ONTOLOGY AND INTEGRATION OF FORMAL AND LEXICAL SEMANTICS

Vladimir Borschev & Barbara H. Partee

University of Massachusetts, Amherst

Dialog 2014

June 4-8, 2014, Bekasovo, Russia

Acknowledgements

- Thanks to Elena Paducheva, Ekaterina Rakhilina, Louise McNally, Susan Rothstein, Fred Landman, Per Anker Jensen, Carl Vikner, Georgy Moroz, and Maria Kholodilova, and to audiences and participants at an Ontology Workshop in Copenhagen, 2006, the Conference on Concept Types and Frames in Language, Cognition, and Science in Düsseldorf in 2007, in Paris and Nancy in 2010, and at workshops at Bar Ilan University in 2011 and 2014.
- This material is based upon work supported in part by the National Science Foundation under Grants No. BCS-9905748 and BCS-0418311 to B.H. Partee and V. Borschev for projects on the Integration of Lexical and Compositional Semantics.

1. Introduction: Thesis, background, examples

Our main thesis: Formal and lexical semantics can be integrated if they speak the same language. A substantial part of lexical semantics can be incorporated into formal semantics without adding to the latter any new mechanisms.

- This talk continues the authors' work on the ontology and the semantics of measure constructions in Russian.
- The work concerns expressions like *dva stakana moloka* 'two glasses of milk', *polkorziny gribov* 'half a basket of mushrooms', *tri meshka muki* 'three bags of flour', etc., describing various kinds of *containers*, or corresponding measures based on them, which we will call *container-measures*, and their contents—*portions of substances*.

Introduction, *cont'd*. Background.

- In our previous works, describing ontological information, including *sorts* of things and the words and expressions that designate sorts, we did not include those sorts in our formal semantic analyses.
- We do that in the present work, declaring *sorts* as *types* and thereby significantly expanding Montague's system of types.
- On the one hand this gives us the means for specifying various aspects of the ontology, and on the other hand it lets us more fully specify the semantics of the constructions under consideration.
- The substantive goals of this research are, in part, to be able to describe and explain co-occurrence constraints and ideally to be able to formally distinguish well- formed from ill-formed expressions in this domain.

Introduction, *cont'd*. Examples.

- **Examples. A fragment of ontology for the expression of measure.**
- (1) *On vypil dva stakana moloka.*
  He drank two glass-GEN.SG milk-GEN.SG
  He drank two glasses of milk.
- (2) *Voz'mite poltora stakana muki.*
  Take one-and-a-half glass-GEN.SG flour-GEN.SG
  Take one and a half glasses of flour.
- (3) *On prines polkorziny gribov.*
  He brought half-basket mushroom-GEN.PL
  He brought half a basket of mushrooms.
- (4) *dva puchka rediski*
  two bunch-GEN.SG radish-GEN.SG
  two bunches of radishes

Introduction: Examples, *cont'd*.

- Those examples contain the *genitive measure construction.* In the first three, the measure is constituted by *containers*—in this case glasses and baskets; these are *container-measures*.
- Jars, bags, boxes, etc., can also be used as measures. They can be *filled*—completely or to some *degree*—with various *substances*—milk, water, flour, mushrooms, etc., and so can serve as a measure of quantity of such substances.
- The contrasting example (4), *dva puchka rediski,* is a genitive measure construction but does not use a container-measure.
- Substances are of various sorts—*liquids, granular substances,* and others. In our examples we are concerned with the measuring of *portions of substances.*

Introduction: Examples, *cont'd*.

- We note that formal semanticists have discussed ontological and semantic commonalities in the description of plural entities and mass stuff in natural language; see Parsons 1970, Link 1983, Landman 2004.
- The normal quantity measure for finite pluralities of entities is their cardinality—the number of elements in the corresponding set, which is a whole number (*five boys*), and a normal measure for portions of sub-stances is their volume, measured in terms of certain standard portions (*two liters of milk* or *one and a half cups of flour*), and in this case we find fractional as well as whole numbers.

Introduction: Examples, *cont'd*.

- Having taken the quantity of a certain portion as a unit, that is, as a unit of measure, we can measure portions of substances in those units, for instance in liters or in the volume of particular glasses, baskets, and the like, determining how many liters (glasses, baskets) or parts thereof are contained in a given portion of substance.
- Therefore with each unit of measure there is a correlated *function*, defined on portions of substances. For example, corresponding to the *liter* unit we can assign the function LITER: if ***m*** is a portion of milk, then LITER(***m***) is a number giving the volume of that portion in liters (cf. Landman 2004).

