Annotating Superlatives

Silke Scheible

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

Abstract

This paper describes a three-part annotation scheme for superlatives: The first identifies syntactic classes, since superlatives can serve different semantic purposes. The second and third only apply to superlatives that express straight-forward comparisons between targets and their comparison sets. The second form of annotation identifies the spans of each target and comparison set, which is of interest for relation extraction. The third form labels superlatives as facts or opinions, which has not yet been undertaken in the area of sentiment detection. The annotation scheme has been tested and evaluated on 500 tokens of superlatives, the results of which are presented in Section 5. In addition to providing a platform for investigating superlatives on a larger scale, this research also introduces a new text-based Wikipedia corpus which is especially suitable for linguistic research.

1. Introduction

The use of superlative forms like *largest* or *most beautiful* represents an important means of expressing comparison in English. However, so far there has been no large-scale investigation of their syntactic and semantic properties. In theoretical linguistics, most studies of superlatives have focused on particular semantic ambiguities which may only rarely occur in everyday language (Szabolcsi, 1986; Heim, 1999). Two recent studies by Bos and Nissim (2006) and Jindal and Liu (2006) illustrate the usefulness of superlatives for language technology, but neither study exhaustively deals with the different constructions in which superlatives occur (Scheible, 2007). This paper introduces a corpus in which tokens of superlatives have been annotated with syntactic and semantic information. In addition to providing a platform for investigating superlatives on a larger scale, this research also introduces a new text-based Wikipedia corpus which is especially suitable for linguistic research.

2. Motivation

In text books, superlatives are usually introduced alongside comparatives as special forms of adjectives or adverbs which are used to compare two or more things, as for example in:

[1] Bill is **taller** than Sue.   [*comparative*]
[2] Joe is the **tallest** boy at school.   [*superlative*]

Superlative constructions like the one in [2] express a comparison between a target entity (*Joe*) and its comparison set (*the other boys at school*). An initial investigation of superlative forms showed that there are two types of relation that hold between a superlative target and its comparison set (Scheible, 2007):

Relation 1: *Superlative relation*
Relation 2: *IS-A relation (hypernymy)*

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 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 Scheible (2007) discusses their use in applications such as Question Answering (QA) and Sentiment Detection and Opinion Extraction.

In natural language superlatives occur in a variety of different constructions, which differ not only in their syntactic structure, but also in the way in which they express a comparison (Scheible, 2007). The aim of the annotation scheme described in this paper is to provide a platform for a thorough investigation of the different types of superlatives. Three forms of annotation are proposed: The first identifies different syntactic classes of superlatives which serve different semantic purposes. The second and third forms of annotation only apply to superlatives that express straight-forward comparisons between targets and their comparison sets. The second form of annotation identifies the spans of each target and comparison set, while the third one labels superlatives as facts or opinions, an undertaking that has not yet been investigated in the area of sentiment detection (Wilson *et al.*, 2005).

3. Annotation of superlatives

(The complete annotation guidelines are available at http://homepages.inf.ed.ac.uk/s0454417/guidelines.pdf.)

3.1 Classification

Superlatives occur in a variety of syntactic structures which usually represent different types of comparisons. From a computational point of view, it is worth dealing with the different structural classes separately. The proposed classification is therefore mainly based on syntactic features.

Each occurrence of a superlative form in a given corpus is classified as belonging to one of eight predefined classes, following a binary decision tree given to annotators. Each node represents a particular question and each leaf stands for one of the eight classes. (The tree is displayed in the

appendix.) Depending how each successive question is answered, the token is labeled with the leaf that is reached. The eight classes of superlatives are:

Class	Example
[ISA]	**ISA-1:** {The blue whale} is <u>the **_largest_** mammal</u>.
	ISA-2: {The blue whale} is considered <u>the **_largest_** mammal</u>.
[DEF]	The **_largest_** mammal weighs around 200 tons.
[FREE]	The human foot is **_narrowest_** at the heel.
[ADV]	First Class mail usually arrives the **_fastest_**.
[INDEF]	It's a **_most interesting_** book.
[IDIOM]	We stayed at the **_Best_** Western.
[PP]	I will arrive at 8 at the **_earliest_**.
[PROP]	**_Most_** successful bands are from the U.S.

