

On the Representation of Physical Quantities in Natural Language Text

Sven E. Kuehne (skuehne@northwestern.edu)
Qualitative Reasoning Group, Northwestern University
1890 Maple Avenue, Ste. 300
Evanston, IL 60201, USA

Abstract

In this paper we investigate the forms in which quantity information can appear in written natural language. Our focus is on physical quantities found in descriptions of physical processes, such as expansion, movement, or transfer. Using Qualitative Process Theory as our underlying formalism, we show how information extracted from natural language text corresponds to the five constituents of physical quantities. The results of this analysis can be used for the creation of interpretation rules and extraction patterns in NL systems.

Introduction

Ordinary people know a lot about the physical world around them. They know that water will eventually boil if you heat it on a stove, that a ball placed at the top of a steep ramp will roll down, and that a cup will overflow if you continue to pour coffee in it. When people talk and write about such phenomena in everyday language, references to continuous properties are usually part of these descriptions. From simple utterances like “*The coffee is hot*” to a more complicated comparison like “*The velocity of gas molecules is higher than the velocity of molecules in a liquid.*” being able to identify and extract the information about physical quantities is essential to understand these sentences. Using Qualitative Process Theory (Forbus, 1984) as the underlying formalism, we investigate the forms in which continuous properties can appear in written natural language. Our focus is on *physical* quantities found in descriptions of *physical* processes, such as expansion, movement, or transfer.¹

The way in which continuous parameters and processes are described in natural language is not accidental. Since Qualitative Process Theory is a formalism of how people reason about the physical world, the basic ideas of the

theory should be reflected in the language that people use to communicate their understanding of physical phenomena. This paper shows that the natural language descriptions of physical processes contain abundant information about the constituents of physical quantities. Moreover, the results of this study can be used in a variety of applications, such as grammatical rules of a parser or in the design of information extraction algorithms.²

Physical quantities

In Qualitative Process Theory, all physical changes in *continuous properties* are caused by physical processes. The identification of continuous parameters is therefore an essential step in the extraction of information about physical processes from natural language text. In an earlier analysis (Kuehne & Forbus, 2002) we presented a scheme for the extraction process that uses FrameNet-compatible representations (Baker, Fillmore, & Lowe, 1998; Fillmore, Wooters, & Baker, 2001) to capture information about physical processes. The examples presented draw from the same corpus material (Buckley, 1979; Maton et al., 1994; Moran & Morgan, 1994) used in our previous analysis. Our goal here is to show how information about continuous parameters can appear in natural language, and the ways in which this information corresponds to the following five constituents of physical quantities:

- The *Entity* is a uniquely named object or an instance of a process associated with the quantity. For example, the word ‘brick’ in the noun phrase ‘the temperature of the brick’ denotes an entity.³
- The *Quantity Type* specifies the kind of parameter. The word ‘temperature’ in the noun phrase ‘the temperature of the brick’ is a reference to a quantity type.
- The *Value* specifies the numerical or symbolic value of the property. The number ‘3’ in ‘3 liters of water’ or

¹ The findings of this analysis are applicable to other types of quantities as well. The framework of QP theory determines just determines kind of information we are interested in, i.e. constituents of a physical quantity. Abstract and conceptual quantities are often referred to metaphorically by words with a physical basis and require a different semantic interpretation. ‘The price is hot.’ is does not have anything to do with temperature, unlike ‘The water is hot.’ However, the techniques for the extraction of information about such quantities are essentially the same.

² Although we use the results of the analysis for exactly these purposes, the findings are presented in a general way and not limited to any particular type of grammar or pattern language.

³ The noun ‘brick’ actually refers a particular individual, maybe ‘brick32’, not the collection of all bricks.

the adjective ‘hot’ in ‘the hot ground’ are values associated with a quantity.

- The *Unit* specifies the physical units of the property. Example: The word ‘kilograms’ in ‘3 kilograms of lead.’ Units usually appear in combination with a numerical value or with a quantifier.
- The *Sign of the Derivative* specifies how the parameter is changing. In the sentence “The temperature is increasing.” the sign of the derivative is expressed by the word ‘increasing’, which indicates that the parameter is changing in a positive direction.

