Towards a Computational Treatment of Superlatives

Silke Scheible
Institute for Communicating and Collaborative Systems (ICCS)
School of Informatics
University of Edinburgh
S.Scheible@sms.ed.ac.uk

Abstract

I propose a computational treatment of superlatives, starting with superlative constructions and the main challenges in automatically recognising and extracting their components. Initial experimental evidence is provided for the value of the proposed work for Question Answering. I also briefly discuss its potential value for Sentiment Detection and Opinion Extraction.

1 Introduction

Although superlatives are frequently found in natural language, with the exception of recent work by Bos and Nissim (2006) and Jindal and Liu (2006), they have not yet been investigated within a computational framework. And within the framework of theoretical linguistics, studies of superlatives have mainly focused on particular semantic properties that may only rarely occur in natural language (Szabolcsi, 1986; Heim, 1999).

My goal is a comprehensive computational treatment of superlatives. The initial question I address is how useful information can be automatically extracted from superlative constructions. Due to the great semantic complexity and the variety of syntactic structures in which superlatives occur, this is a major challenge. However, meeting it will benefit NLP applications such as Question Answering, Sentiment Detection and Opinion Extraction, and Ontology Learning.

2 What are Superlatives?

In linguistics, the term "superlative" describes a well-defined class of word forms which (in English) are derived from adjectives or adverbs in two different ways: Inflectionally, where the suffix *-est* is appended to the base form of the adjective or adverb (e.g. *lowest*, *nicest*, *smartest*), or analytically, where the base adjective/adverb is preceded by the markers *most*/*least* (e.g. *most interesting*, *least beautiful*). Certain adjectives and adverbs have irregular superlative forms: *good* (*best*), *bad* (*worst*), *far* (*furthest*/*farthest*), *well* (*best*), *badly* (*worst*), *much* (*most*), and *little* (*least*).

In order to be able to form superlatives, adjectives and adverbs must be *gradable*, which means that it must be possible to place them on a scale of comparison, at a position higher or lower than the one indicated by the adjective/adverb alone. In English, this can be done by using the comparative and superlative forms of the adjective or adverb:

[1] (a) Maths is ***more difficult*** than Physics.
(b) Chemistry is ***less difficult*** than Physics.
[2] (a) Maths is the ***most difficult*** subject at school.
(b) History is the ***least difficult*** subject at school.

The ***comparative*** form of an adjective or adverb is commonly used to compare two entities to one another with respect to a certain quality. For example, in [1], Maths is located at a higher point on the difficulty scale than Physics, and Chemistry at a lower point. The ***superlative*** form of an adjective is usually used to compare one entity to a set of other entities, and expresses the end spectrum of the scale: In [2], Maths and History are located at the highest and lowest points of the difficulty scale, respectively, while all the other subjects at school range somewhere in between.

3 Why are Superlatives Interesting?

From a computational perspective, superlatives are of interest because they express a comparison

Proceedings of the ACL 2007 Student Research Workshop, pages 67–72,
Prague, June 2007.

between a target entity (indicated in bold) and its comparison set (underlined), as in:

[3] **The blue whale** is the *largest* mammal.

Here, the target *blue whale* is compared to the comparison set of *mammals*. Milosavljevic (1999) has investigated the discourse purpose of different types of comparisons. She classifies superlatives as a type of *set complement comparison*, whose purpose is to highlight the uniqueness of the target entity compared to its contrast set.

My initial investigation of superlative forms showed that there are two types of relation that hold between a target and its comparison set:

Relation 1: *Superlative relation*
Relation 2: *IS-A relation*

The *superlative relation* specifies a property which all members of the set share, but which the target has the highest (or lowest) degree or value of. The *IS-A* (or *hypernymy*) *relation* expresses the membership of the target in the comparison class (e.g. its parent class in a generalisation hierarchy). Both of these relations are of great interest from a relation extraction point of view, and in Section 6, I discuss their use in applications such as Question Answering (QA) and Sentiment Detection and Opinion Extraction. That a computational treatment of superlatives is a worthwhile undertaking is also supported by the frequency of superlative forms in ordinary text: In a 250,000 word subcorpus of the WSJ corpus[1] I found 602 instances (which amounts to roughly one superlative form in every 17 sentences), while in the corpus of animal encyclopaedia entries used by Milosavljevic (1999), there were 1059 superlative forms in 250,000 words (about one superlative form in every 11 sentences).[2] These results show significant variation in the distribution of superlatives across different text genres.

