

A Tradeoff between Compositionality and Complexity in the Semantics of Dimensional Adjectives

Geoffrey Simmons

Graduiertenkolleg Kognitionswissenschaft

Universität Hamburg

Bodenstedtstr. 16

D-W-2000 Hamburg 50

Germany

e-mail: simmons@bosun2.informatik.uni-hamburg.de

Abstract

Linguistic access to uncertain quantitative knowledge about physical properties is provided by **dimensional adjectives**, e.g. *long-short* in the spatial and temporal senses, *near-far*, *fast-slow*, etc. Semantic analyses of the dimensional adjectives differ on whether the meaning of the differential comparative (*6 cm shorter than*) and the equative with factor term (*three times as long as*) is a compositional function of the meanings the difference and factor terms (*6 cm* and *three times*) and the meanings of the simple comparative and equative, respectively. The compositional treatment comes at the price of a meaning representation that some authors ([Pinkal, 1990], [Klein, 1991]) find objectionally unparsimonious. In this paper, I compare semantic approaches by investigating the complexity of reasoning that they entail; specifically, I show the complexity of constraint propagation over real-valued intervals using the Waltz algorithm in a system where the meaning representations of sentences appear as constraints (cf. [Davis, 1987]). It turns out that the compositional account is more complex on this measure. However, I argue that we face a tradeoff rather than a knock-down argument against compositionality, since the increased complexity of the compositional approach may be manageable if certain assumptions about the application domain can be made.

TOPIC AREAS: semantics, AI-methods in computational linguistics

1 Introduction

In the past decade, the field of knowledge representation (KR) has seen impressive growth of sophistication in the representation of uncertain quantitative knowledge about physical properties in commonsense reasoning and qualitative physics. The input to most of these systems is entered by hand, but some of them, especially those with commonsense domains involving spatial and temporal knowledge, are amenable to interaction by means of a natural language interface. Linguistic access to knowledge about properties such as durations, rates of change, distances, the sizes of the symmetry axes of objects, and so on, is provided by **dimensional adjectives** (e.g. *long-short* in the spatial and temporal senses, *fast-slow*, *near-far*, *tall-short*). In this paper, I will investigate two aspects of their semantics that have an impact on the quality of a KR system with an NL interface. One aspect is the **complexity of reasoning** entailed by their semantic interpretations. As an example, suppose that we have a text about the installation of new kitchen appliances that contains the following sentences:

- (1) a. The refrigerator is about 60 cm wide.
- b. The cupboard is about as deep as the refrigerator is wide.
- c. The kitchen table is about 5 cm longer than the cupboard is deep.
- d. The oven is about twice as high as the table is long.

We may view the relations expressed by these sentences as constraints on the measurements of the object axes (the width of the fridge, the depth of the cupboard, and so on), which are represented as parameters in a constraint system. Then constraint propagation, along with some knowledge about the

sizes that are typical for object categories, should allow us to derive the following sentences (among others) from (1):

- (2) a. The cupboard is about 60 cm deep.
- b. The kitchen table is longer than the refrigerator is wide.
- c. The kitchen table is short (for a kitchen table).
- d. The oven is about 70 cm higher than the cupboard is deep.
- e. The oven is high (for an oven).

The inferences from (1) to (2) are rather simple, but reasoning can become very complicated if a large number of parameters and constraints must be accounted for. As we will see below, the computational properties of this kind of reasoning are dependent on the types of relations that appear in the knowledge base. Thus in the present paper, I investigate the kinds of relations that appear in formal theories of the meanings of the following morphosyntactic constructions of dimensional adjectives:

- (3) a. **Positive**
 The board is long/short.
- b. **Comparative**
 The board is (6 cm) longer/shorter than the table is wide.
- c. **Equative**
 The board is (three times) as long as the table is wide.
- d. **Measurement**
 The board is 50 cm long.

This brings us to the second issue: the **compositionality** of meaning representations proposed for the sentences in (3). It is appealing from the viewpoint of theoretical linguistics to regard each of the morphosyntactic categories (positive, etc.) as lexical items with their own semantics, and to assume that the semantics of each sentence in (3) is a compositional function of the semantics of the morphosyntactic category and the semantics of the adjective stem. Compositional meaning representations may also be computationally more advantageous, since they can be computed very efficiently from syntactic representations (e.g. in unification-based formalisms).

Most formal theories of the meanings of adjectives attempt to fulfill this criterion of compositionality, but as we will see, they differ on a more far-reaching criterion: whether the meaning of the differential comparative (*6 cm shorter than*) and the equative with factor term (*three times as long as*) is a compositional function of the meanings the difference and factor terms (*6 cm* and *three times*) and the meanings of the simple comparative and equative, respectively. Although compositionality is generally regarded as a virtue in and of itself, some authors ([Pinkal, 1990], [Klein, 1991]) have objected to compositional treatments of difference and factor terms on the grounds

that they introduce an excessive amount of mathematical structure into our linguistic models.

In section 3, I will compare semantic representations that do and do not foresee a compositional treatment of difference and factor terms by analyzing the complexity of reasoning that they entail. In particular, I will investigate the complexity of constraint propagation in a system where the meaning representations appear as constraints. In this paradigm, uncertain quantitative knowledge is accounted for with real-valued intervals, a popular choice in KR systems, and constraint propagation is performed by the **Waltz algorithm** (which gets its name from David Waltz [1975]). Ernest Davis [1987] shows in his detailed analysis that the Waltz algorithm is one of the best choices for this task, for reasons that I will explain in section 3.¹

It turns out that the constraint propagation with the Waltz algorithm under the compositional approach is more complex; thus, we apparently face a **tradeoff between compositionality and complexity**. I argue in section 4 that this is indeed a tradeoff, since the non-compositional formation of meaning representations may be expensive, and the increased complexity of the compositional approach may be manageable, especially if certain assumptions can be made about the domain of physical properties being represented.

