуации. Ситуация выражена в первую очередь глаголом.

Е.Г. СОКОЛОВА: Я хочу сказать про наш опыт. Мы занимаемся описанием картинок, и вот на картинках мы ни одного глагола не видим. Мы видим дома, мы видим деревья, мы видим людей.

А.А. БОНЧ-ОСМОЛОВСКАЯ: Это статическое. А если вы мультики описываете?

И.А. БОЛЬШАКОВ: А если люди лежат?

Е.Г. СОКОЛОВА: Люди лежат, но мы все равно не видим «лежат», мы видим человека в определенной плоскости.

И.А. БОЛЬШАКОВ: Человек бежит на картинке.

Е.Г. СОКОЛОВА: Человек бежит на картинке — мы видим человека в определенной позе или человека, совершающего определенное движение.

С.А. КРЫЛОВ (Ин-т востоковедения РАН; Ин-т системного анализа РАН): Просто человека, у которого одна нога поднята над землей, а другая касается земли.

Е.Г. СОКОЛОВА: Мы не видим глаголов, глаголы мы вносим. Глагол в языке — это действительно обязательное средство для формирования предложений. Язык так устроен, и мы вынуждены этот глагол везде вводить. Но совершенно необязательно он что-то отражает. Если мы говорим, что дом отражается в озере, то только потому, что мы дом воспринимаем как зеркало, а не потому что там есть какой-то процесс.

РЕПЛИКА ИЗ ЗАПА: Из трубы идет дымок.

Е.Г. СОКОЛОВА: Из трубы идет дымок — ну вот «идет» что такое? Пустое, ничего. А дым — это вещественная вещь, мы ее видим. Мы описываем картинки, но этот глагол — мы его вносим.

И.М. КОБОЗЕВА (МГУ им. М.В. Ломоносова): Что нам мешает онтологически описать это свойство озера, помимо энциклопедических сведений о том, что это разновидность водоемов, мы прекрасно записываем свойство, что оно имеет часть, поверхность, поверхность имеет свойство зеркальное, отражать.

Е.Г. СОКОЛОВА: А в следующий раз ты в нем искупался.

И.М. КОБОЗЕВА: А неважно, если это свойство не будет задействовано контекстом. Надо описать все свойства, которыми эта сущность поворачивается к нам в жизни.

Е.Г. СОКОЛОВА: Я как раз призываю к тому, что всё это надо внести.

Н.В. ЛУКАШЕВИЧ: Может быть, тогда мы будем завершать. Мне показалось, что Ирина Михайловна на такой оптимистической ноте подвела черту, что многие наши знания о мире можно описать, и некие более применимые в разных контекстах знания, и, может быть, то, что интересно в каком-то конкретном контексте, в принципе можно описать в таких ресурсах, как онтологии. Но с другой стороны, чем больше свойств, чем больше разнообразия мы свяжем с той или иной конкретной сущностью, тем сложнее нам описать с той же подробностью любое количество сущностей, которое нам нужно по крайней мере для реальных приложений, скажем, автоматической обработки текстов, и всё время эта дихотомия, что, с одной стороны, как бы сделать достаточно подробное описание, с другой стороны, чтобы это не ограничилось слишком небольшим количеством элементов, которое не даст эффекта, — всё время, мне кажется, исследователь всегда борется с этой проблемой — как сделать так, чтобы, с одной стороны, описание было достаточно интересным, давало достаточно много знаний о мире, с другой стороны, чтобы это было масштабируемое описание, и всегда, мне кажется, каждый разработчик онтологии пытается найти свой путь соотношения подробности описания и насколько он это описание может масштабировать. Это проблема, и никуда от нее не деться.