Formal and Computational Semantics: a Case Study

Stephen G. Pulman
Computational Linguistics Group,
Oxford University Computing Laboratory,
sgp@clg.ox.ac.uk

May 18, 2007

Abstract

The literature in formal linguistic semantics contains a wealth of fine grained and detailed analyses of many linguistic phenomena. But very little of this work has found its way into implementations, despite a widespread feeling (among linguists at least) that this can't be very difficult in principle: you just fix a grammar to produce the right logical forms and hook them up to a theorem prover, don't you? In this paper I take a representative analysis of adjectival comparatives and ask what steps one might actually have to go through so as to use this analysis in a realistic computational setting.
[1]

1 Introduction

Formal semantics and computational semantics are not as closely linked as they should be. Formal semanticists produce detailed descriptions of small, interesting, areas but without any guarantee that these descriptions are closed under union. Coverage is limited and fragmentary, on the whole. In this respect, formal semantics is in something like the state that syntax and parsing were in ten or more years ago. But these days, we have parsers that are based on very sophisticated grammatical theories (Clark and Curran, 2004), and that are capable of quite accurate performance with relatively wide coverage, certainly enough to support a certain degree of robust semantic interpretation (Bos et al., 2004). It would be nice to be able to draw on the substantial amount of potentially relevant work in formal semantics so as to improve the coverage and depth of our semantic interpretation systems.

Of course, there are many obstacles in the way. Formal semanticists hardly ever give any thought to how their analyses could be mechanised, either for the purposes of testing the analysis to ensure that it does what it is supposed to do (and does not do what it is not supposed to do), or how it could be used in the context of some practical computational application like database query or question answering. Consequently, very little of their work is framed in a computational way, and not much has found its way into implementations of any kind, despite a widespread feeling (among linguists at least) that this can't be very difficult in principle: you just fix a grammar to produce the right logical forms and hook them up to a theorem prover, don't you?

But things are not quite so straightforward. The 'right' logical forms for most formal semanticists will be some kind of higher order logic: to derive logical forms compositionally

[1]Appeared in Jeroen Geertzen, Elias Thijsse, Harry Bunt and Amanda Schiffrin (eds.) *Proceedings of the Seventh International Workshop on Computational Semantics: IWCS-7*, Tilburg, The Netherlands, 2007, ISBN 90-74029-31-0, pp 181-196.

we usually need to associate higher order meanings to constituents even where the final result is first order, or nearly so. However, many natural language constructs are known not to have a natural formalisation (perhaps not any formalisation at all) in first order logic and so we seem to be stuck with an ineradicable higher order component. For practical automatic theorem proving purposes, this is bad news, for the undecidability of higher order logic is not going to go away. Moving to first order is not a complete solution, either, since of course even first order logic is only semi-decidable, but it is probably the best we can do at present. It would be nice to hope that someone in the theorem proving community will come up with efficient special purpose and limited (probably) inference mechanisms for some useful fragments of higher order logic, but in the meantime the best route for being able to use the results of work in formal semantics seems to be to get some kind of translation of the analyses into first order logic and then use some of the quite efficient (though still not efficient enough) theorem provers that have been developed over the last few years.

In this paper I attempt a case study of this type, taking a semantic analysis off the shelf, and trying to cast in in an implementable form. My example is Ewan Klein's analysis of adjectival comparatives (Klein, 1980; Klein, 1982). I will develop a fragment that assigns Klein-style logical forms (in an intension-free version of the higher order logic he is using) and then show how to translate these to first order forms which preserve the various inferential relationships we will be concerned with. The resulting inferences (or lack of them) have all been verified using the SPASS theorem prover (`http://spass.mpi-sb.mpg.de/`).

Note that this paper is not itself intended as a contribution to the linguistic literature on comparatives (and there is far too much of it to discuss adequately here). However, the choice of Klein's analysis is not entirely arbitrary: although it has been criticised in the literature (often inaccurately, in my view) and although some viable alternatives have appeared, I believe his analysis stands up rather well, and few of these alternatives approach the level of coverage and detail that Klein achieved.

