

# **Be Brief and Vague! And how Bidirectional Optimality Theory allows for Verbosity and Precision.<sup>1</sup>**

**Manfred Krifka**

## **1. The Sad Story of the Metric System in America**

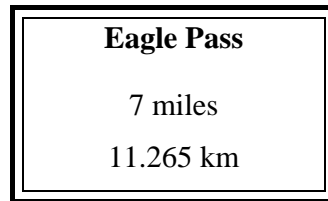
Given the beginnings of the United States of America, its sympathy with the French revolution and its rationalist attitude towards the institutions of society, one would have expected that it would have been one of the first nations to adopt the new metric system that was introduced in France in 1800. But the history of the attempts to do so is decidedly mixed. American Congress authorized the use of the metric system in 1866. In 1959, American measurements were defined in relation to the metric system. In 1968, the government ordered a study which was published three years later under the title “A Metric America: A Decision Whose Time Has Come”. The year 1975 then saw the Metric Conversion Act, leading to the establishment of the US Metric Board. Amended in 1988, it resulted in the Metric Program, an organization founded to support the various federal agencies, which are required since 1991 to file an annual report on their efforts to change to the metric system.

In spite of all these attempts, the United States of America are still the one major industrial nation that does not use the metric system as the predominant one. To this day, American schoolchildren have to count with miles that contain 1760 yards, yards that contain 3 feet, and feet that contain 12 inches. They have to memorize that an acre is 4840 square yards, and that a gallon contains 231 cubic inches. The costs of this are undoubtedly huge – they include, for example, the Mars mission of 1998 that failed because measurements were converted wrongly. Still, the general attitude of the American public towards the metric system is largely negative. There are websites with telling addresses like

[www.metricsucks.com](http://www.metricsucks.com), and it is not uncommon that common people suspect a secret communist, catholic or Jewish plot behind the attempts to go metric.

Why did the metric system not catch on? There are many reasons<sup>2</sup>. But one that cannot be taken lightly is that certain well-intended public relation attempts intended to familiarize the American people with the metric system just did not work. Since the Metric Conversion Act, road distances in National Parks are often given in miles and kilometers. And since then, travelers encounter signs like the following one:

(1)



It is not hard to see why road signs like (1) suggest that the metric system is something for intellectuals, or “rocket scientists”, far too unwieldy for everyday purposes.

## 2. Problems with precision

The problem with (1) leads us to the question: How much precision is enough? When can we stop being precise, relax and be a little vague? I won't have much to say about this in general, but I will have to say something about the relation of precision level and linguistic form.

Assume that the distance between Amsterdam and Vienna, measured as usual from city border to city border along the shortest connecting road path, is 965 kilometers. Now consider the following examples:

(2) A: *The distance between Amsterdam and Vienna is one thousand kilometers.*

B: *#No, you're wrong, it's nine hundred sixty-five kilometers.*

B's reaction strikes us as inadequate; what he says is true but pedantic. But not so in the following exchange, where A's utterance is actually closer to the truth.

- (3) A: *The distance between Amsterdam and Vienna is nine hundred seventy-two kilometers.*  
B: *No, you're wrong, it's nine hundred sixty-five kilometers.*

The road sign example (1) showed that the phenomena to be talked about here constitute a problem for translation. And indeed, if we translate (2) into the American measurement system, the oddness of the exchange vanishes (one has to know that 600 miles are 965 kilometers, and 621 miles are 1000 kilometers).

- (4) A: *The distance between Amsterdam and Vienna is six hundred and twenty-one miles.*  
B: *No, it's six hundred miles.*

It is quite obvious that the oddness of (2) cannot simply be stated in terms of truth conditions. Otherwise, the following exchange should be odd as well, but it isn't (if we disregard the fact that it is odd to render the distance of two cities with a precision of  $\pm 50$  meters in the first place!).

- (5) A: *The distance between Amsterdam and Vienna is one thousand point zero kilometers.*  
B: *No, you're wrong, it's nine hundred sixty-five kilometers.*

Examples like (2) can be multiplied at will. Consider the following:

- (6) A: *The number  $\pi$  is 3.14159.*  
B: *#No, it's 3.1415926535.*  
(7) A: *The number  $\pi$  is 3.1415926534.*  
B: *No, it's 3.1415926535.*

Clearly, the reaction of B in (6) is pedantic, but it is not pedantic in (7).