2 Compositionality in the Semantics of Adjectives

There is a vast amount of linguistic data on which a formal semantics of adjectives can be evaluated, such as the interaction of comparative and equative complements with scope-bearing operators: quantifiers, logical connectives, modal operators and negative polarity items (e.g. *John is taller than I will ever be*). A good theory must also account for the phenomenon of markedness, i.e. the semantic asymmetry of the antonyms (see [Lyons, 1977, Sect. 9.1]). However, I will ignore these issues in order to focus on the matter of compositionality. Thus I classify the existing theories of adjective meaning very coarsely as 'compositional' or 'non-compositional'. Note that these labels indicate *only* whether or not the treatment of difference and factor terms is compositional (in other respects, all of the theories mentioned below are compositional).

To begin with, I presuppose a component of dimensional designation that determines which property of an object is described by an adjective, thus

¹I have only recently become acquainted with Eero Hyvönen's "tolerance propagation" (TP) approach to constraint propagation over intervals (see [Hyvönen, 1992]), which in some circumstances can compute solutions that are superior to those of the Waltz algorithm, but at the price of increased complexity. I comment on this briefly in section 3.2.

Semantic Analyses of Dimensional Adjectives

Formal interpretations of (3)

a. Positive	$amount(length(board))\{\sqsupset / \sqsubset\}N_C(length(board))$
b. Comparative	$amount(length(board))\{\sqsupset / \sqsubset\}amount(width(table))$
c. Equative	$amount(length(board)) \sqsupseteq amount(width(table))$
d. Measurement	$amount(length(board)) = \langle 50, cm \rangle$

Table 1: Non-compositional approach

a. Positive	$amount(length(board))\{\sqsupset / \sqsubset\}D \pm N_C(length(board))$
b. Comparative	$amount(length(board))\{\sqsupset / \sqsubset\}D \pm amount(width(table))$
c. Equative	$amount(length(board)) \sqsupseteq n \times amount(width(table))$
d. Measurement	$amount(length(board)) = \langle 50, cm \rangle$

Table 2: Compositional approach

determining that *short conference* describes a duration but *short stick* describes the length of the stick's elongated axis. Each class of properties (duration, length, etc.) is assumed to be associated with a set of degrees reflecting their magnitudes. I will simply use the function expression $amount(p(x))$ to denote the degree to which entity x exhibits property p . Each set of degrees is assumed to be ordered, and I will use the symbols \sqsupset and \sqsubset for the ordering relation. Most authors assume measurement theory ([Krantz *et al.*, 1971]) as the axiomatic basis in the formal semantics of linguistic measurement expressions (cf. [Klein, 1991]). For measurement expressions such as *3 cm*, I simply use a tuple $\langle 3, cm \rangle$ denoting a degree. Finally, I follow [Bierwisch, 1989] in using the symbol $N_C(a)$ for the 'norm' expected for amount a in context C . This reflects the usual assumption that the positive expresses a relation to a context-dependent standard. In this paper, I will restrict my attention to norms that are typical for the categories named in the sentence, such as *tall for an adult Dutchman*, *slow for a sports car*, etc.²

The class of theories that I am referring to as 'non-compositional' include those of [Cresswell, 1976], [Hoeksema, 1983] and [Pinkal, 1990], who propose formulas similar to those in Table 1 as interpretations of the sentences in (3). The relation used in place of the expression $\{\sqsupset / \sqsubset\}$ is \sqsupset for the unmarked case (e.g. *tall*) and \sqsubset for the marked case (*short*).³

²Clearly, there are many other kinds of norms. *Jan is tall* may mean tall for his age, taller than I expected, etc. [Sapir, 1944] is still one of the best surveys of the norms employed in natural language, while Bierwisch has a more modern analysis.

³Of course, Tables 1 and 2 are strong simplifica-

I call this approach non-compositional because interpretations of the differential comparative (*6 cm longer than*) and of the equative with factor term (*three times as long as*) are not derivable from the formulas shown in lines (b) and (c) (the same can be said of [Kamp, 1975] and [Klein, 1980]).

The compositional approach is taken by [Hellan, 1981], [von Stechow, 1984] and [Bierwisch, 1989], whose renderings of (3) are, in simplified form, something like those in Table 2. The symbol ' \pm ' is + in the unmarked case and - in the marked case, and ' \times ' stands for scalar multiplication.⁴

In the case of the positive and the ordinary comparative, the difference term D is existentially quantified, as is the factor term n in the case of the ordinary equative (with the additional condition that n is greater than or equal to one). But if the difference or factor term is realized in the sentence surface, then its contribution to (b) and (c) in Table 2 is embedded compositionally.⁵

tions that fail to reflect important differences between the authors mentioned that are unrelated to the issue of compositionality.

⁴In measurement theory, the '+' operation is interpreted as concatenation in the empirical domain, and scalar multiplication is interpreted as repeated concatenation. Krantz *et al.* [1971] show that under proper axiomatization, concatenation is homomorphic to addition on the reals.