Table 1: Classes of superlatives

Superlatives belonging to the [ISA] class are incorporated in a definite NP and contain a clear-cut comparison between a target item and its comparison set. In the example provided in Table 1, the blue whale is compared to all other mammals with respect to its size. The difference between the ISA-1 and ISA-2 subclasses lies in the way in which the relation between target and comparison set is expressed. In the case of ISA-1 superlatives, the verb "to be" or appositive form is used, while ISA-2 superlatives involve other forms. While superlatives classified as [DEF] are also incorporated in a definite NP, they differ from members of the [ISA] class in that the target of comparison is not explicitly mentioned in the context. The example in Table 1 illustrates that even though we are dealing with exactly the same superlative NP as the one in the [ISA] examples, the comparison remains implicit as there is no explicit mention of the target item.

When superlative forms are incorporated in an *indefinite* NP they are classified as [INDEF]. Members of this class are often used as intensifiers. In the [FREE] class, on the other hand, superlative forms are not incorporated in a noun phrase but occur freely in the sentence. This often makes the comparison less easy to pinpoint: What is being compared in the example above is not the human foot and a set of other entities, but rather different parts of the human foot. Superlatives that are derived from adverbs form their own class, [ADV].

Finally, the [IDIOM], [PP], and [PROP] classes contain superlatives which do not express proper comparisons: [IDIOM] contains superlative that occur as part of an idiom, [PP] contains so-called PP superlative constructions (Corver and Matushansky, 2006), and [PROP] includes uses of *most* as a proportional quantifier (Huddleston and Pullum, 2002).

A pilot annotation of 500 tokens of superlatives with respect to these eight classes is described in Section 5.

3.2 Target and Comparison Set Identification (ISA class)

The second annotation applies only to superlatives which are classified as [ISA] members in the previous step. For each [ISA] instance, the strings representing the {target} and the <u>comparison set</u> of the superlative are marked up (cf. Bos and Nissim, 2006). For example:

[3] **Sentence**: *Philadelphia Zoo is the **oldest** zoo in America.*
  → **Target:** {*Philadelphia Zoo}*
  → **Comparison set:** *the **oldest** <u>zoo in America</u>.*

Of special interest here is that both target and comparison set can have restrictive and non-restrictive modifiers (e.g. postmodifiers like PPs and various kinds of clauses). Compare for example:

[4] {VW} is <u>Europe's **_largest_** car maker with this product range</u>.
　　　　　　　　　　　　　　　　　　　　[*restrictive*]

[5] {VW} is <u>Europe's **_largest_** car maker</u> with an impressive product range.　　　　　[*non-restrictive*]

[6a] {VW} is the largest <u>car maker in China</u>.

[6b] <u>In China</u>, {VW} is the largest <u>car maker</u>.

The "with" PP phrases in [4] and [5] both occur in postmodifying position, but differ in that the one in [4] is involved in the comparison, while the one in [5] is non-restrictive. In addition, restrictive modifiers of the comparison set can also occur elsewhere in the sentence, as shown by the PP phrase "in China" in [6a] and [6b]. It is important that all and only modifiers defining the target or comparison set be identified.

3.3 Subjectivity Labelling (ISA class)

A third form of annotation deals with the fact that like adjectives and adverbs, superlatives can express facts or opinions. Compare for example:

[7] {Trier} is the *oldest* <u>town in Germany.</u>

[8] {Trier} is the *most beautiful* <u>town in Germany</u>.

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

ISA superlatives are labelled as fact if the ranking expressed by the superlative (of the target entity with respect to the other members of the comparison set) is based on an objective comparison that doesn't involve an evaluation. Otherwise, the sentence is subjective and should be labelled as opinion.

In a second step, ISA superlatives labelled as facts have been further annotated according to their reliability on a scale between +2 (reliable) and -2 (unreliable). Reliability in this context refers to how much the annotator would trust the ranking expressed by the superlative comparison (without knowledge of the speaker). ISA superlatives labelled as opinions have been further annotated according to their polarity. Polarity here refers to whether a positive or a negative opinion is expressed, and is rated on a scale between +2 (positive) and -2 (negative). This annotation is also described in Section 5.

4. Corpus and Annotation Procedure

4.1 *TextWiki* Corpus

Previous experiments have shown that superlatives are particularly frequent in encyclopedia text (Scheible, 2007). For this reason I decided to use the Wikipedia as a knowledge base. As existing Wikipedia corpora such as the Wikipedia XML corpus by Denoyer et al. (2006) are primarily aimed at Information Retrieval tasks such as INEX, they have several shortcomings with respect to studying linguistic phenomena like superlatives. Firstly, the XML conversion in Denoyer's corpus retains most of the original wiki-markup, and thus includes information that is redundant for linguistic investigations (e.g. formatting). Furthermore, database structures such as tables, lists, figures, galleries, and templates are included, which usually do not contain full sentences and may therefore skew experimental results. Finally, the corpus includes empty or incomplete articles ('stubs') which are of little interest from a linguistic point of view.

