Computing the Semantics of Plurals and Massive Entities Using Many-Sorted Types

Bruno Mery, , Christian Retoré

Computing the Semantics of Massive Entities using Many-Sorted Types

Bruno Mery, Richard Moot, and Christian Retoré

LaBRI-CNRS & LIRMM-CNRS
Université de Bordeaux & Université Montpellier 2

Abstract We demonstrate how the specifics of the semantics for mass nouns can be integrated in a recent type-theoretical framework with rich lexical semantics, similarly to collective plural readings. We also explore the significance of an higher-order type system for gradable predicates and other complex predications, as well as the relevance of a multi-sorted approach to such phenomena. All the while, we will detail the process of analysis from syntax to semantics and ensure that compositionality and computability are kept. [1]
Keywords: Lexical Semantics, Plural Nouns, Mass Nouns, Higher-Order Logic, Syntax and Semantics Analysis, New Type Theories.

The distinction between *massive* and *countable* entities is similar to a classical type/token distinction — as an example of the type/token distinction, "the bike" can refer both to a single physical bicycle (as in the sentence "the bike is in the garage") but also the the class of all bicycles (as in the sentence "the bike is a common mode of transport in Amsterdam"). However, linguists such as Brendan Gillon warn against such a generalisation (long made in the literature) and remark that, as far as the language is concerned, mass nouns are more alike to the collective readings of pluralised count nouns. Among the many similarities is, for instance, the identical behaviour of plurals and mass nouns with cumulative readings: "Both the pens on the desk and the pens in storage use black ink, so I only have black pens" and "There is red wine on display and red wine in the back, so we only have red" are logically similar (see [5] for discussion). Several different approaches have been proposed to account for the specific semantic issues of mass nouns, from Godehard Link's augmented mereological approach in [12] to David Nicolas' revision of plural logic in [25], all remarking upon this similarity.

Many different formalisms, using advanced type theories for the purpose of modelling semantics, have been recently proposed, see e. g. [1,3]. Among those, we proposed a semantic framework based on a multi-sorted logic with higher-order types in order to account for notoriously difficult phenomena pertaining

[1] *This work is supported by the CNRS with the PEPS CoLAN, and by the ANR ContInt Polymnie. The authors are indebted to all the LENLS 11 committee and organisers, to the reviewers for their comments, as well as to Nicholas Asher, Daisuke Bekki, Stergios Chatzikyriakidis, Robin Cooper, Thomas Icard, Robert Levine, Zhaohui Luo and Yusuke Kubota (among many others) for many helpful discussions.*

to lexical semantics – see [30] for a recent synthesis. We recently integrated a semantic representation of the various readings of plurals in [15].

The aim of the present paper is twofold: to demonstrate that the semantics for mass nouns can be integrated in our framework, and that having a multi-sorted logic is an advantage in tackling related issues. Our approach emphasises the *computable* aspect of such a framework. We detail the analysis from syntax to semantics (using state-of-the-art categorial grammars in the tradition of Lambek [9]), with a sound logical framework for meaning assembly and the computation of logical representations, while keeping compositionality as a basic principle, as in Montague's original semantic program [17]. As [19,33] have shown, this process can be fully and transparently implemented.

1 On Classical Issues in New Type Theories

1.1 Compositional Lexical Semantics and New Type Theories

Type-theoretical framework for computational semantics based on "new" type theories (that mostly stem from Martin-Löf Type Theory and a practical-oriented understanding of the Curry-Howard Correspondence) are relatively recent additions to the formal semantics scene, beginning with Aarne Ranta's seminal work, [28]. They can be used to present a solution to issues of lexical semantics such as polysemy, deriving from James Pustejovsky's Generative Lexicon – [26]. These formalisms have matured during the last 20 years to become a set of logically sound compositional frameworks that can be inserted in the Montagovian chain of analysis: see [1], [3], [13], and many others.

1.2 Our Formal and Computational Model, ΛTY_n

Our system, ΛTY_n, is based on System-F, using higher-order types with many sorts and coercive sub-typing for modelling different phenomena. Detailed in [2,14], it has been constantly upgraded and aggregates many different phenomena such as deverbals ([29]) or the narrative of travel ([27,10]). See [30] for a recent, complete, open-access synthesis.

Summarising, our system is formally based on a higher-order version of λ-calculus in the tradition of Girard's System-F (see [6]), in which types can be abstracted and applied like standard λ-terms: for example, $\Lambda\alpha\lambda P^{\alpha\to\mathtt{t}}.(\iota\ P)$ is a functional term that requires both a type α and a λ-term P, a predicate of type $\alpha \to \mathtt{t}$. This example is the *selection* operation, used to model definite articles such as *the*.

