Journal of the Text Encoding Initiative

Issue 12 | July 2019 - May 2020

Selected Papers from the 2017 TEI Conference

Encoding Cryptic Crossword Clues with TEI

Martin Holmes

Electronic version

URL: <http://journals.openedition.org/jtei/2222>

DOI: 10.4000/jtei.2222

ISSN: 2162-5603

Publisher

TEI Consortium

Electronic reference

Martin Holmes, « Encoding Cryptic Crossword Clues with TEI », *Journal of the Text Encoding Initiative* [Online], Issue 12 | July 2019 - May 2020, Online since 19 August 2019, connection on 10 December 2020. URL : <http://journals.openedition.org/jtei/2222> ; DOI : <https://doi.org/10.4000/jtei.2222>

For this publication a Creative Commons Attribution 4.0 International license has been granted by the author(s) who retain full copyright.

Encoding Cryptic Crossword Clues with TEI

Martin Holmes

ABSTRACT

The cryptic crossword is a highly sophisticated and challenging type of intellectual puzzle that has been a daily feature of British newspapers¹ for nearly a century, and yet the culture and traditions surrounding it have received little scholarly attention. This article outlines a short history of the cryptic crossword and explains how cryptic clues work. I argue that cryptic crossword clues have a great deal in common with poetry, and that we have much to learn from their structure. Many cryptic clues depend for their effect on confusing the solver through the use of overlapping syntactic and semantic hierarchies, so they serve as evidence that overlapping hierarchies are not merely an unfortunate limitation afflicting XML languages but are psychologically and linguistically real. Finally, I present a TEI schema and an approach to encoding the components

of cryptic crossword clues in a way that enables algorithmic analysis of trends, features, and clue types, with a view to creating a historical corpus of encoded clues which will illuminate the evolution of the tradition.

INDEX

Keywords: cryptic crosswords, analysis and interpretation

1. How Do Cryptic Crossword Clues Work?

- 1 Whereas a “simple” crossword clue is merely a definition, a cryptic clue is a more sophisticated puzzle typically consisting of two parts: a definition and a set of codified instructions for building the solution. These components are woven into a phrase or sentence which has its own internal logic usually unconnected with the actual answer, intended to mislead the solver. This is a recent example from a *Guardian* crossword:

Amending pub sign, add in Cook’s vessel (7,5)

(Nutmeg² 2017)

The answer is *PUDDING BASIN*, an anagram of “pub sign add in,” defined by “Cook’s vessel.” The instruction to decode an anagram is provided by the word “amending.” The complete clue suggests perhaps the addition of HMS Endeavour to a pub signboard. The words comprising the “constructor” (my term for instructions for constructing the answer) are purposely obscured by their distribution across a phrasal boundary. Here is a second example:

Guard dog kept within sight (6)

(*Daily Telegraph* 2002)

The definition here is the first word alone. The “constructor” can be explained thus: put a word for “dog” (CUR) inside a word for “insight” (SEE) to create *SECURE*, which can be glossed as “guard.” The coherence of the familiar phrase “guard dog” confounds the solver’s attempt to separate the constructor from the definition, as does the fact that the word “sight” is a noun in the context of the whole clue, but a verb when used as part of the constructor. Other types of cryptic clues exist, but this type (definition with instructions for construction) is the most common in modern cryptic crosswords and will be the focus of this article.

2. A Brief History of the Cryptic Crossword in Britain

- 2 Word-square puzzles such as the Sator Square (figure 1) have existed since ancient time (Austin 1939), and riddles such as those found in the *Exeter Book* (figure 2) are among the earliest English-language documents we have. However, the first modern crossword did not appear until 1913, when Arthur Wynne authored a word-square puzzle which was published in the *New York World* (figure 3).

Figure 1. Sator Square at Oppède. Taken by M Disdero at Oppede, Luberon, France [via Wikimedia Commons].

Figure 2. Riddle 24 on folio 106v, the *Exeter Book*, Exeter Cathedral Library MS 3501 (tenth century). Image reproduced with permission of the University of Exeter, Digital Humanities and the Dean & Chapter, Exeter Cathedral.

Figure 3. Recreation of the first crossword puzzle, created by Arthur Wynne, published in the *New York World* on December 21, 1913. [Via Wikimedia Commons].

Crosswords featured in British newspapers from 1923, and, within a few years, some began to include clues which were more than “plain definitions,” such as “elusive definitions,” anagrams, and “hints” (Macnutt 1966, 19). The wholly cryptic crossword evolved by the early 1940s, since which time cryptic crosswords have appeared daily in all major British newspapers.

