Integrating QR Quantity Representations with the Semantic Web: A Progress Report

Kenneth D. Forbus & Walker Demel
Qualitative Reasoning Group, Northwestern University
2233 Tech Drive, Evanston, IL, USA 20208
{forbus, walker.demel}@northwestern.edu

Abstract

The rise of the Semantic Web opens new opportunities for qualitative reasoning research. It provides massive amounts of factual data that complement the ability of qualitative representations to express general principles and causal laws. This paper summarizes how we are bringing together three kinds of resources that represent quantities. The first is the traditional QR notion of quantities as fluents. The second is the OpenCyc ontology's notion of quantities as values. The third is Wikidata's massive collection of ground facts. We argue that the OpenCyc ontology provides a useful intermediate representation for bridging the other two representations. We show the potential utility for these bridges by examining data coverage for Science Olympics back of the envelope questions and language coverage with the QuaRel dataset.

1 Introduction

Quantities are central to the representation of continuous world. For example, we think about the physical world in terms of mass, size, temperature and pressure, among many others. Similarly, we think about the social world in terms of how friendly people are, how big a favor we are asking, and how well we know someone, among many others. Continuous aspects of our mental life are also characterized by many quantities, such as how enjoyable we believe an experience will be, how hard a problem is, and how much energy we currently have for different types of activities. We think about these quantities using a range of representations. While our cultures have quantitative values for many physical properties, no one would claim that we have accurate quantitative scales for most quantities in the social and mental worlds, but we do have the ability to use them in ordering and comparative analysis. Thus qualitative reasoning provides an essential component of reasoning about quantities across a broad range of human cognition. This has been somewhat obscured by the early framing of the field as "qualitative physics", but now we have qualitative modeling of human social and mental life as new frontiers for the field. What resources can we bring to bear to help us explore these new frontiers? Moreover, the Semantic Web has radically increased the available amount of formally represented knowledge. Can these resources enable us to scale qualitative reasoning research in new ways?

This paper brings together three resources for the formal representation of quantities. The first of course are notions of quantity developed in qualitative reasoning research. QR has developed models that are excellent at capturing human causal reasoning about relationships between continuous quantities[1]. These models have mostly been used to capture domain theories for aspects of the physical world. While in some cases they have been combined with quantitative models (e.g. Forbus et al. 1992,1999; Struss et al. 2014), much effort has gone into exploring how much can be done with purely qualitative models (e.g. Bredeweg et al. 2009; de Kleer, 1984; Kuipers, 2003). Meanwhile, the Cyc project developed a broad-scale knowledge base and ontology, the OpenCyc subset of which is the second resource we discuss. Cyc draws upon some ideas from the QR community, but also has many interesting representation conventions of its own that go far beyond what the QR community has done. Cycorp has encoded a broader range of quantity types, including aspects of the social and mental worlds, as well as developed schemes for units and qualitative values that have proven useful for qualitative reasoning (e.g. Paritosh 2004). The third resource we discuss is Wikidata which focuses on providing ground facts. For studying everyday and professional reasoning involving qualitative knowledge for framing numerical problems, such as back of the envelope reasoning (Paritosh & Forbus, 2005; Bundy et al. 2013), these compendia look to be quite valuable.

This paper describes these three resources, focusing on their complementary strengths. Since automatic encoding of problems from natural modalities and providing explanations to users is a crucial part of creating useful AI systems, we also describe how we have been integrating them into natural language resources for English. Some analyses of coverage

[1] And also spatial properties, of course, but this paper only focuses on quantities.

with respect to existing datasets is also provided. We close with some conclusions and future work.

2 Quantities in Qualitative Reasoning

We briefly summarize the key properties of quantities in QR that are needed to set the stage for discussing the other two resources. Readers who want more details might consult (Forbus, 2019).

The default representation of quantities in QR is that of a kind of *fluent*, i.e., a thing that can change over time. Quantities have values that can vary based on time (or in some cases, space as well as time). These values are often considered to be numerical values, albeit unknown, with the nature of qualitative representations being to provide partial information about them, such as ordinal information and/or signs of derivatives. In other cases, values are drawn from a finite set of symbols, totally or partially ordered (e.g. *large*, *medium*, *small*). Systems of qualitative representation have been connected to quantitative representations, but typically the details of quantitative equations, units, and numerical properties have been outside the formalisms (but see Klenk et al. 2015).

