Grading on the Fly

Steffen Staab & Udo Hahn
Computational Linguistics Lab
Freiburg University, Werthmannplatz 1
D-79085 Freiburg, Germany
URL: http://www.coling.uni-freiburg.de

Abstract

We specify a model for the conceptual interpretation of relative adjectives (like *"big"*), which covers a crucial aspect of the underlying comprehension process – the comparison to a norm that is associated with a comparison class. Building on an elaborate domain ontology and knowledge about intercorrelations, comparison classes are dynamically created depending on the context in which adjectival utterances occur.

Introduction

The conceptual description of relative adjectives[1] differs significantly from those of other word classes. Unlike concepts denoted by nouns and verbs, such degree expressions have no canonical, self-contained concept representation. They are rather dependent on a reference point or class norm that is associated with a *comparison class*. For instance, *"Peter"* in (1b) should not be referred to as *"tall"* in a general sense, but as *"tall in comparison to the class norm of a comparison class C"*, where *C* is constrained by the *context* in which *"tall"* occurs. This becomes immediately evident in example (1) where the context of the utterance (1a) crucially determines the valid comparison class for *"tall(Peter,C)"*.

(1) a. Peter is 4 years old.
  b. Peter is tall.

While linguists (Sapir, 1944; Klein, 1980; Bierwisch, 1989) have already agreed upon modeling relative adjectives like *"tall"* by a binary predicate that relates a degree to a comparison class (respectively, its class norm), many challenging problems from a natural language understanding perspective are still left open. These fall into two main categories, *viz.* representational and computational issues. From the *representational* perspective one may ask:

- How are comparison classes represented?
- What kind of knowledge is needed to select the most suitable comparison class in a given discourse context?

From the *computational* point of view one may ask:

- How are comparison classes actually determined given a degree expression?
- Is there a flexible, on-the-fly assembly process for comparison classes or are they just accessed from a static, precompiled class inventory?

In order to answer these questions, we introduce a model for the representation and proper selection of comparison classes. These representational prerequisites are then used in a computational model which accounts for a variety of linguistic phenomena. We claim that this joint model constitutes an improvement upon proposals which cannot account for dynamically created classes, like "4 year old boy". We start, however, with the presentation of experimentally grounded cognitive evidence which lays the framework on which we build our model of the comprehension of degree expressions.

A Cognitive Framework for Grading

Early on it was recognized that the determination of the class norm or the respective comparison class[2] against which the graded property of the object is compared is hard to determine (cf., e.g., Sapir (1944)). The only explicit proposal with respect to this problem we are aware of assumes that the comparison class is given by a superordinate class of the semantic subject[3] of the adjective, e.g., a superordinate of "Peter" in example (1) (Bierwisch, 1971; 1989). Though this proposal is on the right track, it is, nevertheless, insufficient to account for frequently occurring expressions which refer to uncommon or *ad hoc categories*.

Rosch et al. (1976) showed that people are aware of correlational structures by which attributes are linked. Also, they prefer to use categories that take maximal advantage of these linkages. For instance, "feather" and "flying" are strongly intercorrelated with each other and these attributes, as well as their intercorrelations, are strongly indicative for the *common*

---

[0]The first author is now affiliated with Fraunhofer IAO, Institute for Industrial Engineering, Nobelstr. 12, D-70569 Stuttgart — `http://www.swt.iao.fhg.de`.

[1]Relative adjectives like *"tall"* or *"fast"* are opposed to absolute adjectives such as *"married"* or *"rectangular"*. Though this distinction can be further refined (cf., e.g., Bierwisch (1989)), the particular relevance of relative adjectives is commonly agreed upon (Klein, 1980; Hutchinson, 1993).

[2]The terms *class norm* and *comparison class*, cf. Bierwisch (1971), can be interchanged with *reference point* and *reference class*, respectively, as used by Rips & Turnbull (1980).