I therefore decided to create a new Wikipedia corpus which is especially suitable for linguistic research (referred to as *TextWiki*). Although marked up in XML, it is primarily text excluding information irrelevant for linguistic investigations:

- All tables, lists, figures, galleries, and templates have been deleted.
- Document structure markup is reduced to title, body, paragraph and sentence tags.
- Formatting markup is reduced to a small set of tags (e.g. <b> for bold text, <i> for italics)
- Only hyperlinks to other Wikipedia articles are retained.

When completed, the *TextWiki* corpus will yield one million words, and all superlative forms in the corpus will be marked up and annotated according to the annotation scheme introduced in Section 3. To date, superlatives associated with the words *most* and *least* have been identified, as have adjectives and adverbs graded with the suffix *-est* (including hyphenated superlatives like *kindest-hearted*) and irregular superlative forms (*best*, *worst*, *furthest* and *farthest*).

TextWiki aims to be a balanced corpus and draws equally from all main Wikipedia categories. Due to Wikipedia's complex categorisation structure and its open-endedness, an exhaustive and even coverage of the whole encyclopedia is problematic. However, the same random sampling technique of articles was applied to each of the top level categories, resulting in a corpus containing articles from a broad number of areas.

The corpus is approximately half completed (#W, #A, and #S are the number of words, articles and superlative forms in each of the categories):

Top-level Category	# W	# A	# S
Culture and the arts	90025	100	292
Geography and places	90406	107	361
Health and fitness	90174	91	233
History and events	90691	98	234
Natural and physical sciences	90369	96	265
TOTAL:	451,665	492	1385

The table shows around two to three superlatives on average per Wikipedia article.

4.2 Pilot annotation study

The following sections describe the results of a pilot annotation study that was carried out prior to the *TextWiki* compilation on a sample drawn from the Wikipedia XML corpus by Denoyer et al. (2006), with sentence mark-up added by Jijkoun et al. (2006). This pilot corpus consists of 142 articles randomly selected from Part-0 of the Wikipedia XML corpus (excluding articles with less than 50 words and all database structures). The corpus contains 500 tokens of superlatives, with (on average) one superlative per 14 sentences.

Superlative instances were identified and annotated as described in Section 3. In addition to myself, a second annotator was recruited and trained to test the validity of the proposed annotation scheme. Errors were discussed and resolved after each set of 100 superlative-containing sentences. All annotations were carried out with a tool specifically designed for the task. The following section summarises the results of this pilot study.

5. Results and discussion

5.1 Classification task

The classification task (Section 3.1) achieved an overall inter-annotator agreement of 89% (444/500 instances). Disagreements were discussed after each set of 100 instances, and were, with the exception of three cases, resolved in favour of Annotator 1 (the author). The feedback sessions caused the performance curve to rise steadily, with 76% agreement after the first 100 cases and 95% agreement after the last set:
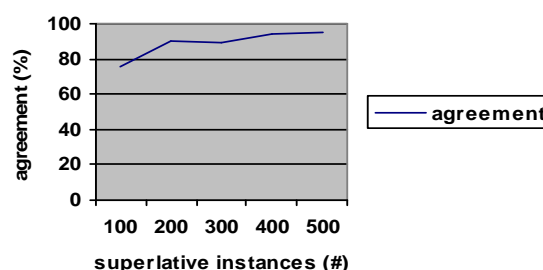

Figure 1: Inter-annotator agreement of superlative classification

Surprisingly, the adverbial class [ADV] was the source of most disagreement, in particular with adverbial superlatives in ISA constructions, as for example in:

[9] There is not even complete consistency to be found between *The Lord of the Rings* and *The Hobbit*, the two **most** closely related works, because Tolkien was never able to fully integrate all their traditions into each other.

[Article: *J. R. R. Tolkien*]

925

Despite this structural resemblance to ISA comparisons, these adverbial superlatives do not modify the comparison set head, but rather (one of) its modifiers. In [9], *most closely* modifies the deverbal adjective *related* and not the NP head *works*. Another problem was in cases of the form '*superlative + deverbal adjective*', as for example:

[10] An Egyptian scribe named Ahmes wrote the **oldest** known text to give an approximate value for π.　　　　[*Pi*]

Although similar to *longest-known*, which should be classified as [ADV], the superlative *oldest* in this context modifies *text* rather than *known*. [10] should therefore be labeled as [DEF].