2 Klein's analysis

Klein analyses gradeable adjectives like 'tall' as partial functions which presuppose a contextually determined comparison class, and a scale:

```
<-----------------|------|------------------->
     not-A           ext gap        A
```

6ft 6in is definitely tall (for a person), 4ft 6in definitely not tall, and 5ft 8in, in the extension gap, is neither tall nor not tall. Klein distinguishes between 'linear' adjectives like 'tall', and 'non-linear' adjectives like 'clever' or 'generous'. The distinction between them is that with linear adjectives, a sentence like 'John is taller than Bill and Bill is taller than John' will always be a contradiction, whereas 'John is more generous than Bill and Bill is more generous than John' can be interpreted in such a way that it is not contradictory, if the respect in which John is more generous than Bill (with money, say) is different from the respect in which Bill is more generous than John (with his time, for example).

Adjectives can combine with degree modifiers as in 'very tall', 'quite tall', or with measure phrases like '5ft 6in tall'. Degree modifiers can be seen as the kind of thing you are asking for in a question like 'How tall is John?'. On Klein's account, when an adjective is combined with an explicit modifier of these types the interpretation is no longer that of a partial but a total function. While this is plausible for the measure phrases, many people have found it less plausible for modifiers like 'very' and 'quite', since it appears easy to imagine that there would be clear cases of 'very tall', clear cases of 'not very tall' and some where it is difficult to decide.

Comparative forms of the adjectives are analysed as containing an abstract degree modifier. So 'John is taller than Bill is' means "there is some degree modifier d that makes 'John is d-tall' true and 'Bill is d-tall' false":

(1) $\exists \mathsf{d}.\mathsf{d}(\mathsf{tall})(\mathsf{john}) \wedge \neg \mathsf{d}(\mathsf{tall})(\mathsf{bill})$

The types of adjectives and degree modifiers are:

(2) Type(tall)= (e→t), type(d)= type(very,quite,5-ft-6in, etc.) = (e→t)→(e→t)

Note that without further stipulation, the logical form given in 1 will be true not just under the intended circumstances but also when Bill is taller than John. This is because there is no restriction on the kinds of degree modifiers that 'd' can range over, and although it is complex we can construct a measure phrase something like '5 inches shorter than Bill' which will be true of John's tallness but not of Bill's, in the circumstances just described. Note that the fact that the sentence 'John is [5 inches shorter than Bill] tall' is not grammatical is irrelevant, just the fact that if it were, then its logical form would be a substitution instance of the formula in 1. But also note that '5 inches shorter than Bill' would in fact be a perfectly good answer to the question 'How tall is John?', suggesting that it is indeed a genuine degree modifier.

In order to fix this, Klein imposes a 'Consistency Postulate':

(3) $\forall \mathsf{x},\mathsf{y},\mathsf{Q}.[\exists \mathsf{d}.\mathsf{d}(\mathsf{Q})(\mathsf{x}) \wedge \neg \mathsf{d}(\mathsf{Q})(\mathsf{y})] \rightarrow \forall \mathsf{e}.\mathsf{e}(\mathsf{Q})(\mathsf{y}) \rightarrow \mathsf{e}(\mathsf{Q})(\mathsf{x})$

If we let Q be instantiated by 'tall', this says that if x is d-tall and y is not d-tall, then every value of e which makes it true that y is e-tall is also a value of e making it true that x is e-tall. More informally, if x is taller than y, then x has at least every degree of tallness that y has. In fact, as far as I can see, this postulate still does not solve the problem. For if John is actually taller than Bill is, then there will be a complex degree modifier 'X inches less than John's height' which is a degree of tallness of Bill but not of John, contrary to the postulate. It seems clear that Klein had some other restrictions in mind on the range of possible degree modifiers: that they should correspond to valid linguistic forms, or perhaps that they should be 'primitive' in some way (i.e. not constructed using logical connectives or the comparative construction itself). We will assume some such restriction from now on and accept the Consistency Postulate at face value.