We note that here we are only concerned with quantities which can be considered continuous. The discrete states of an automobile transmission are sometimes viewed as a type of quantity for modeling purposes, but since one can shift from one gear to any other, it is not continuous in the usual sense used in QR. By contrast, describing age in terms of childhood versus adolescence versus adulthood does have that continuous character, since they always occur in sequence.

Quantities are quantities of something. That something might be a simple entity (e.g., temperature of coffee in a mug), or a situation (e.g., difficulty of a problem to be solved). Being explicit about such connections is important for fully capturing knowledge about quantities. If there is no coffee in a mug, its temperature is moot, and a textbook problem that would be hard to solve with just pencil and paper can become much easier with a graphing calculator, for example. Quantities are often defined with respect to several dimensions in complex situations. For example, the pressure of the ocean depends on the depth at which it is measured, while the temperature when analyzing a smartphone might be measured at many points within it. We handle such complexities by defining abstract entities which bundle together the multiple properties needed to fully specify what each quantity is about. Thus without loss of generality, domain theories can consider just a single entity for every type of quantity. The definition of such complex entities can sometimes be expressed via model fragments (e.g. a depth parameter of an entity) but may require mechanisms outside QP theory (e.g. the relevant measurement points in a smartphone thermal analysis). The key thing is the clean separation of what defines an entity from quantities of that entity[2].

Multiple representations of time have been used in QR. Hayes' (1985) *histories* notion combines spatially bounded and temporally extended entities that serve as the locus of description. Other representations use time-points, including numerical time-stamps, to distinguish different times. We note that there is a subtle type/instance distinction in QR, namely the idea of a qualitative state as a template for a class of behaviors, versus a specific instance of that class. In QSIM (Kuipers, 2003), for example, qualitative states were instances of behaviors, whereas in envisioners such as QUAL (de Kleer & Brown, 1984) and QPE (Forbus 1989) qualitative states represented categories of behaviors.

To summarize, the QR notion of quantities revolves around fluents, each of which is distinguished by

- Quantity type: What kind of quantity it is, which helps determine what causal laws it participates in.
- Entity: What it is a quantity of, be it concrete object, constellation of objects, or situation.
- Value: What can be said about its value for some specific time or category of behavior.
- Context: The spatial-temporal scope of statements concerning value.

The QR community has mostly focused on constructing domain theories that are broadly applicable, so that they can cover a wide variety of examples and systems. Even within the physical world, the range of quantity types and situations has been small compared to what people have experience with in everyday life. Very few studies have involved social reasoning and we are currently unaware of any studies involving qualitative reasoning about mental operations. OpenCyc, as demonstrated below, has a wide range of quantity types, and Wikidata has a sea of ground facts which can potentially be used for quantitative reasoning, including distilling generalizations to construct more portable knowledge, including QP models (Forbus & Gentner, 1997; Gentner & Medina, 1998). We turn to OpenCyc next.

3 Quantities in OpenCyc

The discussion of OpenCyc here is based on the distillation of material from several versions, now incorporated as the ontological framework for Northwestern's open-license NextKB[3]. We have made considerable extensions, including new resources for natural language understanding, representations and support for visual reasoning, spatial reasoning, and qualitative reasoning along with support for analogical reasoning and learning.

Much of the knowledge about quantities in OpenCyc concerns values. The anchoring collection is `ScalarQuantity`, which currently has 4,852 instances. These include a number

[2] Early QP theory implementations allowed arbitrary arity quantity types. This has been streamlined in current implementations to simply a single entity as suggested here.

[3] https://www.qrg.northwestern.edu/nextkb/index.html contains downloadable files in various formats, browsers, and reasoning systems. We use Creative Commons Attribution 4.0 licensing, compatible with OpenCyc, FrameNet, and other resources.

of qualitative values, e.g. `Rarely` for low-frequency of occurrence, values that are implicit comparisons to everyday values (e.g. `HouseSized`), values that represent ranges (e.g. `ColdToBitterlyCold`), including natural scales (e.g. `AFewMinutesDuration`, `AFewHoursDuration` and other instances of `OrderOfMagnitudeInterval`). It also includes a compositional system for constructing qualitative values of any type of quantity, out of sets of ordered symbolic ranges, e.g. `(HighAmountFn MentalEffortLevel)` is more than `(MediumAmountFn MentalEffortLevel)`.