The following table shows the (agreed) distribution of superlative classes over the 500 superlative sentences:

	TOTAL	%
PROP	124	24.8
PP	31	6.2
IDIOM	28	5.6
ADV	52	10.4
FREE	8	1.6
INDEF	14	2.8
DEF	93	18.6
ISA	150	30
of which ISA-1	*108*	*21.6*
and ISA-2	*42*	*8.4*

The ISA class is clearly the most populous, containing 30% of all superlative instances (150/500), which justifies the focus on this class in the other two annotation tasks. The high number of proportional quantifiers (24.8%) can be explained by the fact that encyclopedia entries usually define classes, and proportional quantifiers are a good way of describing properties that do not apply to all members of a class (but to most of them).

With 99.4% agreed accuracy, Annotator 1's labeling can be reliably used as gold standard for the classification of superlatives in the *TextWiki* corpus.

5.2 Target and Comparison Set Identification

The results of the Target and Comparison Set Identification for superlatives classified as ISA also look very promising (Section 3.2). Of 116 superlatives classified as ISA-1 (89 cases) or ISA-2 (27 cases) by both annotators, there was full agreement for target and comparison set spans in 108 cases (93%). Of the eight disagreements between the two annotators, five concerned the span of ISA-1 targets.

Considering the fact that of 89 ISA-1 superlatives over two thirds (60) have a comparison set with at least one postmodifier, the results look very impressive. However, a closer study reveals that only four of these (4/60, around 7%) have a postmodifier that is marked non-restrictive by the annotators. This implies that given a postmodified comparison set, there is a chance of approximately 93% that the postmodifier is restrictive. (The probability may actually be even higher since some comparison sets have more than one postmodifier.)

Compared to ISA-1 comparison sets, ISA-1 targets are less likely to be postmodified: Only 32 out of 89 ISA-1 targets have at least one postmodifier (around 36%). However, the proportion of non-restrictive postmodifiers among them is much higher, with 16 out of 32 ISA-1 targets having a non-restrictive postmodifier (50%). In most cases these are postmodifying clauses (such as relative and non-finite clauses).

According to English comma rules, it should be possible to distinguish between non-restrictive (usually referred to as "non-defining") and restrictive ("defining") relative clauses by the presence or absence of a comma. In the following example, the comma after *Ceres* indicates that the following relative clause is to be considered non-defining:

[11] The **biggest** asteroid belt member is {Ceres}, which is about 1000 km across.　　　　[*Asteroid Belt*]

However, commas are not used reliably, as the following sentence illustrates:

[12] The **most famous** diesel-hydraulic locomotive is {the german V200} which were built from 1953 in a total number of 136.　　　　[*Locomotive*]

If the relative *which*-clause was a defining one, this would imply that there are at least two separate versions of *the german V200*, which is very unlikely.

Interestingly, it seems that distinguishing between defining and non-defining relative clauses also causes problems for the annotators, even when correct comma rules have been applied:

[13] {The temporary exhibition Treasures of Tutankhamun}, held by the British Museum in 1972, was the ***most successful*** in British history, attracting 1,694,117 visitors.　　　　[*British Museum*]

Annotator 2's target string includes the relative clause "held by...", which should be considered non-restrictive in this context.

The unreliability of commas in practice means that annotators must rely on their world knowledge in order to identify target and comparison set spans correctly.

With an inter-annotator agreement of 93%, we can conclude that the identification of target and comparison set spans is a fairly straightforward task. However, for interpreting an ISA comparison identifying spans is not enough. The following issues will also have to be taken into account:

1.) Around 20% (18 out of 89) of ISA-1 targets or comparison sets contain a pronoun and require anaphora resolution. For example:

[14] Its **most populous** city is {Vancouver}, which is in the southwest corner of the mainland of the Province of BC.　　　　[*British Columbia*]

2.) Around 19% (17 out of 89) of ISA-1 comparison sets contain a "fused head", where the superlative and the NP head "merge" into one unit (Huddleston and Pullum, 2002). The NP head is implied in the context (usually, but not necessarily in the same sentence) but has not yet been considered in the annotation, as for example *works* in the following sentence:

[15] He was the author of several <u>works</u>, the **_most important_** being {*The Principles and Practice of Surgery*} (1878-1883). [*David Hayes Agnew*]