4 Elements of a Computational Treatment of Superlatives

For an interpretation of comparisons, two things are generally of interest: What is being compared, and with respect to what this comparison is made. Given that superlatives express set comparisons, a computational treatment should therefore help to identify:

a) The **target** and comparison set
b) The type of ***superlative relation*** that holds between them (cf. Relation 1 in Section 3)

However, this task is far from straightforward, firstly because superlatives occur in a variety of different constructions. Consider for example:

[4] The pipe organ is the ***largest*** instrument.
[5] Of all the musicians in the brass band, Peter plays the ***largest*** instrument.
[6] The human foot is ***narrowest*** at the heel.
[7] First Class mail usually arrives the ***fastest***.
[8] This year, Jodie Foster was voted ***best*** actress.
[9] I will get there at 8 at the ***earliest***.
[10] I am ***most tired*** of your constant moaning.
[11] ***Most*** successful bands are from the U.S.

All these examples contain a superlative form (bold italics). However, they differ not only in their syntactic structure, but also in the way in which they express a comparison. Example [4] contains a clear-cut comparison between a target item and its comparison set: The pipe organ is compared to all other instruments with respect to its size. However, although the superlative form in [4] occurs in the same noun phrase as in [5], the comparisons differ: What is being compared in [5] is not just the instruments, but the musicians in the brass band with respect to the size of the instrument that they play. In example [6], the target and comparison set are even less easy to identify. What is being compared here is not the human foot and a set of other entities, but rather different *parts* of the human foot. In contrast to the first two examples, this superlative form is not incorporated in a noun phrase, but occurs freely in the sentence. The same applies to *fastest* in example [7], which is an adverbial superlative. The comparison here is between First Class mail and other mail delivery services. Finally, examples [8] to [11] are not proper comparisons: *best actress* in [8] is an idiomatic expression, *earliest* in [9] is part of a so-called PP superlative construction (Corver and Matushansky, 2006), and [10] and [11] describe two non-comparative uses of *most*, as an intensifier and a proportional quantifier, respectively (Huddleston and Pullum, 2002).

Initially, I will focus on cases like [4], which I call IS-A superlatives because they make explicit the IS-A relation that holds between target and comparison set (cf. Relation 2 in Section 3). They

[1] www.ldc.upenn.edu/Catalog/LDC2000T43.html
[2] In the following, these 250,000 word subcorpora will be referred to as SubWSJ and SubAC.

are a good initial focus for a computational approach because both their target and comparison set are explicitly realised in the text (usually, though not necessarily, in the same sentence). Common surface forms of IS-A superlatives involve the verb "to be" ([12]-[14]), appositive position [15], and other copula verbs or expressions ([16] and [17]):

[12] **The blue whale** is the ***largest*** mammal.
[13] **The blue whale** is the ***largest*** of all mammals.
[14] Of all mammals, **the blue whale** is the ***largest***.
[15] The ***largest*** mammal, the blue whale**,** weighs...
[16] **The ostrich** is considered the ***largest*** bird.
[17] **Mexico** claimed to be the ***most peaceful*** country in the Americas.

IS-A superlatives are also the most frequent type of superlative comparison, with 176 instances in SubWSJ (ca. 30% of all superlative forms), and 350 instances in SubAC (ca. 33% of all superlative forms).

The second major problem in a computational treatment of superlatives is to correctly identify and interpret the comparison set. The challenge lies in the fact that it can be restricted in a variety of ways, for example by preceding possessives and premodifiers, or by postmodifiers such as PPs and various kinds of clauses. Consider for example:

[18] **VW** is [Europe's ***largest*** **maker** of cars].
[19] **VW** is [the ***largest*** European car **maker** with this product range].
[20] **VW** is [the ***largest*** car **maker** in Europe] with an impressive product range.
[21] In China, **VW** is by far [the ***largest*** car **maker**].

The phrases *of cars* and *car* in [18] and [19] both have the role of specifying the type of *maker* that constitutes the comparison set. The phrases *Europe's*, *European* and *in Europe* occur in determinative, premodifying, and postmodifying position, respectively, but all have the role of restricting the set of car makers to the ones in Europe. And finally, the "with" PP phrases in [19] and [20] both occur in postmodifying position, but differ in that the one in [19] is involved in the comparison, while the one in [20] is non-restrictive. In addition, restrictors of the comparison can also occur elsewhere in the sentence, as shown by the PP phrase and adverbial in [21]. It is evident that in order to extract useful and reliable information, a thorough syntactic and semantic analysis of superlative constructions is required.

