Generalizing Syntactic Collocates for Creative Language Generation

David Hardcastle
Birkbeck College, University of London, London, UK
Open University, Milton Keynes, UK

`d.w.hardcastle@open.ac.uk`

Abstract

This paper presents the construction of a data source that supports the automatic generation of cryptic crossword clues in a system called ENIGMA. Cryptic crossword clues have two layers of meaning: a surface reading that appears to be a fragment of English prose, and a puzzle reading that the solver must uncover to solve the clue. The content expressed by the clue, and the input to the generation process, is a word play puzzle, such as an anagram, perhaps. In expressing this puzzle ENIGMA must choose language creatively, so that a separate, surface reading of the text is also generated – in effect translating a semantic input via a layered text to a new semantic output. To ensure that this surface text is meaningful, ENIGMA uses corpus data to determine which words can be combined meaningfully and which cannot.

1   Introduction

In a typical natural language generation (NLG) system lexical choice is tightly constrained by content, its goal being to select "words that adequately express the content that is to be communicated" (Stede, 1993:1). ENIGMA is unusual in that it must communicate two different expressions of content within the same text. This problem is similar to the problem faced in computational humour or poetry generation, but the approach taken is rather different. Instead of computing a multi-layered semantic representation and then generating from it (Ritchie, 2005; Binsted, 1996; Attardo et al, 2002) or using

reflexive generation to consider multiple combinations of content (Manurung et al, 2000), ENIGMA assembles the clue based on the content of the puzzle reading and explores connections between possible lexicalisations that would result in a grammatical and meaningful surface text. I term this natural language creation (NLC) as it involves the creation of new meaning alongside the generation of text, and it integrates natural language understanding (NLU) tasks into the process of generating a text (see Hardcastle, 2007).

At a high level of abstraction, we can think of ENIGMA as translating from a semantic representation of some content (the puzzle reading) via an intermediary, multi-layered text (the clue) to a semantic representation of some novel, created content (the surface reading) that belongs to a different domain from the semantic input.

For this process to work, ENIGMA needs to be able to make lexical choices that are informed not just by grammatical considerations but also by judgements about meaning. This paper presents a process whereby sparse data relating to occurrences of dependency relations is extracted from the British National Corpus (BNC), generalized using WordNet (Miller, 1990) and marshaled into a *collocational semantic lexicon* – a structured data source that determines if two words can be attached **meaningfully** through a particular relation. For example, if ENIGMA wants to express the idea that the solver must think of an anagram of the word *lionesses* it can do so by juxtaposing *lionesses* alongside one of around 400 crossword convention keywords that indicate anagram, such as *wild*, *out*, *letters*, or *confuse*. The grammar rules

tell the system how to link the words together grammatically into fragments of text such as *wild lionesses, lionesses confuse, letters and lionesses* and so on., but only some of these fragments will be meaningful. Using the semantic collocational lexicon presented here ENIGMA is able to semantically constrain the grammar rules and construct fragments such as *wild lionesses, lionesses wander* or *lionesses are lost* whilst avoiding alternatives such as *scrambled lionesses, rebuild lionesses* or *lionesses are reordered*.

This lexicon is not specific to the generation of cryptic crosswords; it defines the terms that can participate, in English, in a given dependency relation with a given colligand. For example, the set of things that can be red, that can drive or that can one can cook.

2   Cryptic Crossword Clues

Cryptic crossword clues of the sort generated by ENIGMA present a wordplay puzzle to the solver - such as an anagram, writing a word backwards, writing one word inside another, and so on - disguised as a fragment of English text. There are many conventions that determine what words can be used to indicate a given wordplay and the order in which the elements of the puzzle must be presented. The most difficult, and entertaining, clues are usually those that present the most natural surface reading to the solver, since the solver must look beyond this surface reading in order to read the clue as a puzzle. Consider for example the following clue generated by ENIGMA:

Strangely tiny scale drawing (5)

The surface text is grammatically correct (an adverbially qualified adjective modifying a compound noun phrase) and appears to mean something. To solve the clue, though, the reader must reinterpret the text as a wordplay puzzle in which an anagram of *tiny* – *tyin* - is combined with a (musical) *scale* – *g* – to give a word that can mean *drawing*, namely *tying*[1]. Note that the surface text requires *drawing* to be interpreted as a noun, whereas for the puzzle it must be interpreted as a

verb. The use of homographs in this way is a common cryptic feature.