Logical forms for other types of comparatives are:

(4) a. John is less tall than Bill is.

b. $\exists \mathsf{d}.\ \neg \mathsf{d}(\mathsf{tall})(\mathsf{john}) \wedge \mathsf{d}(\mathsf{tall})(\mathsf{bill})$

(5) a. John is as tall as Bill is

b. $\forall \mathsf{d}.\mathsf{d}(\mathsf{tall})(\mathsf{bill}) \rightarrow \mathsf{d}(\mathsf{tall})(\mathsf{john})$

(6) a. John is taller than he is wide

b. $\exists \mathsf{d}.\mathsf{d}(\mathsf{tall})(\mathsf{john}) \wedge \neg \mathsf{d}(\mathsf{wide})(\mathsf{john})$

It is a particularly nice feature of Klein's analysis is that it predicts that when different adjectives are compared, as in the example just given, the result will only be sensible if there is a measure appropriate for both adjectives: compare:

(7) John is taller than Bill is clever

which can only be interpreted by setting up a special context in which the two properties are commensurable, or by using some Gricean reasoning.

Klein extends his analysis to also deal with modifiers like 'much' and fully specified measure phrases. He analyses 'much' as a generalised quantifier meaning roughly 'many' with respect to some contextually specified comparison class.

(8) John is much taller than Bill is

is analysed as:

(9) $\mathsf{much}(\lambda \mathsf{d.d(tall)(john)} \wedge \neg \mathsf{d(tall)(bill)})$

where'much' has truth conditions something like 'much(P)' is true with respect to comparison set Q iff the cardinality of P is more than (some fraction) of the cardinality of Q. 'Much' has the type $(((\mathsf{e}{\to}\mathsf{t}){\to}(\mathsf{e}{\to}\mathsf{t})){\to}\mathsf{t}){\to}\mathsf{t}$, although it is interpreted as if it had the type $((((\mathsf{e}{\to}\mathsf{t}){\to}(\mathsf{e}{\to}\mathsf{t})){\to}\mathsf{t}){\to}(((\mathsf{e}{\to}\mathsf{t}){\to}(\mathsf{e}{\to}\mathsf{t}))){\to}\mathsf{t}){\to}\mathsf{t}$,

Klein's analysis of measure phrases like 'one foot' is a little complex. He regards noun phrases like 'one foot' as denoting an equivalence class of objects that are exactly as long as some prototypical one-foot long object. He allows formulae of the form 'Q(one-foot)' which are true iff Q is true of at least one element in the set 'one-foot'. His examples make it clear that 'one-foot' is to be of type e, and Q of type $\mathsf{e}{\to}\mathsf{t}$. Now the logical form for:

(10) John is one foot taller than Bill is

becomes (I have renamed the variables that Klein originally used to avoid confusion):

(11) $\forall \mathsf{d}.[\mathsf{d(much)}(\lambda \mathsf{e.e(tall)(one\text{-}foot)}) \to \mathsf{d(much)}(\lambda \mathsf{f.f(tall)(john)} \wedge \neg \mathsf{f(tall)(bill)})]$

Now 'd' has to have the type of a modifier of 'much', and the types of e and f are those of ordinary degree modifiers. The intuitive interpretation of 10 is quite complex, I find. The sub-expression $\lambda \mathsf{e.e(tall)(one\text{-}foot)})$ denotes the set of degree modifiers e such that e-tall is true of something one foot long. The second lambda expression denotes the set of degree modifiers that are true of John's tallness but not Bill's. The whole thing says that every modifier d of *much* that makes it true that there are d-much degree modifiers e such that e-tall is true of something one foot long, is also a modifier that makes it true that there are d-much degree modifiers that are true of John's tallness but not Bill's. Klein points out that this logical form could be paraphrased as 'John is as much as one foot taller than Bill'.