[3]The semantic subject of an adjective is its head noun if it is in attributive position. If the adjective is in predicative position, the subject of the predicate is the semantic subject.

category "bird". In contrast, *ad hoc categories*,[4] e.g., "things to take for a camping trip", are defined by Barsalou (1983) as *"sets that (1) violate correlational structure and (2) are usually not thought of by most people"*. He finds that ad hoc categories generate typicality ratings very similar to the way common categories do. This is an important observation for any model for determining comparison classes that accounts for ad hoc categories like "4 year old boy". Only the existence of typical degrees for a gradable property allows the division of such a comparison class into groups of "more", "less" and "equal" with regard to the relevant graded property. In contrast to common categories, he also finds that ad hoc categories lack any strong category-instance and instance-category links.[5] He suggests that because ad hoc categories are so specialized, the perception of an entity should not activate all the ad hoc categories to which it belongs. Furthermore, he concludes that ad hoc categories should come to mind only when primed by current goals. Considering the apparent complexity of the task of constructing appropriate comparison classes, this raises the question as to why people are still so versatile at understanding graded attributes even when they encounter ad hoc categories such as "4 year old boy" in example (1).

Starting from Bierwisch's (1971) proposal, Rips & Turnbull (1980) investigate the reference class determination problem. They let subjects verify sentence pairs like (2) and (3). Whereas reaction time decreases from (2a) to (2b), no such change can be observed between (3a) and (3b).

(2) a. An insect is small.
    b. An insect is a small animal.

(3) a. An insect is six-legged.
    b. An insect is a six-legged animal.

Rips & Turnbull conclude that the determination of reference classes/points ought to be considered a dynamic process, one that uses information available from the discourse context. Given the assets from Bierwisch (1971), Rosch et al. (1976), and Barsalou (1983), our basic idea for determining the comparison class is to use the correlational structure between a grading attribute and the properties which are attached to the major category. With regard to size then, the property of being 4 years old has a different knowledge status than, e.g., the property of being of a fair complexion. But, how much correlational structure is available to humans?

Kersten & Billman (1992) have investigated the correlational structure of complex events. Subjects observed events in an artificial world rendered on a computer screen. Agents, patients and the environment in this scenario each had a set of attributes which correlated with the displayed events. An agent, e.g., with one body color approaching the patient made it flee, another agent triggered a color change in the patient. In fact, Kersten & Billman (1992) found that correlations between attributes and events were learned. Simulations with richly intercorrelated attribute-behavior patterns generated a rather high learning accuracy. Richly correlated settings generated higher accuracy rates than lowly correlated ones. Hence, these findings support the availability of complex intercorrelations like *"an expensive printer produces good output"* in richly intercorrelated world descriptions.

We may now summarize the cognitive framework underlying our model of determining comparison classes: Barsalou's conclusion indicates that only information given in the *discourse context* should play a role for determining reference classes/points. Also, the existence of typicality effects for ad hoc categories fosters the assumption that *reference points* do exist *for ad hoc categories*. The results from Rips & Turnbull (1980) yield support to a dynamic reference class/point model. Finally, we will exploit rich *intercorrelation* knowledge to guide the computation of comparison classes.

Representation of Comparison Classes

Our efforts directed at the conceptualization of degree expressions are rooted in a text understanding system (Hahn et al., 1996), which operates in two domains, *viz.* test reports from information technology magazines and medical finding reports. In each domain, understanding the evaluative portions of the discourse is vital for adequate comprehension results. A typical example from our IT corpus is given by (4).

(4) The picture has good quality for a <u>picture printed by a laser printer.</u>

In this example, a relative adjective occurs with an explicitly given comparison class. Hence, its computation merely boils down to a parsing problem and the associated knowledge base operations for the generation of a conceptual interpretation of the utterance. In our system, the representation of a comparison class is *dynamically* created from the utterance and the concepts available in the knowledge base.