Introduction, *cont'd*.

- What we have said so far could be called a 'dotted-line outline' of a fragment of ontology for expressions of container-measure.
- Ontology, within philosophy, is a branch of metaphysics that studies what there is and the nature of the basic categories of the things that make up the world.
- The task of **natural language metaphysics** (Bach 1986a) is to understand what presuppositions a language makes about how the world is structured, and **natural language ontology** is a part of that task.
- Ontology studies the various kinds of existents, and usually includes some classification of their sorts and types. *Glasses (cups), baskets, bags, containers, water, milk, flour, liquids, granular matter, substances*—these are examples of *sorts* of entities which we have considered in our previous works.

Introduction, *cont'd*.

- The semantics of expressions of natural languages rests on ontology. The examples mentioned so far are semantically well-formed, because they rest on an ontologically correct picture of the world.
- Thus in example (1), *On vypil dva stakana moloka*, milk is a liquid, a substance, it can fill a glass, and all glasses are containers. And since milk is a liquid, a portion of substance constituting two glasses of milk is something that can be drunk.
- In an analogical manner the well-formedness of expressions (2), *Voz'mite poltora stakana muki,* and (3), *On prines polkorziny gribov*, is based on ontology. And the bunch of radishes in example (4) is a natural 'portion' of radishes; one can measure radishes in bunches and count the bunches.

Introduction, *cont'd*.

- But examples (5–7) below are ontologically ill-formed, or at least doubtful.

(5) # *On vypil dva stakana muki.*
He drank two glass-GEN.SG flour-GEN.SG
He drank two glasses of flour.

(6) # *dva puchka moloka*
two bunch-GEN.SG milk-GEN.SG
two bunches of milk

(7) ??*On uronil s podnosa poltora stakana moloka.*
He dropped from tray 1½ glass-GEN.SG milk-GEN.SG
He dropped from the tray one and a half glasses of milk.

- *Dva stakana muki* 'two glasses of flour' is well-formed: flour, like other particulate matter, can be measured in glasses and one can count the corresponding portions. So two glasses of flour is a quantity of flour. But flour cannot (normally) be drunk; one can drink only liquids.

Introduction, *cont'd*.

- And example (6), *dva puchka moloka,* is ill-formed because portions of liquid aren't the kind of thing that can occur in bunches.
- Things are somewhat more complex with example (7), *On uronil s podnosa poltora stakana moloka*. *Poltora stakana moloka* is a well-formed expression denoting a portion of milk measuring one and a half glasses.
- But the verb *uronit'* ('drop') does not apply to portions of substance, but to solid objects. For a liquid an appropriate verb would be *prolit'* ('spill').
- And on a tray, we carry objects, not (directly) portions of liquid. And while portions of matter can be measured in fractional container-measures, solid objects are counted with whole numbers. *Poltora stakana moloka* is fine when it refers to milk, but unlike the ambiguous *dva stakana moloka* 'two glasses of milk', it can't easily refer to glasses. Hence one cannot drop one and a half glasses of milk from a tray.

Introduction, *cont'd*.

- Our task here is to formally describe a fragment of the ontology of natural language on which the semantics of measure expressions depends. We aim to do that by giving a semantics that assigns suitable meanings for semantically well-formed expressions and accounts for the anomaly of semantically ill-formed expressions.
- Our larger goal is to show that this can be done with the tools of formal semantics if we include in formal semantics at least a certain part of lexical semantics.
- The main idea will be to enrich the notion of *type* with *sortal* information.

2. Formal semantics

- Formal semantics of natural language is historically associated with the name of R. Montague. Montague showed that the syntax and semantics of natural language can be described using the tools developed by logicians for the formal description of their formal languages.
- These methods give a model-theoretic semantic interpretation of syntactic structures, obeying the principle of compositionality. The tools for such formal description have been greatly extended in the last forty years by the cooperative efforts of linguists, logicians, and philosophers of language.

Formal semantics, *cont'd.*

- Over the last forty-plus years formal semantics has become (especially in the West) the mainstream approach to semantic research.
- But especially in the beginnings, formal semantics by no means described the whole semantics of natural language. Montague did not try to describe lexical semantics, considering that a more empirical task. Montague's semantics can be reasonably characterized as *the semantics of syntax* (Paducheva's term).