I think Klein's intuition here is that the degrees of muchness characterising one foot are as many as the degrees of muchness characterising the difference in height between John and Bill. It's difficult to find evidence one way or the other for this intuition, but notice that since adjectives are interpreted relative to a comparison class, here presumably the set of men, there can't be many (naturally expressible) degree modifiers e that make e-tall-for-a-man true of something that is one foot long. ('Not very' is the only one that comes readily to mind!) The only way I can make a logical form like this seem right is to think not of sets of degree modifiers of tall, but of sets of degrees of tallness that you can do a modest amount of arithmetic with. Then it seems plausible that however you characterise the units of tallness in one foot, that will also be a characterisation of the number of units of tallness difference between John and Bill, but this is not quite the same as the number of degree modifiers true of that degree of tallness. However, if we do reconceptualise things in this way (we began to do so when discussing the Consistency Postulate earlier) then there is a much simpler way of characterising the logical form of 9, and one which is more consistent with the usual type for NPs: interpret 'one foot' as a kind of summing quantifier, and represent the logical form as:

(12) $\mathsf{one\text{-}foot}(\lambda \mathsf{d.d(tall)(john)} \wedge \neg \mathsf{d(tall)(bill)})$

meaning that the number of units or degrees of tallness that John has but Bill doesn't add up to one foot. This still leaves open some further questions about the ontology of units and degrees that we are presupposing here, some of which we will return to later.

Note that these logical forms are almost identical to the basic comparative forms if we give the latter their un-sugared forms:

(13) $\exists(\lambda d.d(tall)(john) \wedge \neg d(tall)(bill))$

Deriving them compositionally is reasonably straightforward: we have to arrange for the basic comparative forms of AdjP to have an existential or universal as appropriate if there is no explicit 'much' or measure quantifier present. The complement sentence will be of the form $\lambda P.(...P(bill))$, of type $(e \rightarrow t) \rightarrow t$ representing the contribution of the AdjP gap which we follow standard practice in assuming to be there in [... taller than [Bill is __]]. The general form of the comparative AdjP will then be (operators associate left):

(14) $\lambda y.Comparative\_complex(Adj)(y)(\lambda P.(...P(bill)))$

where the various comparative complexes will be (R has type $(e \rightarrow t) \rightarrow t$):

-er than	$= \lambda QzR.\exists(\lambda d.d(Q)(z) \wedge R(\lambda x.\neg d(Q)(x)))$
6in/much -er than	$= \lambda QzR.6in/much(\lambda d.d(Q)(z) \wedge R(\lambda x.\neg d(Q)(x)))$
An example:	
taller than	$= \lambda zR.\exists(\lambda d.d(tall)(z) \wedge R(\lambda x.\neg d(tall)(x)))$
taller than Bill is	$= \lambda z.[\lambda R.\exists(\lambda d.d(tall)(z) \wedge R(\lambda x.\neg d(tall)(x))](\lambda P.(...P(bill)))$
	$= \lambda z.\exists(\lambda d.d(tall)(z) \wedge [\lambda P.(...P(bill))](\lambda x.\neg d(tall)(x)))$
	$= \lambda y.\exists(\lambda d.d(tall)(y) \wedge \neg d(tall)(bill))$

3 Degrees vs. intervals

Several recent analyses (Kennedy, 2001; Schwarzschild and Wilkinson, 2002) have criticised degree-based treatments of comparatives, arguing that there are various phenomena which remain unexplained on such an account. There is not space here to discuss these issues in any detail, other than to observe that Klein's analysis stands up rather well to at least some of these criticisms. For example, Schwarzschild and Wilkinson, assuming an analysis on which 'John is taller than Bill is' has the logical form:

(15) $\exists d,e.tall(john,d) \wedge tall(bill,e) \wedge d > e$

point out that a sentence like 'John is taller than everyone else is' will get the wrong logical form:

(16) $\exists d,e.tall(john,d) \wedge [\forall x.x \neq john \rightarrow tall(x,e)] \wedge d > e$

16 requires everyone else to be the same height, which is a stronger requirement than that imposed by the sentence itself. Similarly, the Russell-type sentences like:

(17) a. John's yacht is longer than Bill expected it to be

b. $\exists d,e.long(JY,d) \wedge expect(bill,long(JY,e)) \wedge d > e$

if analysed as shown, require Bill to have had a specific expectation about the length of John's yacht, which is not very plausible.