There are a great many alternative renderings of this puzzle that ENIGMA could have chosen, and many that would be grammatically correct, such as this example:

Drawing addles tiny golf (5)

This clue works as a puzzle (*tying* can be constructed from an anagram of *tiny* followed by the letter *g*), and it is grammatical - a simple verb clause with a noun as subject and a noun phrase, consisting of an adjectivally modified noun, as direct object. However, it is nonsense; drawings don't addle things, golf can't be tiny, and it doesn't make sense for golf to be the direct object of the verb to addle. Supplementing syntactic constraints with semantic selectional constraints is critical to ENIGMA's performance, and so the system needs to be able to determine what can be tiny, what can addle, and so on.

3   Some Related Work

Choices about the fit between pairs or groups of words can be informed by distributional information from corpus analysis using a variety of statistical techniques (Church and Hanks, 1990; Dunning, 1993; Hardcastle, 2005), and this data has been used to inform lexical choice in NLG. For example, Smadja and McKeown (1990) extract likely "binary lexical relations" from a corpus using cooccurrence information and statistical analysis and use it to assist lexical choice; Langkilde and Knight (1998) use statistical information about bigrams to support determiner-noun, subject-verb and other collocational lexicalization decisions, and Inkpen and Hirst (2002) use a variety of statistical methods to determine lexical choices between near-synonyms in collocations.

In language understanding tasks, the use of n-gram language models (Brown et al, 1992) or collocational statistics (Golding and Roth, 1999) can assist in ranking a closed set of alternatives highly effectively. In an NLG context though, the choices are more fine-grained. Consider, for example, the difference between choosing from a list of near-synonyms (Inkpen and Hirst, 2002) as opposed to a list of orthographic confusables (Golding and Roth, 1999). Because generation requires such fine

---

[1] *Tying* and *drawing* are both polysemous, the clue is using the sense of coming in equal position.

granularity this also precludes the use of existing comparable resources such as PropBank (Kingsbury and Palmer, 2002), FrameNet (Johnson and Fillimore, 2000) or VerbNet (Kipper et al, 2000). For example, Shi and Mihalcea (2005) link FrameNet to WordNet and use subsumption under the WordNet hierarchy to generalize the data and increase coverage. However, the granularity remains very coarse; although there are many more lexemes in each category there are still very few categories defined. In an Information Retrieval context this increases the effectiveness of the resource, but for lexical choice in generation the granularity needs to be much finer.

A further problem that is, perhaps, particular to ENIGMA is the wealth of options available to the system to express components of the clue. For example, there are around four hundred terms known to the system that can indicate an anagram, including adjectives, adverbs, verbs and adverbial prepositions. This provides the system with the flexibility to locate a workable solution, but for the system to exploit the range of available options, it needs to know about the interaction of a wide range of different words through a variety of different relationships.

Another drawback with statistical data from corpora is the bias toward typical, or even prototypical, usage. Many of the collocations that prove to be statistically significant will not be fully compositional (see Manning and Schutze, 2002:151), meaning that the bigram itself carries more meaning than the sum of its parts. But in an NLG context we don't necessarily want this additional level of meaning that arises from the collocation; we may want to make lexical choice decisions based simply on compositional meaning, regardless of the frequency of the terms.

I address these issues by using data based on dependency relations (such as *subject of verb*, *direct object* or *adjective modifier*) evidenced in the corpus text, rather than raw collocation or n-gram data. The benefit in this approach is that it incorporates the wealth of knowledge and experience invested in state-of-the-art parsers into the system, providing ENIGMA with data not just about the company that words keep (which is important), but about the other words that a word can interact with,

and the manner of those interactions – crucial information for generating language creatively.