Only the first two of these five constituents are required to identify a physical quantity. The quantity type together with the entity are sufficient to talk about quantities like ‘the temperature of a brick’ or the ‘the flow rate of heat’. Values, units, and information about changes are optional and often not explicitly stated.

Entities and quantities types

We begin with a look at the forms commonly used in natural language descriptions to express information about the two required constituents of physical quantities, the entity and the associated quantity type. The remaining three constituents, i.e. values, units, and changes, will be discussed in the subsequent sections.

Explicitly referenced quantities

Natural language text can refer to physical quantities either directly or indirectly, depending on whether the type of the quantity is explicitly mentioned in the sentence. *Explicit references* can be found in nouns, verbs, and adjectives that are morphologically related to quantity types.

Nouns

The quantity type can be explicitly mentioned as a noun, together with one or more entities that it is associated with.

- (1) VOLUME flows from the *can* to the *ground*.
- (2) The TEMPERATURE of the *brick* is rising.

Sentence 1 contains information about two physical quantities, the volume of some substance in the can and on the ground. The quantity type ‘volume’ is associated with both locations, i.e. the ‘can’ and the ‘ground’. In (2) the quantity type ‘temperature’ is associated with a single entity.

The quantity type can also appear as the head of a compound noun. The remaining constituents of the compound noun can be treated as information about a specialization of the quantity type. For example, in (3) the quantity type ‘radiation heat’ is a specialization of ‘heat’; in (4) ‘heat energy’ is a type of ‘energy’.

- (3) RADIATION HEAT flows from the *heater*.
- (4) The HEAT ENERGY of the *water* increases.

Verbs

Verbs can refer to physical events as well as to quantity types associated with these events.⁴ The verb in sentence 5 appears as a direct reference to the quantity type ‘length’. Sentence 6 is slightly more complicated, because it allows two different interpretations. The obvious interpretation is to treat the verb as an explicit reference to a quantity, as it is in (5). In this case, the quantity type ‘heat’ is tied to both entities, the stove as the source of the heat flow and the kettle as the destination of the heat flow.

- (5) The press LENGTHENS the *iron beam*.
- (6) The stove HEATS the *kettle*.

Alternatively, (6) can be interpreted as an increase in temperature of the kettle caused by the stove. Even though the quantity type ‘temperature’ is not mentioned in the sentence, we might infer that heating the kettle also increases the temperature of the kettle. This is an inference that most readers of such a descriptions will readily draw, and it coincides with the kind of conclusions that are supported by QP Theory.

Adjectives

Certain adjectives can refer to quantity types directly, if the adjective is morphologically related to a quantity type. For example, in (7) the adjective ‘denser’ refers to the quantity type ‘density’. The quantity type in this sentence is associated with both entities, i.e. ‘iron’ and ‘wood’.

- (7) *Iron* is DENSER than *wood*.

Implicitly referenced quantities

While the quantity types in explicitly referenced quantities are usually easy to identify and extract, *implicit references* to quantities are more difficult to figure out. Implicitly referenced quantities do not mention a quantity type. Instead, the reader has to use the contextual information provided by the sentence as well as available background knowledge. The following examples show how nouns, verbs, adjectives, and adverbs can determine a quantity that is not explicitly mentioned in a sentence.

Verbs

A quantity type can be implicitly referenced by a verb that describes a physical process, e.g. movement, expansion, or transfer. The sentence in which the verb occurs usually

⁴ Events such as the increase or decrease of a parameter, e.g. the temperature of a brick, can be involved in an instance of a physical process. For one linguistic perspective on actions, processes, and events, see (Parsons, 1990).

provides additional contextual information for the interpretation of the implicitly referenced quantity.

(8) As the temperature rises, the *liquid* EXPANDS.

The verb 'expand' in (8) indicates that something is changing in different physical dimensions, i.e. in length, area, or volume. For the three-dimensional entity 'liquid' the appropriate quantity type is therefore 'volume'. The verb also includes implicit information about a positive change in the quantity, i.e. an increase in volume of the liquid, which we will address later.