Formal semantics, *cont'd.*

- Formal semanticists focus on compositionality, how the meaning of a sentence is built up from the meanings of its parts.
- This requires having some ideas about the meanings of the smallest parts—words and morphemes—because they form the starting point for semantic composition. So formal semantics needs some kind of lexical semantics to start from.
- The bare minimum is to make some assumptions about the nature of lexical meanings and not make any specific claims about any particular lexical meanings— that was Montague's strategy, since he had neither the interest nor the competence to address empirical matters of lexical semantics. He limited himself to trying to figure out the "semantic type" of various classes of lexical items, and the actual semantics for certain key 'logical words'.

Formal semantics, *cont'd.*

- Montague's framework uses two basic types: **e** and **t**, and every model includes two basic domains, $\mathbf{D}_e$ and $\mathbf{D}_t$, the set of all 'entities' of the given model and the set of truth values (normally1 and 0.) 'Entity' here is considered in the broadest possible way, including ordinary objects as well as numbers, colors, wars—anything a language has names for.
- (Semanticists have sometimes added additional basic types, for instance for events or situations, for moments or intervals of time, for degrees (used in the semantics of comparatives and other degree modification), for numbers.)
- Starting from the basic types, a hierarchy (tower) of functional types is constructed: <**e**,**t**>, <**e**,**e**>, <<**e**,**t**>, **t**> etc.
- The type <**a,b**> corresponds to the domain $\mathbf{D}_{<a,b>}$, the set of all functions *f* from domain $\mathbf{D}_a$ to domain $\mathbf{D}_b$.

Formal semantics, *cont'd.*

- With this hierarchy of types, Montague has a framework for an important part of the ontology of natural language, namely providing semantic values for expressions of all sorts of syntactic categories, including everything from nouns, verbs and adjectives to adverbs, declarative and interrogative sentences, embedded propositions, etc.
- And since the most basic way that expressions combine semantically is by function-argument application, this simple type structure characterizes which expressions can combine with which others. Within this relatively simple ontology he is thus able to capture the "semantics of syntax."

Formal semantics, *cont'd.*

- Since its beginnings around 1970, there has been a great deal of work in formal semantics, including work that brings formal semantics and lexical semantics together.
- We have already mentioned some of the work of Parsons, Link, and Landman. We also note the work of E. Bach (Bach 1986a, 1986b) on natural language metaphysics.
- There are many other works which have extended formal semantics by including more lexical semantics and making use of ontological specifications, including Dowty 1979, Kamp and Partee 1995, Pustejovsky 1995, and many more in recent years.

3. Incorporating sorts into formal semantics

- Building on our previous work, we refine our earlier semantics for measure constructions in Russian, adding ontological information.
- Technically this is accomplished by unifying the notions of *type* and *sort*: the sorts to which words and phrases belong become types and are added to the hierarchy of types.
- This radically increases the collection of types, and a significant part of lexical semantics immediately becomes part of formal semantics.
- So for the measure expressions which are considered in this work, we introduce new basic types for different kinds of containers and portions of substance: **glass**, **basket**, ..., **container**, and also **milk**, **water**, **flour**, ... , **liquid**, **granul_subst**, **pourable_subst**.

Incorporating sorts, *cont'd.*

- For these types we introduce the corresponding domains $\mathbf{D}_{glass}$, $\mathbf{D}_{basket}$, $\mathbf{D}_{container}$, and likewise $\mathbf{D}_{milk}$, $\mathbf{D}_{water}$, $\mathbf{D}_{flour}$, …, $\mathbf{D}_{liquid}$, $\mathbf{D}_{granul\_subst}$, $\mathbf{D}_{pourable\_subst}$, etc.

- The domain $\mathbf{D}_{glass}$ is the set of all glasses, $\mathbf{D}_{basket}$ is the set of all baskets, and $\mathbf{D}_{container}$ is the set of all containers, including all glasses, baskets, etc.
- So every glass is included in the domain corresponding to the type **glass**, every basket similarly corresponds to the type **basket**, and both glasses and baskets also correspond to the type **container**; $\mathbf{D}_{container} = \mathbf{D}_{glass} \cup \mathbf{D}_{basket} \cup \mathbf{D}_{bag} \cup \ldots$.

Incorporating sorts, *cont'd.*

- Formally all the domains we have just identified are subsets of the domain $\mathbf{D}_e$ and are picked out by characteristic functions from $\mathbf{D}_e$ to $\mathbf{D}_t$, the same kinds of functions that correspond to one-place predicates of entities.
- Thus the domain $\mathbf{D}_{glass}$ is formally defined by the predicate ***glass*** of type <e,t>, whose characteristic function from $\mathbf{D}_e$ to $\mathbf{D}_t$ yields the value **1** (true) for all glasses and **0** (false) for all other entities in $\mathbf{D}_e$.
- In some cognitive sense the opposite may be true: in our linguistic consciousness the domain $\mathbf{D}_e$ may well be an abstraction and generalization that is derived from more familiar subdomains.