The first generalization that we can draw from examples (2) to (7) is that one should not increase the level of precision that was set by the first speaker. To say that the distance between Amsterdam and Vienna is one thousand kilometers sets the level of precision to  $\pm 50$  km. The speaker indicates with the choice of words that the only distance values that should be mentioned are 800 km, 900 km, 1000 km, 1100 km, and so on. Changing this level in the reaction is pedantic. Similarly, to say that the value of  $\pi$  is 3.14159 sets the level of precision to 5 decimal

points; again changing this level is pedantic. If the first speaker starts out with a higher level of precision, as in (3), (5) or (7), the reactions of the second speaker at the same level are not considered pedantic. The translation problems we encountered with (1) and (5) show that the level of precision does not translate under a “precise” translation of terms. A better translation of *7 miles* into the metric system would have been *11 kilometers*, and there is no conceivable translation of (2) into mile measurements that would preserve the oddity of the exchange.

The precision level of an expression can be marked explicitly by modifiers like *roughly* or *exactly*. We virtually have to employ *exactly* or *precisely* if we want to have a round number understood in a precise way:

- (8) *The distance between Amsterdam and Vienna is exactly one thousand kilometers.*

We can also use these modifiers to support an interpretation that an expression would have had anyway:

- (9) a. *The distance between Amsterdam and Vienna is exactly nine hundred sixty five kilometers.*  
 b. *The distance between Amsterdam and Vienna is roughly one thousand kilometers.*

But it is ludicrously pedantic to utter (10), because it suggests that one could be even more precise:

- (10) *The distance between Amsterdam and Vienna is roughly nine hundred sixty five kilometers.*

The notion of precision level is applicable only to the values of measurement terms, and not in cases in which numbers are used in an ordinal way. There is nothing pedantic in B’s reaction in the following example:

- (11) A: *Her phone number is sixty-five one thousand.*  
 B: *No, her phone number is sixty-five one-thousand and one.*

I will refer to the phenomenon we are after by the term *precision level choice*. The principle of proper precision level choice can be formulated as follows:

- (12) *Precision level choice*: When expressing a measurement of an entity, choose a level of precision that is adequate for the purpose at hand.

In (1), the adequate level of precision for motorists is on the mile or kilometer level; it is simply irrelevant for their concerns, like driving times or the amount of gas in one's tank, to give distances to a precision down to meters. In (2), the questioner A indicates a particular precision level as adequate for the purpose at hand (roughly,  $\pm 50$  km), and B's reaction is appropriate only at another precision level (roughly,  $\pm 0.5$  km). Notice that B's reaction is fine as a rejection of the precision level selected by A. For example, if B is the manager of a Dutch transportation company that wants to pay as little as possible to their truck drivers, the exchange sounds quite idiomatic.

How do we recognize precision levels? The empirical generalization is easy enough: Round Numbers suggest Round Interpretations in measure expressions. I will call this the *RN/RI principle*:

- (13) RN/RI principle:
- a. Short, simple numbers suggest low precision levels.
  - b. Long, complex numbers suggest high precision levels.

This is not necessarily so. In the mathematical language of science we can combine any measurement with any error margin. But it is an evident pragmatic principle in the interpretation of natural language.

The RN/RI principle is, as I said, an empirical generalization only. We should be able to derive it from more general principles of language use, that is, of linguistic pragmatics, which in turn should be motivated as general principles of human behavior. This is the main topic of this article. After a short review of the few previous attempts to deal with the phenomenon, I will try to identify simpler principles that are independently motivated and lead to the RN/RI principle. We will see that two such principles can be identified that give us the correct result provided they interact in a way that is suggested by Bidirectional Optimality Theory, a theory of the interaction of interpretative principles developed in Blutner (1998, 2000) and subsequent work.

### 3. Previous work

The phenomenon we are after is, of course, well known. It belongs to our everyday experience with language, and it has been noted in linguistic work at various places, for example in Sadock (1977), a response to Lakoff (1972), and in Pinkal (1995: 271), a comprehensive theory of the semantics of vagueness. A somewhat more detailed treatment can be found in Wachtel (1980), who develops a theory of *rounded numbers* for examples like the following:

- (14) a. *Sam is approximately six feet tall.*  
 b. *Odessa has approximately 1,000,000 inhabitants.*

Rounded numbers are marked by indicators of precision levels, like *approximately* or *roughly*. Wachtel also observes that we get the approximate interpretation without overt marking with “round” numbers like 1.000.000, but does not explain why this is so.