OpenCyc has a flexible representation for quantitative values. Units are explicitly represented, e.g. `(DegreeCelsius 20)` denotes 20 degrees Celsius. Two numerical arguments indicate a range, e.g. `(DegreeCelsius 15 25)` denotes the range of temperatures from [15, 25] degrees Celsius. Higher-order functions permit composing units, e.g. `(PerFn Meter SecondsDuration)` is a logical function, such that `((PerFn Meter SecondsDuration) 30)` denotes 30 m/s. Similarly, multipliers commonly used in unit systems are also implemented by higher-order functions, e.g. `(Mega Watt)` for power in the millions of watts range. Such unit-denoting functions are instances of `UnitOfMeasurementConcept`, and in NextKB there are 632 such instances currently. Numerical computations involving units automatically do any necessary conversions in the FIRE reasoning engine (Forbus et al. 2010). FIRE also traps errors in unit computations, e.g. units combined by addition or subtraction have to be of the same dimension, one cannot add a length and a temperature, for example.

Entities of quantities are specified via relationships between the entity and values, with the specific relationship providing the type of quantity involved. For example, `(accountBalance FredMattress (Dollar-United-States 22))` says that Fred has $22 hidden in his mattress. These predicates are instances of `ScalarIntervalSlot`, of which there are currently 1,267 more specific predicates.

For performing qualitative reasoning or textbook problem solving (e.g. Klenk & Forbus, 2009; Crouse et al. 2018), fluents are needed as constituents of equations and as conceptual entities representing a quantity which takes values over time. Given the extensive set of concepts (aka *collections*) in Cyc for values of quantities, we found it useful to define a logical function, `QPQuantityFn`, that denotes a 2nd order function whose domain is entities and whose range are fluents corresponding to the type of quantity. For example, `(QPQuantityFn Temperature)` denotes a logical function which denotes the fluent representing the temperature of its argument. Thus `((QPQuantityFn Temperature) Stove)` denotes the temperature of the stove. Another relationship, `valueOf`, relates fluents to values. For example, if the stove's temperature is 200° C, we would say

```
(valueOf ((QpQuantityFn Temperature) Stove)
         (DegreesCelsius 200)).
```

Of course, the values of fluents change over time. There are multiple ways to express such differences. The first is `holdsIn`, a modal operator which takes a time interval and a statement, indicating that the statement is true within that time interval. For example,

```
(holdsIn CookingDinner
  (valueOf ((QpQuantityFn Temperature)
             Stove)
    (DegreeCelsius 200)))
```

indicates that the temperature of the stove is 200° C while dinner is being cooked, whenever that is/was/will be. The first argument can be anything that is a `TemporalThing`, e.g. a time interval, but also a complex event (such as cooking dinner).

The second way to scope the values of fluents is by using *microtheories*. Microtheories are local contexts. For example, another way to describe the temperature of the stove during dinner would be to use a microtheory to describe the specific dinner, e.g.

```
(ist-Information
  (MealCookingMtFn Dinner
   (DayFn 20 (MonthFn May (YearFn 2022))))
    (valueOf ((QPQuantityFn Temperature)
               Stove)
         (DegreeCelsius 200)))
```

indicates that for May 20th, 2022, the temperature of the stove was 200° Celsius. The first argument to `ist-Information` is always a microtheory, the second is a statement that holds in that microtheory.

There are multiple relationships that can hold between microtheories. The most fundamental is `genlMt`, which indicates that one microtheory inherits from a more general microtheory. Inheritance is monotonic, e.g., if a microtheory inherits from the `NaivePhysicsMt` microtheory, then, for example, the primary object moving in any earthquake is the ground. Every reasoning operation occurs with respect to some microtheory, which, along with the set of microtheories visible from it via `genlMt` relations, recursively, define the *logical environment* for that operation. Logical environments should typically be logically consistent, but the knowledge base as a whole need not be. For example, Cyc includes knowledge of Marvel comics characters and DC comics characters, but they are distinct microtheories and should not be combined[4], and they are also kept out of logical environments used in applications outside the entertainment industry. In qualitative reasoning systems built using NextKB, microtheories are used to implement incompatible aspects of domain theories as required for compositional modeling (Falkenhainer & Forbus, 1991). Microtheories are also used to represent qualitative states. The common assumptions underlying an analysis that are constant are represented in a microtheory for that analysis, with specific qualitative states represented by microtheories containing statements particular to that state. State transitions are represented via relationships between microtheories.