- 3 Several different types of cryptic clue emerged in the first forty years of the tradition, and the “rules” for setting clues were codified by the influential early setters Afrit (Alistair Ferguson Ritchie) and Ximenes (Derrick Somerset Macnutt). In his seminal work *Ximenes on the Art of the Crossword* (1966), Ximenes presents a taxonomy of clue types and principles for setters to adhere to in the interests of fairness. In the decades since, crossword setters have largely conformed to these core principles. Although some have been more rigorously “Ximenean” than others, it is fair to say that the tradition of “cluesmanship” (Macnutt 1966, 42) has been remarkably consistent; a crossword solver doing a regular puzzle in a daily newspaper over the last 50 years would not have experienced much change in the form and style of clues. Some clue types, such as those based on literary quotation, appear to be less common in recent years, while some new conventions and clue types have developed. At least one species of clue, in which the answer comes from describing the clue itself, seems to be more common recently. This type is not covered by Macnutt’s taxonomy, presumably, it would fall into his miscellaneous “various” category, and might be characterized as an “embodiment” clue. An excellent example from the master setter Araucaria (John Graham) is this:

Of of of of of of of of of (10) ³

(quoted in Hoggart 2013)

3. Crossword Grids

- 4 Ordinary daily newspaper crosswords (as opposed to themed special crosswords, which appear on holidays or special occasions) have symmetrical grids consisting of 15 x 15 squares, and ideally at least half of the “lights” (white squares) are “checked” (meaning that they form part of two separate answers). Macnutt (1966, 32ff) provides detailed advice on the construction of the grid and discusses different arrangements. However, grids are not in themselves particularly interesting and are not the focus of this article.

4. Why Study Cryptic Crosswords?

- 5 The cryptic crossword tradition stretches back nearly a century and has been stable for over fifty years. The best cryptic crossword clues exhibit the allusive compression, elegance, and wit that characterize good poetry, and they can elicit similarly delighted responses from solvers. Cryptic crosswords also make common use of a literary technique which is of particular interest to the XML encoding community. Philip Larkin's poem "Myxomatosis" (figure 4) illustrates the technique ([1955] 1966, 31):

Figure 4. Philip Larkin, "Myxomatosis," *The Less Deceived*, p.31.

This poem imagines an encounter between the speaker and a rabbit stricken by the disease myxomatosis, which was spreading through the UK at the time Larkin wrote the poem in the mid-1950s. The middle of the poem contains a brief imaginary conversation: "What trap is this? Where were its teeth concealed? / You seem to ask. / I make a sharp reply ..." ([1955] 1966, 31). The reader is initially disconcerted by the realization that the reply is in fact the act of killing the rabbit.

⁴ The "conversation" and the killing overlap; this technique is purposely deployed by the poet to increase the shock of the reader in encountering the sudden and brutal (but sadly necessary) act. The overlap is shown in figure 5.

Figure 5. Overlap of imagined conversation (in red) with killing (in blue) in “Myxomatosis.”

We see the same technique applied in the cryptic clues exemplified above. In the example of the “guard dog” from the *Daily Telegraph*, for instance, the phrasal hierarchy of the overall clue and the coherence of the phrase “guard dog” disguise the overlapping hierarchy of the parsed clue.

- 6 This overlap presents a scenario familiar to most XML encoders; attempting to encode these two components of the poem would require the use of workarounds such as anchors and pointers rather than conventional wrapping tags. The issue of overlapping hierarchies has long been presented as a “problem” for the “Ordered Hierarchy of Content Objects” approach which is inherent to XML encoding. Robinson (2007), for example, provides a good summary of the issue, particularly as it relates to the competing hierarchies of the conceptual work and its physical realization in a paginated book. However, in Larkin’s poem and the cryptic clues, we see two overlapping conceptual hierarchies. It is also important to notice that, since these devices “work”—we are (initially at least) fooled by the clue and jarred by the killing in the poem—they must in fact represent something psychologically and linguistically real.⁵ In other words, rather than being an irritating and merely technical problem for a markup language, the workarounds required to encode such phenomena reflect the mental discomfort we experience as readers.
- 7 In the paragraphs above, I have discussed a single literary technique which is manifested both in well-written cryptic clues and in poetry. There are many other techniques which deserve similar investigation—exploitation of homonyms and homophones, ambiguity in word categories, and

suggestive use of imagery are some of the more obvious. I argue that cryptic crossword clues should be studied as a distinct, albeit tiny, form of literary text. In addition to examining the use of such techniques, and how they relate to similar devices in other literary forms, it would also be illuminating to examine the evolution of clue types and conventions over the last eighty years, and investigate how the content and themes of cryptic clues reflect the changing world of the compilers and solvers. To my knowledge, no such investigation has ever been published. To do such work, a systematic method of encoding clues is a fundamental requirement.