Adjectives

The quantity type can be implicitly referenced by certain adjectives. For example, the quantity type described by the adjective 'hot' in (9) is 'temperature'. The comparative also encodes the ordinal relationship between the quantities associated with the two entities, i.e. the fact that the temperature of the stone is greater than the temperature of the water. Similarly, the quantity type expressed by 'lighter' in (10) is 'weight'.

(9) The *stone* is HOTTER than the *water*.

(10) The *upper air masses* are LIGHTER than the *lower air masses*.

For a correct interpretation the relationship between the adjective and the associated quantity type has to be known. The fact that the adjective 'hot' is associated with 'temperature' is a fact learned by a human reader and is typically provided as background knowledge in NL systems.

Verb/Adverb combination

Quantity types can also be determined by combining verbs and adverbs. The quantity type referenced in (11) is the rate of movement, or 'velocity'. The adverb alone is not sufficient to determine the quantity type. Although 'faster' is generally associated with velocity, it just qualifies the rate of change, i.e. that something is happening in less time. There are cases in which the quantity type referenced by 'faster' is not velocity. For example, 'expanding faster' in (12) refers to the rate of expansion.

(11) The *gas molecules* are MOVING FASTER than *molecules in a solid*.

(12) *Liquid A* is EXPANDING FASTER than *liquid B*.

All these cases have one thing in common: the referenced quantity is a rate, most likely associated with a process referenced by the verb ('movement', 'expansion', 'decay').

Noun/Verb combination

This type of implicitly referenced quantity uses a noun/verb combination to refer to the rate of change of a quantity.

(13) The less heat is supplied, the slower the *temperature* RISES.

The quantity type in (13) is not 'temperature' but the rate of change in temperature, resulting from a change in the amount of heat. The combination of 'rises' and 'temperature' determines the quantity type, while the combination of the verb 'rises' and the adverb 'slower' gives the direction of change.

Noun/Adjective combination

The quantity type is only implicitly referenced by a combination of a noun and an adjective.

(14) The BIGGER the *surface* [is], the more heat is absorbed.

The quantity type in (14) is the size of the surface (not the surface itself) associated with an unnamed participant or the size of a participant 'surface'. The adjective 'bigger' refers to the quantity type 'size' (or 'area'). Since 'big' can also refer to the quantity type 'volume', the dimensionality of the entity determines the appropriate quantity type in this case.

Representation of values in physical quantities

Knowing the type of a quantity and the entity it is associated with enables us to talk and reason about it. A simple noun phrase such as 'the depth of the water' contains enough information to recognize it as a physical quantity, even without having any information about a particular value the quantity might have, the unit of that value, or the direction in which the quantity is changing. The following two sections examine how values and units of quantities appear in natural language text, and how changes in quantities can be identified.

There are three common types of references to values and units that can be found in natural language text: in the context of comparisons, as symbolic labels, and as quantitative information. We will discuss values and units together because units usually appear in combination with values.⁵

⁵ Units can appear separately from values in definitional statements, like "Length is measured in Meters." or, even more explicit, "The unit of power is the Watt."

Comparison

Values in the context of a *comparison* appear in sentences like “The brick is warmer than the plate.” The comparison orders the quantities, i.e. the temperature of the brick is greater than the temperature of the plate. However, it does not contain exact information about the possible values of the quantities. Even though the comparative ‘warmer’ might refer to a specific range of temperature, the exact values cannot be known or even guessed from the information provided by the sentence. The brick might be red hot, while the plate is frosted with ice.

It is impossible to determine how far the values associated with the two compared quantities are apart from each other. The only information that can be extracted from this sentence is the fact that the value of one quantity is greater than the other. With several of these comparisons along the same dimension, it is possible to identify the potential ranges of the values for particular quantities. For example, the temperature of a coffee is greater than the temperature of an ice cube, and it is lower than the temperature at the tip of a lit cigarette.

Symbolic labels

Values can also take the form of a *symbolic label* associated with an entity, e.g. “The brick is hot.” Even though the exact temperature of the brick is unknown, the adjective ‘hot’ suggests a certain temperature range. The range might be different depending on the context of the sentence. In refrigeration ‘hot’ might be in a very different range of temperatures than in the context of metallurgy.

