

The Computer Generation of Cryptic Crossword Clues



Project staff

David Hardcastle
www.davidhardcastle.net

Other Project Partners

Project Details
Self-funded part-time PhD
2003-2007

Project Web Site
www.davidhardcastle.net

Keywords

Natural Language
Processing

Natural Language
Generation

Creative Language
Generation

Corpus Linguistics

Project Aims

The aim of this project [1] was to build a system which could automatically generate cryptic crossword clues based on word-play puzzles. Each clue consists of a word which defines the solution, and a puzzle – such as an anagram, one word inside another, or a combination of such puzzles.

To be valid, the clue must also appear to be a plausible fragment of English prose, and so the system must determine for each possible clue not just if it could be grammatically correct, but also if it might appear to mean something.

Key details

The search space for each clue is very large; typically there are between 10^7 and 10^{12} different ways in which a particular puzzle could be expressed. Only a tiny fraction of these will also appear to be sensible English clauses.

ENIGMA uses language understanding techniques to explore this search space, aggregating progressively larger phrases until it has created a whole clue, and rejecting phrases which break syntactic rules and semantic constraints [4].

The data sources for the semantic constraints were mined from a corpus, using a word-distance based algorithm to determine strength of thematic association [5] [6] and generalized syntactic dependencies to impose selectional restrictions on grammatical constructs [3].

ENIGMA was evaluated using various metrics, and also domain expert commentary. I also conducted a Turing-style test in which subjects had to spot the human-authored clues, on average the subjects guessed correctly 72% of the time [2].

Sample Generated Clues

1. Char holds battered tin delicately (8)
2. Slip plate above boiled peas (5)
3. Tipped scale down below broken bier (6)
4. Glaring line in rough going (6)
5. Still wild lionesses (9)
6. Damaged reactors contain chemical band (9)
7. Vitiating force returns call (7)
8. Consider spoken style (5)
9. Notice that question covers time (6)
10. Inquiry regarding quest (8)

Sample clues generated by ENIGMA

Answers to Sample Clues

1.DAINTILY 2.LAPSE 3.BRIBED 4.OGLING 5.NOISELESS
6.ORCHESTRA 7.MARRING 8.WEIGH 9.POSTER 10.RESEARCH

Key publications

- [1] Hardcastle, D. (2007). *Riddle posed by computer (6): The Computer Generation of Cryptic Crossword Clues*. PhD Thesis, Birkbeck College, University of London
- [2] Hardcastle, D. and Scott, D. (2008). *Can We Evaluate the Quality of Generated Text?* LREC08, Marrakech, Morocco.
- [3] Hardcastle, D. (2007). *Generalizing Syntactic Collocates for Creative Language Generation*. UCNLG07, Copenhagen.
- [4] Hardcastle, D. (2007). *Cryptic Crossword Clues: Generating Text with a Hidden Meaning*. EWNLG07, Schloss Dagstuhl.
- [5] Hardcastle, D. (2005). *An examination of word association scoring using distributional analysis in the British National Corpus: what is an interesting score and what is a useful system?* Corpus Linguistics 2005, Birmingham.
- [6] Hardcastle, D. (2001). *Using the BNC to produce dialectic cryptic crossword clues*. Corpus Linguistics 2001, Lancaster.