Incorporating sorts, *cont'd.*

- The situation with substances is analogous. The domain $\mathbf{D}_{\boldsymbol{milk}}$, corresponding to the type **milk**, consists of all portions of milk; the corresponding characteristic function is ***milk***: $\mathbf{D}_{\boldsymbol{e}} \rightarrow \mathbf{D}_{\boldsymbol{t}}$, yielding the value **1** for all portions of milk.
- And to the type **liquid** there corresponds the domain $\mathbf{D}_{\boldsymbol{liquid}}$, consisting of all portions of liquid, and of course **D*milk*** $\subseteq$ $\mathbf{D}_{\boldsymbol{liquid}} \subseteq \mathbf{D}_{\boldsymbol{pourable\_subst}}$.
- In exactly the same way we have $\mathbf{D}_{\boldsymbol{flour}} \subseteq \mathbf{D}_{\boldsymbol{granul\_subst}} \subseteq \mathbf{D}_{\boldsymbol{pourable\_subst}}$, since both liquids and granular substances are pourable sub- stances. The domains $\mathbf{D}_{\boldsymbol{liquid}}$ and $\mathbf{D}_{\boldsymbol{granul\_subst}}$ do not intersect.
- These are all parts of the naive ontology that makes up part of the naive metaphysics of every language user.

Incorporating sorts, *cont'd.*

- Following Landman 2004 we also introduce type **r** as the type of real numbers, and the corresponding domain $\mathbf{D}_r$.
- From the "new" basic types, and the basic types **e** and **t**, we build the hierarchy of types. We will consider functions whose arguments and values fall within these domains.
- Now we can say that the function constant LITER belongs to the type <**pourable_subst, r**>, denoting a function in the domain $\mathbf{D}_{<pourable\_subst,\ r>}$, i.e., $\mathbf{D}_{pourable\_subst} \rightarrow \mathbf{D}_r$.
- Given the natural partial order among sorts of substances, the function LITER is defined not only on for arguments in the domain $\mathbf{D}_{pourable\_subst}$, but also on the domains $\mathbf{D}_{milk}$ and $\mathbf{D}_{liquid}$, and also on the domain $\mathbf{D}_{granul\_subst}$, but it is not defined on bunches of radishes.

Incorporating sorts, *cont'd.*

- In general when we assign a type <**a,b**> to a lexical item, the function that is the semantic value of that lexical item will be defined only for arguments in domains that have a non-empty intersection with $\mathbf{D_a}$.

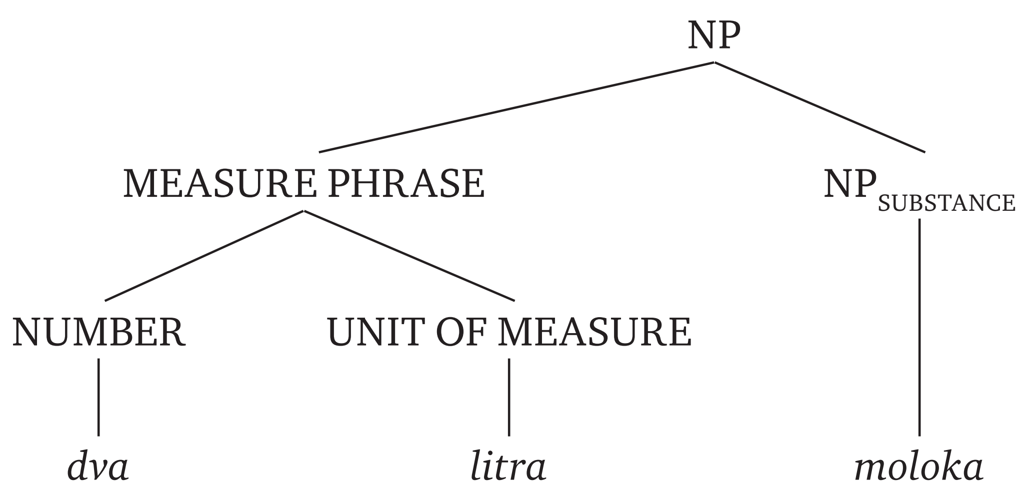
Two observations

- 1. In introducing new types, relations among them, and certain functions defined on them, we obtain the means for describing a fragment of the ontology of natural language.
- 2**.** Our modifications to the system create *multi-sorted models*, that is, models in which there are many basic domains. Multi-sorted systems are introduced in order to delimit the domains of definedness of functions, as for instance in our example function LITER.