In addition, we use a functional logic with n base types in addition to propositions, TY_n, described in [24]. We will use the following conventions in the present paper :

Semantic Type	Implied meaning
$\alpha, \beta, \gamma, \tau$	Type variables.
t	The type for propositions (or "truth values")
e	The Montagovian entities
	All subsequent sorts are subtypes of e.
φ	The sort for *physical* objects.
H	The sort for individual human beings.
Pl	The sort for *places*, locations.
$Container, Animal, Food\ldots$	Other specific sorts are given explicitly.
$\mathbf{g}_\tau$	The sort for groups of individuals of sort τ.
$\mathbf{m}_\tau$	The sort for masses of sort τ.

$\mathbf{g}$ and $\mathbf{m}$ are constructors for groups of individuals and masses of measurable quantities of other, pre-defined sorts. The lexicon defines for what sort τ there exists a sort $\mathbf{g}_\tau$ and $\mathbf{m}_\tau$; group types typically denote countable individuals ($\mathbf{g}_H$ is the type for *committee* and *team*) and mass types denote measurable quantities ($\mathbf{m}_\varphi$ is the type for *water* and *sand*, while *stone* is ambiguous between types φ and $\mathbf{m}_\varphi$). Details in section 2.1.

The use of multiple sorts (including many others not used here) allows us to correctly model complex phenomena such as co-predications, "dot"-types, qualia, while preserving a compositional model based on well-understood Montagovian principles. We use an analysis based on Categorial Grammars that provides the syntactic structure of the sentence and parse the semantics according to that structure, in a very classical way that happens to be used in the implementation of our parser for syntax and semantics, Grail (see [19]). In the rest of this paper, we will present the syntactic categories in the usual fashion: a/b yields an a given a b on its right while $b\backslash a$ yields an a given a b on its left, we use n for nouns, np for noun phrases, pp for prepositional phrases and s for sentences.

Our system uses the syntactic structure provided by such grammars in the tradition of Lambek, substitute the lexically-provided semantic terms and assemble their meanings to form the meaning of the sentence using β-reduction. The difference with the usual approaches is that using multiple sorts makes it possible to detect many lexical phenomena in the event of a typing mismatch. Higher-order types, together with *lexical transformations* – optional λ-terms that provide fine-tuned type-shifting operations to specific lexical entries, gives us the necessary control and flexibility to integrate lexical semantics in this process without sacrificing computationality and compositionality.

1.3 Plural Readings in System-F

In [15], we demonstrated an implementation of Link's classical account of plurals. To sum up, plural readings are obtained by enriching the single-sorted Montagovian type system (e, t) with a sort for group individuals $\mathbf{g}$, introducing operators that indicate the membership of an individual in a group, and modelling sets of individuals as predicates of type $\mathsf{e} \rightarrow \mathsf{t}$. We demonstrated that *lexical transformations* are well-suited to express the polysemy between collective, distributive and hybrid readings.

For any group modelled as a (generic) predicate $\Lambda\alpha\lambda x.P(x)$, we can define its cardinality by the means of an operator $|\_|^{(\alpha\to\mathbf{t})\to\mathbb{N}}$, where the types of natural numbers can be define as Church integers (as is in native System-F), or more simply as a primitive type such as is used in Gödel System-T, Martin-Löf Type Theory or, indeed, any reasonable computer implementation. This operator is only defined in context, and restriction of selection and a finite domain ensure that the statisfiability of the predicate is decidable and cardinality is finite. Similar operators have also been implemented in functional programming in [33].

The details of the previous paper are omitted for brevity, except a few technical points needed for the present work that are given in section 2.1.

1.4 Some Issues with Mass Nouns

Mass nouns, much like collections and group nouns, have distributive, collective and hybrid readings: compare (examples adapted from [12]):

(1) The foliage was uniformly red.

(2) The foliage was of all kinds of bright colours.

(3) The foliage was bleak and creepy.

(1) is distributive (every individual leaf is red) while (2) is collective (individual leaves are of mostly different colors). *Foliage* is a mass noun that might be considered as denoting a group of entities, but human language does not associate the same status to leaves on a tree and members of a committee: the individuation condition is clearly different.

A more clear-cut example is the following (paraphrased from textbook entries on the water cycle):

(4) The water gathers in the lake.

(5) The water feeds several rivers.

(4) has a collective reading, (5) a distributive one, but the individuation conditions are even less clear that with *foliage*. In [12], Link proposes to distinguish as *atoms* any terms that denote individual objects; atoms comprising "the water" in the above examples are thus *any portion* of the water that is denoted here. In (4), the mereological sum of every portion of the water is considered (thus aggregating any source of water that might have been previously mentioned), and in (5), the salient "atoms" are the volumes of water being contributed to every river.

Our operators for plurals are easily adapted for the purpose of deriving the correct semantics in that fashion, as we will now demonstrate.