⁵Bierwisch [1989] differs from the other authors advocating a compositional approach in that he does not assume the interpretation of the equative shown in Table 2. He points out (p. 85) that this analysis does not account for the fact that the equative is norm-related in the unmarked case: *Fritz is as short as Hans* presupposes that Fritz and Hans are short. Moreover, it is not clear whether this approach can capture the duality of

For the computational analysis, we will need to classify the relations shown in Tables 1 and 2, since these relations form the input to a knowledge base. But to do so, we must first decide what sorts of entities the difference and factor terms denote. I assume that they do not denote constants, since we may be just as uncertain of their magnitudes as we are of the other magnitudes mentioned in the sentences. Thus it should be possible to treat each of the mini-discourses in (4)-(6) in a similar fashion:

- (4) a. The board is 90 to 100 cm long.
b. In fact, it is about 95 cm long.
- (5) a. The board is longer than the table is wide.
b. In fact, it is about 6 cm longer.
- (6) a. The board is five to ten times as long as the table is wide.
b. In fact, it is about seven times as long.

The information given in (b) in (4)-(6) can be accounted for by simply modifying the terms introduced in (a). Hence, the difference and factor terms, like the 'amount' terms in Tables 1 and 2, denote uncertain quantities whose magnitude may be constrained by sets of sentences. I will refer to these terms generally as 'parameters'.

With this assumption, we can classify the relations in Tables 1 and 2 as follows:

- (7) **Non-compositional**
 - a. Ordering relations
(Positive, Comparative, Equative)
 - b. Linear relations
of the form $amount(x) + D \sqsupset amount(y)$
(Differential Comparative)
 - c. Product relations
of the form $n \times amount(x) \sqsupseteq amount(y)$
(Equative with factor term)
- (8) **Compositional**
 - a. Linear relations
(Positive, Comparative, Differential Comparative)
 - b. Product relations
(Equative with & without factor term)

In both approaches, measurements simply serve to identify the degree to which an object exhibits the property in question.

Under the compositional approach, it is possible to assume a single semantic representation in the lexicon for each adjective stem and each morphosyntactic category such that the formulas in Table 2 are generated from those lexical entries. Bierwisch [1989], for example, proposes lexical entries of the following form for each dimensional adjective:

$$\lambda c \lambda x [amount(p(x)) = (v \pm c)]$$

comparatives and equatives: *Fritz is taller than Hans* should be semantically equivalent to *Hans is not as tall as Fritz*. However, Bierwisch does assume a representation like this for equatives with realized factor terms.

where c is a difference value and v is a comparison value (see [Bierwisch, 1989] for details).

But the elegance of the compositional approach comes at the price of lexical semantic representations that include addition and multiplication operators, which is precisely what Pinkal [1990] and Klein [1991] have criticized: they find the assumption of mathematical operations as basic constituents of lexical meaning uncomfortably strong. This is one of the reasons why Pinkal proposes separate lexical entries for each morphosyntactic form of an adjective.

3 The Complexity of Constraint Propagation

The objection to the complexity of the lexical meaning representations required for the compositional approach appeals to intuitions of parsimony, and is in part a matter of philosophical opinion that may be difficult to resolve. Perhaps a decision could be made on the basis of psycholinguistic experimentation, but I will pose a more utilitarian question in this section by examining whether the increase in representational complexity in the transition from Table 1 to Table 2 entails an increase in the computational complexity of reasoning for a knowledge base containing those representations. The reasoning paradigm to be investigated is constraint propagation (sometimes called constraint satisfaction) over real-valued intervals.

Intervals are intended to account for uncertainty in quantitative knowledge. For example, the measurement of a parameter at 20 units on some scale with a possible measurement error of ± 0.5 units is represented as [19.5, 20.5], to be interpreted as meaning that the unknown measurement value in question lies somewhere in the set $\{x | 19.5 \leq x \leq 20.5\}$. Additional knowledge about the relations that hold between parameters constrains their possible values to smaller sets (hence the term 'constraints' for the propositions in a knowledge base expressing such relations).

Constraint propagation over intervals has been applied in spatial reasoning ([McDermott and Davis, 1984; Davis, 1986; Brooks, 1981; Simmons, 1992]), temporal reasoning (e.g. [Dean, 1987; Allen and Kautz, 1985]) and in systems of qualitative physics (see [Weld and deKleer, 1990; Bobrow, 1985]). Intervals have a very obvious weakness in that the highly precise choice of endpoints can rarely be well-motivated in natural domains such as these. In particular, the reasoner may draw very different inferences, e.g. about whether two intervals overlap, if the endpoint of some interval is changed by what seems to be an insignificant amount. Thus, as McDermott and Davis [1984] note, such a system must not only be able to report *whether* they overlap, but also "how close" they come to overlapping.

If they do come close ..., then ...[the

reasoner] must decide whether to act on the suspect information or work to gather more, which is really the only interesting decision in a case like this. Eventually, when all possible information has been gathered, if things are still close to the borderline then a decision maker must just use some arbitrary criterion to make a decision. We don't see how anyone can escape this. [McDermott and Davis, 1984, p. 114]

A formalism such as fuzzy logic attempts to alleviate the problem of sharp borderlines by using infinitely many intermediate truth values for vague predicates. I happen to have reservations about the adequacy of fuzzy logic for this task⁶, but I have chosen to study constraint propagation mainly because its computational properties are well-researched and are attractive for applications in which the potential overprecision of endpoints can be tolerated. Thus it provides a sound basis for comparing the semantic analyses presented in section 2.

