

A GRAMMAR OF PUNCTUATION IN POETRY

Modern Computational Linguistic Perspectives

DOYLE CALHOUN

THE FIELD OF LINGUISTICS FOCUSES ON THE SPOKEN INCARNATION OF LANGUAGE AND CONSIDERS WRITTEN PHENOMENA AS AN IMPERFECT APPROXIMATION OF THE FORMER. THE AUTHOR ARGUES THAT PRESENT UNDERSTANDING OF PUNCTUATION AS A SUBSYSTEM OF HUMAN LANGUAGE IS INADEQUATE, AND THAT EVEN WHEN STUDIES EXAMINE PUNCTUATION FROM A LINGUISTIC STANDPOINT, THEY EMPHASIZE MODERN WRITTEN LANGUAGE AND AMERICAN ENGLISH IN PARTICULAR. IT IS, HOWEVER, QUITE POSSIBLE TO DESCRIBE THE BEHAVIOR OF PUNCTUATIONS WITH THE SAME METHODOLOGY THAT WAS UNTIL RECENTLY RESERVED FOR SPOKEN LANGUAGE. JUST AS LANGUAGE DEVELOPS OVER TIME AND CHANGES FROM PLACE TO PLACE, SO TOO DOES PUNCTUATION. THIS PAPER ARGUES FOR THE COMPATIBILITY BETWEEN RECENT DEVELOPMENTS IN COMPUTATIONAL LINGUISTICS' APPROACHES TO PUNCTUATION AND TEXTUAL CRITICISM THAT ATTENDS TO PUNCTUATION IN POETRY. A FULL UNDERSTANDING OF ITS CONVENTIONS IS CRITICAL FOR A PROPER INTERPRETATION OF MEANING. FURTHERMORE, THE AUTHOR SHOWS THAT THE RELEVANCE OF THESE STUDIES TO LITERARY ANALYSIS REMAINS UNEXPLORED. NEVERTHELESS, WITH THE DEVELOPING TECHNIQUES MADE POSSIBLE THROUGH COMPUTATIONAL LINGUISTICS, THE POTENTIAL FOR IN-DEPTH STUDY IS EXPANDING.

While the punctuation of a poem may not always figure into its critical analysis, literary critics and textual scholars seem to agree that punctuation constitutes at least one of the non-verbal tools of poetic expression available to the poet. As such, it is one of the features to which the critic could attend. It would be unusual, for instance, for an analysis of E. E. Cummings's poem "l(a)" to fail to mention the twisting parentheses, or for a critical edition of Dickinson's verse to not make some ruling on typesetting the length of her dashes. The extent to which scholars perceive punctuation as relevant to their text-interpretive or text-critical endeavor varies. Nonetheless, punctuation has its discernible place in literary studies: John Lennard gives a book-long treatment to the use of parentheses in English verse (1991); Christopher Ricks attributes an almost metaphysical aspect to punctuation in *The Force of Poetry*—it is "at once a uniting and a separating" (1984); Robert L. Kellogg and Oliver L. Steele dedicate an entire article to alternative punctuations of the last two lines of *The Faerie Queene* (1963).

Punctuation has not benefitted from the same sustained attention in linguistics. Only in the past thirty years has punctuation become an object of empirical linguistic study, apart from prescriptive accounts of its usage. This is probably because formal linguistics, unlike literary studies, tends to view written language as a byproduct of spoken language or its imperfect image, a stance that reduces punctuation to a deficient and inadequate mode of representing or registering prosodic features of speech stream, like intonation. Or, perhaps, the reluctance of modern generative linguistics to see punctuation as a worthwhile focus for linguistic inquiry stems from the reality that it is an extra-verbal phenomenon. Punctuation is not a necessary (or very common) feature of human language: not all speech communities have written language, and not all written languages have punctuation. Typological studies on punctuation systems vary in different linguistic subfields: psycholinguistic work on how punctuation is realized in the lexicon; large-scale empirical investigation into punctuation at various synchronic stages of a single language; context-specific analyses of punctuation (particularities of its use in literary or legal language, as an example). All these areas remain under-examined.