A range of different methods have been used to extract dependency relations from text: Smadja and McKeown (1990) post-process concordance data to infer dependency relations; Kilgarriff (2004) effectively uses regular expressions for the Sketch Engine; Velardi et al. (1991) apply heuristics to chunk the text and then parse those chunks, and Hindle (1990), Lin (1997) and Zinsmeister et al. (2003) turn to statistical parsers.

I experimented with my own regular expression searches, since they are a fast and efficient means of analyzing a large corpus such as the BNC, but decided to use the Stanford parser (Klein and Manning, 2003) due to its superior accuracy. I also considered using a broad coverage parser such as MiniPar (Lin, 1998) and note that since performing the analysis the CCG parser described in (Clark and Curran, 2004) has been made available. Future developments of the research presented in this paper could include a comparison of the trade-off between accuracy and efficiency in retrieving broad coverage typed dependencies from large corpora.

4   Generalization

Collocational data extracted from corpora is notoriously sparse, since the data relies not just on the frequency of the words in question, but on the frequency of their use in combination. The data relating to dependency relations suffers even more from sparsity than cooccurrence information based on distributional analysis. While almost all occurrences of a word in the corpus have some surrounding context, and thus co-occur with some other words, few occurrences may participate in the dependency relations mined from the corpus.

This sparsity problem is mitigated by generalizing the sets of nouns that participate in each recovered relation (for example the set of nouns that are evidenced in the BNC as being *red*) by mapping them into WordNet and applying a minimal arc-distance algorithm to group them into sub-trees from which generalizations can be inferred. A coarse-grained sense disambiguation is applied as the data is mapped into WordNet to reduce noise,

and the resulting mappings are grouped and generalized into sub-trees that are then filtered for coverage and compiled into a lexicon.

4.1 First-Pass Disambiguation

To see why disambiguation matters consider, for example, the noun *chicken*, which is a member of the set of nouns that are found to be modified by the adjective *grilled* in the BNC. It has four senses in WordNet: a fowl, a foodstuff, a coward and a game. The hope is that there will be sufficient evidence of foodstuff being grilled for the system to generalize safely, but, because of polysemy, there may also be evidence of birds, people or games being grilled. Unlike typical approaches to word sense disambiguation (see Ide and Véronis, 1998: 3f) this algorithm does not disambiguate the words in their original context, but rather attempts to cluster all the nouns evidenced as being grilled using the WordNet hierarchy.

When the lexicographers responsible for WordNet tackle a new entry, they first classify it into a broad semantic class known as a lexicographer file number (hereafter referred to as a *lexnum*). There are forty-four such classes for WordNet 2.1, of which twenty-nine relate to nouns. ENIGMA performs the first-pass disambiguation using these noun classes as quasi-domains, allowing it to attempt a coarse-grained disambiguation that seeds the allocation of each term in the argument list to a particular WordNet sense.

First, the system spreads the frequency of each collocation (as evidenced in the BNC) over all of the WordNet senses available; so for example since *grilled chicken* is evidenced 3 times in the BNC each of the 4 synsets for *chicken* receives a starting score of 0.75. I also experimented with adding bias for the depth of the synset, since depth equates to specificity of meaning in WordNet, but the topology of WordNet is very uneven and this bias seemed unhelpful.

The synsets are then aggregated by lexnum into quasi-domains, with each score being first normalized by the relative size of the domain in WordNet – i.e. the percentage of WordNet synsets that belong to that lexnum. These scores are then re-expressed as percentages. Figure 1 shows the initial allocations for *grilled*; although the food do-

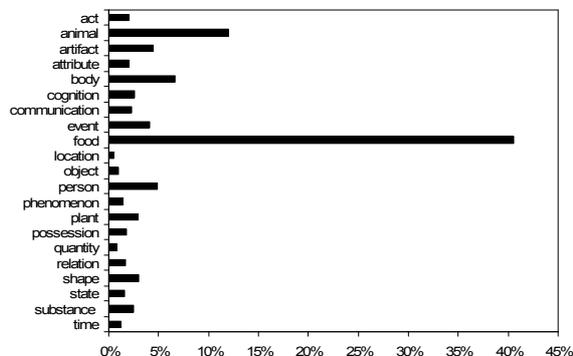main has the highest score (41%), the domains animal, body and person are also well-represented.