But Klein's analysis doesn't suffer from this problem, unless I have missed something. Klein's version of a degree-based analysis only involves one degree, and the logical forms assigned to the sentences in question will be:

(18) a. $\exists d.d(tall)(john) \wedge [\forall x.x \neq john \rightarrow \neg d(tall)(x)]$

b. $\exists d.d(long)(JY) \wedge expect(bill, \neg d(long)(JY))$

The logical form in a is unproblematic: in b, presumably the second conjunct will be true provided that the length that Bill expected John's yacht to be is not d, which is a much less specific expectation. There may be other reasons to use intervals rather than degrees (and in fact our later interpretation of Klein's ontology moves in this direction), and there is more to be said about the interaction of quantifier scope with comparatives, but these examples are not compelling arguments against Klein's analysis.

4 Implementation

Let us begin with the basic type of adjectival predication: 'John is tall'. On Klein's theory this is formalised as 'tall(john)' and it is true iff John is above the norm for the relevant comparison class (say, men) and scale (say, height). The non-linear adjectives like 'clever' or 'generous' require also a contextually determined respect in order to fix the scale: to keep things simple, I will not discuss them further, but I believe the analyses below can easily be extended so as to accommodate them.

So in order to interpret such adjectives we need to represent the comparison class, determine an appropriate scale, find the 'norm' or extension gap in the scale, and find a first-order equivalent to the degree modifiers. When we know the class and scale in advance, we can simply reify degrees and comparison classes, and for each predicate like 'tall' or 'man' introduce a new predicate 'tall*' and constant 'man*' respectively:

(19) $\exists \text{d.d(tall)(john)} \wedge \text{man(john)} \rightarrow \exists \text{e.tall*(john,man*,e)}$

'If there is some d such that John is d-tall, then there is some e such that John is tall* (for a man*) to extent e'. (I will suppress the * hereafter, since it will be clear from context whether we are using the introduced or the original predicates).

To illustrate, assume a scenario in which we have group of men, some of whom are basketball players. Notoriously the standard of height for basketball players is above that for men, so that some of these individuals may count as tall men but not tall basketball players. Since we have to assume that non-linguistic knowledge about heights and norms is available to speakers, we can just axiomatise these notions by brute force, assuming a height predicate and a predicate norm(Adj,ComparisonClass,Min,Max) with the arguments indicated. Min and Max refer to the lower and upper points of the extension gap associated with the comparison class. (The numbers represent feet, not metres!).

```
                        man    bbplayer
height(kim,6.67).        +       +           norm(Adj,ComparisonClass,Min,Max).
height(sandy,6.25).      +       ?           norm(tall,man,5.67,5.83).
height(chris,6.0).       +                   norm(tall,bbplayer,6.25,6.42).
height(andy,5.92).       +       -
height(alex,5.67).       ?       -
height(jo,5.42).         -
height(jan,5.25).        -       -

     man: kim, sandy, chris, andy, alex, jo, jan
bbplayer: andy, sandy, alex, kim, jan
```

Now we need various axioms: the first one connects up height and tallness, saying that you are as tall (for anything) as your height:

(20) $\forall \text{x,p,d.height(x,d)} \rightarrow \text{tall(x,p,d)}$

We need a version of the consistency postulate for every adjective:

(21) $\forall \text{x,y,q.}[\exists \text{d.tall(x,q,d)} \wedge \neg \text{tall(y,q,d)}] \rightarrow [\forall \text{e.tall(y,q,e)} \rightarrow \text{tall(x,q,e)}]$

This will imply that if you are d-tall, then you are also e-tall for e < d. But so far, nothing prevents you being f-tall for f > d. We also need to ensure that you are no taller than your height!

(22) ∀x,y,z,p.height(x,y) ∧ z > y → ¬tall(x,p,z)

Notice that our 'tall' predicate does not completely correspond to English tall: we have to translate 'John is tall/John is a tall man' as:

(23) tall(john,man,above_norm)

and 'John is not tall' as:

(24) tall(john,man,below_norm)

where above_norm and below_norm are abstract degree modifiers defined by axioms like:

(25) ∀x,y,z,p,d.norm(tall,p,x,y) ∧ tall(z,p,d) ∧ d > y → tall(z,p,above_norm)

With respect to the situation described by the axioms above, the following will now be provable:

(26) a. Andy is a tall man
  b. tall(andy,man,above_norm)