2 Our Account of Plurals and Mass Nouns with Multiple Sorts

2.1 Including Multiple Sorts

Our system is multi-sorted, i.e. it distinguishes various sorts of entities that are differentiated as various sub-types of $\mathtt{e}$. For the sake of clarity, we did not use that differentiation in the previous account of the semantics of plurals. However, having multiple sorts is invaluable for the semantics of mass nouns and for complex predications.

This is a short summary of constructions introduced in [15], adapted to a multi-sorted framework that applies to mass nouns and plural readings:

Lexical item	**Example**	**Syntactic Category**	**λ-Term**
Individual Nouns	student	n	$\lambda x^H.student(x)$
(Abbreviated. Selection operators will yield the H (human) sort.)			
*The **group** types*			$\mathbf{g}_\tau$
For some singular sort τ.			
Group nouns	committee	n	$\lambda x^{\mathbf{g}_H}.committee(x)$
***Member** function*	[Committee] member	n/n	$\lambda y^{\mathbf{g}_\tau} x^\tau.member\_of(x,y)$
Group nouns are associated with predicates that we can build upon to create phrases such as *everyone within the committee* as well. The generic transformation specialises for any sort i.			
*The **mass** types*			$\mathbf{m}_\tau$
For some singular sort τ.			
Mass nouns	Water	n	$\lambda x^{\mathbf{m}_\varphi}.water(x)$
***PartOf** function*	Some volume of [water]	n/n	$\lambda y^{\mathbf{m}_i} \lambda x^i part\_of(x,y)$
Mass nouns are also associated with partitive transformations that can be specialised to any sort i.			
Collective verbs	Met	$np\backslash s$	$\lambda Q^{H\to\mathbf{t}}.\ (\lvert Q\rvert > 1) \wedge\ meet(Q)$

– *Continued on next page* –

Operators for mass and plural readings – continued from previous page

Lexical item	**Example**	**Syntactic Category**	**λ-Term**
	Gathered	$np\backslash s$	$\Lambda\alpha\lambda Q^{\alpha\to\mathsf{t}}\ .\ (\lvert Q\rvert > 1)$ $\wedge\ \ gather(Q)$
# transformation			$\Lambda\alpha\lambda R^{(\alpha\to\mathsf{t})\to\mathsf{t}}$ $\lambda S^{(\alpha\to\mathsf{t})\to\mathsf{t}}$ $\forall P^{\alpha\to\mathsf{t}}.S(P) \Rightarrow R(P)$
Coerced forms	Met$^{\#}$	$np\backslash s$	$\lambda R^{(H\to\mathsf{t})\to\mathsf{t}}\forall S^{H\to\mathsf{t}}.R(S) \Rightarrow$ $\lvert S\rvert > 1 \wedge meet(S)$
	Gathered$^{\#}$	$np\backslash s$	$\Lambda\alpha\lambda R^{(\alpha\to\mathsf{t})\to\mathsf{t}}\forall S^{\alpha\to\mathsf{t}}.R(S) \Rightarrow$ $\lvert S\rvert > 1 \wedge meet(S)$
Collective verbs such as *meet* or *gather* apply to sets, modelled as predicates, of non-singular cardinality, including groups or massive terms. The # transformation, available in the lexical entries for those verbs, provides the necessary machinery to turn sets of sets into a set. The transformed (coerced) terms for those verbs can apply to group or mass terms by the means of the *Member* or *PartOf* transformations (designated as lexical transformations for group and mass lexical items).			

Thus example (4) can be analysed as

$$|\lambda x^{\varphi}.part\_of(x, \text{the\_water})| > 1 \wedge \text{gather}(\lambda x^{\varphi}.part\_of(x, \text{the\_water}))$$

This reads as: the water, as a mass, is made of several (more than one) distinguishable parts, and these parts gathered (in a lake).

Of course, the individuation conditions are different between group nouns and mass nouns, and it is unreasonable to hope to account exhaustively for all possible physical sub-volumes of water gathering into a lake; the *part_of* relation selects portions that have been introduced in the discourse or context. In the absence of such, we argue that $|\lambda x^{\varphi}.part\_of(x, \mathrm{y}^{\mathbf{m}_i})| > 1$ always holds for any mass noun.

Distributive and hybrid readings are obtained in the same fashion.

2.2 The Pertinence of Multi-Sorted Semantics

Having a multi-sorted semantics with sub-typing gives us an edge in taking into account difficult lexical phenomena. First, we differentiate count and mass sorts (τ and $\mathbf{m}_\tau$), assuming that the lexicon includes the pertinent transformations from one to the other. We therefore can block infelicitous phrases such as **the water shattered* (requiring a physical, countable object) or **the bottle gathered* (requiring a mass or plural noun), and disambiguate between count and mass readings.