Aficionados of Hayes' (1985) notion of histories will be pleased to know that representations for his notion of slices is available in NextKB as well. `AtFn` has as its domain `Tem-`

[4] At least to avoid licensing issues 😉.

poralThing and Situation, with its range being TemporalThing. Situation, in turn, is a generalization of Event and StaticSituation, hence it is extremely general. More often used for quantities is MeasurementAtFn, which specifies a temporal slice of the fluent, e.g.

```
(MeasurementAtFn
   ((QPQuantityFn Temperature) Chicago)
   Summer)
```

denotes the temperature of Chicago in the summer.

Temporal information can also be specified as arguments to functions or predicates when needed to handle common cases. For example, the value of money changes over time, so USDollarFn takes an argument corresponding to a year. For example, Walter P. Murphy donated ((Mega (USDollarFn 1939)) 6) to Northwestern to create its engineering school, which is roughly ((Mega (USDollarFn 2022)) 125), i.e. $6M in 1939 dollars is equivalent to $125M today.

The OpenCyc ontology provides a rich set of quantity types. Many of them are physical quantities, ranging from frequently used types in qualitative reasoning research (e.g. Mass, Temperature, Charge) but also geographical quantities (e.g. Biodiversity, DensityOfPopulation, CrimeRate). It also includes social quantities (e.g. Glamor, Arrogance, Popularity), mental quantities (e.g. Curiosity, Hope, Doubt) and physiological quantities (e.g. Endurance, Strength, LevelOfSweat). This breadth provides conceptual representations that can be used in natural language semantics (see below). The cost of using OpenCyc is that many of the axioms which provide the inferential power found in the commercial version of Cyc are not available in OpenCyc. That is understandable, given that Cycorp wants to encourage people to use Cyc instead. We have decided that, for replicability, it will be better to start with OpenCyc and use learning by reading, conversation, and other methods of learning to build out the inferential knowledge that is needed[5].

To that end, we have constructed broad lexical resources for English that tie to the NextKB ontology. These are divided into two components. The Nulex lexicon provides grammatical features for words, based on Allen's (1994) grammatical features for English. For example,

```
(definitionInDictionary Nulex temperature
    Temperature-TheWord Noun
  (TheSet (root temperature)
   (agr (TheSet 3s)) (mass +)
   (countable +)))
```

is the singular form of the word temperature. The second argument is the token form of the word, with the formal representation of the word (Temperature-TheWord) being the 3rd argument. Part of speech (here, Noun) is next, followed by a set of grammatical features. This statement indicates that it is a mass noun as well as a count noun, and that it can be used in 3rd person singular (agr), with another statement, not shown, indicating it can also be used for 3rd person plural.

The second component are semantic translations, mapping a word and a given part of speech to expressions in the NextKB ontology. We use FrameNet information as an intermediary, e.g. FN_Temperature refers to the FrameNet frame Temperature. For instance,

```
(fnSemtrans (TheList) Temperature-TheWord
            Noun
  (and (isa :NOUN Temperature))
  (frame FN_Temperature)
  (bindingTemplate
    […] temperatureOfObject […])
```

This defines the word temperature as a noun, within the Temperature FrameNet frame. The and statement is the direct entailment, where the keyword :NOUN is replaced during parsing with the discourse variable representing the token in the sentence. The bindingTemplate, details not shown for space reasons, indicates that, depending on what other grammatical relationships are found in the sentence, the quantity slot temperatureOfObject can be used to relate the value to the entity, among other things.

4 Quantities in Wikidata

Wikidata is a collaboratively edited knowledge graph hosted by the WikiMedia foundation. Like its sibling Wikipedia, Wikidata utilizes the distributed-community model of editors—as of this writing, thousands of editors and bots have made over 1.6 billion edits to over 97 million items. This model allows Wikidata to serve as the downstream aggregate of otherwise siloed structured data sources. We describe how information about quantities is represented in Wikidata.

Wikidata is organized around items, with each having a unique identifier (QID) and a set of statements about it. Each statement is an RDF triple of <subject, property, value>. For example, "milk is white" can be expressed as

<*milk* (Q8495), *color* (P462), *white* (Q23444)>

Where the terms in italics are the English rendition of the objects whose ids are in parentheses. In QR terminology, items are entities and values are quantity values. Properties are analogous to OpenCyc's ScalarIntervalSlots, in that they link an entity to a value, with the type of quantity implicit in the relationship. In some cases the quantity type is obvious (e.g. *Area*, *Color*), while in others it is more opaque (*Statistical Population*). Any property can specify constraints on its value. Certain properties specify that their values must be a string, number, date, URL, media file, or another Wikidata entity. Other properties, like *capital (P36)* enforce no more than one value since most states have only one capital. Since Wikidata consists of RDF triples, it can be queried via a SPARQL endpoint (query.wikidata.org).

