

Pre-modifying -ing participles in the parsed BNC

In our presentation we report on an ongoing study on premodifying -ing participles, their distribution across different genres and their syntactic complexity. The data for the study comes from an exciting new resource, a parsed version of the British National Corpus (BNC), which is currently being developed at the Computer Laboratory of the University of Cambridge (Andersen et al. 2008).

Premodifying -ing participle constructions (e.g. “*the running boys*”, “*the following examples*”) have thus far been very hard to study using corpus methodologies. The constructions are not very frequent to start with, but the most significant challenge is that without a substantially large parsed corpus, it is virtually impossible to find the relevant constructions amid all other -ing forms, such as progressives and gerunds; forms that are much more frequent than the premodifying -ing participles.

Moreover, there are problems with categorization. The premodifying -ing participle is a theoretically debated class: some linguists consider these forms to be adjectives (e.g. Borer 1990, Bresnan 1996), while others (e.g. Quirk et al. 1985, Laczkó 2001) maintain that at least some -ing participles should be regarded as verbs. For a corpus linguist, this means that -ing participles may have been tagged either as inflected verbs or adjectives (or both) in different corpora.

This problem is overcome in the parsed BNC. By using both the POS tags and the dependency information in the parsed corpus, we can find premodifying -ing participles quite accurately among the set of all -ing forms. Moreover, all words in the parsed BNC have two POS tags: the original tags have been retained and the new POS tag, associated with the parse tree, is added. By looking at the differences between the two POS tags we can increase the selection accuracy even further.

Extending this result, we show that both tag sets still have their flaws and cannot provide us with all the information we need. As automated tagging systems always produce probabilities for each possible tag, we investigate how using the probabilities associated with the tags can help us find exactly those -ing forms we are looking for and compare our findings to the results obtained from the parsed BNC with double tagging. We believe that this method may be useful for other research questions as well.

References

Andersen, Øistein, Julien Nioche, Ted Briscoe & John A. Carroll 2008. *The BNC Parsed with RASP4UIMA*. <http://www.lrec-conf.org/proceedings/lrec2008/pdf/218_paper.pdf>

Borer, Hagit 1990. "V + ing: It Walks like an Adjective, It Talks like an Adjective". *Linguistic Inquiry* 21: 95-103.

Bresnan, Joan 1996. *Lexicality and Argument Structure*. Invited paper given at the Paris Syntax and Semantics Conference. <<http://www.stanford.edu/~bresnan/paris.pdf>> Accessed November 30, 2009.

Laczkó, Tibor 2001. Another Look at Participles and Adjectives in the English DP. Miriam Butt and Tracy Holloway King (eds.). *Proceedings of the LFG01 Conference*. CSLI Publications.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.