(27) a. Andy is not a tall basketball player
  b. tall(andy,bbplayer,below_norm)

(28) a. Jo is a taller man than Jan
  b. ∃d.tall(jo,man,d) ∧ ¬tall(jan,man,d)

(29) a. But Jo is not a tall man, and Jan is not a tall man
  b. tall(jo,man,below_norm)
  c. tall(jan,man,below_norm)

(30) a. There is a tall basketball player (Kim)
  b. ∃ x.tall(x,bbplayer,above_norm)

There are some inferences that should not go through, of course. We introduce the fact that John is a taller man than Bill is, but we do not know anything about their heights, so neither is known to be tall or not tall in any respect:

(31) a. assert: ∃d.tall(john,man,d) ∧ ¬tall(bill,man,d)
  b. tall(john,man,below_norm) (no proof found)
  c. tall(john,man,above_norm) (no proof found)

The following purely linguistically based inferences should go through based just on the previous assertion:

(32) a. Bill is less tall than John
  b. ∃d.¬tall(bill,man,d) ∧ tall(john,man,d)
  c. or: Bill is not as tall as John
  d. ¬(∀d.tall(john,man,d) → tall(bill,man,d))

(33) a. John is as tall a man as Bill is: (= at least as tall)
  b. ∀d.tall(bill,man,d) → tall(john,man,d)

We also correctly capture the transitivity of 'Adj-er than':

(34) a. Bill is a taller man than Fred is

b. assert: ∃d.tall(bill,man,d) ∧ ¬tall(fred,man,d)

c. So John is a taller man than Fred is, too

d. ∃d.tall(john,man,d) ∧ ¬tall(fred,man,d)

The extra linguistic information we have just added interacts correctly with the non-linguistic information axiomatised earlier:

(35) a. John is a taller man than Andy (who we know to be definitely tall)

b. assert: ∃d.tall(john,man,d) ∧ ¬tall(andy,man,d)

So we should now be able to prove that John is in fact also a tall man:

(36) tall(john,man,above_norm(tall))

5 Some wrinkles

During the course of developing and testing this axiomatisation, I stumbled over various things which I mention here in order that others may avoid my mistakes. Firstly, note that since the above_norm/below_norm device is essentially implementing a kind of partiality within ordinary (i.e. non-partial) FOPC, we have to be very careful with how we translate natural language negation. The tautological expression:

(37) tall(john,man,above_norm) ∨ ¬tall(john,man,above_norm)

will of course always come out true, but this is **not** the translation of the sentence 'Either John is a tall man or John is not a tall man'. That sentence should be formalised as:

(38) tall(john,man,above_norm) ∨ tall(john,man,below_norm)

and this will come out as true, false or unprovable depending on what the facts are about John.

Also notice that our initial axiomatisation will not let us prove 'Kim is taller (for a man) than everyone else'.

(39) ∃d.tall(kim,man,d) ∧ ∀x.x ≠ kim → ¬tall(x,man,d)

For that to be provable we need to add a 'closed world' axiom, saying that the known individual men are the only men we are dealing with:

(40) ∀x.man(x) → x=kim ∨ x=andy ∨ x=sandy ... etc

It turns out that there are some serious limits to the strategy of reification that I have used to replace higher order constructs by first order proxies. For example, this approach cannot easily deal with multiple modification by gradeable adjectives: e.g.

(41) a. This is a pretty expensive cheap hotel

b. John is a tall old man

Ignore the conjunctive readings of these: the one we want is the reading on which the hotel is expensive by the standards of cheap hotels, and that John is tall by the standards of old men. But to formalise this on the pattern of our earlier examples, we would need to be able to construct scales and norms on the fly as combinations of those we already know about, or anticipate all the combinations we are likely to get. The latter is probably impossible in principle, and not very compositional even if it were possible. I would have no idea how to set about the former, which would involve combining something like norm(old,man,50,60) and norm(tall,man,5.67,5.83) to get something that was a norm of tallness for old men. Note that examples like this push our standard notions of formal semantics to the limit: our usual idea is that we set up a model to fix the interpretations of our non-logical constants, and then work out the denotations of complex expressions on this basis. But examples like this show that in interpreting natural language in this way at least some elements of models have to be constructed on the fly.