5. Encoding Cryptic Clues

- 8 Computing methods have been applied to cryptic crosswords to auto-generate grids and clues (Berghel and Yi 1989), as well as to parse clues (Hart and Davis 1992). P. W. Williams and D. Woodhead (1979) proposed a formal notation called LACROSS, in which they represent clues as sequences of simple and compound components linked by operators. However, as far as my research has shown, no systematic approach to encoding cryptic clues in XML has been developed. After I presented this work at the TEI 2017 conference, Bethan Tovey made a presentation at Balisage on encoding crossword clues in XML (Tovey 2018). Tovey uses a custom (non-TEI) schema, whereas this article proposes a TEI schema and guidelines for encoding the components of clues and solutions using `<taxonomy>`, `<seg>`, and `@ana`. I initially created this schema for a personal project aiming to encode a representative sample of puzzles from British newspapers over the last eighty years, enabling algorithmic analysis of trends, features, and clue types. Currently, I am in the process of developing two taxonomies, one of clue types (starting from the lists in chapters 6, 7, and 8 of Macnutt 1966), and the other of clue components. The objective is to assign each clue to one or more categories and to break down its structure to clarify the way it works, showing how it misleads the solver. The following example of my proposed encoding is from the setter Picaroon (2017) in *The Guardian*. The clue is “Four card players wrapping party gifts (6)”:

```
<item ana="crs:ctpContainerContents">
  <seg ana="crs:ccpForm">
    <seg ana="crs:ccpConvention">Four card players</seg>
    <seg ana="crs:ccpSignal">wrapping</seg>
    <anchor xml:id="item_003_1"/>party </seg>
```

```

<seg ana="crs:ccpDef" xml:id="item_003_2">gifts</seg><anchor
xml:id="item_003_3"/>
<seg ana="crs:ccpLength">(6)</seg>
<span ana="crs:ccpMisdirection" from="#item_003_1" to="#item_003_3"> The phrase
  <mentioned>party gifts</mentioned> crosses the definition/form
  boundary.</span>
<span ana="crs:ccpMisdirection" target="#item_003_2"> The definition
  <mentioned>gifts</mentioned> is a noun in the context of the complete
  clue, but needs to be read as a verb to function as the
  definition.</span></item>

```

The answer, *ENDOWS*, is defined by “gifts.” The “four card players” are the points of the compass, ENWS (East, North, West, and South), as used in writing on bridge and other four-player card games; inside these four letters is a common British word for a party, *do*. Note that the coherent phrase “party gifts” spans the boundary between the form (constructor) component and the definition component. This is another instance of the overlapping hierarchy phenomenon discussed above, undermining the solver’s ability to parse the clue correctly.

9 The encoding may be analyzed as follows:

- Throughout the encoding, values from the two taxonomies are referenced by using a private URI scheme with the prefix *crs* (A private URI scheme is a machine-processable method of encoding potentially long or complex URIs in a shortened form, documented through the use of the TEI `<prefixDef>`.) These abbreviated pointers are supplied in the global `@ana` attribute.
- The entire clue is an `<item>` with the clue type “*crs:ctpContainerContents*”, derived from Macnutt’s “Container and contents” clue type.
- The `<item>` is divided into three `<seg>`s, the form or constructor (“*crs:ccpForm*”), the definition (“*crs:ccpDef*”), and the solution length (“*crs:ccpLength*”).
- The constructor contains a call to a well-known convention (“*crs:ccpConvention*”), the use of points of the compass to identify card players in four-player games.
- Following this is a signal word (“*crs:ccpSignal*”), “wrapping,” which acts as an instruction to include one component inside another.

- The final part of the constructor is the word “party,” the component whose synonym is to be wrapped.
- Two `<anchor>`s are used to delimit the beginning and end of the phrase whose coherence disguises the parsed clue hierarchy.
- Two `<span>`s describe instances of misdirection (`"crs:ccpMisdirection"`) employed by the compiler to make the clue more difficult. (A TEI `<span>` element, unlike the HTML element of the same name, is a standoff element used to point to a range within a block of XML, and is useful particularly when such a range is not amenable to direct tagging due to overlapping hierarchy issues.)