Curtin (1995), in an unpublished MA thesis, introduces the notion of *granularity* that can be used in the linguistic representation of reality. He develops the idea that fewer entities and distinctions exist on a coarse-grained level than on a fine-grained level, but that these levels are related to each other by homomorphic mappings. He observes that, with measuring terms, expressions on the coarse-grained level are generally shorter (e.g., *one hundred, two hundred, three hundred*) than on the fine-grained level (e.g., *one hundred, one hundred and one, one hundred and two*), but he does not give an explanation why this is so.

Most recently, Lasersohn (1999) developed a theory of imprecise interpretations in terms of pragmatic “slack” of expressions. But the relation to the complexity of expressions is not addressed at all.

### 4. A preference for short expressions

It is quite obvious that, everything else being equal, shorter and simpler expressions are preferred. This is a well-known law of the economics of language, and has found a well-known formulation in Zipf (1949). It is also behind a submaxim of manner in Grice’s William James Lectures of 1967, published as Grice (1975), which states, briefly, “Be brief!”. In the Neo-Gricean accounts of Atlas & Levinson (1981), Horn (1984),

and Levinson (2000) it corresponds to the I-principle (or Horn's R-principle), which says: "Produce the minimal linguistic information sufficient to achieve your communicational ends!" (cf. Levinson 2000: 114). The principle is typically motivated as one that is rooted in the interest of the speaker; speakers are lazy, hence like to be brief. However, experience with chatterboxes will make one doubt that (i) this is true of speakers in general in all situations, and (ii) that it is solely based on the interest of speakers. In any case, one cannot deny that it is a powerful pragmatic principle.

The principle plays an obvious role in precision level choice. It explains why, if the distance between Amsterdam and Vienna is 965 kilometers, one can quite truthfully utter (15.a) instead of (b).

- (15) a. *The distance between Amsterdam and Vienna is one thousand kilometers.*  
 b. *The distance between Amsterdam and Vienna is nine hundred sixty-five kilometers.*

Whatever (15.a) is lacking in precision is compensated by the gain in brevity, at least in typical situations. In general, in a decimal-based system, number words referring to the powers of ten, like 10, 100 or 1000, or multiples thereof, like 30, 400 or 7000, are shorter than number words referring to adjacent natural numbers.

This suggests that principle (a) of the RN/RI (cf. (13)) can be explained by a preference of brevity. In many situations, brevity is a worthier goal to aim for than precision. But can preference for brevity also explain principle (b) of the RN/RI? Perhaps it can: There is no gain when we interpret (15.b) vaguely, hence it is interpreted in a precise way.

I will argue that this line of reasoning is right, and I will develop it more precisely below. But before we do that, a somewhat longish detour concerning brevity is in order.

### 5. A closer look at brevity

Consider the following, where we talk about different distances.

- (16) a. *The distance A is one thousand one hundred kilometers.*  
 b. *The distance B is one thousand and one kilometers.*

Clearly, (16.a) can be interpreted in a less precise way even though it is more complex in terms of numbers of syllables and perhaps even in its morphosyntactic structure. We can counter such examples by saying that *in general*, multiples of powers of ten are expressed in a simpler way, and therefore allow for a less precise interpretation, even though there might be exceptions of shorter expressions that are not multiples of powers of 10.

The following examples also show an apparent exception to brevity:

- (17) a. *The train will arrive in five / fifteen / twenty-five minutes.*  
 b. *The train will arrive in four / sixteen / twenty-four minutes.*

We allow for more lax interpretations in the cases of (17.a) than in the corresponding cases of (17.b). The reason cannot be greater brevity of expression in (17.a), as the expressions are equally complex on all counts (morphosyntactic complexity, number of syllables). Obviously, the number or digit “5” is treated in special ways in our decimal-based system, just like the numbers that are powers of ten or multiples thereof. The reason is that they represent the half point between a multiple  $n$  of a power of 10 and the multiple  $(n+1)$  of the same power of 10, and that halves are the conceptually most prominent fractions. The prominence of half points can also be observed in cases like the following:

- (18) a. *The train will arrive in half a minute.*  
 b. *The train will arrive in thirty seconds.*  
 c. *The train will arrive in forty seconds.*

Here, (18.a) allows for an imprecise interpretation. This also holds, interestingly, for (18.b), perhaps to a somewhat lesser extent, in contrast to examples like (18.c), which are interpreted at a higher precision level. It appears that *thirty seconds* is just another standardized way to refer to one half of a minute, just as *sixty seconds* is a standardized way to refer to a minute. We find similar phenomena for other measures of time that



are not based on the decimal system, cf. e.g. *fifteen seconds*, *fifteen minutes*, *forty-five minutes*, *twelve hours* and *twenty-four hours*.