5 Previous Approaches

5.1 Jindal and Liu (2006)

Jindal and Liu (2006) propose the study of comparative sentence mining, by which they mean the study of sentences that express "an ordering relation between two sets of entities with respect to some common features" (2006). They consider three kinds of relations*: non-equal gradable* (e.g. *better*)*, equative* (e.g. *as good as*) and *superlative* (e.g. *best*). Having identified comparative sentences in a given text, the task is to extract comparative relations from them, in form of a vector like *(relationWord, features, entityS1, entityS2)*, where *relationWord* represents the keyword used to express a comparative relation, *features* are a set of features being compared, and *entityS1* and *entityS2* are the sets of entities being compared, where *entityS1* appears to the left of the relation word and *entityS2* to the right. Thus, for a sentence like "*Canon's optics is better than those of Sony and Nikon*", the system is expected to extract the vector (*better*, {*optics*}, {*Canon*}, {*Sony*, *Nikon*}).

For extracting the comparative relations, Jindal and Liu use what they call *label sequential rules* (LSR), mainly based on POS tags. Their overall F-score for this extraction task is 72%, a big improvement to the 58% achieved by their baseline system. Although this result suggests that their system represents a powerful way of dealing with superlatives computationally, a closer inspection of their approach, and in particular of the gold standard data set, reveals some serious problems.

Jindal and Liu claim that for superlatives, the *entityS2* slot is "normally empty" (2006). Assuming that the members of *entityS2* usually represent the comparison set, this is somewhat counterintuitive. A look at the data shows that even in cases where the comparison set is explicitly mentioned in the sentence, the *entityS2* slot remains empty. For example, although the comparison set in [22] is represented by the string *these 2nd generation jukeboxes (ipod , archos , dell , samsung)*, it is not annotated as *entityS2* in the gold standard:

```
[22] all reviews i 've seen seem to indicate that the creative mp3 jukeboxes have the best sound quality of these 2nd generation jukeboxes ( ipod , archos , dell , samsung ) .
```

(*best*, {*sound quality*}, {*creative mp3 jukeboxes*}, {--})
Jindal and Liu (2006)

Furthermore, Jindal and Liu do not distinguish between different types of superlatives. In constructions where the superlative form is incorporated into an NP, Jindal and Liu consistently interpret the string following the superlative form as a "feature", which is appropriate for cases like [22], but does not apply to superlative sentences involving the copula verb "to be" (as e.g. in [4]), where the NP head denotes the comparison set rather than a feature. A further major problem is that restrictions on the comparison set as the ones discussed in Section 4 and negation are not considered at all. Therefore, the reliability of the output produced by the system is questionable.

5.2 Bos and Nissim (2006)

In contrast to Jindal and Liu (2006), Bos and Nissim's (2006) approach to superlatives is explicitly semantic. They describe an implementation of a system that can automatically detect superlatives, and determine the correct comparison set for attributive cases, where the superlative form is incorporated into an NP. For example in [23], the comparison set of the superlative *oldest* spans from word 3 to word 7:

[23] `wsj00 1690 [...] Scope: 3-7`
`The` ***`oldest`*** <u>`bell-ringing group in the country`</u> `, the Ancient Society of College Youths , founded in 1637 , remains male-only , [...] .`

(Bos and Nissim 2006)

Bos and Nissim's system, called DLA (Deep Linguistic Analysis), uses a wide-coverage parser to produce semantic representations of superlative sentences, which are then exploited to select the comparison set among attributive cases. Compared with a baseline result, the results for this are very good, with an accuracy of 69%-83%.

The results are clearly very promising and show that comparison sets can be identified with high accuracy. However, this only represents a first step towards the goal of the present work. Apart from the superlative keyword *oldest*, the only information example [23] provides is that the comparison set spans from word 3 to word 7. However, what would be interesting to know is that the target of the comparison appears in the same sentence and spans from word 9 to word 14 (*the Ancient Society of College Youths*). Furthermore, no analysis of the semantic roles of the constituents of the resulting string is carried out: We lose the information that *the Ancient Society of College Youths* **IS-A** kind of *bell-ringing group*, and that the set of bell-ringing groups is restricted in location (*in the country*).

6 Applications

The proposed work will be beneficial for a variety of areas in NLP, for example Question Answering (QA), Sentiment Detection/Opinion Extraction, Ontology Learning, or Natural Language Generation. In this section I will discuss applications in the first two areas.

6.1 Question Answering

In open-domain QA, the proposed work will be useful for answering two question types. A superlative sentence like [24], found in a corpus, can be used to answer both a factoid question [25] and a definition question [26]:

[24] A: **The Nile** is the ***longest*** <u>river in the world</u>.
[25] Q: **What** is <u>the world's</u> ***longest*** <u>river?</u>
[26] Q: **What** is **the Nile**?