3.1 Syntax and Semantics

In the following, I briefly review some definitions from [Davis, 1987, Appendix B] (with slight modifications)

Syntax

Assume a set of symbols $X = \{X_1, \dots, X_p\}$ called **parameters**. A **label** is written $[x_-, x_+]$ with real numbers $0 \leq x_- \leq x_+$; the symbol ∞ may also be used for x_- and x_+ .⁷ A **labelling** L for X is a function from parameters to labels. If L is understood, we write $X_i \doteq [x_-, x_+]$ for $L(X_i) = [x_-, x_+]$. A **constraint** is a formula over parameters in X in some accepted notation (e.g. $X_1 \times X_2 = X_3$ or $p \leq -X_1 + X_2 + X_3 \leq q$). A **constraint system** $C = \langle X, C, L \rangle$ consists of a set X of parameters, a set C of constraints over X , and a labelling L for X .

Semantics

A **valuation** V for X is a function from the parameters to reals. The **denotation** of a label $[x_-, x_+]$ is the set $D([x_-, x_+]) = \{x | x_- \leq x \leq x_+\}$ if $x_+ \neq \infty$, $D([x_-, \infty]) = \{x | x_- \leq x\}$ if $x_- \neq \infty$, $D([\infty, \infty]) = \{\infty\}$ otherwise. A labelling L is interpreted as restricting the set of possible valua-

⁶This is not because I object to the notion of truth measurement, but rather because I believe that the fuzzy logicians' assumption that the connectives of a logic of vagueness are truth functional is contradicted by the facts of human reasoning about vague concepts (as argued by [Pinkal, to appear]). In my opinion, a formalism for truth measurement would have to be more like probability theory.

⁷I assume the non-negative reals for simplicity, because most of the physical properties mentioned in the examples have non-negative measurement scales. Even some of the exceptions, such as the common temperature scales, are in fact equivalent to a scale of non-negative values.

tions for X to those V such that for all $X_i \in X$, if $L(X_i) = [x_-, x_+]$, then $V(X_i) \in D([x_-, x_+])$. Thus we may view L as denoting a set of valuations on the parameters; we refer to this set as $V(L)$.

A constraint C_j denotes the largest set of valuations that are consistent with the relation expressed by C_j ; call this set $V(C_j)$.

3.2 Constraint Propagation Algorithms

The task of a constraint propagation algorithm (CPA) is to tighten the interval labels in an attempt to either (1) find a labelling that is just tight enough to be consistent with the constraints and initial labelling, or (2) signal inconsistency. Constraint propagation separates a stage of **assimilation**, during which intervals are tightened, from **querying**, during which the tightened values are reported. It is also possible to infer previously unknown relations between the parameters in the querying stage by inspecting the tightened intervals. This method of reasoning may be applied in the linguistic application under study, for example to derive the sentences in (2) above from (1).

A CPA is **sound** if $V(C_1) \cap \dots \cap V(C_n) \cap V(L_1) \subseteq V(L)$ for every labelling L returned by the algorithm, where $\{C_1, \dots, C_n\}$ is the set of constraints in the system and L_1 is the initial labelling. It is **complete** if $V(L) \subseteq V(C_1) \cap \dots \cap V(C_n) \cap V(L_1)$ for every L that it returns. In other words, the algorithm is sound if it does not eliminate any values that are consistent with the starting state of the system, and complete if it returns only such values.

As we will see, CPA's for intervals can only be complete under very restricted circumstances. Thus Davis defines a weaker form of completeness for the assimilation process. A CPA is **complete for assimilation** if every labelling L that it returns assigns labels $X_i \doteq [x_-^i, x_+^i]$ such that if $V_i(X_i) \in D([x_-^i, x_+^i])$, then $V_i \in V(C_1) \cap \dots \cap V(C_n)$. That is, the label assigned to each parameter accurately reflects the range of values it may attain given the constraints in the system.

The Waltz algorithm, which is stated below, is superior to many other CPA's in these respects. It is a sound algorithm, unlike the Monte Carlo method used by [Davis, 1986] and the hill-climber used by [McDermott and Davis, 1984]. Moreover, for constraint systems containing restricted types of constraints, the Waltz algorithm is complete for assimilation and terminates very quickly. In contrast, Davis reports that the hill-climbers used by [McDermott and Davis, 1984] were prohibitively slow and unreliable.

The algorithm is based on an operation called **refinement**, defined as follows. Given a constraint C_j , a parameter X_i appearing in C_j , and labelling L define:

$$REFINE(C_j, X_i, L) = \{V'(X_i) | V' \in V(L) \cap V(C_j)\}$$

Relation	Time Complexity	Completeness	Complexity of Complete Solutions
Order	$O(pc)$	Assimilation	$O(p^3)$
Unit Linear Inequality	$O(pS)^*$	Incomplete	As hard as linear programming
Product	$O(pS)^\dagger$	Incomplete	NP-hard

Table 3: Complexity of the Waltz algorithm for various systems of relations (from [Davis, 1987] and [Simmons, 1993])

p = number of parameters, c = number of constraints
 S = size of the system (the sum of the lengths of all of the constraints)
 $*$ May not terminate if the system is inconsistent
 † Terminates in arbitrarily long (finite) time if the system is inconsistent
 ‡ May not terminate if the solution is inadmissible (see text)

This is the set of values of X_i that consistent with both the labelling and the constraint.