Another case in which brevity all by itself does not explain everything is the following. In English there are traces of competing numbering systems, based on 12 (*dozen*) or 20 (*score*). I found that people share the intuitions that (19.a) is more easily interpreted in a lax way than (19.b):

- (19) a. *A dozen / two dozens bandits were approaching.*  
 b. *Twelve / twenty-four bandits were approaching.*

Notice, that the complexity of *two dozen bandits* and *twenty-four bandits* is roughly the same: The first expression is syntactically more complex, but the latter one has fewer syllables.

It seems that brevity of expression is related to another factor that is the critical one. The idea is the following: When we want to express certain meanings, we select a particular set of expressions, which I will call the *expression-choice space*, out of which we choose the expressions that identify the semantic objects. The expressions of the set that we do not choose are possible alternatives, and the fact that they were not chosen may lead to pragmatic implicatures. Expressions that are not in the set do not count in the semantic or pragmatic interpretation. The notion of expression-choice space is similar to Curtin's notion of granularity. The difference is that granularity concerns meanings, whereas expression-choice space concerns expressions.

To illustrate this idea, consider again the following examples (we assume as above that the "real" distance is 965 kilometers):

- (20) a. *The distance between Amsterdam and Vienna is one thousand kilometers.*  
 b. *The distance between Amsterdam and Vienna is nine hundred seventy-two kilometers.*

In (20.a), the expression-choice space for the number word are number words based on the multiples of 100, that is, *eight hundred*, *nine hundred*, *one thousand*, *one thousand one hundred*, etc. In (b), the expression-choice space for the number word consists of number words based on multiples of 1. In the expression-choice space of (a) there doesn't even exist a number word *nine hundred sixty-five*; the number word *one thousand* is the closest one to the real distance. In the expres-

sion-choice space of (b), on the other hand, this number word exists, and hence the choice of *nine hundred seventy-two* is not the best one.

How do we know which items are contained in the expression-choice space? The expression that we are actually using tells us. If, in a measuring context, we use the number word *thirty*, then the expression-choice space will probably contain the number words that are multiples of 10, such as *twenty*, *thirty*, *forty*. If we use a number word like *three hundred*, the most likely alternatives are *two hundred*, *three hundred*, *four hundred*, and so on. If we use a number word like *one thousand*, we can either assume an alternative set like *one thousand*, *two thousand*, *three thousand* etc., or an alternative set of number words that are factors of 100, such as *nine hundred*, *one thousand* or *one thousand one hundred*. If we use a number word that refers to the mid number between two adjacent multiples of powers of 10, such as *thirty-five*, then the alternatives are the multiples of this power of 10 and their mid values, such as *twenty-five*, *thirty*, *thirty-five*, *forty*, *forty-five*, etc. If we evoke numbering systems with different bases, then similar principles apply to the powers of these bases; for example, the number word *two dozens* invokes alternatives like *one dozen*, *two dozens*, *three dozens*, etc. If we refer to dimensions that are measured by measure functions that are not based on the decimal system, such as time, then the expression-choice space may be determined by prominent halves and quarters of the basic measurement units of these dimensions. For example, choosing *fifteen minutes* will evoke alternatives such as *thirty minutes*, *forty-five minutes* and *one hour* as alternatives.

One important point to be mentioned is that a particular expression may allow for different expression-choice spaces; for example, *one thousand* is, in principle, compatible with number words that express multiples of 100 or multiples of 1000. In general, I will assume that expression-choice spaces of number words based on multiples of lower powers of 10 are possible. The number word *one thousand* then allows for the following expression-choice spaces:

- (21) a. *one thousand, two thousand, three thousand, ...*  
 b. *... nine hundred, one thousand, one thousand one hundred, ...*  
 c. *... nine hundred ninety, one thousand, one thousand and ten,*  
 ...  
 d. *... nine hundred ninety-nine, one thousand, one thousand and one, ...*

There is a systematic relationship between these expression-choice spaces: The upper ones are more coarse-grained, the lower ones are more fine-grained. We can model this relationship by set inclusion of sets of expressions.