Nouns that are associated with the adjective can impose restrictions on the range of the value in certain cases. For example, (Bierwisch, 1967) compares two simple sentences, “*The room is tall.*” and “*The space is tall.*”. In the first sentence the noun ‘room’ might restrict the average range of values for the height to those for a typical room, e.g. between 8 and 10 feet. Without further information, this kind of assumption is more difficult to make for the second sentence. Is the space a small compartment or a crawl space? Or is it the inside of a cathedral? The range of typical values would be quite different for these two cases.

Adjectives that represent a value are generally quantity-specific, i.e. they refer to a particular type of quantity as in the sentence “The brick is hot.” Alternatively, a quantity-neutral form could be used to express the same fact, e.g. “The temperature of the brick is high.”⁶

While adjectives and adverbs such as ‘hot’ or ‘slow’ generally refer to a range of values along a dimension, natural language also uses symbolic labels to refer to concrete values, i.e. particular points along a dimension. The noun phrase ‘boiling point of water’ usually refers to

the point where liquid water turns into steam and the value of approximately 212 degrees Fahrenheit. The noun phrase provides a label for this particular point.⁷

The structure for labels that describe limit points is not arbitrary. Usually the head of the noun phrase refers to a point on a scale (e.g. ‘point’, ‘barrier’), while the noun modifier is associated with a process, a dimension, or a quantity type (i.e. ‘boiling’, ‘sound’). These two parts are mandatory components of the label. Determining the quantity type and the dimension is difficult in many cases, e.g. we have to know that ‘boiling point’ is associated with ‘temperature’ and that ‘sound barrier’ actually refers to the speed of sound or velocity. Additionally, the label can take an optional complement phrase that restricts the compound noun. For example, the complement phrase ‘of water’ restricts the interpretation of boiling point to a particular substance. The key idea here is that the underlying mechanisms for handling limit points are essentially the same as those for symbolic references to intervals on a particular dimension.

Concrete numeric values and units

The most explicit form in which values can appear is *quantitative information*, i.e. by using concrete numeric information and units. For example, in (15) the quantity type (‘temperature’) is explicitly stated, together with detailed information about the numeric value (‘120’) and the unit (‘degrees Fahrenheit’).

(15) The temperature of the brick is *120 degrees Fahrenheit.*

Sentences that contain concrete numeric values and units typically do not use quantity-specific adjectives or adverbs instead of explicit references to the quantity type.

(16) *The water is 80 degrees Celsius *hot.*

(17) The water has 80 degrees Celsius.

Sentence 16 should be considered anomalous, because the adjective ‘hot’ provides at best redundant information in the form of a symbolic value. Units can refer indirectly to the quantity that they are associated with, as shown in (17). The association between units and quantity types (degrees Celsius as a unit for temperature) is a learned fact and has to be encoded as background knowledge.

Representations of changes in physical quantities

The values of physical quantities cannot always be treated as static information, because they can change while

⁶ The Cyc knowledge base (Lenat & Guha, 1989) handles values in a similar way. For example, the value # $\$$ Hot is the result of # $\$$ HighAmountFn of # $\$$ Temperature.

⁷ Note that the compound noun ‘boiling point’ would be an underspecified symbolic label because different substances have different boiling points. Other labels such as ‘sound barrier’ may not need the additional complement.

physical processes are active. The sign of the derivative indicates whether a quantity is changing and in which direction.⁸

The most obvious choice to express changes in the physical world is the use of verbs. For example, if water is flowing from one container into another, there are several ways of expressing the change of the amount of water in each container.

- (18) The amount of water in container A is decreasing, while the amount of water in container B is increasing.
(19) Water flows from container A to container B.

Although (18) and (19) might describe the same scenario, they are not equivalent. For example, (19) only implies a decrease of the amount of water in location A. It does not state this information explicitly. On the other hand, (18) implies a flow, without actually mentioning it. These distinctions are important for a semantic interpretation process, because the information that is directly available from the sentences is different.