Wikidata uses the expressiveness of RDF to encode higher-order information about triples. For example, units and precision (expressed as a margin of error) can be attached to some properties that convey continuous properties, like

[5] Our extensions are all CC-Attribution, so that they can be reused by anyone, including commercial organizations.

Area (P2046) or *Mass (P2067)*. Other properties convey continuous values with the unit implicitly left to the understanding of the property itself, like *vehicles per thousand people (P5167)*. For example, the *United States (Q30)* has *area* 9,826,675±1 square kilometer and *vehicles per thousand people* 778. While less common, other properties do reference qualitative representations of continuous dimensions. *Cold sweet soup (Q14914664)*—the superclass of *gazpacho (Q202677)*—has statement *serving temperature (P7767) cold (Q270952)*, where cold is described as "relative or subjective state of low temperature."

Qualifiers[6] are another tool used in Wikidata for more descriptive statements. They provide a method for preventing local inconsistency, akin to the role of microtheories in OpenCyc. In Wikidata, predicates like *point in time (P585)* can be used to qualify statements like *population (P1082)*, for which there may be several different assertions that hold in different years. In the case where a country's *capital (P36)* may have changed, values can be associated with a *start time (P580)* and *end time (P582)*. Since Wikidata is crowd-sourced, some items may have disputed values depending on their source for a fact. For example, the statement that *Earth (Q2)* has *Creator (P170) God in Christianity (Q825)* is *disputed by (P1310) Athiesm (7066)* and *supported by (P3680) Creationism (Q130352)* and *Christianity (Q5043).*

The correspondence between Wikidata items, properties, and values to entities, quantity slots, and values provides a strategy for mapping between Wikidata and NextKB, thereby potentially making Wikidata a large-scale resource of ground facts. The differences between the Wikidata and OpenCyc ontologies make importing knowledge a challenge, however. One conflict is the lack of distinction between instances and collections in Wikidata—an item is considered an instance only by virtue of having an *instance of* property, and a collection if another item has a fact with an *instance of* or *subclass of* property. In contrast, NextKB strictly distinguishes between entities and collections, relying on this structure for monotonic inheritance and reasoning. Consider for example the OpenCyc predicate `ColorOfType`, whose first argument is a collection, and whose second argument is an instance of `Color`, e.g. `(ColorOfType Milk White)`. Since `ColorOfType` describes a Collection, it is referred to as a *type-level predicate.* Other predicates describe instances themselves, like `cityInState` which takes `City` and a `State-Geopolitical` entities as its arguments, e.g. `(cityInState CityOfChicagoIL Illinois-State)`. Different levels of predicates as well as strict constraints on argument types makes the assimilation of Wikidata facts a difficult task. Another source of difficulty comes from the differing specificity of relationships themselves. While Wikidata has one `causes` property, Cyc has 16 different causal relation predicates of varying levels of specificity. We are currently exploring strategies for automatic assimilation from Wikidata.

[6] For more on Wikidata Qualifiers, see www.wikidata.org/wiki/Wikidata:List_of_properties/Wikidata_qualifier

5 Coverage

The open-ended nature of quantity types and entities makes characterizing the coverage of a knowledge base challenging. In our research we use two measures: (a) Does it provide the facts needed for a task and (b) Can its contents be accessed via natural language? We present two examples of such analyses here.

5.1 Science Olympics Questions

Back of the envelope (BotE) reasoning (Paritosh & Forbus, 2005; Bundy et al. 2013) involves estimating values of parameters from everyday experience and data. A classic example, due to Enrico Fermi, is "How many piano tuners are there in Chicago?" An exact number isn't the point—the goal is to come up with a reasonable estimate. Assuming pianos are kept in tune, one might start by asking how many pianos there are in Chicago and how often they need to be tuned, to estimate demand for such services. How many pianos might in turn be estimated by what fraction of households have pianos, and how many households there are in Chicago. In other words, it is a recursive process: If one knows a number, done. If not, then figure out a reasonable model in terms of other parameters, and then try to estimate those.

Among the examples used in Paritosh (2007) is a small corpus of 35 back of the envelope questions from the Science Olympics, a particularly challenging set of such questions. Paritosh provided the data by hand, in order to show that the families of strategies he had identified sufficed to solve a broad range of BotE questions. Here, we ask how much of the numerical data needed by these questions can instead be supplied by Wikidata, as one estimate of its coverage.