Figure 1. Initial lexnum allocations for grilled.

To remove this noise the system now repeats the process but instead of spreading the scores equally over all of the synsets, the percentage scores for each lexnum are used. So, for example, when the 3 occurrences of *grilled chicken* are processed the foodstuff meaning of *chicken*, which has lexnum food, receives 41% of 3, the fowl receives 12%, the coward 5% and the game 4%. Note that the percentages do not total 100% since not all lexnum domains are represented. Once this process is complete the system recalculates the percentages for each domain and restarts the process, resulting in a positive feedback loop that accentuates the shape of the data.

Once the positive feedback cycle settles and no changes to the proportions are detected, each entry is now allocated to the synset whose lexnum has the highest weighting score. So, for example, chicken is allocated to the synset belonging to the lexnum food. Figure 2 shows the allocation of the 66 entries for grilled by lexnum domain.
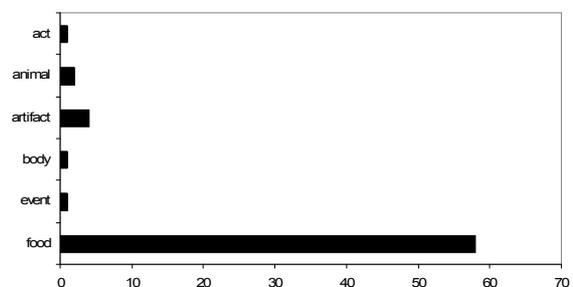

Figure 2. Disambiguated allocations for grilled.

Using the lexicographer file numbers as stand-in domain annotation provides an axis orthogonal to the hyponymy hierarchy of WordNet, and also provides a granularity that is sufficiently coarse-grained that a simple pooling algorithm can succeed – see Ciaramita et al (2003) who also use lexicographer file numbers as coarse-grained sense markers. I considered using the finer grained WordNet Domains described in Magnini et al (2002), although these are annotated against an earlier version of WordNet[2] and so considerable mapping effort would have been required.

The graph in Figure 2 also demonstrates how first-pass disambiguation reduces noise. Many of the nouns that are found to be modified by *grilled* in the BNC have a sense that means the flesh of an animal, but can also mean the animal itself (such as *lamb*, *fish*, *chicken* or *sardine*). In a language understanding context, blurring this distinction may be helpful, as it would enable the application to deal with figurative language, or unexpected collocates such as *grilled dog*. For generation, however, using positive feedback to reduce polysemy reduces noise, and this in turn reduces the chance of generating peculiar sounding collocations.

4.2   Generalization by Hyponymy

This coarse-grained disambiguation associates each noun for an entry such as *grilled* with a specific WordNet sense, and therefore a specific WordNet synset. The next step in the process is to use the hyponymy structure of WordNet to construct *sub-trees*. A sub-tree is composed of synsets that share a common cohypernym that is within a parameterized arc-distance from all of its members. Each sub-tree is then filtered for *coverage* by counting every synset of each member of the entry set (such as things that can be *grilled*) that is a hyponym of the sub-tree's cohypernym root and dividing this number by the total number of nodes in WordNet that are hyponyms of that root. Frequency is not used here as coverage is measuring the *representativity* of the evidence for the generalization that will be made if the sub-tree is used. In particular, this prevents very large sub-trees of WordNet being selected on the basis of a small number of synsets that occur in very common idioms.

For example, in the set for *broken* the sub-tree under the root node *adornment* is retained, as there is direct evidence in the BNC for 20 of its 75 hyponyms. By contrast the sub-tree under the root node *cognition* is discarded as only 20 of its nearly 4,000 are directly evidenced as being *broken* in the BNC.