Secondly, this allows us to define polysemous predicates such as covered_with$^{\mathbf{m}_\mathsf{e}\to\mathsf{t}}$ and specialise their meaning so that *covered with rock* transforms the typing of the argument to $\mathbf{m}_\varphi$, while *covered with shame* produces the type $\mathbf{m}_{Abstract}$. The latter type is common in language, though not ontologically valid (compare *drinking from the fountain of glory*, *a man of much presence*, etc.). We can then correctly parse co-predicative sentences such as *[...] covered themselves with dust and glory* (from Mark Twain – this kind of zeugmatic expressions require specific care, as discussed and detailed by [4]).

2.3 Detailed Example: Partitive Quantification and Comparisons

Simple quantifications and comparative examples with massive entities include phrases such as *some water*, *some water is on the table*, *More water is on the table than wine is in the glass*. Our full analysis is shown in the table below.

The table gives a lexicon with syntactic and semantic types and semantic terms for a number of words in our lexicon. Some pertinent phrases have been given types and terms as well; these have been computed automatically from the lexical entries to give an idea of how complex expressions are derived from their parts.

The syntactic categories have been deliberately kept simple. The given type for the quantifier "some" is valid only for subject positions (but several solutions exist which generalise this type) and "more ... than" is treated as a unit and given a schematic type $n \Rightarrow (np\backslash s) \Rightarrow n \Rightarrow (np\backslash s) \Rightarrow s$, indicating it takes as its arguments first a noun n then a verb phrase $np\backslash s$ then a second noun and a second verb phrase to produce a sentence; the reader can consult [7,22] for the technical details of how we can produce the word order "More n_1 vp_1 than n_2 vp_2" and more complicated comparative constructions. For a more detailed account of the syntactic aspects of modern categorial grammars, see [18,20,21].

For the semantic types of polymorphic terms (namely *is* and *the*, $\Pi\alpha$ denotes a universally quantified type variable; this is implicitly included in the λ-term by the use of the abstraction $\Lambda\alpha$.

Input text	**Syntactic Category**	**Semantic Type**	**λ-Term**
Some	$(s/(np\backslash s))/n$	$(\mathbf{m}_i \rightarrow \mathsf{t}) \rightarrow (\mathbf{m}_i \rightarrow \mathsf{t}) \rightarrow \mathsf{t}$	$\Lambda\mathbf{m}_i \lambda P^{\mathbf{m}_i \rightarrow \mathbf{t}} \lambda Q^{\mathbf{m}_i \rightarrow \mathbf{t}}.\exists \lambda x^{m_i} \lambda a^{s_i}.(\wedge(\wedge\ (P\ x)\ (Q\ x))(\text{amount}\ x\ a))$
In addition to the usual existential quantification and conjunction of the argument and a predication over it, we add a predicate *amount* associated to an existentially quantified measure of the sort of the argument.			