The two refinement operators for a constraint C_j and parameter X_i are functions from labellings to labellings, written $R^-(X_i, C_j)$ and $R^+(X_i, C_j)$. If $L(X_i) = [x_-^i, x_+^i]$, then $R^-(X_i, C_j)(L)$ is formed by replacing x_-^i in L with the lower bound of $REFINE(C_j, X_i, L)$, and $R^+(X_i, C_j)(L)$ is formed by replacing x_+^i in L with the upper bound of $REFINE(C_j, X_i, L)$. We say that these refinements are based on C_j . If the upper and lower bounds of $REFINE$ are computable, then refinement is by definition a sound operation.

For a constraint system $\mathcal{C} = \langle X, \{C_1, \dots, C_n\}, L \rangle$, L is quiescent for a set of refinement operators $R = \{R_1, \dots, R_n\}$ if $R_1(L) = \dots = R_n(L) = L$. The solution to \mathcal{C} (if it exists) is the labelling L' denoting the largest set of valuations $V(L') \subseteq V(L) \cap V(C_1) \cap \dots \cap V(C_n)$ such that L' is quiescent for any set of refinements based on the constraints in the system. If no such solution exists, then \mathcal{C} is inconsistent.

The Waltz algorithm repeatedly executes refinements until the system is quiescent, and returns the solution (or signals inconsistency) if it terminates (cf. [Davis, 1987, p. 286]).

```

procedure WALTZ
  L ← the initial labelling
  Q ← a queue of all constraints
  while Q ≠ ∅ do
    begin remove constraint C from Q
    for each Xi appearing in C
      if REFINE(Xi, C, L) = ∅
        then return INCONSISTENCY
      else L ← the result of executing
        R-(Xi, C) and R+(Xi, C) on L
    for each Xi whose label was changed
      for each constraint C' ≠ C in which Xi appears
        add C' to Q
  end

```

Since refinement is a sound operation, the Waltz algorithm is sound. The completeness, termination and time complexity of the algorithm depends on

what kinds of relations appear as constraints in the system, and on the order in which constraints are taken off the queue. The results for systems consisting exclusively of one of the three kinds of relations mentioned in (7)-(8) in section 2 are given in Table 3, under the assumption that constraints are selected in FIFO order or a fixed sequential order (other orderings lead to worse results). Time complexity is measured as the number of iterations through the main loop of the algorithm. For comparison, Table 3 also gives the best known times for complete solutions to systems of such relations.⁸

In the linguistic application proposed here, the term S in Table 3 (the sum of the lengths of all of the constraints) is proportional to c (the number of constraints), since there are no more than three parameters in each constraint. Hence, $O(pS)$ is $O(pc)$ in this application.

Note that Table 3 gives results for linear inequalities with unit coefficients (of the form $p \leq \sum_i X_i - \sum_j X_j \leq q$, where no coefficients differ from 1 or -1). These are the only kind of linear inequalities under consideration in the linguistic application. In general, the Waltz algorithm breaks down if the system contains more complex relations, such as linear inequalities with arbitrary coefficients or product relations, since it may go into infinite loops even if the starting state of the system was consistent. Consider, for example, the set of constraints $\{n1 \times X = Y, n2 \times X = Y\}$ with the starting labels $n1 \doteq [1, 1]$, $n2 \doteq [2, 2]$, $X \doteq [0, 100]$ and $Y \doteq [0, 100]$. The system continually bisects the upper bounds of X and Y without ever being able to reach the solution, which

⁸Hyvönen's [Hyvönen, 1992] tolerance propagation (TP) approach is similar to the Waltz algorithm, but it uses a queue of solution functions from interval arithmetic [Alefeld and Herzberger, 1983] rather than refinement operations. The "global TP" method computes complete solutions, but at the price of increased complexity. In the "local" mode, tolerance propagation is very similar to the Waltz algorithm in its computational properties.

is $X \doteq [0, 0]$ and $Y \doteq [0, 0]$. Similarly, if the starting labels are $X \doteq [1, \infty]$ and $Y \doteq [1, \infty]$, then the lower bounds are continually doubled without reaching the solution $X \doteq [\infty, \infty]$ and $Y \doteq [\infty, \infty]$.

However, it is shown in [Simmons, 1993] that this happens *only* if the solution contains labels of this kind. Define a label as **admissible** if it is not equal to $[0, 0]$ or $[\infty, \infty]$; otherwise, it is **inadmissible**. A labelling L is admissible if it only assigns admissible labels; otherwise, L is inadmissible. Then it can be shown that if a system of product constraints is consistent and its solution is admissible, then the Waltz algorithm terminates in $O(pS)$ time. Moreover, if the system is inconsistent, the algorithm will find the inconsistency in finite but arbitrarily long time. Unfortunately, the proof is too long to include in the present paper, but a brief outline of the argument is given in the Appendix.

Systems with linear inequalities or product constraints are liable to enter infinite or very long loops if the starting state is inconsistent (or if the solution is inadmissible in the case of products). Davis [1987, p. 305-306] suggests a strong heuristic for detecting and terminating such long loops: stop if we have been through the queue p times (for p parameters). He is not clear on what he means by "having been through the queue x times", but I interpret him as meaning that we should stop if any constraint has been taken off the queue more often than p times. The rationale is the observation that in practice, most systems that do terminate normally seem to do so before this condition is fulfilled, much sooner than the worst-case time predicted by the complexity analysis. The reliability of such a heuristic is one of the topics of the next subsection.