An expression like *one thousand* presumably will prefer an expression-choice space like (21.a) or (b), and not (c) or (d). In general, we can state that the use of an expression  $\alpha$  that comes with expression-choice spaces introduces the most coarse-grained, or a more coarse-grained, expression-choice space in the actual pragmatic evaluation.

What is the relation of brevity of expressions to expression-choice space? Put simply, the average complexity of expressions in more fine-grained expression spaces is greater than the average complexity of expressions in less fine-grained expression spaces. Take the following example:

- (22) a. *one, two, three, four, five, six, seven, ... one hundred*  
 b. *five, ten, fifteen, twenty, twenty-five, thirty, ... one hundred*  
 c. *ten, twenty, thirty, ... one-hundred*

If we measure complexity by number of syllables, the average complexity of (22.a) is  $273/100 = 2.73$ ; of (22.b) it is  $46/20 = 2.3$ , and of (22.c) it is  $21/10 = 2.1$ . Measurement by morphological complexity would yield a similar result. Hence, expressions in less fine-grained expression-choice spaces are, in general, shorter.

We can then qualify the tendency for brevity as one for expression-choice spaces: Speakers prefer expression-choice spaces with expressions of low average complexity. I will call this principle or constraint BRIEFEXPRESSION:

- (23) BRIEFEXPRESSION:  
 Expression-choice spaces with shorter, less complex expressions are preferred over expression-choice spaces with longer, more complex ones.

There actually may be systematic pressures that make expressions of coarse-grained expression-choice spaces shorter. Notice that in English, *fifteen* is a phonologically and orthographically reduced form (*\*fiveteen* but *fourteen*, *sixteen*). The shortness of the form *twelve* obviously relates to the duodecimal system found in forms like *dozen*. Such phonological simplification may be mediated by frequency of use. Of course, more comprehensive studies would be necessary to establish any such connection between expression-choice spaces and phonology.

## 6. A preference for precise interpretations?

After this somewhat longish digression on brevity, let us return to the general theme of preference for short expressions. We can now phrase it as follows: The pragmatic principle BRIEFEXPRESSION is responsible for our interpretation of a measure term like *one thousand kilometers* in a vague way. Even if the speaker knows the distance more precisely, he may choose a term that is true only under a vague interpretation, because of the resulting gain in brevity.

It is unclear whether this also explains why we interpret a measure term like *nine hundred sixty-five kilometers* in a precise way. Intuitively, in this case nothing is to be gained by a vague interpretation, as the expression is long anyway, and therefore we settle on a precise interpretation. In order for this argumentation to go through, we have to assume that, as a matter of principle, precise interpretations of measure terms are preferred over imprecise, vague ones:

- (24) PRECISEINTERPRETATION:  
Precise interpretations of measure functions are preferred over vague ones.

We can think of vague interpretations of measure functions as measure functions that allow for errors. For this we can introduce the following precise notation:

- (25) If  $\mu$  is a measure function, and  $p$  is a positive real number, then  $\mu_{\pm p}$  is the vague measure function defined as: For all  $x$ ,  $\mu_{\pm p}(x) = [\mu(x) - p\mu(x), \mu(x) + p\mu(x)]$

For example, if  $\text{km}(x) = 1000$ , then  $\text{km}_{\pm 0.05}(x) = [950, 1050]$ .

Why should precise interpretations of measure terms be preferred over vague ones? The obvious reason is: Because they are more informative! This is a general pragmatic principle; witness Grice’s first sub-maxim of Quantity, which states that a speaker should make a contribution as informative as required by the current purpose of information exchange.

With BRIEFEXPRESSION and PRECISEINTERPRETATION as two independent constraints we can explain the preference patterns as in the following OT-tableau. We assume, as before, that the true, precise distance is 965 kilometers.

(26)

Expression	BRIEF- EXPRESSION	PRECISE- INTERPRETATION
☞ (a) <i>The distance between A and V is one thousand kilometers.</i>		*
☞ (b) <i>The distance between A and V is nine hundred sixty-five kilometers.</i>	*	
(c) <i>The distance between A and V is nine hundred seventy-two kilometers.</i>	*	*

Here, sentence (a) is the shortest candidate, but it violates PRECISEINTERPRETATION. Sentence (b) violates BRIEFEXPRESSION, but it allows for a precise interpretation. If we do not rank the two constraints (as indicated by the dotted line), then it is predicted that both sentences are fine. In contrast, (c) violates both constraints, and hence is dis-preferred over the other two candidates.