4. Examples

- In Partee, Borschev 2012b we described the semantics of the expressions *dva litra moloka* ‘two liters of milk’ and *dva stakana moloka* ‘two glasses of milk’. Here we show how that description is modified when we make use of the new types introduced here.
- ***Dva litra moloka* ‘two liters of milk’**
- First we present the syntactic structure of the expression in tree form.

(8)



Examples, *cont'd.*

- Let us begin with *litra*. We introduced above a constant LITER of type <**e**,**r**>, more precisely <**pourable_subst**, **r**>, denoting a function that maps a portion of any 'pourable substance' onto a number that gives its volume in liters.
- For the use of *litr* in the genitive of measure construction, as it occurs in *dva litra moloka* in tree (8), we make use of a derived function constant LITER2 defined in terms of LITER. LITER2 takes a number *n* as argument and returns a predicate modifier, a function that can apply to the semantic value of the NP *moloka* to return a predicate true of anything which is a portion of milk and has a volume of *n* liters, letting us express *so-and-so many liters of such-and-such substance*.
- **Definition**: LITER2 = λn [λP [λx [(LITER(x) = n) & P(x)]]].
- On the next slide we repeat the definition and work through its parts.

Examples, *cont'd.*

- Repeating:
- **Definition**: LITER2 = λn [λP [λx [(LITER(x) = n) & P(x)]]].
- Here the first argument, n, of LITER2 is a variable of type **r** over numbers; the second argument, P, is of type <**e,t**>. Because the original function LITER is defined only for entities of type **e** that are furthermore of type **pourable_subst**, and since according to the formula in the definition of LITER2, P must apply to x, which is also an argument of LITER in the same formula, the only admissible values for P will be properties that can hold of an entity x which is of the type **pourable_subst**.
- In terms of Montague's basic type structure (assuming that numbers are a subset of entities), the type of the variable n above is **e**, the type of P is <**e,t**>, and the type of x is **e**. The type of the whole formula to the right of the = sign, and hence of LITER2, is <**e**, <<**e,t**>, <**e,t**>>>: it maps an entity (a number) onto a function from properties to properties.

Examples, *cont'd.*

- Using our enriched type system that includes sortal information, we can specify the types of n, P, and x more narrowly.
- We have already noted that n is of type **r**.
- As a result of the constraints imposed by LITER, we can determine that any admissible value for x in the formula must be of the type **pourable_subst** (or some subtype thereof; what LITER tells us is that any value of x must be of a type that is compatible with the type **pourable_subst**).
- And since, as noted above, P must apply to that x, any admissible value for P must be of type **<pourable_subst, t**>.
- We thus derive that the more fine-grained type of LITER2, is <**r**, <<**pourable_subst, t**>, <**pourable_subst, t**>>>: it maps a number onto a function from properties of pourable substances to properties of pourable substances, letting us express *so-and-so many liters of such- and-such substance*.

Examples, *cont'd.*

- Below we spell out the semantic derivation for the expression *dva litra moloka.*

(i) *litr: litr2:* Type <**r, <<pourable_subst, t>, <pourable_subst, t>>>**
Meaning: LITER2 = λn [λP [λx [(LITER(x) = n) & P(x)]]]

(ii) *dva*: Type **r**. Meaning: 2.

(iii) *dva litra*: Type **<<pourable_subst, t>,<pourable_subst, t>>**

Meaning: λP [λx [(LITER(x) = 2) & P(x)]]

(iv) *moloka*: Type <**milk**,**t**>. Meaning: ***milk***

(v) *dva litra moloka*: Type <**milk**,**t**>.

Meaning: λP [λx [(LITER(x) = 2) & P(x)]] (***milk***) =λx [(LITER(x) = 2) & ***milk*** (x)]

- Note that according to line (iii), any admissible argument of *dva litra* must be of the type **<pourable_subst, t**>. Since **milk** is a subtype of **pourable_subst,** *moloka* is an admissible argument for *dva litra*.

Examples, *cont'd.*

- How do we determine that the sort, or type, of the result has the more narrow specification <**milk**,**t**> rather than the more inclusive sort **<pourable_subst, t**>?
- That follows from the fact that the interpretation of the result includes the subformula ***milk*** (x); therefore any admissible value of x in the interpretation of *dva litra moloka* must satisfy the more restrictive constraint that it be of sort **milk** and not merely the constraint imposed by the subformula "LITER(x) = 2" that it be of sort **<pourable_ subst, t**>.
- And note, as we will illustrate below, that if the sortal restrictions imposed by the two subformulas were not compatible, the whole expression would be semantically ill-formed.