It is important to note that while only the disambiguated synsets are used to *construct* the sub-trees using the arc-distance constraint, the filtering by coverage uses *all* synsets for each member of the entry set. So, for example, if there were enough grilled things that were disambiguated as birds to form a sub-tree under the synset bird, then the fowl synset for chicken would count toward the coverage of this sub-tree, **and** the foodstuff meaning would count toward the coverage of the food sub-tree. This allows the system to handle collocations such as *to throw a ball* where the whole collocation can be read polysemously, and also to handle occasions where multiple senses of a noun could reasonably be modified by some adjective, or governed by some verb (consider *large bank*, for example).

I parameterised a tight arc-distance threshold (3) and a high threshold for coverage (20%). In an information retrieval context these parameters could be relaxed, allowing wider generalizations to be made on the basis of less evidence. For ENIGMA a cautious approach is more appropriate, since any member of the generalization could be used in a clue, even if a directly evidenced alternative exists, to maximize lexical variety in generation.

Finally the generalized data is written to a 'collocational semantic lexicon' which lists against each headword the indices of the cohypernym roots of each sub-tree. ENIGMA can use this lexicon to determine the semantics of a proposed relation by checking to see if any synset for the proposed noun is a hyponym of any of the root nodes listed against the headword. The lexicon also contains all of the remaining nouns that were evidenced in the BNC but have not been allocated to a sub-tree, in other words that are not part of any generalization. Collocations with unusually high log-likelihood

---

[2] Version 1.7 rather than version 2.1.

(Dunning, 1993) are flagged[3]; these are likely to be non-compositional collocations, and so will add weight to a clue's ranking for idiomaticity if used.

4.3 Sample Output

Table 2 presents a small proportion of the entry in the lexicon for red, by way of example. The generalizations are synsets in WordNet, and the lexicon asserts that any hyponym of each of these nodes can be red. The entries listed as evidenced could not be generalized, yet were found modified by red in the BNC. The asterisked members of this group are flagged as likely non-compositional collocations.

Generalizations	vegetable, coat, furniture, merchandise, flower, injury
Evidenced	alligator, blister, belly, phosphorus, stone, flame, sauce, crescent*, admiral*, squirrel*, label*, meat*

Table 2. Sample data from the entry in the collocational semantic lexicon for the adjective red.

4.4 Evaluation

I performed a task-based evaluation of the lexical choice component of ENIGMA using a forced choice questionnaire to test the collocations chosen by the system for a set of sixty adjective-noun pairs generated for nouns known to be anagrams of other words. In each case the adjective was chosen by the system as an apposite indicator of an anagram, but was accompanied by two control adjectives selected at random from the pool of anagram keyword adjectives not thought to be apposite in the particular case. Subjects were asked to choose the adjective-noun pairing that they imagined they would be most likely to encounter in spoken English. Figure 3 presents a sample of the forced choices offered to subjects; in the figure the adjective chosen by ENIGMA is the first of the three, for

---

[3] The system measures -2logλ as in Dunning's paper (Dunning, 2003) but we cannot just use χ^2 significance as the members of the set were extracted on the basis of a syntactic relation and so they are not independent. Instead a top-slice is taken, since these could plausibly be non-compositional collocations and so it makes sense for the system to flag them.

the experiment the ordering was of course randomised.

```
mixed/ordered/modified spice
broken/corrected/blended anvil
awkward/varied/untrue teenager
```
Figure 3. A sample of the adjective-noun choices presented to the subjects.

Thirty subjects participated in the experiment. I calculated the P-value for each entry as the probability under the binomial distribution of getting more than the number of matches observed if choices were made randomly, and used a confidence threshold of p<0.01. The null hypothesis was rejected in favour of ENIGMA in 51 (85%) of the 60 entries. Overall, agreement between subjects was very high: in 6 of the 9 negative results, the null hypothesis was rejected, i.e. the choice was not random, but not in ENIGMA's favour. This implies that where the chosen adjective differed from ENIGMA's selection, there may be other circumstantial factors that made these alternatives seem appealing in their own right.