Following the terminology introduced by Bierwisch (1989) and Klein (1980), we say that a relative adjective a is related to a *class norm*,[6] which is a degree of the same type (e.g., QUALITY) as the one described by the adjective (e.g., *"good"*). The class norm belongs to a *comparison class* (e.g., the set of pictures printed by laser printers), which is a class C with instances o_i. If the degree of such an instance o_i of C is above the class norm, one may assert that "o_i is a for C".

We use a terminological knowledge representation system (cf. Woods & Schmolze (1992) for a survey) as a framework for the specification of domain knowledge. It allows, e.g., to create a comparison class COMP-CLASS-1 for example (4) on the fly. COMP-CLASS-1 is defined by restricting the class, PICTURE, to pictures printed by a LASER-PRINTER, which is

[4] This notion is extended by Barsalou (1985) toward goal-derived categories. This (roughly) subsumes true ad hoc categories as well as categories which once have been ad hoc but meanwhile turned into conventionalized expressions.

[5] Note that these results, in particular the lack of category-instance links, predict that subjects should have difficulties in determining this reference point, though they should have less problem with categorizing instances into the "high", "medium", or "low" group for the relevant graded property.

[6] We abstract here from different graded properties which are associated with different class norms.

a subconcept of PRINTER (cf. Fig. 1, with COMP-CLASS-1 $\doteq$ PICTURE $\sqcap$ $\forall$PRINTED-BY.LASER-PRINTER). As a necessary result, the instance of picture, O-1, is classified not only as belonging to PICTURE, but also to COMP-CLASS-1. In a metarelation (CLASS-NORM-OF) the comparison class is associated with CLASS-NORM-1 which is related to the quality Q-1 of the picture O-1 by the relation EXCEEDS.

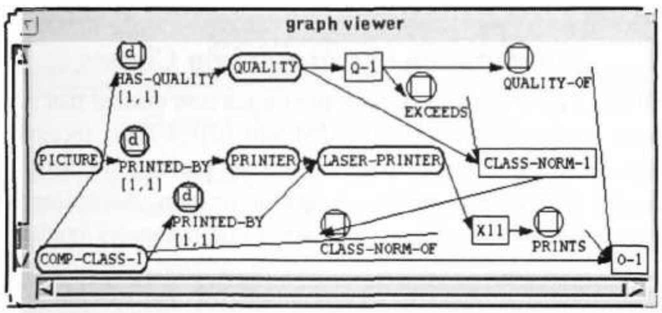

Figure 1: Representing Comparison Classes and Class Norms

Knowledge about Intercorrelations

In a discourse setting, various linguistic expressions can be formed to associate an adjective with a comparison class.

(5) Paul is 4 years old. He is tall.

(6) Paul celebrated his 4th birthday yesterday. He is tall.

(7) Paul is tall for a 4 year old boy.

These examples indicate that purely linguistic criteria, e.g., Bierwisch's proposal mentioned earlier, are insufficient to restrict the comparison class of an adjective. Computations that rely only on static knowledge structures fail to determine the proper interpretations, too. As an alternative, we use *(meta)knowledge about intercorrelations* that describes how a class subhierarchy may influence the relations of class norm instances on a scale *or* how two degrees of a given concept are correlated. As an example, consider the sentences (8) and (9). In both of these the comparison classes are stated explicitly, and, thus, they elucidate the distinction between a proper comparison class restriction and an improper one:

(8) Peter is tall for a gymnast.

(9) ? Peter is tall for a flute player.

The intercorrelation between "hasHeight" and "practicesGymnastics" describes gymnasts to be usually smaller than average people. So, being tall for a gymnast does not necessarily imply being tall for the comparison class of all people. It is exactly the absence of corresponding intercorrelations between "hasHeight" and "playsFlute" that renders the restriction of the comparison class to flute players awkward.