Verbs with direct references to a quantity change

Verbs can directly refer to a change in a quantity and its direction, i.e. whether the quantity is increasing or decreasing, when the verb alone contains all the information about the change and the direction and we can therefore distinguish between verbs for positive and negative changes in quantities. For example, *gain*, *increase*, and *add* are verbs for positive changes, while *lose*, *decrease*, and *leak* are associated with negative changes.⁹ Some verbs belonging to this class also allow prepositional phrase as a complement, which is restricted to the particular direction of change indicated by the verb itself (e.g. 'add to' vs. '*add from').

- (20) The *brick* LOSES *heat* to the *room*.
(21) The *temperature of the water* is INCREASING.
(22) The *brick* GIVES OFF *heat*.

Some otherwise 'neutral' verbs can also fall into this class if they use specific particles to indicate a change in a quantity, as in (22).¹⁰

Verbs with directional prepositional phrases

Verbs associated with Transfer and Motion event do not contain a direct reference to changes in quantity. For

⁸ Information about changes in quantities can support other aspects of QP theory, e.g. in determining relationships between continuous parameters such as direct and indirect influences.

⁹ Another distinction could be made between verbs that can only be used with extensive quantities. For example, heat can be *added*, while temperature cannot.

¹⁰ The particle has to agree with the complement structure of verb. For example, the verb phrase '*gives in' cannot take 'heat' as its argument.

example, verbs like *flow* or *move* indicate a transfer of something between two physical or conceptual locations, but they do not contain information about the actual direction of the change. Instead, this information is provided by directional prepositional phrases attached to the verb. The description of the transfer can be complete when both the source and the destination are identified by prepositional phrases, as in (23), or partial when only one of the directional prepositional phrases is attached, as in (24) and (25).

- (23) Heat is *transferred* FROM inside the house TO the outdoors.
(24) Energy is *moved* TO a new location.
(25) The *fan* moves heat away FROM the processor.

Verbs in combination with quantity-specific adverbs

Quantity-specific adverbs can determine the change in a quantity in conjunction with a verb. Analogous to verbs with direct reference to a quantity change, the combination of verbs and quantity-specific adverbs can be associated with a decrease in a quantity, as in (26) or with an increase, as in (27).

Similar to the interpretation of the quantity type from verb/adverb combinations, there are cases in which the same adverb can refer to an increase (or a decrease) of a particular quantity type, depending on the verb with which it is used. For example, in the context of (27), the adverb 'faster' would indicate a positive change in the velocity of the molecules, while in (28) it will indicate an increase in the rate at which a substance dissolves.

- (26) The *glass* is COOLING FASTER.
(27) The *molecules* are MOVING FASTER.
(28) The *substance* DISSOLVES FASTER.

Nouns with direct references to change

Nouns provide another way of describing changes in physical quantities. They can be divided into similar classes as verbs, i.e. nouns with direct references to a change in a quantity, and nouns that use directional prepositional phrases.

Nouns can directly refer to a change in a quantity, and analogous to verbs they can be divided into nouns that refer to positive, as in (29), and negative changes, as in (30).

- (29) The INCREASE in *temperature* is significant.
(30) The DECREASE in *pressure* caused a failure.

Nouns with directional prepositional phrases

Similar to verbs of the Transfer and Motion domain, the corresponding nouns will also need directional prepositional phrases to describe changes in a quantity.

Again, the information about the transfer can be complete, as in (31) or partial as in (32).

- (31) The *flow* of oxygen FROM the tank TO the capsule is blocked.
(32) The *transfer* of heat TO the kettle has been completed.

Discussion

Parts of our current research are concerned with the design of a controlled language for describing physical phenomena. One important aspect in the development of such a language is the goal to reduce possible syntactic and semantic ambiguity. The identification of patterns used for references to continuous parameters in natural language is an essential part of the semantic interpretation process, which must include the detection of directly referenced quantities as well as indirect references.

Research on the lexical semantics of adjectives has tried to establish taxonomies for the different semantic categories of adjectives (see Raskin & Nirenburg (1995) for an overview). Several of these taxonomies focus on the class of adjectives that we are most interested in for extracting information about physical quantities, i.e. qualitative (scalar, gradable) adjectives (Dixon, 1991; Frawley, 1992). From our perspective, using the semantics of Qualitative Process Theory, the taxonomies suggested by Dixon and Frawley are inconsistent. The breakup of types and subtypes appears to be arbitrary, because several of the types of quantities can be collapsed into a single type. In Dixon's taxonomy the adjectives of the 'speed' and 'physical property' types are separated from those classified as 'dimension'. Similarly, 'age' and 'value' are listed as separate types.