The home for this project is [Github](#), and the two taxonomies are encoded in the [ODD file](#) which is available from the repository. The first phase of the project, which is just getting started, involves the collection and encoding of a small representative sample of clues taken from the entire history of the British cryptic crossword, a process facilitated by the increasing availability of historical newspaper archives online. This initial stage will test and develop the two taxonomies. When the taxonomies are fully developed and tested, I propose to begin encoding entire puzzles from the major British newspapers which have published cryptic crosswords since the 1920s.

BIBLIOGRAPHY

- Afrit [Alistair Ferguson Ritchie]. 1949. *Armchair Crosswords*. By Afrit. London: Frederick Warne & Co.
- Austin, R. P. 1939. “Across and Down.” *Greece and Rome* 8 (24): 129–138. <http://www.jstor.org/stable/641256>.
- Berghel, H., and C. Yi. 1989. “Crossword Compiler-Compilation.” *The Computer Journal* 32 (3): 276–80. <https://academic.oup.com/comjnl/article/32/3/276/331540>.
- Daily Telegraph. 2002. *The Daily Telegraph Big Book of Cryptic Crosswords*. 9. London: Pan.
- Hart, M., and R. H. Davis. 1992. “Cryptic crossword clue interpreter.” *Information and Software Technology* 34 (1):16–27. [https://doi.org/10.1016/0950-5849\(92\)90091-3](https://doi.org/10.1016/0950-5849(92)90091-3).
- Hoggart, Simon. 2013. “Araucaria’s last puzzle: crossword master dies.” *The Guardian*, November 26, 2013. <https://www.theguardian.com/crosswords/2013/nov/26/araucaria-puzzle-crossword-master-dies>.

- Larkin, Philip. (1955) 1966. "Myxomatosis." *The Less Deceived: Poems*. 6th edition. 31. Hesse, Yorkshire: The Marvell Press.
- Macnutt, D. S. 1966. *Ximenes on the Art of the Crossword*. London: Methuen.
- Nutmeg [pseud.]. 2017. "Guardian Cryptic Crossword No. 27,183." April 28, 2017. <https://www.theguardian.com/crosswords/cryptic/27183>.
- Picaroon [pseud.]. 2017. "Guardian Cryptic Crossword No. 27,192." May 9, 2017. <https://www.theguardian.com/crosswords/cryptic/27192>.
- Robinson, Peter. 2017. "Some principles for making collaborative scholarly editions in digital form." *Digital Humanities Quarterly* 11 (2). <http://www.digitalhumanities.org/dhq/vol/11/2/000293/000293.html>.
- Tovey, Bethan S. T. 2018. "The Cryptic Crossword Corpus Project: first steps in establishing a markup vocabulary." *Markup UK 2018 Proceedings*. 29–35. <http://markupuk.org/Markup-UK-2018-proceedings.pdf>.
- Williams, P. W., and D. Woodhead. 1979. "Computer assisted analysis of cryptic crosswords." *The Computer Journal* 22 (1): 67–70. <https://academic.oup.com/comjnl/article/22/1/67/353483>.

NOTES

- 1 Cryptic crosswords exist in other languages, including Italian, Finnish, and Hebrew, and they are also common in English-language periodicals outside Britain, but this article deals only with the British tradition.
- 2 "Cruciverbalists" (crossword compilers) are customarily known by pseudonyms, a tradition which goes back to the earliest days. The same compiler will often publish under distinct pseudonyms in different publications.
- 3 The answer is *OFTENTIMES*.
- 4 It is not clear to me whether the killing is done by a blow from an actual stick, or whether the "stick" is actually a rifle, which would need to be cleaned after shooting a creature at close range.
- 5 I use these terms rather loosely to mean that the reader experiences a noticeable effect at the same point in the text where we can identify an overlap between the conceptual hierarchy and the syntactic hierarchy, and therefore that the hierarchical overlap in our model plausibly represents something similar in the reader's mental model of the text.

AUTHOR

MARTIN HOLMES

Martin Holmes is a programmer in the University of Victoria Humanities Computing and Media Centre. He served on the TEI Technical Council (2010–2015) and was Managing Editor of the *Journal of the TEI* (2013–2015).