Several important aspects of intercorrelations should be noted here. First, knowledge about intercorrelations is part of humans' common-sense knowledge (Malt & Smith, 1984; McRae, 1992). Second, these intercorrelations need not be symmetrical.[7] Third, it need not be the case, of course, that all possible intercorrelations one may conceive of are also encoded by people – only the particularly salient ones are available (cf. Malt & Smith (1984), p. 264). This is not an argument against, but rather one in favor of our proposal, since it conforms with observations we made about the formation of comparison classes for utterances from our text corpus. Finally, it is not necessary that knowledge about intercorrelations be overly fine-grained. The need for constructing a conceptually more specific comparison class arises only due to the *strength* of the intercorrelation.[8] As a consequence, the specification of intercorrelations and, thus, the construction of new comparison classes are subject to a principle of parsimony, since only the *most relevant* intercorrelations have to be accounted for. One could think of means to provide a strength indicator in the representation of intercorrelations. For reasons of simplicity, we limited our approach such that an intercorrelation is either represented or it is not.

In order to exploit knowledge about intercorrelations, we first specify what they describe and how they are represented. The intercorrelations we consider characterize local *restriction classes* which will later be gathered to define the comparison class. We categorize intercorrelations along two dimensions. Considering the symbolic representation layer, one is the length of the intercorrelation structure. This roughly corresponds to the distinction between intercorrelations within object categories and across event structures as made by Kersten & Billman (1992). The other dimension is given by the type of the property (whether it is gradable or not) that correlates with the degree that is interpreted.

Sentence (10) illustrates a simple case of a degree-hierarchy intercorrelation (relevant comparison classes are underlined; for a description of the relevant relations in the knowledge base, cf. Fig. 2). In this example, the relevant comparison class (LASER-PRINTER) is the concept NOISE-LEVEL is *directly* associated with. Therefore, the path from the relevant degree NOISE-LEVEL to the relevant restriction class LASER-PRINTER has the unit length *1* (inheritance links are not counted). Example (11) refers to the same type of intercorrelation, but it takes effect across the relation PRINTS which represents printing events in our knowledge

[7] Common-sense knowledge tells us that though gymnasts tend to be smaller than average people, small people do not tend to do gymnastics very much. Assume that a population consists of 50% small and 50% tall people, respectively, 1% being gymnasts, and 90% of the gymnasts being small. Then the probability that a gymnast is a small person is 90%. However, the probability that a small person is a gymnast is only 1.8%. Thus, restricting a comparison class from all people to gymnasts, in fact, decreases the class norm for height considerably, while the reverse is not true.

[8] For instance, for *"a small gymnast"*, it is necessary to define the comparison class GYMNAST (as opposed to the more general class HUMAN) in order to assure that proper assessments about the property HEIGHT can be derived. For *"a small iceskater"*, however, the construction of a corresponding comparison class ICESKATER could possibly be justified, but is not necessary at all. This is due to the fact that iceskaters can still be compared relative to the general class of humans with respect to their height, even though a weak intercorrelation might hold between HEIGHT and ICESKATERS, *viz.* a preference for being small. This case of a weak correlation can further be distinguished from one in which actually no intercorrelation seems reasonable as in the case of SPRINTERS, whose average heights do not seem to differ from those of other persons.

base. Thus, it differs in the length of the distance (two relations have to be passed) that lies between one of the relevant restrictions, 300DPI-LASER-PRINTER, and the degree QUALITY (of picture). Finally, (12) shows an example where the intercorrelation differs with regard to the types that are engaged, *viz.* in contrast to (10) and (11) an intercorrelation between two degrees holds here.

(10) *Degree-hierarchy intercorrelation (with distance 1):* The noise level of the 300dpi laser printer X11 is high for a laser printer.

(11) *Degree-hierarchy intercorrelation (with distance 2):* The picture of the X11 has a good quality for the picture of a 300dpi laser printer.

(12) *Degree-degree intercorrelation (with distance 2):* The X11 offers very good quality for a laser printer that costs $800.