Here is an example from the Science Olympics:

"To what height could loose-leaf paper be stacked if you possessed Avogadro's number of sheets?"

If a sheet of paper is 0.1 millimeters and Avogadro's number is 6.02×10^{23} – both pieces of information from Wikidata – and we assume the paper is incompressible, then the model of a linear stack of paper yields

$0.1\text{mm} \times 6.02\times10^{23} = 6.02\times10^{19}\text{m}$

	DATA FOUND	ESTIMABLE
# QUESTIONS	22	29
% COVERAGE	63%	83%

Table 1: Analysis of Science Olympics BotE coverage

Of the 35 questions in the Science Olympics dataset, 22 (63%) can be answered by using BotE strategies with data readily available in Wikidata (Table 1). There are two other strategies based on analogy that can extend the set of problems solved. The first is the KNACK algorithm (Paritosh & Klenk, 2006) which uses analogy to combine information across several instances to build an estimator for new instances of that category. For example, when asked to estimate the weight of the human population of Earth, it is hard to do without an estimate of how much a person might weigh. While this information is not available for generic humans in Wikidata, weights for various celebrities are known, and KNACK could be used to estimate how much a generic person weighs by using the weights of celebrities as examples. A second strategy is to use a sibling example, e.g. in a question about hydrogen atoms, where Wikidata doesn't have size information, the system could still succeed in providing an estimate (while cruder) by using the size for oxygen atoms, which Wikidata knows the size of. With these two strategies, seven more problems can be solved (83%) (Table 1). The remaining six questions require data that is both unavailable directly and for which no reasonable model that uses Wikidata-available data currently available exists. In all cases, a small number of additional ground facts would suffice to solve them. Given that Wikidata is being built totally independently of this task, we find it encouraging that it already covers so much, and of course its crowd-sourced nature makes it easier for humans (or bots) to extend it.

5.2 The QuaRel Dataset

QuaRel (Tafjord et al. 2019) is a dataset intended to test the ability of ML systems to learn to process comparative analysis problems (Weld, 1990) stated in natural language. For example,

> "Rod was rolling a ball to his dog. He was able to roll the ball much farther on the cement than on the grass. This is because the cement is (A) rougher (B) smoother?"

Here the two events being compared are rolling events, varying in the surfaces (antecedent quantity: smoothness) leading to a variation in outcomes (consequent quantity: distance). The underlying causal laws, here a qualitative proportionality over rolling episodes that distance depends on smoothness, were provided to the systems.

The best system performed at 77% accuracy (Mitra et al. 2019). An earlier version of our lexical resources and NL system were used with analogical learning to achieve 60% accuracy (Crouse & Forbus, 2020), which is equivalent to what AI2 was originally able to achieve on this dataset. We believe the difference between the best system and ours is that we used AI2's provided semantic representation as a target for our system, instead of the more expressive constructs found in NextKB. None of the higher-performing systems used AI2's semantic representations. Humans are at 96% accuracy, so there is plenty of headroom in this task. Consequently, we are looking again at QuaRel, to see if using our more expressive representations with analogical learning can provide more data-efficient learning and inspectability. Thus the QuaRel dataset provides one data point about NL coverage. In QP terms, every QuaRel question implies at least one ordinal relation, perhaps conveyed by a superlative (e.g. "which is hotter, …" versus "Which is hottest, …").

To examine coverage, we looked at the meanings of the words used in questions to see if any of them mapped to a quantity (either a value or a quantity type), an ordinal relation, or a superlative. This bears most directly on NextKB coverage, factoring out grammar and semantic interpretation rules and strategies. We used the QuaRel training set, which consists of 1,940 questions. On average, there were 3.2 quantities per question and 1.8 ordinals per question. This is what one would expect, since comparative analysis questions are often formulated as a comparison between two antecedent quantities, with the question concerning a comparison between a consequent quantity with respect to the comparison between the two situations already introduced.