– *Continued on next page* –

Complete example – continued from previous page

<table>
<tr><th>Input text</th><th>Syntactic Category</th><th>Semantic Type</th><th>λ-Term</th></tr>
<tr><td>Some water</td><td>$s/(np\backslash s)$</td><td>$(\mathbf{m}_\varphi \to \mathsf{t}) \to \mathsf{t}$</td><td>$\lambda Q^{\mathbf{m}_\varphi \to \mathsf{t}}.\exists \lambda x^{\mathbf{m}_\varphi} \lambda a^{s_\varphi}.$
$(\wedge(\wedge\ (\text{Water}\ x)\ (Q\ x))$
$(\text{amount}\ x\ a))$</td></tr>
<tr><td colspan="4">Water is of the physical sort (φ), associated with a massive sort $\mathbf{m}_\varphi$ and an appropriate measure s_φ (physical mass).</td></tr>
<tr><td>Table</td><td>n</td><td>$(\varphi \to \mathsf{t})$</td><td>$\lambda x^\varphi.(\text{table}\ x)$</td></tr>
<tr><td colspan="4">Table also provide us with an optional transformation ${f_{Telic}}^{\varphi \to Pl}$, as the use of tables as places to present or store things is lexical.</td></tr>
<tr><td>The</td><td>np/n</td><td>$\Pi\alpha.(\alpha \to \mathsf{t}) \to \alpha$</td><td>$\Lambda\alpha\lambda P^{\alpha \to \mathsf{t}}.(\iota\ P)$</td></tr>
<tr><td colspan="4">The salient selection operator (see our earlier work on determiners for details). The specialises to any type α.</td></tr>
<tr><td>On</td><td>pp/np</td><td>$Pl \to (\varphi \to \mathsf{t})$</td><td>$\lambda x^{Pl}.(\lambda y^\varphi.(\text{located\_on}\ x)\ y)$</td></tr>
<tr><td colspan="4">On constructs a predicate from locations.</td></tr>
<tr><td>Is</td><td>$\Pi(np\backslash s)/pp$</td><td>$\Pi\alpha.(\alpha \to \mathsf{t}) \to$
$(\alpha \to \mathsf{t})$</td><td>$\Lambda\alpha\lambda P^{\alpha \to \mathsf{t}}\lambda x^\alpha.\ ((\text{is}\ P)\ x))$</td></tr>
<tr><td colspan="4">The polymorphous copula (specialised for any type α) constructs a syntactically valid predicate.</td></tr>
<tr><td>Is on the table</td><td>$np\backslash s$</td><td>$\varphi \to \mathsf{t}$</td><td>$\lambda x^\varphi.(((\text{is}\ (\text{located\_on}\ ({f_{Telic}}^{\varphi \to Pl}$
$(\iota\ \text{table})^\varphi)^{Pl}))^{\varphi \to \mathsf{t}}\ x)$</td></tr>
<tr><td colspan="4">(As is usual in our system.)</td></tr>
<tr><td>Some water is on the table</td><td>s</td><td>t</td><td>$\exists \lambda x^{\mathbf{m}}{}_\varphi \lambda a^{s_\varphi}.(\wedge(\wedge\ (\text{Water}\ x)\ ((\text{is}$
$(\text{located\_on}\ (f_{Telic}\ (\iota\ \text{table})))$
$(Pack^{\mathbf{m}_\varphi \to \varphi}\ x)))$
$(\text{amount}\ x\ a))$</td></tr>
<tr><td colspan="4">This is the final, intuitive reading. Note the use of the universal packager, instantiated for physical massive entities, that was used in order to represent the (implied) container of the water.
Another possible reading (not detailed here), not using this packager nor the telic of the table as a storage, would simply be that some water has spilled on the table.</td></tr>
<tr><td>More...
than...</td><td>$n \Rightarrow$
$(np\backslash s) \Rightarrow$
$n \Rightarrow$
$(np\backslash s) \Rightarrow s$</td><td>$((\mathbf{m}_i \to \mathsf{t}) \to$
$(\mathbf{m}_i \to \mathsf{t}) \to$
$(\mathbf{m}_i \to \mathsf{t}) \to$
$(\mathbf{m}_i \to \mathsf{t})) \to \mathsf{t}$</td><td>$\Lambda \mathbf{m}_i \lambda P^{\mathbf{m}_i \to \mathsf{t}} \lambda Q^{\mathbf{m}_i \to \mathsf{t}} \lambda R^{\mathbf{m}_i \to \mathsf{t}}$
$\lambda S^{\mathbf{m}_i \to \mathsf{t}}.\exists \lambda x^{\mathbf{m}_i} \lambda y^{\mathbf{m}_i} \lambda a^{s_i} \lambda b^{s_i}.$
$(\wedge(\wedge\ (\wedge(P\ x)\ (Q\ x))$
$(\wedge(R\ y)\ (S\ y)))$
$(\wedge(\wedge(\text{amount}\ x\ a)\ (\text{amount}\ y\ b))$
$(>a\ b)))$</td></tr>
</table>

– Continued on next page –

Complete example – continued from previous page

Input text	Syntactic Category	Semantic Type	λ-Term
The comparative asserts the existence of two massive entities, with pertinent discriminating predicates, as well as the inequalities between their measures according to their sort. (We consider common-sorted terms, so that direct comparison is possible – metalinguistic comparison is the other option.)			
Glass	n	$(\varphi \rightarrow \mathsf{t})$	$\lambda x^{\varphi}.(\text{glass } x)$
As is well-known in lexical semantics, *glass* has the telic of a *container*, that is modelled by $f_{\,Telic}{}^{\varphi \rightarrow Container}$ and $f_{\,Content}{}^{Container \rightarrow \mathbf{m}_{\varphi}}$.			
In	pp/np	$Pl \rightarrow (\mathbf{m}_{\varphi} \rightarrow \mathsf{t})$	$\lambda x^{Container}.$ $(\lambda y^{\mathbf{m}}{}_{\varphi}.(\text{contained\_in } x)\ y)$
In this case, *in* constructs a predicate that applies to physical masses that can be inside a container.			
Is in the glass	$np\backslash s$	$\mathbf{m}_{\varphi} \rightarrow \mathsf{t}$	$\lambda x^{\mathbf{m}}{}_{\varphi}.(((\text{is (contained\_in}$ $(f_{\,Telic}{}^{\varphi \rightarrow Container}\ (\iota \text{ glass})^{\varphi}$ $)^{Container}))^{\mathbf{m}_{\varphi} \rightarrow \mathbf{t}}\ x)$
More water is on the table than wine is in the glass	s	t	$\exists \lambda x^{\mathbf{m}_{\varphi}} \lambda y^{\mathbf{m}_{\varphi}} \lambda a^{s_i} \lambda b^{s_i}.(\wedge\ (\text{Water } x)$ $((\text{is (located\_on } (f_{\,Telic}\ (\iota \text{ table}))))$ $(Pack\ x))$ $\wedge(\text{Wine } y)\ (((\text{is (contained\_in}$ $(f_{\,Telic}\ (\iota \text{ glass})))) \ y)\ (\text{amount } x\ a)$ $(\text{amount } y\ b)(> a\ b))$
Note that the universal packager has to be used for the water in order to reify its container; not so for the wine, the glass being an explicit container.			