I also ran a word association measure (described in Hardcastle, 2005) based on a statistical analysis of cooccurrence data from the BNC against the test data set used in the evaluation as a baseline comparison. The cooccurrence algorithm correctly identified 11 of the 60 pairings, but it only returned a result for 13. As one would expect it performed well for idiomatic collocations such as *mixed spice* or *awkward teenager*, but could not deal with purely compositional meaning in pairings such as *broken anvil*, *curious sanction* or *abnormal footprint*. This comparison underlines the benefit in data gain that the generalization step provides.

A full description of the evaluation experiment including results and further discussion is presented in Hardcastle, 2007b.

I also conducted an end-to-end evaluation of ENIGMA comprising a qualitative and quantitative component. Sixty subjects with a wide range of experience in solving cryptic crosswords took part in a Turing style test, in which they were presented with thirty pairs of clues, each of which consisted of a clue from a national

newspaper and a generated clue for the same word, chosen at random from the top-ranked clues generated by the system. On average subjects chose correctly 72% of the time, and this finding was supported by their comments with 25 of 40 subjects who commented saying that they found it hard to tell the clues apart. As a population the subjects performed better and only two generated clues were preferred by the population as a whole. There was substantial disagreement with a sizeable group of generated clues chosen regularly; in over half of the pairs more than a quarter of the subjects mistook the generated clue for the human-authored one.

I presented the generated clues used in the quantitative evaluation to a small group of domain experts – professional compilers and editors and online commentators – and invited them to comment. In general they found the clues rather sterile and lacking in humour. However, they did find some of the surface readings convincing and most thought that at least some of the clues would be acceptable in a published crossword in a broadsheet. Most of the experts and many of the Turing style test subjects who commented reported that they found the connections between words in the clues convincing in the better clues, often citing as examples syntactic collocates extracted and generalized by the component described in this paper.

5   Limitations

To address the issue of data sparsity the dependency relations extracted from the BNC are generalized using WordNet, as described above. This implies that some isomorphism exists between the hyponymy hierarchies defined in WordNet, and the domain of nouns that can be modified by particular adjectives, or be the subjects or objects of particular verbs. This is supported in the literature by Lin (1997) for whom WordNet functions as a point of comparison in evaluating a machine-generated thesaurus based on a collocational similarity measure; Green, Dorr and Resnik (2004) who combine LDOCE verb senses with WordNet synsets to infer high-level semantic frames for SemFrame and by Shi and Mihalcea (2005) who use WordNet to unpack the selectional restrictions defined in FrameNet and VerbNet. On the other hand, Kilgarriff (1997) proposes that word senses amount to clus-

ters of collocations that are large and distinct enough to be salient, for some purpose or in some context. Similarly, Hindle (1990) presents "an approach to classifying English words according to the predicate-argument structures they show in a corpus of text", as opposed to a static classification in a dictionary or thesaurus. Rather than sharing some isomorphism with WordNet, it could be argued that senses grouped according to their role as participants in relationships such as adjective-noun or subject-verb will belong to many different groupings depending on register, domain, context and other factors. This criticism is particularly salient for the work described in this paper due to the fine granularity of the lexicon. Looking at the data in more detail, there are many examples of situations in which collocations evidenced in the BNC do not map straightforwardly onto WordNet groupings, for a number of different reasons.

- Ranged data. Since the colour spectrum is continuous all colour distinctions are arbitrary in nature, and although some objects that share the same colour may also share other features, this need not be the case.
- Synecdoche. One might expect that the entry for *broken* would include a generalization about limbs, or parts of the skeleton. In practice the BNC lists some actual bones that are broken but also includes loci such as *ankle*, *shoulder*, *finger* or *leg* that are in a different part of WordNet.
- Figurative speech. Many of the collocations that could not be generalized are idiomatic or figurative in nature, for example *red mist, broken heart* or *new potato*. Being non-compositional in meaning they cannot tell the system anything that can be generalized. Captured as single collocations they provide useful data to the system, but during disambiguation and generalization they simply introduce noise.
- Sub-domain vocabulary. In addition to *eggs* we find that *goals* and *passages* can also be *scrambled* in the BNC. This occurs because the BNC includes sports coverage, an idiolect with its own peculiar grammar and its own bespoke collocations. Sub-domain usages such as these defy attempts to systematise collocational relationships.
- Predicate polysemy. In this paper I only try to resolve polysemy at the level of the arguments,