Examples, *cont'd.*

- ***Dva stakana moloka*** **'two glasses of milk'**

- In Partee, Borschev 2012b we examined several variants of the semantics of this expression, relating them by some container-specific meaning-shifting rules. Here, because the shifting rules are not our center of interest, we will apply our sortal approach to just one of them—what we called the "Ad Hoc Measure" reading, in which a concrete glass of arbitrary size is used to provide a unit of measure ($stakan_{AHM1}$ in the terminology of the cited work).
- In that interpretation, the word *stakan* has undergone a lexical shift —it denotes not some concrete glass ***c***, but a unit of measure of substances, corresponding to the volume of a portion of substance that would fill this concrete glass ***c*** and analogous to other units like *liter* and *pint*.

Examples, *cont'd.*

- Just as with LITER, and with any term for a unit of measure for measuring volumes of substances, we will have both a basic function denoted by STAKAN_{AHM} and a derivative term STAKAN_{AHM2} that will be used in the genitive of measure construction.
- STAKAN_{AHM}: Type <**pourable_subst**,**r**>. Meaning: the denotation of STAKAN_{AHM} is a function from $\mathbf{D}_{\boldsymbol{pourable\_subst}}$ to $\mathbf{D}_{\boldsymbol{r}}$ corresponding to some concrete glass ***c***, such that if ***m*** is a portion of substance, then $\text{STAKAN}_{AHM}(\boldsymbol{m})$ is the volume of ***m*** measured in terms of glass ***c***.
- The derived STAKAN_{AHM2} used in the genitive construction has an argument structure like that of LITER2, letting us express *so-and-so many glasses of such-and- such substance:*
- $\text{STAKAN}_{AHM2} = \lambda n\ [\lambda P\ [\lambda x\ [(\text{STAKAN}_{AHM}\ (x) = n)\ \&\ P(x)]]]$.

Examples, *cont'd.*

- And the semantic derivation for an expression like *dva stakana moloka* for an arbitrary glass is completely analogous to the semantic derivation for the expression *dva litra moloka*.

(i) *stakan: $stakan_{AHM2}$:* Type <**r, <<pourable_subst, t>, <pourable_subst, t>>>** Meaning: $STAKAN_{AHM2}$ = λn [λP [λx [(СТАКАН$_{AHM}$ (x) = n) & P(x)]]].

(ii) *dva*: Type **r**. Meaning: 2.

(iii) *dva stakana*: Type **<<pourable_subst, t> ,<pourable_subst, t>>**

Meaning: λP [λx [($STAKAN_{AHM}$ (x) = 2) & P(x)]].

(iv) *moloka*: Type <**milk**,**t**>. Meaning: ***milk***

(v) *dva stakana moloka*: Type <**milk**,**t**>.

Meaning: λP [λx [($STAKAN_{AHM}$ (x) = 2) & P(x)]](***milk***) = λx [($STAKAN_{AHM}$(x) = 2) & ***milk*** (x)].

Examples, *cont'd.*

- ***Dva stakana (kakoj-to) gadosti* 'two glasses of (some) filth'**
- *Gadost'* is an evaluative word, and like English *filth*, or *nasty stuff*, it can be applied to things, stuff, happenings. Someone can speak or do *gadost'* to us; a movie can be called *gadost'*, etc. In the genitive of measure in example (9), the 'filth' must be some substance which can be measured by the glassful and can be drunk, hence some liquid; that follows from the sortal requirements of the other parts of the construction and the semantics of the construction.

(9) *On vypil dva stakana (kakoj-to) gadosti.*
He drank two glass-GEN.SG (some-kind-of-GEN.SG) filth-GEN.SG
He drank two glasses of (some sort of) filth.

- We start with a minimal representation of *gadost'* as a predicate **filth'**, with the inclusive predicate type <e,t>, since evaluative predicates do not generally have sortal restrictions. All we need for this example is that the type for *gadost'* has a non-empty intersection with the types **<pourable_subst, t>** and **<liquid, t>**.

Examples, *cont'd.*

- Then the semantic derivation for the expression *dva stakana* (*kakoj-to*) *gadosti* will be similar to that of the expression *dva stakana moloka*, differing only in when and how the semantic type of the result is determined.
- The first three lines, (i)-(iii), which just concern *dva stakana*, will be identical, so we just provide steps (iv) and (v) below.