The problem of this view, however, is that an addressee will preferably interpret an expression like *The distance between Amsterdam and Vienna is one thousand kilometers* as saying that the distance is precisely one thousand kilometers. Under this interpretation, the expression is both brief and precise. But this is clearly not what we find – see the RN/RI-principle, (13.a). If we indeed want a precise interpretation of round numbers, we have to mark this explicitly with expressions that indicate the precision level, such as *exactly*:

- (27) *The distance between Amsterdam and Vienna is exactly one thousand kilometers.*

How can we avoid this problem? Perhaps we should change our line of reasoning by roughly 180 degrees and assume instead a preference for vague interpretations.

## 7. A preference for vague interpretations

The opposite of *PreciseInterpretation* is *VagueInterpretation*:

- (28) **VAGUEINTERPRETATION:**  
Measurement terms are preferably interpreted in a vague way.

How could such a tendency be motivated? It clearly seems to run against Grice's first submaxim of quantity, which states that information content should be maximized. But already Grice added a second submaxim, namely, that the speaker should not give more information than is necessary for the current purpose of information exchange.

Preference for imprecise expressions has been noticed by linguists working in empirical pragmatics. Ochs Keenan (1976), in her seminal study of the pragmatics of rural speech communities in Madagascar, describes conversational maxims that prefer expressions with less informative content over those with more content. One of the reasons, Ochs Keenan claims, is that vague expressions help to save face: One is not as easily proven wrong if one stays vague. This preference is by no means only to be found in exotic rural communities. The philosopher of science Pierre Duhem, noted, famously, that there is a balance between precision and certainty: One cannot be increased except to the detriment of the other (cf. Duhem 1904: 178f., cited after Pinkal 1995: 262).

So, it appears that a tendency for vague interpretation can be motivated. But does it give us what we want? Consider the following tableau:

(29)

Expression	BRIEF- EXPRESSION	VAGUE- INTERPRETATION
(a) <i>The distance between A and V is one thousand kilometers.</i>		
(b) <i>The distance between A and V is nine hundred sixty-five kilometers.</i>	*	(*)

Now (a) is preferred, and we also predict that this sentence is preferably understood under a vague interpretation. But (b) is dispreferred; I put the star for a violation of VAGUEINTERPRETATION in parentheses to indicate that the sentence could be seen either under a vague or a precise interpretation.

There are very obvious problems with this approach. First, (b) is not really dispreferred over (a); it just suggests a precise interpretation. And more importantly, if (b) is actually uttered, a way to reduce the violations on the hearer’s side is to assume not a precise, but a vague interpretation! This is because this would reduce the violation from two stars to once. But of course this is exactly the opposite of what we find.

Something must be fundamentally wrong. Nevertheless, I will argue that BRIEFEXPRESSION and VAGUEINTERPRETATION are indeed the two constraints that explain the RN/RI principle. But we have to consider the way how these constraints interact. Put informally, we must find a way to express, and motivate, the idea that one constraint can be violated just in case the other is violated.

### 8. The interaction of brevity and vagueness

The theoretical framework that gives us what we need is *Bidirectional Optimality Theory*, as developed by Blutner (1998) and Blutner (2000); cf. also Dekker & van Rooy (2000) for a game-theoretic interpretation working with the notion of Nash equilibrium.

In classical Optimality Theory (OT), the input to a rule expressed in a tableau of ranked or unranked constraints is a set of expressions, and the output is the one expression, or the set of expressions, that violate the constraints the least. In Bidirectional OT, the input are pairs of ob-

jects, constraints are independently specified for the members of the pair, and the output are those pairs that violate the constraints the least. The constraints are formulated in a modular fashion, for the members of the pairs. But finding the optimal solution(s) requires optimization for both members.

In semantic and pragmatic applications of Bidirectional OT, the pairs are pairs  $\langle E, I \rangle$  of an Expression and its Interpretation.