6 Intensifying degree modifiers

On Klein's analysis, 'very' shifts the norm for its adjective towards the top of the scale, and absorbs the extension gap: 'quite' does something similar:

```
<-----------------|------|------------------->
     not-A       extn gap        A
<------------------------------|------------->
     not-very-A                   very-A
<-----------------|-------------------------->
     not-A          quite A
```

We can implement this in a rather crude arithmetical fashion[2] by assuming that the amount that the extension gap moves is about the size of the gap itself:

(42) $\forall$x,p,d,y,z.tall(x,p,d) $\wedge$ norm(tall,p,y,z) $\wedge$ d $>$ z+(z-y) $\rightarrow$ tall(x,p,very(p))

And like Klein, we simply have to axiomatise the inferences:

(43) X is very Adj $\rightarrow$ X is Adj $\rightarrow$ X is quite Adj

(44) a. $\forall$x,p. tall(x,p,very(p)) $\rightarrow$ tall(x,p,above_norm(p))

b. $\forall$x,p. tall(x,p,above_norm(p)) $\rightarrow$ tall(x,p,quite(p))

Now it will correctly follow from 'John is very tall' and 'Bill is not very tall' that we get the various inferences like:

(45) a. John is taller than Bill

b. Bill is not as tall as John

Notice that the inferences in 45 will not follow from 'John is very tall' and 'Bill is quite tall', unless we implement some kind of 'maxim of quantity'. But this would be difficult to do consistently given the axioms above.

7 Much

As we saw, Klein treats 'much' as a quantifier:

(46) a. John is much taller than Bill

b. much(λd.d(tall)(john) $\wedge$ $\neg$d(tall)(bill))

[2]Since SPASS does not do 'procedural attachment' inferences involving arithmetic need to be hard-wired - this is what I did with the heights.

46b is true if the set denoted by the λ-expression is large relative to the context. But what is that set, and what is the context? As we saw earlier, we cannot just quantify over every possible degree modifier. We must restrict ourselves in some way. Intuitively, it is reasonably clear what is wanted. In the simple comparative case, we imagine the tallness of John measured on some scale, and the tallness of Bill measured on the same scale. The elements of the scale could be either points or intervals: we will assume that there are some primitive units, not further divisible, and that they are arranged in a sequence.

```
John  -------------------------------------
Bill  ----------------------------
```

The degree of tallness that John has and that Bill doesn't have needs to be one of those units on the part of the scale of John's tallness that stretches beyond Bill's tallness. For tallness, we have some conventional scales, for other properties, we may not, but we have to assume that they would behave in the same way. (This is tantamount to saying that it is a cultural rather than a linguistic fact that we happen not to say things like 'John is five einsteins cleverer than Bill'.)

Now we can interpret the set denoted by the λ-expression in 46b as the set of units on the part of the scale of John's tallness that stretches beyond Bill's tallness. (This is why we need some primitive units, otherwise there will be an infinite number of units along any stretch of the scale.) This set is large, but compared to what contextually determined comparison class? It can't just be the set of units in the heights of John or Bill, because even large height differences will always be small relative to these.

One possible interpretation is that the relevant context consists of the set of height differentials themselves, between the people we are concerned with. If the difference in height between John and Bill is, say, larger than the median height differential between any two individuals, then we would perhaps be tempted to say that it is large. At any rate, we will assume that this is so in order to proceed with our implementation.