Let us look at the extremes to gain some insights. First, where is there missing quantity information? There are only 14 questions where no quantity, ordinal, or superlative information is detected, i.e. less than 0.7% of questions. There are two missing lexical items responsible for most of them. (1) "further" did not have an appropriate semantic translation as an ordinal comparison over distance, and (2) "first" did not have a translation corresponding to earliest in time, a missing superlative. The rest are due to words found only in a particular question (e.g. "bright" as in degree of luminesce) or more complex constructions that imply quantity-specific ordinals or superlatives. Examples of the latter include "is passed by" which implies a speed difference, and "gets the jump on the other", e.g. when accelerating from a stop light, which vehicle has higher acceleration. Sometimes implicit limit points are used, e.g. two objects thrown at a wall, if one breaks and the other doesn't, then the fragility of one is higher than the other. Second, this analysis does not take into account polysemy, e.g. there are words with senses that are quantities but also that aren't (e.g. "work" in the physics sense versus the job sense). How often might this analysis over-estimate the

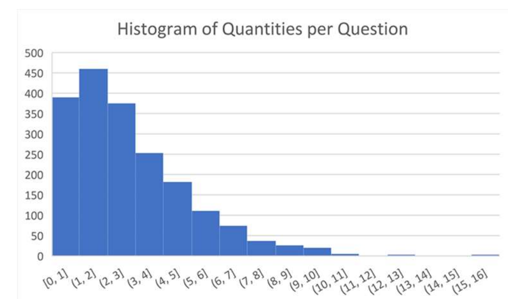

Figure 1: Estimating Polysemy

amount of available information? Suppose we take four quantity types as the most that would normally appear in a comparative analysis question (one each of the antecedent and consequent quantities, for each of the two situations being compared). Under that assumption, any question with more than four quantities is exhibiting polysemy. (We note that four or less does not guarantee no polysemy, but for automated analysis, we live with the approximation.) Figure 1 shows a histogram of quantities per question. 1,478 questions have four or fewer words with quantity interpretations, suggesting that 462 questions, or 24%, are exhibiting polysemy in quantity terms. Since the NLU system we are using strives for a high-precision understanding of text, the gaps in coverage are more of a concern than polysemy, since the syntactic knowledge in the parser and sematic constraints typically prunes excess alternatives. However, QuaRel is crowd-sourced, and thus is more noisy than curated sources. There are questions with missing words (e.g. “The is very bright and the stars aren’t as bright” where from the alternatives one can infer “sun” is missing), incorrect words (e.g. “Who will get their first?” in reference to comparing arriving events, “there” should have been used), and poor grammar. People typically use context to puzzle out such texts, and we believe that strategy will be productive for AI systems as well. That is, comparative analysis questions involve two situations to be compared with antecedent and consequent properties connected by a qualitative causal relationship, so even if implicit (e.g. increase in strength across a period of weight training), constructing meanings that fit this template might provide a way for systems, like people, to puzzle out the meaning of noisy texts.

7 Discussion and Future Work

Quantities are at the core of qualitative reasoning about continuous systems, be they physical, social, or mental. This paper argues that combining the ideas of quantities from qualitative reasoning research with knowledge of quantities represented in NextKB and Wikidata can provide valuable benefits. NextKB, which incorporates OpenCyc’s quantity representations and extends them in various ways, provides a rich vocabulary of concepts pertaining to quantity types and their values, as well as integrating them into natural language resources for English. Wikidata provides a massive set of ground facts about quantity values. Thus each provides complementary materials that, together, provides a new set of capabilities that can be tapped to take qualitative reasoning research into new frontiers.

We see four lines of future work. First, we have only begun laying out the mappings between NextKB and Wikidata, adding them by hand as we need them for tasks. We would like to make this process more automatic, with systems proposing new mappings and testing them both by reasoning and by asking human collaborators. Second, we plan to use these resources to explore a variety of language-based reasoning tasks (e.g. back of the envelope reasoning, social reasoning, and metacognitive reasoning). Third, we plan to develop strategies for NL-using systems to puzzle out meanings when they have trouble understanding texts, moving to a model of incremental learning during task performance, rather than the train/test/operate distinct cycles prevalent in ML practice today. Fourth, we hope that others will join us in extending these resources, including language resources beyond English. Just as there are FrameNets for other languages[7], extending NextKB to incorporate linguistic resources for other languages would both make it more useful to others, but also help shed light on how languages vary in how they package up conceptual structure (e.g. Gentner & Boroditsky, 2001).

[7] https://framenet.icsi.berkeley.edu/fndrupal/framenets_in_other_languages

Acknowledgments

This research was sponsored by the US Office of Naval Research under grant number N00014-20-1-2447 and the US Air Force Office of Scientific Research under award number FA95550-20-1-0091.

References

Allen, J. (1994) *Natural Language Understanding*, *2nd Edition*. Benjamin/Cummings, Redwood City, CA.

Bredeweg, B., Linnebank, F., Bouwer, A., Liem, J. (2009) Garp3: Workbench for qualitative modeling and simulation. *Ecological Informatics*, 4(5-6):263-281.