There are different possibilities for comparing pairs  $\langle E, I \rangle$  of expressions and interpretations. A particularly interesting one has been suggested by Jaeger (2000); it is called *bidirectional super-optimality*, and it is defined as follows:

- (30) A pair  $\langle E, I \rangle$  of a set of candidate expressions GEN is superoptimal iff:
- i. There is no superoptimal  $\langle E', I \rangle \in \text{GEN}$  such that  $\langle E', I \rangle \gg_E \langle E, I \rangle$
  - ii. There is no superoptimal  $\langle E, I' \rangle \in \text{GEN}$  such that  $\langle E, I' \rangle \gg_I \langle E, I \rangle$

The notion of superoptimal pairs  $\langle E, I \rangle$  is thus restricted to those that have no competitor on the expression level or on the interpretation level that is itself super-optimal. (If you think this definition is circular, please read on.)

It is the notion of super-optimality that we need to capture the interactions between constraints as we find them in precision level choice phenomena. That is, from the notion of superoptimality it follows that sometimes violating one constraint allows for the violation of other constraints, with no punishment. Consider the following cases:

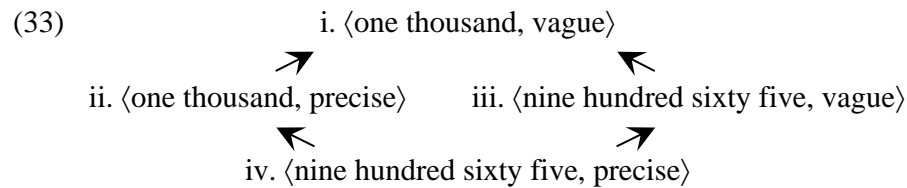
- (31) i.  $\langle$ The distance between A and V is one thousand kilometers, vague $\rangle$   
 ii.  $\langle$ The distance between A and V is one thousand kilometers, precise $\rangle$   
 iii.  $\langle$ The distance between A and V is nine hundred sixty-five kilometers, vague $\rangle$   
 iv.  $\langle$ The distance between A and V is nine hundred sixty-five kilometers, precise $\rangle$



The rankings that we find for these expression/interpretation pairs are the following, where BE stands for BRIEFEXPRESSION and VI for VAGUEINTERPRETATION.

- (32) i.  $\gg_{BE}$  iii.  
 i.  $\gg_{VI}$  ii.  
 ii.  $\gg_{BE}$  iv.  
 iii.  $\gg_{VI}$  iv.

Or, in a Hasse diagram, and in an obvious shorthand:



Notice that (i) and (iv) cannot be compared directly, as they differ both in their expression component and in their interpretation component, and (30) only allows for comparison between pairs that are identical in one component.

Clearly, (31.i) is a superoptimal expression/interpretation pair. There is no superoptimal pair  $\langle E, I \rangle$  such that either  $\langle E, I \rangle \gg_{BE} i$  or  $\langle E, I \rangle \gg_{VI} i$ . Hence we predict that brief measure expressions are preferably interpreted in a vague way. The pair ii clearly is not superoptimal, because it has a better superoptimal competitor, i. Also, the pair iii is not superoptimal, because it also has i as a better superoptimal competitor. But, interestingly, iv is also a super-optimal pair. The reason is that the only competitors ii. and iii., even though they fare better in terms of BRIEFEXPRESSION and VAGUEINTERPRETATION, respectively, are not themselves superoptimal. Hence we predict that complex measure expressions are preferably interpreted in a precise way.

The intuitive reasoning behind this is as follows. Assume that we encounter an expression like (34):

- (34) *The distance between Amsterdam and Vienna is nine hundred sixty-five kilometers.*

We consider a vague vs. a precise interpretation,  $\langle (34), \text{vague} \rangle (= \text{iii.})$  and  $\langle (34), \text{precise} \rangle (= \text{iv.})$ . In general, vague interpretations are pre-

ferred, so we are inclined to impose a vague interpretation, (iii). But under a vague interpretation, we could express the same truth conditions with

(35) *The distance between Amsterdam and Vienna is one thousand kilometers.*

The pair  $\langle (35), \text{vague} \rangle$  (= i) should be preferred over iii because it is shorter. But the speaker has not uttered (35). Hence we can refute the vague interpretation of (34) (= iii), and select the precise interpretation, iv.

This, precisely, is what was meant above how violation of one constraint (here, brevity) allows for violation of another (here, vague interpretation).

This line of argumentation has been used for a number of other phenomena in pragmatics. For example, Blutner employed it to derive the preferred interpretation of double negatives such as *not unhappy* as ‘happy, but not quite so’, or the interpretation of explicit causatives like *cause the sheriff to die* as expressing indirect, non-stereotypical causation. These applications all involve a derivation of the M-principle of Levinson (2000), which states that “marked expressions have marked meanings”. In the case at hand, the “marked” measure expression is the complex one, like *nine hundred and sixty-five*, and the “marked” interpretation is the precise one. With superoptimality, the M-principle is not an axiom anymore, but a theorem. We can explain the Emergence of the Marked.