Recent efforts in computational linguistics, however, make use of computerized corpora to bring attention to the fact that punctuation is a coherent, rule-governed linguistic subsystem in its own right. B. Say and V. Akman survey

extant research on punctuation marks from the point of view of natural language processing (NLP), and propose an information-based framework for describing and understanding punctuation (1997). According to their model, punctuation can be seen as contributing to the information conveyed by a sentence or intrasentential clauses.¹ Taking *information* to mean propositional content which constitutes a contribution of knowledge to [sic] reader's knowledge store, they demonstrate how to use discourse representation theory (DRT) and segmented discourse representation theory (SDRT) to deal with punctuated written texts.² DRT and SDRT use representational boxes with a set of referents (entities in the discourse) and conditions (properties that relate the referents), which are built incrementally over the course of the discourse and which can be resolved into a construction algorithm. The major contribution of this highly technical work is to demonstrate that punctuation can be described systematically, and as they suggest, (eventually) accounted for in a unified theory that computational linguistics can apply to both analyzing and generating written language.

The points raised by Say and Akman and their exercise of formalizing punctuation in terms of DRT and SDRT are relevant to literary studies and textual criticism. The absence of any unified linguistic theory of punctuation and the dearth of research programs aimed at formulating one disadvantage the linguist and literary scholar alike. The result of this gap in language science and literary studies is that the modern understanding of punctuation as a feature of human language is unclear, its use at any given synchronic stage in any given language even less clear, and its marked use in literary contexts totally unexplored.

To complicate the matter, studying punctuation, especially punctuation in poetry, gives rise to a problem of definition: namely, what counts as punctuation? In his Brown Corpus study, Meyer distinguishes between structural punctuation symbols (those that act on units not larger than the orthographic sentence and not smaller than the word, i.e. not hyphens or apostrophes) and other punctuation (text-graphical features like paragraph breaks, font changes, lists) (1986). The distinction is a useful one in linguistics, but obfuscates features of punctuation that might be relevant to literary criticism or poetics: one can imagine a poem in which intra-morphemic punctuation or suprasegmentals, such as diacritics, function in a non-trivial way.

In addition, a purely formal description of punctuation elides certain features of punctuation that are particular to

literature, and poetry in particular. Such particularities of literary punctuation include its: rhetorical force, aesthetic or visual dimension, meta-physical qualities, allusive potential, and effect on meter. The ability of a hyphen to suggest a unique kind of bond that, in Ricks' terms, holds apart at the same time it holds together, is a dimension that is relevant and meaningful to the literary critic but one which the computational linguistic is likely uninterested in capturing.

All this is not to diminish the potential of computational work on punctuation to contribute to a more comprehensive understanding of punctuation in literary studies. Rather, it is to point out that for such research to be truly useful to the literary critic, it must be specific enough to accurately describe punctuation as a subsystem of human language and general enough to handle additional factors based on stylistic register or context.

The work of Nunberg (1990) gets closest to that balance, at least according to Say and Akman (1997).. As they point out, Nunberg's (1990) treatise retains its influential status as the first comprehensive attempt to approach punctuation from within the generative, or Chomskian, paradigm: to decipher it in terms of a *text-grammar* and to conceive of it as a generative system describable by rule-governed operations. Working within this framework, Nunberg builds a collection of rules he uses to explain the distribution of text-categories — the inter-punctuational units (text-adjuncts, text-clauses, text-phrases) dealt with by the lexical grammar.³ An example of such a rule treats the interaction of two English punctuation marks: "The point absorption rule dictates that a period will absorb a comma when they are immediately adjacent."⁴

Nunberg's account of punctuation recognizes that punctuation marks are not distributed randomly or arbitrarily, but like natural language reflect a series of hierarchically ordered abstract rules. What Say and Akman find lacking in Nunberg (1990)— and what they identify as a desideratum in the field of computational linguistics— is a formal

theoretical apparatus that can characterize punctuation not just in terms of its relationship to syntax, but also account for its semantic and discourse-related effects.⁵

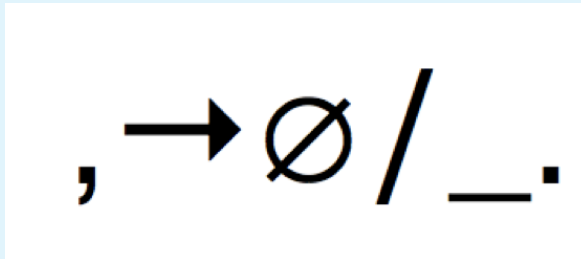
Work in algebraic linguistics underscores the importance of attending to these effects. At least in terms of logical value, punctuation is as much an observable feature for the critic to attend to in a poet's work as it is in his or her own writing on a poet's work. In his article "Punctuation and human freedom", linguist Geoffrey K. Pullum provides the following example of how a simple transposition of punctuation marks alters the truth-conditions of a statement:

- (i) Shakespeare's *King Richard III* contains the line "Now is the winter of our discontent."
- (ii) Shakespeare's *King Richard III* contains the line "Now is the winter of our discontent".⁶

The claim in (i), he argues, is false. For Shakespeare's play does not contain the line "Now is the winter of our discontent." Rather, it contains the line "Now is the winter of our discontent", which is actually the first of two lines that form the sentence: "Now is the winter of our discontent made glorious summer by this sun of York." In the play-text, as Pullum points out, "Now is the winter of our discontent" is not even a syntactic constituent; although, in (i), the convention of placing terminal punctuation before quotes makes it look as if it were. Pullum's point here seems a slight one—or, perhaps, an overly-fastidious one—but it is important to realize that he is not leveling a prescriptive stylistic complaint; he is pointing out the linguistic reality that punctuation marks interact with, and have the ability to alter, the truth-conditions of a claim. The question is one of logical truth, not of style.

An alternative approach to Nunberg's generative-style grammar of punctuation might be a purely distributional one. In his corpus studies, Karlsson (1994) demonstrates the possibility of describing punctuation in terms of statis-

"Recent efforts in computational linguistics, however, make use of computerized corpora to bring attention to the fact that punctuation is a coherent, rule-governed linguistic subsystem in its own right."



THE FORMULATION OF A LINGUISTIC RULE IN STANDARD NOTATION (FONT LICENCED THROUGH ADOBE INDESIGN)

tical frequency, rather than binary tag-features like [±stoppedness], which Nunberg (1990) makes use of. As Karlsson's findings suggest, it is possible to characterize punctuation marks based on their immediate context. For instance, he reports: "A punctuation mark to the left of a finite verb dramatically decreases the probability that the preceding word is a grammatical subject (of all the finite verbs preceded by a punctuation mark, less than 5% had the preceding word as subject)."⁷

Karlsson's work is conducted within and for the milieu of computational linguistics; his concerns are necessarily different from those of the literary critic. But might literary researchers employ a similar distributional approach to more accurately determine statistical patterns of punctuation use? The presence of digitized corpora makes such a task possible and, most importantly, allows researchers to carry out their analyses without reference or subservience to other linguistic features. With the recent understanding of punctuation as a rule-governed system in its own right, rather than as an artifact treated indirectly or filtered through other levels of linguistic inquiry like syntax, comes the ability to claim punctuation as a singular object of study in research. This might, but need not, extend beyond the workings of the suite of punctuation marks.

Also of particular interest is the potential for linguistic research on punctuation to provide insight into cruxes of textual scholarship based on historiographical concerns. Inadequate understanding of punctuation at a particular synchronic stage of the language (herein, English) is not a new problem in literary analysis. But the reality remains that most modern readers— and, to a lesser extent (one hopes), textual scholars— are limited by their own linguistics in that they inevitably read in reference to modern punctuation. Kellogg and Steele (1963) raise this crucial historiographical point when they caution against analyses

built on hasty— or, as they write, "sanguine"— generalizations like 'punctuation marks syntax' or 'punctuation marks intonation':

The critic cannot afford to work on the assumption that Spenser's syntax can be explained by analogy with similar but not identical structures in present-day English. Two considerations must temper our hasty analysis of Elizabethan poetry: first, we do not understand Elizabethan syntax thoroughly; and, second, we do not know what intonation was used to signal syntactical relationships.⁸

One might add to their point a further concession: even if we do achieve an accurate reconstruction of Elizabethan syntax and phonemics, we cannot assume the poet operated within those conventions or norms. Moreover, Kellogg and Steele's stance presupposes that punctuation in a poem is limited to demarcating syntactic constituents and transcribing the prosodic features of inner speech. A critic cannot afford to work on that assumption either.

A more accurate approach would examine punctuation, syntax, and phonemics independently and not as mixed-levels— but, rather, as phenomena that are interrelated but not dependent on or subservient to one another. It would be informed by a reconstruction of Elizabethan syntax and phonemics as well as a reconstruction of a grammar of Elizabethan punctuation. What might such an approach look like? Placing a consideration of Elizabethan syntax and phonemics aside, a series of corpus-based studies on punctuation conducted in the following contexts would yield multiple and increasingly marked perspectives on punctuation as it functioned in the target time-period:

- Elizabethan texts (including newspapers, legal documents, and other non-literary publications);
- Contemporary Elizabethan poetry;
- As well as a more concentrated analysis of punctuation in the poet's own writings (which would consider works in prose and works in poetry separately).