where their grouping within WordNet can support disambiguation. There is no data with which to disambiguate the predicates, but for some entries the different sub-trees represent not just different groupings within some shared overall sense but quite distinct senses. Consider for example the different senses of *broken* in *broken vase, broken beam, broken leg*, and *broken computer*; in each case the noun is broken in a rather different sense, and so any generalization would be based on a different feature set.

- WordNet senses. There is only one entry for *kidney* in WordNet, and that is as an organ. This prevents the collocation *grilled kidney* from being included in the generalization about grilled food.

- WordNet topology. The WordNet topology is very uneven, and this means that constraints such as arc-distance have a different impact in different parts of the structure. For example, the synset *Irish_water_spaniel* is five edges away from the synset for *dog*, too far to be included in the sub-tree. However most modifiers that apply to dogs will likely apply to Irish water spaniels too. Conversely the synsets *bleach* and *deus_ex_machina* have an arc-distance of three, but probably rather less in common when it comes to adjective modifiers.

6 Conclusion

In spite of these limitations, the use of WordNet to generalize the dependency relations that I extracted from the BNC provides ENIGMA with useful data with which it can make informed choices about meaning. A key benefit of generalizing relational data mined from a corpus over manual approaches such as annotating semantic role data is that the granularity is much finer, so the data is suitable for generation. Although there are many occasions where it seems that the axis behind a semantic relation lies in some different dimension to the hierarchy of WordNet, the fine granularity and cautious parameterization of the expansion process prevent these irregularities from dominating the resulting lexicon. Instead, most of the entries contain a large number of small generalizations, preserving enough of the granularity of the original extract to be able to cope with some of the difficulties listed above, whilst still allowing for enough generalization to tackle data sparsity. When things go wrong it is therefore more commonly the case that positive results are missed rather than that false positives are introduced.

References

Salvatore Attardo, Christian Hempelmann and Sara Di Maio. 2002. Script oppositions and logical mechanisms: Modeling incongruities and their resolutions. *Humor – International Journal of Humor Research*, 15(1): 3-46.

Kim Binsted 1996. *Machine Humour: An Implemented Model of Puns*. Ph.D. Dissertation, Department of Artificial Intelligence, University of Edinburgh.

Peter Brown, Peter deSouza, Robert Mercer, Vincent Della Pietra and Jenifer Lai. 1992. Class-based n-gram models of natural language. *Computational Linguistics*, 18(4):467-479.

Kenneth W. Church and Patrick Hanks. 1990. Word association norms, mutual information and lexicography. *Computational Linguistics*, 16(1):22–29.

Massimiliano Ciaramita, Thomas Hofmann and Mark Johnson. 2003. Hierarchical semantic classification: Word sense disambiguation with world knowledge. In *The 18th International Joint Conference on Artificial Intelligence*.

Stephen Clark and James Curran. 2004. Parsing the WSJ using CCG and log-linear models. *In Proceedings of the 42nd Annual Meeting of ACL*, Art. 103.

Ted Dunning. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74.

Andrew Golding and Dan Roth. 1999. A Winnow-Based Approach to Context-Sensitive Spelling Correction. *Machine Learning*, 34(1-3):107-130.

Rebecca Green, Bonnie J. Dorr and Philip Resnik. 2004. Inducing Frame Semantic Verb Classes from WordNet and LDOCE. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics*, pages 375-382.

David Hardcastle. 2005. Using the distributional hypothesis to derive cooccurrence scores from the British National Corpus. In *Proceedings of Corpus Linguistics*.

David Hardcastle. 2007. Cryptic Crossword Clues: Generating Text with a Hidden Meaning. In *Proceedings of the 11th European Workshop on Natural Language Generation*.