Bundy, A., Sasnauskas, G., & Chan, M. (2013). Solving Guesstimation Problems Using the Semantic Web: Four Lessons from an Application. *Semantic Web*, pp1-14.

Crouse, M., McFate, C.J., & Forbus, K. (2018). Learning to Build Qualitative Scenario Models from Natural Language. *Proceedings of QR 2018*, Stockholm.

Crouse, M., & Forbus, K. (2020). Towards High-Precision Understanding of Comparative Analysis Problems Expressed in Natural Langauge, *Proceedings of the 33rd International Workshop on Qualitative Reasoning*

de Kleer, J. (1984) How Circuits Work. *Artificial Intelligence*.

de Kleer, J. & Brown, J. (1984). A qualitative physics based on confluences. *Artificial Intelligence*, 24(1):7-84.

Falkenhainer, B. and Forbus, K. (1991). Compositional modeling: Finding the right model for the job. *Artificial Intelligence*, 51, 95-143.

Forbus, K. (1989). QPE: A study in assumption-based truth maintenance. *International Journal of Artificial Intelligence in Engineering*.

Forbus, K. (2019). *Qualitative Representations: How People Reason and Learn about the Continuous World*, MIT Press.

Forbus, K. and Falkenhainer, B. (1992). Self-Explanatory Simulations: Scaling up to large models. Proceedings of AAAI 1992, San Jose, California, 685-690.

Forbus, K. and Gentner, D. (1997). Qualitative mental models: Simulations or memories? *Proceedings of the Eleventh International Workshop on Qualitative Reasoning*, Cortona, Italy, June 3-6, pp. 97-104.

Forbus, K., Hinrichs, T., de Kleer, J., and Usher, J. (2010). FIRE: Infrastructure for Experience-based Systems

with Common Sense. *AAAI Fall Symposium on Commonsense Knowledge*, Arlington, VA.
Forbus, K.D., Whalley, P., Everett, J., Ureel, L., Brokowski, M., Baher, J. and Kuehne, S. (1999). CyclePad: An articulate virtual laboratory for engineering thermodynamics. *Artificial Intelligence*, 114, 297-347.
Gentner, D., & Boroditsky, L. (2001). Individuation, relativity and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215-256). Cambridge, UK: Cambridge University Press.
Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, 65, 263-297.
Hayes, P. (1985). Naïve Physics 1: Ontology for Liquids. In R. Hobbs & R. Moore, (Eds.). *Formal Theories of the Commonsense World*, Ablex, Norwood, NJ.
Klenk, K. & Forbus, K. (2009). Analogical Model formulation for AP Physics Problems. *Artificial Intelligence*, 173(18), 1615-1638.
Klenk, M., de Kleer, J., Bobrow, D., & Matei, I. (2015) But where do we start? Qualitative initialization problems with quantitative models. *Proceedings of QR 2015*.
Kuipers, B. (2003). *Qualitative Reasoning Modeling and Simulation with Incomplete Knowledge*. MIT Press.
Mitra, A., Baral, C., Bhattacharjee, A., & Shrivastava, I. (2019). A Generate-Validate Approach to Answering Questions about Qualitative Relationships". arXiv:1908.03645.
Paritosh, P. (2004) Symbolizing Quantity. *Proceedings of CogSci 2004*, Chicago.
Paritsh, P. K. (2007). *Back of the Envelope Reasoning for Robust Quantitative Problem Solving.* (Tech. Rep. No. NWU-EECS-07-11). Doctoral dissertation, Northwestern University, Department of Elecrtrical Engineering and Computer Science, Evanston, Illinois.
Paritosh, P. & Forbus, K. (2005). Analysis of Strategic Knowledge in Back of the Envelope Reasoning. *Proceedings of AAAI-2005*.
Paritosh, P. and Klenk, M. (2006). Cognitive Processes in Quantitative Estimation: Analogical Anchors and Causal Adjustment. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, Vancouver.
Struss, P., Sterling, R., Febres, J., Sabir, U., Keane, M. (2014) Combining engineering and qualitative models to fault diagnosis in air handling units. *Proceedings of ECAI-2014*. Pp. 1185-1190.
Tafjord, O., Clark, P., Gardner, M., Yih, W., & Sabharwal, A. (2019). QuaRel: A dataset and models for answering questions about qualitative relationships. *Proceedings of AAAI 2019*.
Weld, D. (1990). *Theories of comparative analysis*. MIT Press, Cambridge, MA, USA