We earlier argued that measure phrases like '5 inches' or 'one foot' could be treated as a kind of summing quantifier, giving a logical form like this:

(47) a. John is 5 inches taller than Bill is

b. 5-inches(λd.d(tall)(john) ∧ ¬d(tall)(bill))

The easiest way to implement this in our first order framework is to introduce a 'measure' predicate:

(48) ∃d.tall(john,man,d) ∧ ¬tall(bill,man,d) ∧ measure(d,5in)

If we can arrange for our axiomatised model to contain the median height differential for the relevant individuals, then we can also deal with 'much' in the same way:

(49) a. John is much taller than Bill

b. ∃d.tall(john,man,d) ∧ ¬tall(bill,man,d) ∧ much(d,man,tall)

- where we define 'much' by something like:

(50) ∀x,y,d,p,q. much(d,p,q) ↔ measure(d,x) ∧ median-diff(p,q,y) ∧ x > y

Simple as this treatment is, it manages to captures the inferences associated with these sentences. Both 48a and 49a correctly entail that John is taller than Bill. 48a entails that Bill is 5in less tall than John. 49a entails that Bill is much less tall than John. But on this type of analysis we will need to be careful how we treat Wh-questions:

(51) a. How tall is John?

b. ∃h,d.tall(john,man,d) ∧ measure(d,h)

can get an answer of h=5in. However, I think we can regard this as just another instance of the violation of quantity implicatures, since it is literally true according to the model we have set up. There are various obvious ways in which this problem might be solved.

8 Conclusions

The implementation of Klein's analysis in first order logic is reasonably successful, I feel, and has not strayed too for from the assumptions made in his analysis. I do not know of any other implemented analysis that manages to capture as many of the relevant inferences as this one. A natural question to ask, though, is how could we use an analysis like this in a realistic application?

The most obvious one that comes to mind (although rather unfashionable these days) would be as part of a database query system. Given that all the information needed to model scales and degrees could be assumed to be present in the database (perhaps not in the raw database, but some linguistically motivated entries derived from the raw data) it is not too difficult to imagine being able to cope sensibly with questions like:

(52) a. Find me a cheap holiday in Slovenia.
b. How much more expensive is it to fly with Air France instead of BA?
c. Is the Hotel Apple further from the beach than the Hotel Orange?

However, it is more challenging to see one might use this analysis of comparatives in augmenting a question answering system so that it could deal with questions like:

(53) a. Is Bill Clinton tall?
b. Is Bill Clinton taller than Bill Gates?

As we have seen, in order to deduce an answer for questions like this we need to determine the relevant scale (= height), establish the likely norms for the relevant Adj + Noun combinations, and find the position of the individuals in question on the scale. All of this needs to be representing in axiomatic form. As a quick *gedankenexperiment* to see whether future work in text mining might be able to discover this information from the web I searched Google (Sept 21st 2006) with the queries above. In the first two hits for the respective queries were contained these sentences:

> Clinton is 6 feet 2 inches tall, or perhaps 6 feet 2.5 inches tall
> Bill Gates's height is 5 ft 10 in (178 cm)

I also searched with the query 'average height for men' and the first hit contained the sentence:

> The average male American mens height is 177 cm, which is 69.7 inches, which is approximately 5 foot - 10 inches tall. (for white males).

So this might have been a lucky example, but the relevant information *was* there: transforming it into the axiomatic form required, however, is certainly a non-trivial exercise.

9 Acknowledgements

This paper is based on an invited talk at the Inference in Computational Semantics conference in Buxton, England, April 2006. I am grateful to the participants there for their comments: particularly Johan Bos, Lauri Karttunnen, Ian Pratt-Hartmann, and Allan Ramsay. I am also grateful to the IWCS referees, and to Ewan Klein for their comments on an earlier version of the paper.

References

Bos, J., S. Clark, M. Steedman, J. Curran, and J. Hockenmaier. 2004. Wide-coverage semantic representations from a CCG parser. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING '04)*, pages 1240–1246, Geneva, Switzerland. COLING.

Clark, Stephen and James R. Curran. 2004. Parsing the WSJ using CCG and log-linear models. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*, pages 103–110, Barcelona, Spain, July.

Kennedy, C. 2001. Polar opposition and the ontology of 'degrees'. *Linguistics and Philosophy*, 24:33–70.

Klein, E. 1980. A semantics for positive and comparative adjectives. *Linguistics and Philosophy*, 4:1–45.

Klein, E. 1982. The interpretation of adjectival comparatives. *Journal of Linguistics*, 18:113–136.

Schwarzschild, R. and K. Wilkinson. 2002. Quantifiers in comparatives: A semantics of degree based on intervals. *Natural Language Semantics*, 10(1):1–41.