3.) Postmodifiers that have been identified as restrictive must be further analysed according to their semantic roles. In particular, one needs to distinguish between *NP-head complements* and *NP-head modifiers* (Huddleston and Pullum, 2002). In [16], the "*in*" PP-phrase is a complement and an obligatory part of the NP. In [17], on the other hand, the "*in*" PP-phrase is a modifier of the NP head. It has the role of restricting the set in location, but is not as such obligatory:

[16] The **newest** <u>technology in trains</u> is {magnetic levitation (maglev)}. [*Locomotive*]
[17] The **most popular** <u>religion in Switzerland</u> is {Roman Catholicism} (43% of the population). [*Switzerland*]

5.3 Subjectivity Labelling

The Subjectivity Labelling of the 116 agreed ISA cases provides further interesting results. While the annotators agree on the subjectivity label (fact vs. opinion) for 85% (76/89) of ISA-1 cases, the number is much lower for ISA-2 superlatives with only 63% (17/27) agreement. The reason for this may lie in the fact that, for ISA-2 superlatives, the ISA relation is often expressed by terms like "claim/believe/consider to be", which introduce a notion of doubt to otherwise factual information. That ISA-2 superlatives are more likely to be classified as opinions is also illustrated by this table, which shows the distribution of facts and opinions for the agreed cases:

	ISA-1 (*76 instances*)	**ISA-2** (*17 instances*)
fact	64 (*84%*)	9 (*53%*)
opinion	12 (*16%*)	8 (*47%*)

The high percentage of facts for ISA-1 cases reflects the fact that encyclopedia articles should be objective rather than subjective.

Interestingly, in cases where the annotators agree on the subjectivity label, they also largely agree on the orientation of the reliability and polarity values: 90% (66/73) agreement for reliability orientation of facts (mostly positive) and 95% (19/20) agreement for polarity orientation of opinions (also mostly positive).

6. Future work

Once the *TextWiki* corpus is fully compiled and annotated for superlatives, I plan to implement a "superlative relation extractor" which automatically identifies ISA superlatives

and can extract their targets, comparison sets, and the relations that hold between them. 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: ISA or not?)
> **TASK 3:** For ISA superlatives, 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), although the reliability of POS taggers on recognising superlative constructions needs to be assessed. For Tasks 2 and 3 I am going to use machine learning techniques based on the gold standard annotations of superlatives in the *TextWiki* corpus. I plan to consider features relying on the output of the MiniPar parser (Lin, 1998). Finally, I will also investigate the ways in which these recognisers can be used to aid Question Answering and Sentiment Detection/Opinion Extraction.

7. Acknowledgements

8. References

Bos, J., Nissim, M. (2006). An Empirical Approach to the Interpretation of Superlatives. In *Proceedings of EMNLP 2006*. Sydney, Australia: pp. 9--17.

Corver, N., Matushansky, O. (2006). At our best when at our boldest. Handout. TIN-dag, Feb. 4, 2006.

Denoyer, L., Gallinari, P. (2006). The Wikipedia XML Corpus. *SIGIR Forum* 40(1): pp. 64--69.

Heim, I. (1999). *Notes on superlatives*. Ms., MIT.

Hu, M., Liu, B. (2004). Mining Opinion Features in Customer Reviews. In *Proceedings of AAAI*. San Jose, California, USA: pp. 755--760.

Huddleston, R., Pullum G. K. (Eds.). (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Jijkoun, V., de Rijke, M. (2006). Overview of WiQA 2006. In A. Nardi, C. Peters, and J.L. Vicedo (Eds.), *Working Notes CLEF 2006*.

Jindal, N., Liu, B. (2006). Mining Comparative Sentences and Relations. In *Proceedings of AAAI*. Boston, USA.

Lin, D. (1998). Dependency-based Evaluation of MINIPAR. In *Proceedings of the Workshop on the Evaluation of Parsing Systems*. Granada, Spain.

Scheible, S. (2007). Towards a Computational Treatment of Superlatives. In *Proceedings of the ACL 2007 Student Research Workshop*. Prague, Czech Republic.

Szabolcsi, A. (1986). Comparative superlatives. In N. Fukui, T. R. Rapoport and E. Sagey (Eds.), *MIT Working Papers in Linguistics* (8).

Wilson, T., Wiebe, J., Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In *Proceedings of HLT/EMNLP* 2005. Vancouver, British Columbia, Canada: pp. 347--354.

Appendix : Decision tree for superlative classification