Here I will focus on the latter. The common assumption that superlatives are useful with respect to answering definition questions is based on the observation that superlatives like the one in [24] both place an entity in a generalisation hierarchy, and distinguish it from its contrast set.

To investigate this assumption, I carried out a study involving the TREC QA "other" question nuggets[3], which are snippets of text that contain relevant information for the definition of a specific topic. In a recent study of judgement consistency (Lin and Demner-Fushman, 2006), relevant nuggets were judged as either 'vital' or 'okay' by 10 different judges rather than the single assessor standardly used in TREC. For example, the first three nuggets for the topic "Merck & Co." are:

[27] Qid 75.8: 'other' question for target Merck & Co.

75.8	1	vital	World's ***largest*** drug company.
75.8	2	okay	Spent $1.68 billion on RandD in 1997.
75.8	3	okay	Has experience finding new uses for established drugs.

(taken from TREC 2005; 'vital' and 'okay' reflect the opinion of the TREC evaluator.)

My investigation of the nugget judgements in Lin and Demner-Fushman's study yielded two in-

[3] http://trec.nist.gov/data/qa.html

teresting results: First of all, a relatively high proportion of relevant nuggets contains superlatives: On average, there is one superlative nugget for at least half of the TREC topics. Secondly, of 69 superlative nuggets altogether, 32 (i.e. almost half) are judged "vital" by more than 9 assessors.

Furthermore, I found that the nuggets can be distinguished by how the question target (i.e. the TREC topic, referred to as T1) relates to the superlative target (T2): In the first case, T1 and T2 coincide (referred to as class S1). In the second one, T2 is part of or closely related to T1, or T2 is part of the comparison set (class S2). In the third case, T1 is unrelated or only distantly related to T2 (S3). Table 1 shows examples of each class:

	T1	**nugget** (T2 in bold)
S1	Merck & Co.	World's ***largest*** drug company
S2	Florence Nightingale	**Nightingale Medal** *highest* international nurses award
S3	Kurds	**Irbil** *largest* city controlled by Kurds

Table 1. Examples of superlative nuggets.

Of the 69 nuggets containing superlatives, 46 fall into subclass S1, 15 into subclass S2 and 8 into subclass S3. While I noted earlier that 32/69 (46%) of superlative-containing nuggets were judged vital by more than 9 assessors, these judgements are not equally distributed over the subclasses: Table 2 shows that 87% of S1 judgements are 'vital', while only 38% of S3 judgements are.

	number of instances	% of "vital" judgements	% of "okay" judgements
S1	46	87%	13%
S2	15	59%	40%
S3	8	38%	60%

Table 2. Ratings of the classes S1, S2, and S3.

These results strongly suggest that the presence of superlatives, and in particular S1 membership, is a good indicator of the importance of nuggets, and thus for answering definition questions. Some experiments carried out in the framework of TREC 2006 (Kaisser et al., 2006), however, showed that superlatives alone are not a winning indicator of nugget importance, but S1 membership may be. A similar simple technique was used by Ahn et al. (2005) and by Razmara and Kosseim (2007). All just looked for the presence of a superlative and raised the score without further analysing the type of superlative or its role in the sentence. This calls for a more sophisticated approach, where class S1 superlatives can be distinguished.

6.2 Sentiment Detection/Opinion Extraction

Like adjectives and adverbs, superlatives can be objective or subjective. Compare for example:

[28] **The Black Forest** is the ***largest*** forest in Germany. [objective]
[29] **The Black Forest** is the ***most beautiful*** area in Germany. [subjective]

So far, none of the studies in sentiment detection (e.g. Wilson et al., 2005; Pang et al., 2002) or opinion extraction (e.g. Hu and Liu, 2004; Popescu and Etzioni, 2005) have specifically looked at the role of superlatives in these areas.

Like subjective adjectives, subjective superlatives can either express positive or negative opinions. This polarity depends strongly on the adjective or adverb that the superlative is derived from.[4] As superlatives place the adjective or adverb at the highest or lowest point of the comparison scale (cf. Section 2), the question of interest is how this affects the polarity of the adjective/adverb. If the intensity of the polarity increases in a likewise manner, then subjective superlatives are bound to express the strongest or weakest opinions possible. If this hypothesis holds true, an "extreme opinion" extraction system could be created by combining the proposed superlative extraction system with a subjectivity recognition system that can identify subjective superlatives. This would clearly be of interest to many companies and market researchers.

