An Information-Based Treatment of Punctuation

Bilge Say and Varol Akman
Department of Computer Engineering and Information Science,
Bilkent University,
Bilkent, 06533 Ankara, Turkey
{say,akman}@bilkent.edu.tr
Fax: +90 312 2664126

Abstract

Punctuation marks have recently attracted attention within the linguistics community mostly from a syntactic perspective. In this paper, we aim to give a preliminary account of information-based aspects of punctuation marks, drawing our points from examples and links with related phenomena such as intonation. We give our initial treatment within the Discourse Representation Theory.

Conference Area: Varia (Punctuation)

1 Introduction

Punctuation marks have not been studied much by linguists apart from a prescriptive aspect, until recently. Similarly, most natural language processing systems do not take punctuation marks into account except for the period and the spacing. Recent linguistic works have attempted to produce systematic characterizations of punctuation marks descriptively. Levinson [7] brings out the distinction between the orthographic sentence and the grammatical one as well as explaining how punctuation marks join entities according to their informational links. Meyer [9] gives a classification of American punctuation marks according to their functions and the realization of those functions. Nunberg [10] shows how punctuation is a linguistic system on its own and devises a "text-grammar" for this purpose using mechanisms of conventional, or "lexical" grammars. Based on his seminal work, several researchers integrated punctuation marks into the NLP systems especially from a syntactic point of view [3, 5, 11]. We want to add on top of the previous work a formal characterization of the information that punctuation marks bring to the discourse, semantically and pragmatically, within or above (grammatical) sentence level.

2 Punctuation in Discourse

Punctuation marks play various roles in natural language discourse. They can have a morphological role such as in *anti-feminist*, a delimiting role such as in *Jones, my brother, came yesterday*, or a separating role such as in *two bottles of wine, three cans of beer*. They can also have distinguishing roles such as usage of capital letters for proper names. These roles sometimes serve to resolve ambiguities, e.g., *new, regular time for Tai-Chi classes* as compared to *new regular time for Tai-Chi classes*. If our intended meaning is to announce classes with a fixed schedule, the second construct would be ambiguous. As in this example, some of these roles of punctuation may have semantic functions. Our claim is that they can even change the analysis of discourse. In fact, various punctuation marks operate above sentence level connecting independent clauses that can function as stand-alone sentences. In addition, these connections result in special effects such as elaboration.

According to Lyons [7, p. 65], punctuation of an utterance can be seen as:

> The marking of boundaries at the beginning and the end of utterance and various points within the utterance to emphasize particular expressions, to segment the utterance into manageable information units, to solicit the listener's permission for the utterance to be continued.

The above definition includes cue phrases, graphical devices in written text, intonation in speech but punctuation marks can also be seen within this enlarged view.

Our aim is to be able to capture the effects of punctuation within a suitable framework. A suitable choice looks like the Discourse Representation Theory (DRT) by Kamp and Reyle [6] which integrates current approaches in a semantic theory. Asher [1] has developed a theory within DRT taking discourse relations into account which will be relevant for some usages. As an example to discourse effects of punctuation, consider the following pair of discourse fragments:

> Jane, and Joe and Sue write books on England. If her books are best-sellers then they are jealous.
>
> Jane and Joe, and Sue write books on England. If her books are best-sellers then they are jealous.

The exact position of the comma in the first sentence of the pair changes the resolution of pronominal anaphora in the second sentence, which is identical in both pieces of discourse. Suitable triggering configurations will lead to different structures within DRT: the first pair of sentences will have *her* attached to Jane and *they* to Joe and Sue, whereas the second pair will have *her* attached to Sue and *they* to Jane and Joe.

Consider now the example below taken from Bolinger [1, p. 183]:

> She looks right, he looks left; she smiles, he frowns; she clasps her hand around her knee, he clasps his around his head.

Here, there is not only a temporal sequence but also a relation of causality between the subsentences separated by commas. Semicolons, on the other hand, bring parallelism to the structure. Such sentences require treatment within Asher's framework.

3 Conclusion

Information cues provided by punctuation marks can be valuable to NLP systems if captured adequately. We aim to give the preliminaries of such a formal and integrated approach in the full version of this paper.

References

[1] Asher, N. *Reference to Abstract Objects in Discourse.* Kluwer Academic Publishers, Dordrecht, Netherlands, 1993.

[2] Bolinger, D. *Intonation and Its Uses: Melody in Grammar and Discourse.* Stanford University Press, Stanford, California, 1989.

[3] Briscoe, T. *Parsing (with) Punctuation*, Technical Report, Rank Xerox Research Centre, France, 1994.

[4] Dale, R. "Exploring the Role of Punctuation in the Signalling of Discourse Structure" in *Proceedings of a Workshop on Text Representation and Domain Modelling: Ideas from Linguistics and AI.* pp. 110–120, Technical University of Berlin, 1991.

[5] Jones, B. "Exploring the Role of Punctuation in Parsing Natural Language" in *Proceedings of COLING '94.* pp. 421–425, Kyoto, Japan, 1994.

[6] Kamp, H. and U. Reyle. *From Discourse to Logic*, Parts 1 and 2, Kluwer Academic Publishers, Dordrecht, Netherlands, 1993.

[7] Levinson, J. P. *Punctuation and the Orthographic Sentence: A Linguistic Analysis*, PhD Thesis, City University of New York, 1985.

[8] Lyons, J. *Semantics*, Volume 1. Cambridge University Press, Cambridge, UK, 1977.

[9] Meyer, C. F. *A Linguistic Study of American Punctuation*, Ph.D. Thesis, University of Wisconsin-Milwaukee, 1983.

[10] Nunberg, G. *The Linguistics of Punctuation*, CSLI Lecture Notes, No. 18, Stanford University Press, Stanford, California, 1990.

[11] White, M. "Presenting Punctuation" in *Proceedings of the Fifth European Workshop on Natural Language Generation*, pp. 107–125, Leiden, Netherlands, 1995.