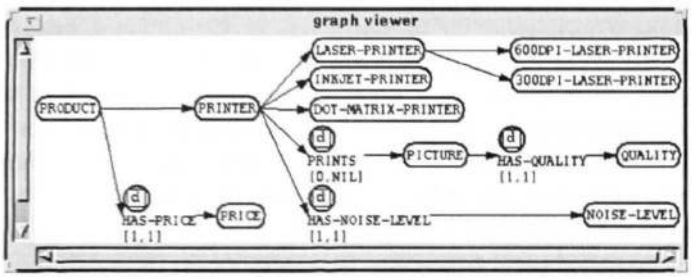

Figure 2: Hierarchy and Definitory Roles

In order to represent the above-mentioned intercorrelations, knowledge is made available about which relations lie between the restricting hierarchy and the correlated degree. For instance, the representation of the intercorrelation appearing in (11) is depicted in Fig. 3:

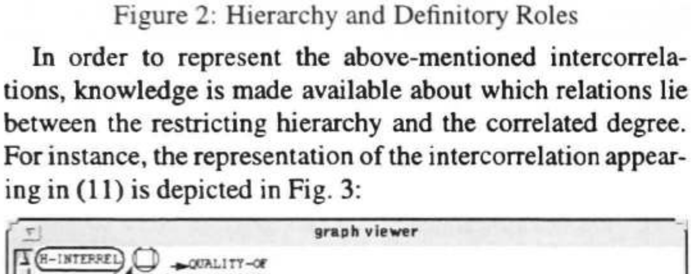

Figure 3: The Intercorrelation Knowledge Needed for (11).

The relations QUALITY-OF and PRINTED-BY connect the relevant degree, QUALITY, with the subhierarchy with which it correlates, *viz.* the hierarchy below PRINTER[9]. This connection is reflected by QUALITY-OF having the domain (D:) QUALITY, by PRINTED-BY having the range (R:) PRINTER, and by both being associated to the hierarchy-degree intercorrelation I-1 through the series of relations: REL-1, REL-2. Moreover, it must be known which common subclasses of PRINTER have a norm attached for NOISE-LEVEL, which is either below or above the class norm associated with their direct superclass. In our example, LASER-PRINTER, INKJET-PRINTER and 600DPI-LASER-PRINTER belong to the set of classes that are associated with class norms above that of their superclass, while DOT-MATRIX-PRINTER and 300DPI-LASER-PRINTER relate to corresponding lower class norms. They are therefore marked as belonging to the POS-CLASS set and to the NEG-CLASS set of I-1, respectively.

[9]We abstract here from the consideration of multi-hierarchies. Furthermore, for each relation (e.g., HAS-QUALITY, PRINTS) we always assume the existence of its inverse, which is then referred to by an intuitively plausible name (e.g., QUALITY-OF, PRINTED-BY).

Computation of Comparison Classes

Often relative adjectives refer to comparison classes that are only implicitly available (cf. (5) and (6)). Their recognition cannot be considered the task of the parsing mechanism proper, but rather constitutes a task on its own. Accordingly, we illustrate here such an algorithm that computes implicit comparison classes by making use of semantic relations, of the knowledge about intercorrelations, of text-specific and world knowledge, and of the representation mechanism for comparison classes. As a starting condition, we presume the completion of anaphora resolution, verb interpretation and the interpretation of prepositional phrases.

The basic idea of the algorithm for computing implicit comparison classes is expressed in Fig. 4 (for a more technical presentation, cf. Staab & Hahn (1997)): A positive adjective a denotes a degree d in the current text fragment. This degree d is related to an object o_1, which itself is related to another object o_p. Of course, there might be no object or several objects related to o_1, and o_p itself might have other relations. Each object o_i has a most specific type $C_{i,1}$.