## 9. Conclusion

I showed that by assuming pragmatic tendencies for short expressions and vague interpretations one can account not only for the fact that short expressions have a preference for vague interpretations, but also that long expressions have a preference for precise interpretations. I have qualified the notion of short expressions as one relativized to the expression-choice space, and I have tried to motivate these pragmatic tendencies. The two tendencies have to interact in a particular way so that the two observed pairings of expressions and interpretations result as stable solutions. This can be achieved within the framework of Bidirectional

Optimality theory, where they appear as the two superoptimal solutions. In essence, long expressions are allowed, provided that they are interpreted precisely, because under a vague interpretation a short expression would have been more optimal. I have indicated that this line of reasoning is just one application of a general type of explanation that applies to a wide range of phenomena of the type “marked expressions express marked meanings” and “marked meanings are expressed by marked expressions”.

## Notes

1. This paper is based on talks given at the *Sinn und Bedeutung* Conference in December 2000 in Amsterdam, and on a talk at the Hebrew University in Jerusalem in April 2001. I thank the audiences for their comments, and in particular Reinhard Blutner, Larry Horn, Gerhard Jäger, Michael Morreau, Carl Posy, Phillipe Schlenker, Henk Zeevat.
2. Including the good one that it is indeed easier to give common fractions like  $1/2$ ,  $1/3$  or  $1/4$  for feet and yards, which suggests that we should give up the decimal system of counting in favor of the duodecimal one, if not the sexagesimal one of the Babylonians.

## References

- Atlas, Jay David and Stephen C. Levinson  
1981 *It*-clefts, informativeness, and logical form: Radical pragmatics (revised standard version). In: Peter Cole (ed.), *Radical pragmatics*, 1–61. New York: Academic Press.
- Blutner, Reinhard  
1998 Lexical pragmatics. *Journal of Semantics* 15: 115–162.  
2000 Some aspects of optimality in natural language interpretation. *Journal of Semantics* 17: 189–216.
- Curtin, Paul  
1995 Prolegomena to a theory of granularity. Unpublished MA thesis.
- Dekker, Paul and Robert van Rooy  
2000 Bi-directional Optimality Theory: An application of game theory. *Journal of Semantics* 17: 217–242.

- Duhem, Paul  
 1904 *La theory physique, son objet et sa structure*. Paris:
- Grice, Paul  
 1975 Logic and conversation. In: Peter Cole and Jerry L. Morgan (eds.), *Syntax and Semantics 3: Speech Acts.*, 41–58. New York: Academic Press.
- Horn, Laurence R.  
 1984 Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. in Deborah Schiffrin (ed.), *Meaning, form, and use in context: Linguistic applications*, 11–89. Washington, D.C.: Georgetown University Press.
- Jaeger, Gerhard  
 2000 Some notes on the formal properties of bidirectional optimality theory. To appear in *Journal of Logic, Language and Information*.
- Lakoff, George  
 1972 Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Papers from the Eighth Regional Meeting of the Chicago Linguistic Society 1972*: 183–228.
- Lasersohn, Peter  
 1999 Pragmatic halos. *Language* 75: 522–551.
- Levinson, Stephen C.  
 2000 *Presumptive meanings*. Cambridge (MA): MIT.
- Ochs Keenan, Elinor  
 1976 The universality of conversational implicature, *Language in Society*, 5: 67–80.
- Pinkal, Manfred  
 1995 *Logic and lexicon*. Dordrecht: Kluwer.
- Sadock, Jerrold M.  
 1977 Truth and approximations. In: In Kenneth Whistler, Jr. Robert D. Van Valin, Chris Chiarello, Jeri J. Jaeger, Miriam Petruck, Henry Thompson, Ronya Javkin, and Anthony Woodbury (eds.), *Proceedings of the 3rd Annual Meeting of the Berkeley Linguistics Society*, 430–439. Berkeley: Berkeley Linguistics Society.
- Wachtel, Tom  
 1980 Pragmatic approximations. *Journal of Pragmatics* 4: 201–211.
- Zipf, George K.  
 1949 *Human behavior and the principle of least effort*. Cambridge: Addison-Wesley.