Computational linguistic techniques allow such corpus searches to be exacting. In the style of Karlsson (1994), it is possible to determine the statistical frequency of, say, an ellipsis separating a verb and its subject, or the probability that the word preceding an em dash is in the nominal case when the word that follows is a finite verb. Admittedly, exhaustive corpora for this type of research will not be avail-

able in all cases. Already, though, digitized corpora for various genres of literary texts or author-specific databases exist. If and when corpora become available, though, applying computational linguistic techniques of the kind detailed by Say and Akman (1967) seems like a useful but unexplored resource for literary studies.

Accommodating a linguistic theory of punctuation—formal, as in Nunberg’s case, or distributional, as in Karlsson’s— to literary analysis might provide the literary critic with surer footing while navigating historically remote punctuation systems. A body of research that examines punctuation in this manner offers to the critic and textual scholar a hereto-unavailable interpretive tool. It would allow for cross-comparison between the poet’s deployment of punctuation and statistical patterns or norms of punctuation in use in the particular socio-cultural milieu in which he or she was writing.

In this way, the critic or scholar could more easily, and more systematically, decenter a reading of a text from his or her own linguistics of punctuation. Historiographer of the language sciences Konrad Koerner (1997) points to the value of a similar process of re-orientation or recalibration when approaching non-contemporary (linguistic) texts: “[M]odern scholars have been misled in their assessment... probably because they could not or did not divest themselves of their own twentieth-century structuralist background.”⁹

A similar concern holds true for an analysis of punctuation in poetry. A critically responsible treatment of a poem’s punctuation recognizes that reading only with reference to contemporary punctuation results in a limited interpretive stance, and also constitutes a critical move based on presupposition. The critic might work around this logical fallacy by drawing on the technical apparatus of computational linguistics to equip him- or herself with descriptions of: contemporary punctuation, punctuation in the target time-period, and punctuation within the body of work being examined. Only by refracting a historically displaced reading of the poem through these multiple lenses can we hope to achieve a more historiographically sound and interpretively accurate assessment of the working of the poem’s punctuation.

This paper does not propose a unified theory of punctuation, or even attempt to define in detail how such a theory might be applied to literary criticism or textual

scholarship; the current lack of research makes this unlikely. Rather, it aims to draw attention to the general paucity of research directed at constructing an empirical theory of punctuation for use in both linguistics and literary studies. And, furthermore, to suggest the utility of such a theory and its compatibility between recent efforts in computational linguistics and the recognition in literary studies, especially among scholars of poetry, that punctuation warrants attention as a non-verbal resource of creative expression.

Present understanding of punctuation as a subsystem of human language is inadequate. Even when studies do examine punctuation from a linguistic standpoint, they privilege modern written language and American English in particular. Moreover, the application of these studies beyond their use in natural language processing systems and punctuation pedagogy is unexplored, their relevance to literary analysis undocumented. What wants in literary studies is a body of research that approaches punctuation linguistically— making use of corpus and computational methods of analysis— in order to attend more accurately to the particularities of its workings in poetry.

ENDNOTES

1. Akman, Current Approaches to Punctuation, 464.
2. Ibid.
3. Akman, Current Approaches to Punctuation, 460.
4. Ibid.
5. Akman, Current Approaches to Punctuation, 467.
- 6 Pullum, Natural Language and Linguistic Theory, 421-422.
- 7 Akman, Current Approaches to Punctuation, 461.
- 8 Kellogg, On the Punctuation of Two Lines, 174.
- 9 Koerner, Einar Haugen, 228.

REFERENCES

Akman, V., and B. Say. "Current Approaches to Punctuation in Computational Linguistics." *Computers and the Humanities*, no. 6 (1996): 457-69.

Kellogg, Robert L., and Oliver L. Steele. "On the Punctuation of Two Lines in *The Faerie Queene*." *PMLA*, no. 78 (1963): 147-148. doi: 10.2307/461235.

Koerner, E. F. K. "Einar Haugen as a historian of linguistics." *American Journal of Germanic Linguistics and Literatures*, no. 9 (1997): 221-38.

Lennard, John. *But I Digress: The Exploitation of Parentheses in English Printed Verse*. Oxford: Oxford UP, 1991.

Pullum, Geoffrey. "Topic... Comment: Punctuation and Human Freedom." *Natural Language & Linguistic Theory*, no. 4 (1984): 419-25.

Ricks, Christopher. *The Force of Poetry*. New York: Clarendon, 1984.