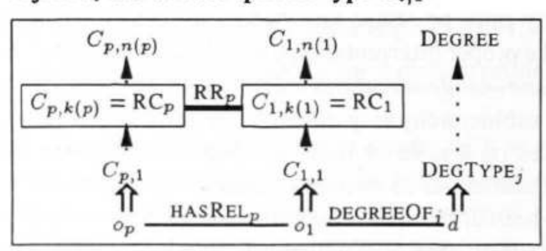

Figure 4: Knowledge Structures for Comparison Classes

The goal of the algorithm is to select all objects o_i that are relevant for the computation of the correct comparison class. Furthermore, for each object o_i it must select its correct intermediate superconcept $C_{i,k(i)}$, which neither restricts the comparison class too narrowly (as $C_{i,1}$ might do) nor too widely (as $C_{i,n(i)}$ might do, since it yields no restriction at all). This goal is achieved by matching the available knowledge on intercorrelations against the semantic structures of the current text fragment. Finally, a comparison class is (recursively) computed by combining all the gathered restrictions. In Fig. 4 this means that the new comparison class is defined by restricting RC_1 to a new class where the role RR_p is restricted to the range RC_p. We illustrate the comparison class determination by considering example (13):

(13) The picture with the giraffe was printed by the fast laser printer X11. It shows good quality.

The information conveyed by this fragment is depicted in Fig. 5. We must now find the proper comparison class for the

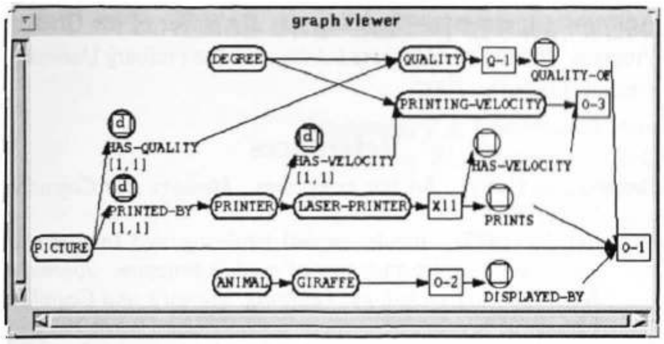

Figure 5: Input for Example Text

graded property *"good quality"*. PICTURE O-1 itself is not in a particular hierarchy that correlates with QUALITY Q-1, but since QUALITY is modeled at the level of PICTURE, the latter is chosen as the first restriction class from which the comparison class is computed. PICTURE O-1 is furthermore related to GIRAFFE O-2 and to PRINTER X11. The former has no correlation whatsoever with the quality of the picture — unless knowledge about such a correlation has been introduced in the preceding discourse. Given that no such correlation is available, there is no reason to consider GIRAFFE for the computation of the comparison class. The object X11, however, shows an intercorrelation with the quality of the picture, because laser printers tend to produce better output than general printers, which include the class of dot matrix printers (cf. Fig. 3). At this point our algorithm always proceeds with the most specific concept for which such an intercorrelation is found. Under this heuristic, LASER-PRINTER is activated here, though in general PRINTER might also be a reasonable alternative.

The algorithm proceeds recursively here, meaning that it considers the objects related to X11. One could possibly imagine that the PRINTING-VELOCITY correlates with the quality of the picture, since high velocity printers tend to be more expensive and more expensive printers tend to produce higher quality. An expert in the field of printers could perhaps produce such a reading which differs from that of a novice. However, since the intercorrelation between PRINTING-VELOCITY and QUALITY is weak, if there is one at all, including or disregarding it will hardly affect the location of the class norm to which "good quality" is compared. Here, for our system, we decided that it was too weak to be included. Hence, we ignore the velocity property for X11 and end up with the restriction classes PICTURE and LASER-PRINTER which are composed to COMP-CLASS-1, *picture printed by a laser printer*, as shown in the section on explicit comparison classes. Fig. 1 depicts the corresponding relations computed by our algorithm for example (13).

Empirical Data

In a preliminary empirical evaluation study we compared our algorithm (henceforth, *c3*) against two simpler, more naive approaches. The first of these, *n1*, constantly uses the most specific concept the semantic subject has. The second one, *n2*, does not select this most specific concept, but its immediate superconcept instead. Both approaches constitute somewhat of a lower bottom line to our approach, since it can revert to one of these simpler approaches if it is unable to identify more selective restrictions.