(iv) *gadosti*: Type **<e, t>**. Meaning: ***filth'***

(v) *dva stakana (kakoj-to) gadosti*: Type **<pourable_subst, t>**.

Meaning: $\lambda P\ [\lambda x\ [(STAKAN_{AHM}\ (x) = 2)\ \&\ P(x)]]$ (***filth'***) =

$\lambda x\ [(STAKAN_{AHM}\ (x) = 2)$ & ***filth'*** $(x)]$

- The type of the whole expression results from the restrictions imposed on the type of admissible values of x by the type of $СТАКАН_{AHM}$. Since ***filth'*** imposes no sortal restrictions of its own, the restrictions imposed by $СТАКАН_{AHM}$ determine the final result.

5. Anomalous examples.

- Let us once more contrast the “well-formed” and “ill-formed” examples from the Introduction. The expression *dva stakana moloka* from example (1) *On vypil dva stakana moloka* can describe some portion of milk and is of type <**milk**, **t**>.
- The verb *vypit’* (*drink*) is defined for direct objects of the type **liquid**. The whole expression will require that what was drunk is both of type **liquid** and of type **milk** (because in the derivation of the meaning some e-type variable x will occur both as an argument of *vypit’* and as an argument of the predicate ***milk***.) And since **milk** is a subtype of **liquid**, that is consistent.
- The expression *dva stakana muki* ‘two glasses of flour’ in example (5) is of type <**flour**, **t**>, the intersection of types <**flour**, **t**> and <**pourable_subst**, **t**>. But <**flour**, **t**>, which is a subtype of <**granul_subst,t**>, is disjoint from the type <**liquid, t**>, and hence inadmissible as an argument of *vypit’*. As a result, example (5) is semantically anomalous.

Anomalous examples, *cont'd*.

- Example (6) (*Dva pučka moloka* 'two bunches of milk') is anomalous because *pučok* 'bunch' is not a unit of measure for portions of liquid, of type <**liquid,t**>.
- Analogously, example (7) (??*On uronil s podnosa poltora stakana moloka* 'He dropped from the tray one and a half glasses of milk') is anomalous, or at least doubtful, since *uronit'* is restricted (in its literal uses) to solid objects and does not apply to arguments of type <**liquid,t**>. And while *stakan* in its most basic use is a solid object, and some uses of *stakan moloka* do refer to the glass together with its contents (see Partee, Borschev 2012b), on those uses glasses can only be counted with whole numbers. When *stakan* has a measure interpretation as in (7), it can be measured in fractional numbers, but then the sort of the whole expression is <**pour_subst, t**>, not <**solid_entity**, **t**>.
- So it is impossible or nearly so to impose a consistent typing on the whole sentence in example (7).

Anomalous examples, *cont'd*.

- One interesting complication, mentioned briefly in a footnote in our printed paper, is that the restrictions we have explored hold for normal non-modal affirmative sentences. But in modal, interrogative, negative, and fictional contexts, these constraints do not always hold. Sentences like (10) and the English example (11) (from Thomason 1972) are fully acceptable.

(10) *Vrjad li on mog vypit' dva stakana muki.*

Hardly he could drink-pf two glass-GEN.SG flour-GEN.SG

It's doubtful that he could drink two glasses of flour.

(11) *It is not true that The Painted Desert is reluctant.*

- These sentences contain subparts which we have analyzed as sortally incorrect. Why doesn't a sortally incorrect subpart make the whole sentence sortally incorrect in these cases, as it usually does?

Anomalous examples, *cont'd*.

- The conclusion should probably be that we want to use sortal information to explain the anomaly of the anomalous examples, but we do not want the semantic derivations to be impossible.
- We need the grammar to be able to generate the anomalous examples and the semantics to interpret them, so that they are available to be embedded under modals, negation, etc.
- Such examples, as Thomason argued, put some constraints on the nature of the explanation of sortal incorrectness.
- Making sort theory an extension of type theory, as we have done here, may in the end not be the best way to incorporate ontological information into the semantics.

* * *

In closing …

- We note in closing that even for the small fragment of ontology that we have considered here, the work is by far not complete. It is obvious that there are many and varied problems that arise.
- Words and expressions can belong to several types at once; one needs a mechanism for describing regular metonymy, metaphor, and other kinds of semantic shifts; the distinctions between words that belong to “ordinary” ontological sorts and words like the evaluative *gadost’* ‘filth’ need to be studied; and there are many other problems.
- These and other problems are beginning to receive greater discussion in works aimed at the integration of lexical and formal semantics. Interesting work of this sort is also going on in the context of advances in computational semantics. One can hope that with solutions to these problems and further such advances, formal semantics can progress from being the semantics of syntax to being a more complete semantics of natural language.