From the syntax (not detailed here) to the final formula, the whole process is computational and compositional.

3 Scales, Measures, and Units with Many Sorts

3.1 Defining Scales in System-F

Gradable adjectives such as *tall*, as studied extensively in e.g. [8], can be represented in our system. In order to model those, as well the semantics of comparatives between quantities that can be explicit or implied by such adjectives, we need to be able to define scales or degrees in our system. Integers can be defined, and are already used for cardinality, by the means of Church numerals.

It is also straightforward to define floating point numbers in scientific notation[2][3]. We thus define the type s for scales, that can be specialised for any pertinent sort. Sorts are typically associated with *measures*, with useful measures including lengths, surfaces, volumes; masses or weights; and frequencies, durations. Pairs of comparable measures can also be used to specify *ranges*, used for the semantics of gradable adjectives.

Gradable adjectives are usually modelled in terms of *ranges*, given as a couple of scales indicating the "typical range" of values for the adjective. Using information provided by the sort, we can provide much more detailed semantics, associating predicates such as tall^{φ} with optional coercions such as $\lambda x^H.(f^{H \to \varphi}\ x) \wedge (\mathbf{R}\ x\ r_H)$; adverbial modifiers such as *very* modifying that range.

Many gradable adjectives are directly linked to mass nouns and generalised quantification in their semantics: *a bit/very wet* corresponds to *some/a lot of water*, etc. The analysis of their semantics is similar, but inferring one from the other is far from trivial.

3.2 Sorts, Units and Classifiers

As having different sorts is useful in order to distinguish semantically different classes of lexical items, they also provide a means to distinguish between *units*, ways of counting and accumulating quantities of items. Terms of individual or group sorts can be counted (using natural integers), terms of massive sorts can be measured (using scales and ranges), but the actual type of the term specifies the means of doing such an operation: comparing volumes for liquids, mass for generic physical massive terms...

Having a different unit for every sort, arranged in a hierarchy provided by sub-typing (all liquids are physical objects and thus can be compared by mass as well as by volume; every countable entity can be counted as "some object" but we might distinguish the number of people, of pets and of cars in a given situation...) is linguistically interesting, as it fits very well with a feature prominent in many languages: the classifier system.

Classifiers are syntactic items common to many spoken and written Asian, Amerindian and West African languages, as well as most variations of Sign Languages. In other linguistic groups, some traces remain present, such as *head* in the phrase "ten heads of cattle". Classifiers are used to count individuals and measure masses.

Our intuition, as expressed in [16], is that classifiers can provide the basis for a system of sorts. They are clearly linguistically motivated, and correspond to

[2] Such a number is a simple data structure comprising l, the list of digits (in base 10) of the mantissa, s, a constant indicating its sign, e, an integer representing the exponent and r, a constant indicating the sign of the exponent. Comparison between such floating point numbers is easy, and the common operations are definable.

[3] This short point illustrates that scales (and operations) can be defined natively in pure System-F; of course, we can take floating-point numbers of sufficient precision to exist as a primitive type, as is the case in any reasonable computer implementation.

an ontological hierarchy. Even if classifiers differ between languages, being influenced by the cultural and social evolution of each language or dialect involved (see [11,32,34] for details and discussions), the main features of the system are shared by most variations.

Of most interest to us is the hierarchisation of classifiers. For instance, in Japanese, some classifiers are generic, and commonly used by people not fluent in the language (denoting things, people, order, and broad "units" of any quantity measurable with a container – *Hai*); some are commonly used for specific categories of items (appliances, small objects, flat objects. . .); and some are very specific to an usage or trade (such as *koma* for panels). The fact that many classifiers (such as the ones counting small objects) are shared between Japanese and French Sign Language, for instance, illustrates the cognitive pertinence of using classifiers to define a system of linguistically different sorts.

Having a different way of counting items or measuring quantities is also needed to solve quantificational puzzles such as posed by Nicholas Asher in [1]. Building a complete system of sorts and units based on common aspects of the various classifier systems requires more resources than we have currently at our disposal, however.

3.3 Comparison Classes

The ranges used for gradable adjectives do not tell the whole story, as the study of comparison classes show. There are many views on what comparison classes expressed in phrases such as *tall for a seven-years-old girl* imply for the semantics of gradable adjectives: while Christopher Kennedy says that the semantics are not changed but that comparison classes modify the set of individuals that can be referred to at the discursive or pragmatic level [8], Stéphanie Solt argues that they give a specific range [31]. We believe that such phrases convey intrinsic (and not just use-specific) information, that might be modelled as predicates modifying the range pertinent to the sort of the argument.