We chose a text which contained 226 sentences with about 4,300 words. 121 positive gradable adjectives were screened, for which a reasonable semantic representation could be determined in 72 cases – and only these were evaluated. The remaining 49 occurrences graded idiomatic expressions, concepts that are hard to model (e.g., *"a good idea"*), or entailed other problems that were not directly related to finding the correct comparison classes. Under the assumption of complete knowledge, *c3* achieved a high success rate (60 cases (83%) were correctly analyzed). *n1* and *n2* performed much worse, as they were only able to properly determine 20 and 15 valid comparison classes (28% *vs.* 21%), respectively.

n1 and *n2* are equivalent to the procedures Bierwisch (1989) suggests for adjectives related with generic and non-generic nouns, respectively (e.g., in *"towers are high"* the related noun *"towers"* is generic, while in *"this tower is high"* it is not). An oracle that tells whether an adjective is related to a generic object and, depending on the result, changes the strategy from *n1* to *n2* would render a mechanism close to the one Bierwisch proposes. However, it would not add much benefit. Since none of the 72 considered adjectives are related to generic nouns, the positive cases of *n2* are not due to any generic use. Our results are still interesting, even though we restricted our approach to *distance-1* and *distance-2* interrelations to keep the modelling problem manageable.

Related Work

Though the notion of comparison class has been around for quite a long time among linguists (Sapir, 1944; Klein, 1980; Simmons, 1993), no comprehensive theory of comparison class formation has been shaped that accounts for ad hoc categories and properly incorporates context information. We build our model for determining comparison classes on considerations by Bierwisch (1971, 1989) and findings by Rosch et al. (1976), Rips & Turnbull (1980), Barsalou (1983, 1985), and Kersten & Billman (1992). In particular, we extend Bierwisch's (1971) approach to cover ad hoc categories. Rips & Turnbull's (1980) findings support our model in that they favor a dynamic process without excluding rich domain knowledge that guides the understanding process. Moreover, Barsalou's (1983) results support the existence of reference points also for complex categories like *"quality of a picture printed by a laser printer"*.

Further supporting evidence for our proposal is available from research that does not directly address the comparison class formation problem, but which is based on experiments that indicate that major assumptions underlying our approach can be traced in empirical findings. First, distance and contiguity effects that are observed in comparative judgments suggest that people categorize dynamically for grading processes (Sailor & Shoben, 1993; Cech et al., 1990). Second, several sources (e.g., McRae (1992), Sailor & Shoben (1993)) maintain the assumption that people encode knowledge about intercorrelations, which lies at the heart of the proposed mechanism, and use this information for categorization processes. In particular, Kersten & Billman (1992) report that intercorrelations are not restricted to simple object categories, but are also learned for complex dependencies, e.g., a more expensive printer produces better output. This holds especially in richly intercorrelated settings such as the commonsense world.

The importance of comparison classes for the semantics of relative adjectives has often been underestimated. Much previous work on the representation of degrees completely abstracts from the problem of comparison class determination. Simmons' interval approach (1993) uses class norms to denote the meaning of relative adjectives, but disregards the comparison class formation problem. Other computational accounts, e.g., Raskin & Nirenburg (1996), Zadeh (1978), neglect the effects a comparison class has at all.

Hutchinson (1993) shows in detail that comparison classes are not an inherently semantic feature, but rather dependent on language use. He also gives examples that go way beyond the capabilities of our comparison class determination method. Example (14) could plausibly mean that Chomsky is famous for a linguist, for a scholar, or even for an American.

(14) Chomsky is a famous linguist.

Though we cannot cope with all the challenges Hutchinson (1993) puts forth, our proposal improves the existing model in a way that makes it interesting for text understanding systems.

Finally, one should note that reference points affect comparative judgements, in general. Holyoak & Mah (1982) observed that explicit reference points strongly increase discriminability in their vicinity. However, for implicit reference points they could produce only inconclusive evidence.