References (incomplete; the printed paper has more.)

Apresjan Ju. D. (1999), Semanticheskaja motivacija nesemanticheskix svojstv leksem, In Die Grammatischen Korrelationen, Graz, 96–116.

Bach E. (1986a) Natural language metaphysics, In Logic, Methodology, and Philosophy of Science VII, North-Holland, Amsterdam, 573–595.

Bach E. (1986b) The algebra of events. Linguistics and Philosophy 9:5–16.

Borschev V. B., Knorina L. V. (1990), Tipy realij i ix jazykovoe vosprijatie, In Language of Logic and Logic of Language, Moscow, 106–134.

Borschev V. B. (2014), Eshche paz o tipax i sortax (ili ob integracii formal'noj i leksicheskoj semantiki). In Jazyk. Konstanty. Peremennye: pamjati Aleksandra Evgen'evicha Kibrika, Aleteja, Saint Petersburg, 38–56.

Borschev V. B., Partee B. H. (1999), Semantika genitivnoj konstrukcii: raznye podxody k formalizacii. In Tipologija i teorija jazyka: Ot opisanija k ob"jasneniju. K 60-letiju Aleksandra Evgen'evia Kibrika, Moscow, 159–172.

Borschev V. B., Partee B. H. (1999b), Semantic types and the Russian genitive modifier construction, In FASL: The Seattle Meeting 1998, Ann Arbor, 39–57.

Borschev V. B., Partee B. H. (2001), Genitive modifiers, sorts, and metonymy, Nordic Journal of Linguistics, Vol. 24, 140–160.

References, *cont'd.*

Borschev V. B., Partee B. H. (2004), Genitives, types, and sorts: The Russian genitive of measure. In Possessives and Beyond: Semantics and Syntax (UMOP 29), Amherst, MA, 29–43.

Borschev V. B., Partee B. H. (2011), Genitiv mery v russkom jazyke, tipy i sorta, In Slovo i jazyk. Sbornik statej k 80-letiju akademika Ju. D. Apresjana, Moscow, 95–137.

Chierchia G. (1982), Bare plurals, mass nouns, and nominalization, WCCFL 1, Stanford.

Chierchia G. (1998b), Reference to kinds across languages, Natural Language Semantics, 339–405.

Dowty D. (1979), Word meaning and Montague grammar. The semantics of verbs and times in Generative Semantics and in Montague's PTQ, Dordrecht.

Kamp H., Partee B. (1995), Prototype theory and compositionality, Cognition 57, 129–91.

Krifka M. (1995), Common nouns: a contrastive analysis of Chinese and English, In The Generic Book, Chicago, 383–411.

Landman F. (2004), Indefinites and the Type of Sets, Explorations in Semantics, Oxford.

Link G. (1983), The logical analysis of plurals and mass terms: A lattice-theoretical approach, In Meaning, use and the interpretation of language, Berlin, 303–323.

Parsons T. (1970), An Analysis of Mass and Amount Terms, Foundations of Language, Vol. 6, 363–385.

References, *cont'd.*

Partee B. H., Borschev V. (2012a), Sortal, relational, and functional interpretations of nouns and Russian container constructions, Journal of Semantics 29, 445–486.

Partee B. H., Borschev V. (2012b), Dva stakana moloka: substances and containers in genitive of measure constructions in Russian, Russkij jazyk v nauchnom osveshchenii 2, 140–166.

Pustejovsky J. (1995), The Generative Lexicon, The MIT Press.

Rothstein S. (2009), Measuring and counting in Modern Hebrew, Brill's Annual of Afroasiatic Languages and Linguistics 1, 106–145.

Schwarzschild R. (2002), The grammar of measurement, Proceedings of Semantics and Linguistic Theory XII, Ithaca, NY, 225–245.

Schwarzschild R. (2005), Measure phrases as modifiers of adjectives, Recherches Linguistiques de Vincennes 34: L'adjectif, Paris, 207–228.

Schwarzschild R. (2006), The role of dimensions in the syntax of noun phrases, Syntax Vol. 9/1, 67–110.

Thomason R. (1972), A Semantic Theory of Sortal Incorrectness, Journal of Philosophical Logic 6, 209–258.