David Hardcastle. 2007b. *Evaluation of ENIGMA's Lexical Choice Component*. Technical Report BBKCS-07-06, Birkbeck College, London.

Donald Hindle. 1990. Noun classification from predicate-argument structures. In *Meeting of the Association for Computational Linguistics*, pages 268–275.

Diana Inkpen and Graeme Hirst. 2002. Acquiring collocations for lexical choice between near synonyms. In *Proceedings of the ACL-02 workshop on Unsupervised Lexical Acquisition - Volume 9*, pages 67-76.

Nancy Ide and Jean Véronis. 1998. Word Sense Disambiguation: The State of the Art. *Computational Linguistics*, 24(1): 1-40.

Christopher R. Johnson, and Charles J. Fillmore. 2000. The FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure. In *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics (ANLP-NAACL 2000)*, pages 56-62

Adam Kilgarriff. 1997. I don't believe in word senses. *Computers and the Humanities*, 31(2):91-113.

Adam Kilgarriff. 2004. The sketch engine. In *Proceedings of Euralex*, pages 105–116.

Paul Kingsbury and Martha Palmer. 2002. From Treebank to Propbank. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*.

Karin Kipper, Hoa Trang Dang and Martha Palmer. 2000. Class Based Construction of a Verb Lexicon. In *Proceedings of AAAI-2000 Seventeenth National Conference on Artificial Intelligence*.

Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *ACL '03: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pages 423–430.

Irene Langkilde and Kevin Knight. 1998. The practical value of N-grams in derivation. In *Proceedings of the Ninth International Workshop on Natural Language Generation*, pages 248–255.

Dekang Lin. 1997. Using syntactic dependency as local context to resolve word sense ambiguity. In *Meeting of the ACL*, pages 64–71.

Dekang Lin. 1998. Automatic retrieval and clustering of similar words. In *The 17th International Conference on Computational Linguistics*, pages 768–774.

Christopher D. Manning and Hinrich Schutze. 2002. *Foundations of Statistical Natural Language Processing*. The MIT Press.

Hisar Manurung, Graeme Ritchie and Henry Thompson. 2000. Towards A Computational Model of Poetry Generation. In *Proceedings of AISB Symposium on Creative and Cultural Aspects and Applications of AI and Cognitive Science*, 79-86.

Bernardo Magnini, Carlo Strapparava, Giovanni Pezzulo and Alfio Gliozzo. 2002. Using domain information for word sense disambiguation. In *Proceedings of Senseval-2 Workshop, Association of Computational Linguistics*, page 111-115.

George A. Miller. 1990. Wordnet: a lexical database for English. *Communications of the ACM*, 38(11):39–41.

Graeme Ritchie. 2005. Computational Mechanisms for Pun Generation. In Proceedings of the 10th European Workshop on Natural Language Generation (ENLG-05), pages 125-132.

Lei Shi and Rada Mihalcea. 2005. Putting Pieces Together: Combining FrameNet, VerbNet and WordNet for Robust Semantic Parsing. In *Proceedings of the Sixth International Conference on Intelligent Text Processing and Computational Linguistics*, pages 100-111.

Frank A. Smadja and Kathleen R. McKeown. 1990. Automatically extracting and representing collocations for language generation. In *Proceedings of the 28th. Annual Meeting of the Association for Computational Linguistics*, pages 252–259.

Michael Small. n.d. *Ongoing Phd Thesis*. Ph.D. thesis, Birkbeck College, London.

Manfred Stede. 1993. Lexical choice criteria in language generation. In *Proceedings of the 6th Conference of the European Chapter of the ACL*.

Paola Velardi, Michela Fasolo, and Maria Teresa-Pazienza. 1991. How to encode semantic knowledge: a method for meaning representation and computer aided acquisition. *Computational Linguistics*, 17(2):153–170.

Heike Zinsmeister and Ulrich Heid. 2003. Significant triples: Adjective+noun+verb combinations. In *Complex 2003*.