Initial searches in Hu and Liu's annotated corpus of customer reviews (2004) look promising. Sentences in this corpus are annotated with information about positive and negative opinions, which are located on a six-point scale, where [+/-3] stand for the strongest positive/negative opinions, and [+/-1] stand for the weakest positive/negative opinions. A search for annotated sentences containing superlatives shows that an overwhelming majority are marked with strongest opinion labels.

7 Summary and Future Work

This paper proposed the task of automatically extracting useful information from superlatives oc-

[4] It may, however, also depend on whether the superlative expresses the highest ('most') or the lowest ('least') point in the scale.

curring in free text. It provided an overview of superlative constructions and the main challenges that have to be faced, described previous computational approaches and their limitations, and discussed applications in two areas in NLP: QA and Sentiment Detection/Opinion Extraction.

The proposed task can be seen as consisting of three subtasks:

TASK 1: Decide whether a given sentence contains a superlative form
TASK 2: Given a sentence containing a superlative form, identify what type of superlative it is (initially: IS-A superlative or not?)
TASK 3: For set comparisons, identify the target and the comparison set, as well as the superlative relation

Task 1 can be tackled by a simple approach relying on POS tags (e.g. JJS and RBS in the Penn Treebank tagset). For Task 2, I have carried out a thorough analysis of the different types of superlative forms and postulated a new classification for them. My present efforts are on the creation of a gold standard data set for the extraction task. As superlatives are particularly frequent in encyclopaedic language (cf. Section 3), I am considering using the Wikipedia[5] as a knowledge base. The main challenge is to devise a suitable annotation scheme which can account for all syntactic structures in which IS-A superlatives occur and which incorporates their semantic properties in an adequate way (semantic role labelling). Finally, for Task 3, I plan to use both manually created rules and machine learning techniques.

Acknowledgements

I would like to thank Bonnie Webber and Maria Milosavljevic for their helpful comments and suggestions on this paper. Many thanks also go to Nitin Jindal and Bing Liu, Johan Bos and Malvina Nissim, and Jimmy Lin and Dina Demner-Fushman for making their data available.

References

Kisuh Ahn, Johan Bos, James R. Curran, Dave Kor, Malvina Nissim and Bonnie Webber. 2005. *Question Answering with QED*. In Voorhees and Buckland (eds.): The 14th Text REtrieval Conference, TREC 2005.

Johan Bos and Malvina Nissim. 2006. An Empirical Approach to the Interpretation of Superlatives. In *Proceedings of EMNLP 2006*, pages 9-17, Sydney, Australia.

Norbert Corver and Ora Matushansky. 2006. At our best when at our boldest. Handout. TIN-dag, Feb. 4, 2006.

Irene Heim. 1999. *Notes on superlatives*. Ms., MIT.

Minqing Hu and Bing Liu. 2004. Mining Opinion Features in Customer Reviews. In *Proceedings of AAAI*, pages 755-760, San Jose, California, USA.

Rodney Huddleston and Geoffrey K. Pullum (eds.). 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Michael Kaisser, Silke Scheible and Bonnie Webber. 2006. Experiments at the University of Edinburgh for the TREC 2006 QA track. In *Proceedings of TREC* 2006, Gaithersburg, MD, USA.

Nitin Jindal and Bing Liu. 2006. Mining Comparative Sentences and Relations. In *Proceedings of AAAI*, Boston, MA, USA.

Jimmy Lin and Dina Demner-Fushman. 2006. Will pyramids built of nuggets topple over? In *Proceedings of the HLT/NAACL*, pages 383-390, New York, NY, USA.

Maria Milosavljevic. 1999. *The Automatic Generation of Comparisons in Descriptions of Entities*. PhD Thesis. Microsoft Research Institute, Macquarie University, Sydney, Australia.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of EMNLP*, pages 79-86, Philadelphia, PA, USA.

Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *Proceedings of HLT/EMNLP-2005*, pages 339-346, Vancouver, British Columbia, Canada.

Majid Razmara and Leila Kosseim. 2007. A little known fact is... Answering Other questions using interest-markers. In *Proceedings of CICLing-2007*, Mexico City, Mexico.

Anna Szabolcsi. 1986. Comparative superlatives. In *MIT Working Papers in Linguistics* (8). ed. by Naoki Fukui, Tova R. Rapoport and Elisabeth Sagey. 245-265.

Theresa Wilson, Janyce Wiebe and Paul Hoffmann. 2005. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In *Proceedings of HLT/EMNLP* 2005, pages 347-354, Vancouver, British Columbia, Canada.

[5] www.wikipedia.org