4 Complex Cases

4.1 Mass-Count Alternations

For many mass nouns, there is a usage as a singular entity – compare *ten wild salmons* and *some raw salmon*. We, in agreement with most of the literature, have used the *grinding* transformation for such sentences. This operation (usually named the *universal grinder*) can take place on terms that have already undergone other grinding transformations, as the following example illustrates:

(6) The salmon we caught was lightning fast. It is delicious, and we preserved some of it. Wild salmon caught from this river could quickly become a source of income.

This shows that, given a suitable discourse structure, types $Animal$, $Food$, $\mathbf{m}_{Food}$ and $\mathbf{m}_{Animal}$ can coexist in referents with related senses.

4.2 Comparisons between Different Units

The difficulties of comparing between different predicates, such as in the comparisons in (7) and (9), has been oft remarked upon. While there have been some discussions of their characteristics and possible semantics such as [23], their integration within a Montagovian framework remains forthcoming. Considering:

(7) The table is longer than it is wide.

(8) The thread is bigger than the table.

(9) He his more dumb than ill-intended.

Our solution would compare different ranges together. The differentiation between sorts appears especially relevant on such examples. Comparing between two lengths in (7) is straightforward. Comparing between a length and a surface in (8) is much more difficult; if the sentence is felicitous, it seems to force the use of comparable units and reads as "The thread is *longer* than it is wide". On the other hand, comparing between different abstract and subjective values in (9) implies a strong intensional component. (9) has been called a *metalinguistic* comparison, as it constitutes a comparison in appearance only, instead establishing a relation between two different social judgements; it is hard to argue that the "degree of dumbness" is higher than the "degree of ill-intention" (or, indeed, to automatically assign any kind of numerical value to such measures).

Conclusion

We demonstrated a straightforward implementation of the semantics of mass nouns in our higher-order computational framework, which is based on Link's classical linguistic analysis, but does not commit to a specific ontological view, and can be adapted to other linguistic formulations of the phenomena, such as David Nicolas.

We continue expanding the coverage of our framework, in order to prove that complex issues of lexical semantics need not be resolved in isolation.

Indeed, having a multi-sorted approach makes the analysis of other complex phenomena easier.

At the present time, our implementation of the syntax-semantics analyser Grail includes an syntactical analysis using variations of Categorial Grammars, and an analysis of the semantics of sentences that is done in $\lambda - DRT$ rather than λ-calculus, with wide coverage grammars (in French) statistically acquired from corpora. We have formulated the semantics of plurals and mass nouns in a fashion similar to the process used in Grail, but in order to complete their implementation, a single step is yet missing: the construction of a wide-coverage system of sorts that can form the base types of a full-fledged semantic lexicon. We are keen to provide such a lexicon, inspired by the classifier system, and are excited to see many researcher moving towards that goal.

References

1. Nicholas Asher. *Lexical Meaning in Context: a Web of Words*. Cambridge University Press, March 2011.
2. Christian Bassac, Bruno Mery, and Christian Retoré. Towards a Type-Theoretical Account of Lexical Semantics. *Journal of Language, Logic, and Information*, 19(2), 2010.
3. Daisuke Bekki and Nicholas Asher. Logical polysemy and subtyping. In Yoichi Motomura, Alastair Butler, and Daisuke Bekki, editors, *JSAI-isAI Workshops*, volume 7856 of *Lecture Notes in Computer Science*, pages 17–24. Springer, 2012.
4. Lionel Clément and Kim Gerdes. Analyzing Zeugmas in XLFG. In *LFG 2006*, Konstanz, Germany, 2006.
5. Brendan S. Gillon. Towards a common semantics for english count and mass nouns. *Linguistics and Philosophy*, 15(6):597–639, 1992.
6. J. Y. Girard. Interprétation fonctionnelle et élimination des coupures de l'arithmétique d'ordre supérieur. Thèse de Doctorat d'État, Université Paris VII, 1972.
7. Petra Hendriks. Ellipsis and multimodal categorial type logic. In Glyn Morrill and Richard T. Oehrle, editors, *Proceedings of Formal Grammar 1995*, pages 107–122, Barcelona, Spain, 1995.
8. Christopher Kennedy. Vagueness and grammar: the semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, 30(1):1–45, 2007.
9. Joachim Lambek. The mathematics of sentence structure. *American Mathematical Monthly*, 65:154–170, 1958.
10. Anaïs Lefeuvre, Richard Moot, and Christian Retoré. Traitement automatique d'un corpus de récits de voyages pyrénéens : analyse syntaxique, sémantique et pragmatique dans le cadre de la théorie des types. In *Congrès mondial de linguistique française*, 2012.
11. Li XuPing. *On the semantics of classifiers in Chinese*. PhD thesis, Bar-Ilan University, 2011.
12. Godehard Link. The logical analysis of plurals and mass terms: A lattice-theoretic approach. In P. Portner and B. H. Partee, editors, *Formal Semantics - the Essential Readings*, pages 127–147. Blackwell, 1983.
13. Z. Luo, S. Soloviev, and T. Xue. Coercive subtyping: Theory and implementation. *Inf. Comput.*, 223:18–42, February 2013.
14. Bruno Mery. *Modélisation de la Sémantique Lexicale dans le cadre de la Théorie des Types*. PhD thesis, Université de Bordeaux, July 2011.
15. Bruno Mery, Richard Moot, and Christian Retoré. Plurals: individuals and sets in a richly typed semantics. In Shunsuke Yatabe, editor, *LENSL'10 - 10th Workshop on Logic and Engineering of Natural Semantics of Language, Japanese Symposium for Artifitial Intelligence, International Society for AI - 2013*, pages 143–156, Hiyoshi, Kanagawa, Japan, October 2013. jSAI-ISAI, Keio University.
16. Bruno Mery and Christian Retoré. Semantic Types, Lexical Sorts and Classifiers. In *NLPCS '10- 10th International Workshop on Natural Language Processing and Computer Science - 2013*, Marseille, France, October 2013.
17. Richard Montague. The proper treatment of quantification in ordinary English. In R. Thomason, editor, *Formal Philosophy. Selected Papers of Richard Montague.* Yale University Press, New Haven, 1974.
18. Michael Moortgat. Categorial type logics. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, chapter 2, pages 95–179. North-Holland Elsevier, Amsterdam, 2011.