Conclusion

Only little evidence has been collected so far concerning the conceptualizations underlying adjectival expressions, relative adjectives in particular — the third major word class of Western languages. We have introduced a model of adjective interpretation that accounts for some of the intriguing complexities of relating degree expressions to a proper conceptual representation. At the center of the model lie comparison classes and their associated class norms to which degree expressions are related. This is not a static linkage. Rather, contextual information together with knowledge about correlations controls the process of selecting the appropriate comparison class on the fly. We extend Bierwisch's (1989) approach thereby, and include ad hoc categories into our model.

Still, some desiderata remain unsolved: A more comprehensive model, e.g, would have to take into account shared beliefs between participants in the discourse, since these may substantially influence the comparison class formation process. Also, granularity effects in the knowledge base are notoriously difficult, but should be solvable along the lines of path-length neutral computations for textual ellipsis resolution as discussed by Markert et al. (1996).

Acknowledgements. S. Staab was a member of the Graduate Program *"Human and Machine Intelligence"* at Freiburg University, which is funded by *DFG*.

References

Barsalou, L. (1983). Ad hoc categories. *Memory and Cognition*, 11:211–227.

Barsalou, L. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11:629–654.

Bierwisch, M. (1989). The semantics of gradation. In M. Bierwisch & E. Lang (Eds.), *Dimensional Adjectives*, pp. 71–261. Berlin: Springer.

Bierwisch, M. (1971). On classifying semantic features. In D. Steinberg & L. Jakobovits (Eds.), *Semantics : An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*, pp. 410–435. Cambridge: Cambridge University Press.

Cech, C., E. Shoben & M. Love (1990). Multiple congruity effects in judgments of magnitude. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(6):1142–1152.

Hahn, U., M. Romacker & K. Schnattinger (1996). Automatic knowledge acquisition from medical texts. In *Proc. of the 1996 AMIA Annual Fall Symposium*, pp. 383–387.

Holyoak, K. J. & W. Mah (1982). Cognitive reference points in judgments of symbolic magnitude. *Cognitive Psychology*, 14:328–352.

Hutchinson, L. (1993). The logic of relative adjectives. In M. Eid & G. Iverson (Eds.), *Principles and Prediction: The Analysis of Natural Language*, pp. 105–117. Amsterdam, Philadelphia: John Benjamins.

Kersten, A. W. & D. O. Billman (1992). The role of correlational structure in learning event categories. In *Proc. of the CogSci '92*, pp. 432–437.

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

Malt, B. & E. Smith (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior*, 23(2):250–269.

Markert, K., M. Strube & U. Hahn (1996). Inferential realization constraints on functional anaphora in the centering model. In *Proc. of the CogSci '96*, pp. 609–614.

McRae, K. (1992). Correlated properties in artifact and natural kind concepts. In *Proc. of the CogSci '92*, pp. 349–354.

Raskin, V. & S. Nirenburg (1996). Adjectival modification in text meaning representation. In *Proc. of COLING-96*, pp. 842–847.

Rips, L. & W. Turnbull (1980). How big is big? Relative and absolute properties in memory. *Cognition*, 8:145–174.

Rosch, E. H., C. Mervis, W. D. Gray, D. Johnson & P. Boyes-Braem (1976). Basic objects in natural categories. *Cognitive Psychology*, 8:382–439.

Sailor, K. & E. Shoben (1993). Effects of category membership on comparative judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(6):1321–1327.

Sapir, E. (1944). Grading: A study in semantics. *Philosophy of Science*, 11:93–116.

Simmons, G. (1993). A tradeoff between compositionality and complexity in the semantics of dimensional adjectives. In *Proc. of the EACL-93*, pp. 348–357.

Staab, S. & U. Hahn (1997). 'Tall', 'good', 'high' — compared to what? In *Proc. of the IJCAI-97*, pp. 996–1001.

Woods, W. A. & J. G. Schmolze (1992). The KL-ONE family. *Computers & Mathematics with Applications*, 23(2-5):133–177.

Zadeh, L. A. (1978). PRUF: A meaning representation language for natural languages. *International Journal of Man-Machine Studies*, 10:395–460.