3.3 Empirical Testing

The analytic results given in the previous subsection have left two important questions open:

- What is the complexity of constraint propagation if the system contains different kinds of constraints?
- How reliable is Davis' heuristic for terminating infinite (or very long) loops?

The first question lends itself to an analytic answer, but the results are not known at present. But we can seek empirical evidence by running the algorithm on mixed systems of constraints to see if the time to termination is significantly greater than the complexity expected for systems containing just the most complex type of relation in the system. If this does not happen for a number of representative systems, we may conjecture that the combination of constraints has not made the problem more complex. The second question can only be answered empirically, by testing whether the heuristic tends to terminate the algorithm too soon (i.e. whether it terminates refinement of systems that might have terminated normally in a short time).

Empirical investigations of these questions are reported in [Simmons, 1993], and described briefly here. To investigate the first question, the algorithm was run on a number of large, consistent constraint systems with admissible solutions in which the three types of constraints shown in Table 3 appeared in approximately equal numbers. On each run, the constraints in the initial queue were permuted randomly to suppress the possible effects of ordering. None of these runs required more time to termination than is predicted by the $O(pS)$ result for systems containing just unit linear inequalities or just product constraints.

To investigate the second question, I attempted to build consistent constraint systems with admissible solutions that are terminated by Davis' heuristic sooner than they would have been normally. It turns out that the algorithm runs to completion on almost all systems that were tested long before any constraint is taken off the queue p times, although there are systems for which refinement is terminated too soon on this heuristic. If the limit is increased by a constant factor, e.g. if assimilation is stopped after some constraint is processed $2p$ times, then the risk of early termination is greatly reduced.

In all, the empirical results on the open questions mentioned above have been encouraging. It is an admitted weakness of these tests, however, that they were performed on systems built by hand, not on constraint systems that occur "naturally" as part of an NL interface to a KR system.

4 Conclusions

The results of the previous section yield Tables 4 and 5 as the complexity of reasoning with the Waltz algorithm under the non-compositional and compositional approaches, respectively. These results depend in part on the fact that there is a maximum number of parameters in each constraint in the linguistic application. Measurements are modelled as predicate constraints, i.e. they simply impose interval bounds on some parameter. Intervals are also assumed to model the range of measurement values for the physical property that is typical for members of a category (e.g. the typical width of refrigerators), thus accounting for the norm used in the interpretation of positives. An important property of such "norm intervals" is that they may not be refined, at least not too much. This may be achieved by adding constraints imposing absolute upper and lower bounds on their ranges (cf. [Simmons, 1992]).

Although the worst-case time complexity in all cases turns out to be the same, the compositional approach is more complex for two reasons. First, the system is prone to enter infinite loops under the compositional approach if the starting state is inconsistent, or if the solution is inadmissible. Consistency cannot generally be guaranteed in the linguistic application under consideration, since the sentences in

Non-compositional

<i>Morphosyntactic Category</i>	<i>Relation</i>	<i>Time Complexity</i>	<i>Completeness</i>
Measurements	Predicate	trivial	Complete
Positive	Order	$O(pc)$	Assimilation
Comparative	Order	$O(pc)$	Assimilation
Equative	Order	$O(pc)$	Assimilation
Differential comparative	Linear Inequality	$O(pc)*$	Incomplete
Equative w/ factor term	Product	$O(pc)†$	Incomplete

Table 4: Complexity of reasoning under the non-compositional approach

Compositional

<i>Morphosyntactic Category</i>	<i>Relation</i>	<i>Time Complexity</i>	<i>Completeness</i>
Measurements	Predicate	trivial	Complete
Positive	Linear Inequality	$O(pc)*$	Incomplete
Comparative	Linear Inequality.	$O(pc)*$	Incomplete
Equative	Product	$O(pc)†$	Incomplete
Differential comparative	Linear Inequality	$O(pc)*$	Incomplete
Equative w/ factor term	Product	$O(pc)†$	Incomplete

Table 5: Complexity of reasoning under the compositional approach

p = number of parameters, c = number of constraints

* May not terminate if the starting state is inconsistent

†Terminates in arbitrarily long (finite) time if the system is inconsistent

‡May not terminate if the solution is inadmissible

a text may contain errors. Second, reasoning under the compositional approach is incomplete in all but the trivial case of measurements, whereas the non-compositional approach guarantees at least assimilation completeness for a subset of the parameters in the system. This means that under the compositional approach, the reasoner does not refine some intervals as tightly as it could have under the non-compositional approach.

These results may be taken as grounds for rejecting the compositional approach to the semantics of dimensional adjectives in the design of an NL interface to a KR system for quantitative knowledge. However, I do not believe that the compositional approach is contraindicated for all conceivable systems. In addition to the general theoretical appeal of compositional semantics, the compositional formation of meaning representations may be computationally more attractive in some cases (e.g. in unification-based formalisms). Thus if the non-compositional formation of semantic representations turns out to be too expensive, it may defeat the computational advantage gained in the reasoning process.

This is especially true if the weaknesses of the compositional approach do not turn out to be highly relevant in the specific application. For example, if the domain of physical properties being represented is such that a set of constraints requiring some parameter to be set to $[0, 0]$ or $[\infty, \infty]$ is unlikely to be encountered, and hence the solution is likely to be admissible, then the risk of infinite loops is reduced. Moreover, if Davis' heuristic for terminating infinite loops turns out to be reliable (which might be determinable by experimentation within the specific application), then inconsistencies need not be very damaging.