Many quantity-specific adjectives and adverbs form opposing pairs for the same quantity type along a single dimension. For example, 'tall' is the opposite of 'short' for the quantity type 'height', and 'wide' the opposite of 'narrow' for the quantity type 'width' (see Bierwisch (1967, 1989) and Kennedy (2001) for a detailed analysis of polar adjectives). For certain quantity types we can identify not just a single opposing pair but a set of quantity-specific adjectives. For the quantity type 'temperature' we can find adjectives such as 'warm', 'cool', 'tepid', and variations such as 'lukewarm' as references besides just 'hot' and 'cold'. It is an interesting question to speculate why this variety of quantity-specific adjectives exists for some quantity types but not for others. Frequent use or familiarity with the concept 'temperature' cannot explain this fact alone.

Understanding the connections between Qualitative Process Theory and natural language is important for understanding the general cognitive plausibility of qualitative models. It will also give us greater insight into how results from qualitative reasoning can be

communicated back to human users in an intuitive way – by using natural language.

Acknowledgements

I would like to thank Ken Forbus and Dedre Gentner for insightful comments on this paper, as well as Praveen Paritosh and Chris Kennedy for interesting discussions on the topic. This research was supported by the Artificial Intelligence program of the Office of Naval Research.

References

- Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998). *The Berkeley FrameNet Project*. In: Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics (COLING-ACL 98), Montreal, Canada.
- Bierwisch, M. (1967). Some semantic universals of German adjectives. *Foundations of Language*, 3, 1-36.
- Bierwisch, M. (1989). The Semantics of Gradation. In M. Bierwisch & E. Lang (Eds.), *Dimensional Adjectives* (pp. 71-261). Berlin, Germany: Springer-Verlag.
- Buckley, S. (1979). *From Sun Up to Sun Down*. New York: McGraw-Hill.
- Dixon, R. M. W. (1991). *A New Approach to English Grammar, on Semantic Principles*. Oxford, England: Clarendon Press.
- Fillmore, C. J., Wooters, C., & Baker, C. F. (2001). *Building a Large Lexical Databank Which Provides Deep Semantics*. In: Proceedings of the Pacific Asian Conference on Language, Information, and Computation, Hong Kong, China.
- Forbus, K. D. (1984). Qualitative Process Theory. *Artificial Intelligence*, 24, 85-168.
- Frawley, W. (1992). *Linguistic Semantics*. Hillsdale, NJ: Erlbaum.
- Kennedy, C. (2001). Polar Opposition and the Ontology of 'Degrees'. *Linguistics and Philosophy*, 24(1), 33-70.
- Kuehne, S. E., & Forbus, K. D. (2002). *Qualitative Physics as a component in natural language semantics: A progress report*. In: Proceedings of the Twenty-fourth Annual Conference of the Cognitive Science Society, George Mason University, Fairfax, VA.
- Lenat, D. B., & Guha, R. V. (1989). *Building large knowledge-based systems : representation and inference in the Cyc project*. Reading, MA: Addison-Wesley.
- Maton, A., Hopkins, J., Johnson, S., LaHart, D., McLaughlin, C. W., Warner, M. Q., & Wright, J. D. (1994). *Heat Energy* (annotated teacher's ed.). Englewood Cliffs, NJ: Prentice Hall.
- Moran, J. M., & Morgan, M. D. (1994). *Meteorology - The Atmosphere and the Science of Weather* (4th ed.). New York, NY: Macmillan College Publishing.
- Parsons, T. (1990). *Events in the Semantics of English*. Cambridge, MA: MIT Press.
- Raskin, V., & Nirenburg, S. (1995). *Lexical Semantics of Adjectives: A Microtheory of Adjectival Meaning* (Technical Report MCCC-95-288). Las Cruces, NM: New Mexico State University.