19. Richard Moot. Wide-coverage French syntax and semantics using Grail. In *Proceedings of Traitement Automatique des Langues Naturelles (TALN)*, Montreal, 2010.
20. Richard Moot and Christian Retoré. *The Logic of Categorial Grammars: A Deductive Account of Natural Language Syntax and Semantics.* Lecture Notes in Artificial Intelligence. Springer, 2012.
21. Glyn Morrill. *Categorial Grammar: Logical Syntax, Semantics, and Processing.* Oxford University Press, 2011.
22. Glyn Morrill, Oriol Valentín, and Mario Fadda. The displacement calculus. *Journal of Logic, Language and Information*, 20(1):1–48, 2011.
23. Marcin Morzycki. Metalinguistic comparison in an alternative semantics for imprecision. *Natural Language Semantics*, 19(1):39–86, 2011.
24. Reinhard Muskens. Combining Montague Semantics and Discourse Representation. *Linguistics and Philosophy*, 19:143–186, 1996.
25. David Nicolas. Mass nouns and plural logic. *Linguistics and Philosophy*, 31(2):211–244, 2008.
26. James Pustejovsky. *The Generative Lexicon.* MIT Press, 1995.
27. L. Prévot R. Moot and C. Retoré. Un calcul de termes typés pour la pragmatique lexicale – chemins et voyageurs fictifs dans un corpus de récits de voyages. In *Traitement Automatique du Langage Naturel – TALN 2011*, pages 161–166, Montpellier, France, 2011.
28. A. Ranta. *Type-theoretical Grammar.* Indices (Claredon). Clarendon Press, 1994.
29. L.-M. Real-Coelho and C. Retoré. A generative montagovian lexicon for polysemous deverbal nouns. In *4th World Congress and School on Universal Logic – Workshop on Logic and Linguistics*, Rio de Janeiro, 2013.
30. Christian Retoré. The Montagovian Generative Lexicon Lambda Ty_n: a Type Theoretical Framework for Natural Language Semantics. In Ralph Matthes and Aleksy Schubert, editors, *19th International Conference on Types for Proofs and Programs (TYPES 2013)*, volume 26 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 202–229, Dagstuhl, Germany, 2014. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
31. Stephanie Solt. Notes on the comparison class. In Rick Nouwen, Robert van Rooij, Uli Sauerland, and Hans-Christian Schmitz, editors, *Vagueness in Communication*, volume 6517 of *Lecture Notes in Computer Science*, pages 189–206. Springer Berlin Heidelberg, 2011.
32. Benjamin K. T'sou. Language contact and linguistic innovation. In Michael Lackner, Iwo Amelung, and Joachim Kurtz, editors, *New Terms for New Ideas. Western Knowledge and Lexical Change in Late Imperial China*, pages 35–56. Koninklijke Brill, 2001.
33. Jan van Eijck and Christina Unger. *Computational Semantics with Functional Programming.* Cambridge University Press, September 2010.
34. Inge Zwitserlood. Classifiers. In Roland Pfau, Markus Steinbach, and Bencie Woll, editors, *Sign Languages: an International Handbook*, pages 158–186. Mouton de Gruyter, 2012.