The incompleteness of reasoning under the compositional approach is unacceptable for an application if it is crucial that the inferred intervals contain *precisely* those values that are warranted by the constraints and the initial labelling. If a superset of those values can be accepted, however, then the compositional approach can be taken. Both approaches suffer a lack of what Davis calls **query completeness**: if the value of a term T is to be determined during the querying stage (i.e. after assimilation),

the system may return a superset of the values for T that are warranted by the constraints.

Thus an engineer building an NL interface to a system for reasoning about uncertain quantitative knowledge of physical properties must make a number of design decisions:

- How important are difference and factor terms in the linguistic material to be processed?

If difference and factor terms are so marginal that they may not occur at all, then the non-compositional approach is probably the better choice, due to its guarantee of termination and assimilation completeness.

- Does the compositional generation of lexical semantic representations have a significant advantage (computational or otherwise) over the non-compositional approach?
- Is it possible or likely for the measurement of some physical property to be exactly zero?

While there is probably no natural application in which the magnitude of some property can be infinitely large, there are different philosophies about the treatment of zero. In a system of temporal reasoning, for example, saying that some event has zero duration may be a way of saying that the event does not exist. But another policy might be to insist that no physical property is represented if it is not exhibited to a positive degree. If this assumption can be made, then the intervals $[0, 0]$ and $[\infty, \infty]$ are truly inadmissible, and hence one weakness of the compositional approach is diminished.

- Is it important that the precise range of permissible measurement values be inferred for each parameter, or can a superset of those values be useful?

If a superset of the possible values is acceptable, then the compositional approach can be chosen. Otherwise, the non-compositional approach must be taken.

By weighing the various answers to these questions, an engineer can stake out a position on the tradeoff and design a system with the power and efficiency most appropriate to his or her needs.

Acknowledgements

Thanks to Carola Eschenbach, Claudia Maienborn, Andrea Schopp, Heike Tappe and the referees for their comments on earlier versions of this paper. Thanks also to Longin Latecki for discussions about constraint propagation, and to Christopher Habel for encouraging me to pursue this work.

Appendix

In the following, the proof of the following theorem (from [Simmons, 1993]) is briefly outlined:

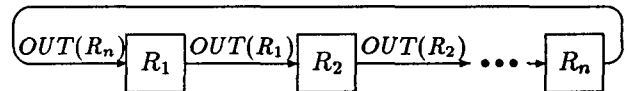
Theorem 1 *If a system of product constraints is consistent and its solution is admissible, then the Waltz algorithm brings it to quiescence in time $O(pS)$.*

Recall that a product constraint is of the form $\prod_i X_i = Y$, and that a labelling is admissible if it does not assign $[0, 0]$ or $[\infty, \infty]$ to any parameter. First we need some terminology defined in [Davis, 1987, Appendix B] (recall the definition of refinement operators in section 3.2 above).

For a refinement operator R , let $OUT(R)$ be the bound affected by R , and let $ARGS(R)$ be the set of bounds other than $OUT(R)$ that enter into the computation of $OUT(R)$. Given a labelling L , R is **active** on L if it changes L , i.e. if $L \neq R(L)$.

A series of refinement operators $\mathcal{R} = \langle R_1, \dots, R_n \rangle$ is active if each refinement in \mathcal{R} is active. We say that R_i is an **immediate predecessor** of R_j in \mathcal{R} if $i < j$, $OUT(R_i) \in ARGS(R_j)$, and for all k such that $i < k < j$, $OUT(R_k) \neq OUT(R_i)$. In other words, some argument of R_i has been set most recently in the series by R_j . We say that R_i **depends on** R_j if either $i = j$ or R_i depends on R_k and R_j is an immediate predecessor of R_k . Thus the dependence relation is the transitive and reflexive closure of the immediate precedence relation. We say that R_i depends on bound B if for some R_j , R_i depends on R_j and $B \in ARGS(R_j)$.

The series of refinements $\mathcal{R} = \langle R_1, \dots, R_n \rangle$ is **self-dependent** if R_n depends on $OUT(R_n)$, its own output bound. In other words, a series is self-dependent if the last bound affected by the series is also an argument to the first refinement in a chain of refinements in the precedence relation, as illustrated below.



Davis shows that such self-dependencies are potential infinite loops:

Theorem 2 *Any infinite sequence of active refinements contains an active, self-dependent subsequence ([Davis, 1987, Lemma B.15]).*

In [Simmons, 1993], it is shown that if any self-dependent sequence \mathcal{R} is active on the labelling of a system of product constraints, then a certain subsequence \mathcal{R}' of \mathcal{R} will be active infinitely many times. Moreover, on the m -th execution of each refinement R_i in \mathcal{R}' , there is a term T_i^m , where each $T_i^m \geq T_i^{m-1} > 1$, such that $OUT(R_i)$ is multiplied by:

$$\begin{cases} (T_i^m)^{-1}, & \text{if } OUT(R_i) \text{ is an upper bound} \\ T_i^m, & \text{if } OUT(R_i) \text{ is a lower bound} \end{cases}$$

It follows that upper bounds are refined so as to become arbitrarily small (asymptotically approaching zero), and that lower bounds become arbitrarily large, up to infinity.

Thus if there is any constraint C_i in the system that imposes a lowest value greater than zero on an

upper bound that is affected by a refinement operator in \mathcal{R}' , that bound will be refined often enough until it becomes inconsistent with C_i . Similarly, if any constraint C_u imposes a largest finite value on a lower bound that is affected by a refinement in \mathcal{R}' , then that bound will be refined until it becomes inconsistent with C_u . In both cases, the system is inconsistent.

If there are no such constraints, then it is consistent for upper bounds affected by \mathcal{R}' to be asymptotically close to zero and for lower bounds affected by \mathcal{R}' to be arbitrarily large. This can only be consistent if, in the case of upper bounds, the solution assigns $[0, 0]$ to the parameter in question, and in the case of lower bounds, the solution assigns $[\infty, \infty]$ to its parameter. Hence, the solution is inadmissible.

But according to Davis' result (Theorem 2), infinite loops must contain an active, self-dependent subsequence such as \mathcal{R} . It follows that if a system of product constraints is consistent and its solution is admissible, then the Waltz algorithm finds its solution in finite time. The time complexity result is a straightforward extension of Davis' analysis of unit linear inequalities (see [Simmons, 1993]).

References

- [Alefeld and Herzberger, 1983] G. Alefeld, J. Herzberger. Introduction to Interval Computations. Reading, MA: Addison-Wesley
- [Allen and Kautz, 1985] J. F. Allen, H. A. Kautz. A Model of Naive Temporal Reasoning. In: J.R. Hobbs, R.C. Moore (ed.): Formal Theories of the Commonsense World. Norwood, NJ: Ablex. 251–268
- [Bierwisch, 1989] M. Bierwisch. The Semantics of Gradation. In: M. Bierwisch, E. Lang (eds.): Dimensional Adjectives. Berlin et al.: Springer-Verlag. 71–261
- [Bobrow, 1985] D. Bobrow (ed.). Qualitative Reasoning about Physical Systems. Cambridge, MA: MIT Press. Reprinted from: *Artificial Intelligence* 24, 1984
- [Brooks, 1981] R. Brooks. Symbolic Reasoning among 3-D Models and 2-D Images. *Artificial Intelligence* 17, 285–348
- [Cresswell, 1976] M.J. Cresswell. The Semantics of Degree. In: B.H. Partee (ed.): Montague Grammar. New York: Academic Press. 261–292
- [Davis, 1986] E. Davis. Representing and Acquiring Geographic Knowledge. London: Pitman
- [Davis, 1987] E. Davis. Constraint Propagation with Interval Labels. *Artificial Intelligence* 32, 281–332
- [Dean, 1987] T. Dean. Large-Scale Temporal Data Bases for Planning in Complex Domains. In: Proceedings of the IJCAI-87. 860–866
- [Hellan, 1981] L. Hellan. Towards an Integrated Analysis of Comparatives. Tuebingen: Narr
- [Hoeksema, 1983] J. Hoeksema. Negative Polarity and the Comparative. *Natural Language and Linguistic Theory* 1, 403–434
- [Hyvönen, 1992] E. Hyvönen. Constraint reasoning based on interval arithmetic. *Artificial Intelligence* 58, 71–112
- [Kamp, 1975] J. A. W. Kamp. Two Theories about Adjectives. In: E. L. Keenan (ed.): Formal Semantics of Natural Language. Cambridge: Cambridge Univ. Press. 123–155
- [Klein, 1980] E. Klein. A semantics for positive and comparative adjectives. *Linguistics and Philosophy* 4, 1–45
- [Klein, 1991] E. Klein. Comparatives. In: A. von Stechow, D. Wunderlich (eds.): Semantics. Berlin: de Gruyter. 673–691
- [Krantz et al., 1971] D. H. Krantz, R. D. Luce, P. Suppes, A. Tversky. Foundations of Measurement. New York, London: Academic Press
- [Lyons, 1977] J. Lyons. Semantics. Vol. 1. Cambridge et al.: Cambridge Univ. Press
- [McDermott and Davis, 1984] D. McDermott, E. Davis. Planning Routes Through Uncertain Territory. *Artificial Intelligence* 22, 107–156
- [Pinkal, 1990] M. Pinkal. On the Logical Structure of Comparatives. In: R. Studer (ed.): Natural Language and Logic. Berlin: Springer. 146–167
- [Pinkal, to appear] M. Pinkal. Logic and Lexicon. On the Semantics of the Indefinite. Dordrecht: Kluwer. Translation by G. Simmons of M. Pinkal (1985): Logik und Lexikon. Berlin: de Gruyter
- [Sapir, 1944] E. Sapir. Grading: A Study in Semantics. *Philosophy of Science* 11, 93–116. Reprinted in: D. G. Mandelbaum (ed.) (1968): Selected Writings of Edward Sapir. Berkeley, Los Angeles: U. Calif. Press
- [Simmons, 1992] G. Simmons. Standardwissen über Normen: Zur konzeptuellen Analyse von Objekten. Master's thesis. Universität Hamburg.
- [Simmons, 1993] G. Simmons. Notes on Product Constraints. Report 22, Graduiertenkolleg Kognitionswissenschaft. Universität Hamburg
- [von Stechow, 1984] A. von Stechow. Comparing Semantic Theories of Comparison. *Journal of Semantics* 3, 1–77
- [Waltz, 1975] D. Waltz. Understanding line drawings of scenes with shadows. In: P.H. Winston (ed.), The Psychology of Computer Vision. McGraw-Hill, New York. 19–91
- [Weld and deKleer, 1990] D.S. Weld, J. deKleer (eds.). Qualitative Reasoning about Physical Systems. San Mateo, CA: Morgan Kaufman