Integrating Verb-Particle Constructions into CCG Parsing

Author: James W. D. Constable, jcon6353@it.usyd.edu.au
Supervisor: Dr James R. Curran
School of Information Technologies

1. Verb-Particle Constructions

Verb-particle constructions (VPCs) are a common type of multiword expression, comprising a verb and a particle (usually a preposition) which together form a complete semantic unit. Sometimes their meaning can be clearly attributed to the component parts (e.g. *pick out*); other times they are idiomatic, with no clear link between words and meaning (e.g. *make out*).

Their ability to manifest in both a 'joined' and 'split' configuration (*gunned down the man* vs. *gunned the man down*) poses a unique challenge to parsers, as it prevents them from being able to treat the construction as a single unit, and demands a system that is able to maintain the semantic bond between the components, even when they are non-adjacent.

2. Aims of the Project

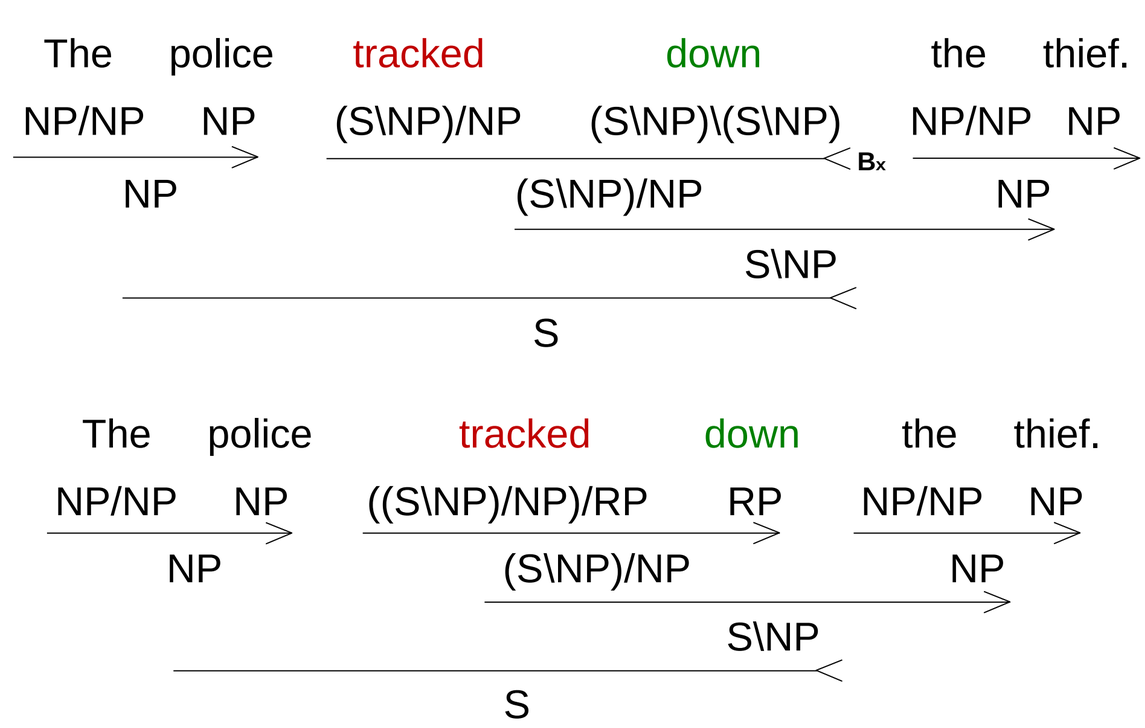
- To improve the parsing of VPCs, both in terms of accuracy and semantic soundness.
- To devise a suitable and consistent system for representing VPCs in CCG (Combinatory Categorial Grammar [4]), and to use this new system to modify the CCGbank corpus [2].
- To assess the impact of these changes on a parser built on the CCG formalism.

3. Combinatory Categorial Grammar (CCG)

CCG[4] is a lexicalised grammar formalism based on combinatory logic, appealing to NLP researchers because its high degree of lexicalisation allows for efficient bottom-up parsing strategies. Examples of CCG derivations are shown opposite. The parser used for this project is built on the CCG formalism [1], and CCGbank, the corpus we are modifying, is the primary corpus for CCG-oriented work.

4. Before and After

The first step in this project was to modify the CCGbank corpus. We did most of this conversion automatically, using the semantic annotations in PropBank[3] to assist in locating VPCs in the corpus. The figure below shows two CCG derivations of the same sentence using (a) the most common CCGbank scheme, and (b) our modified scheme.



a) Before

Note that the particle is given a modifier category (modifiers take the form `X/X` or `X\X`).

b) After

The particle is assigned the new category `RP`.

5. The More Complex Cases

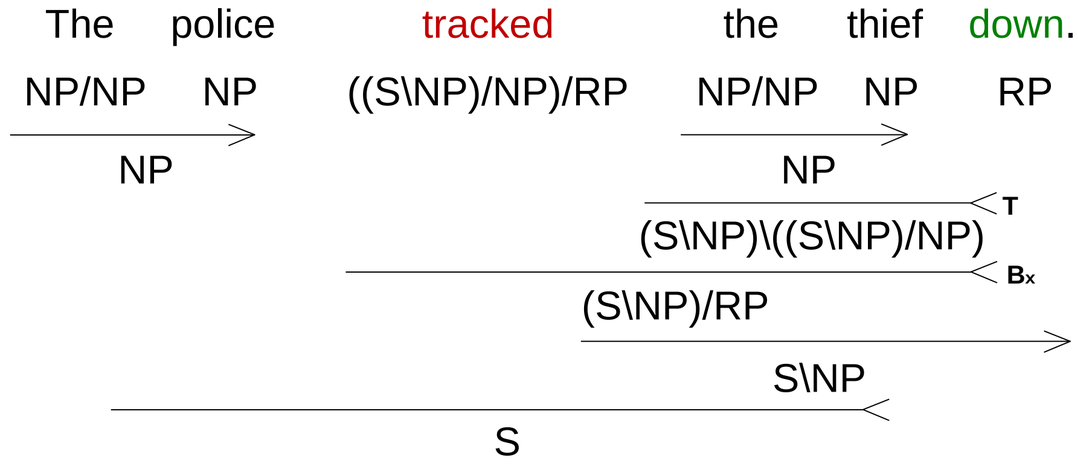
The joined configuration (shown above) is easily handled using functional application. Some of the other behaviours exhibited by VPCs require a slightly more complex approach.

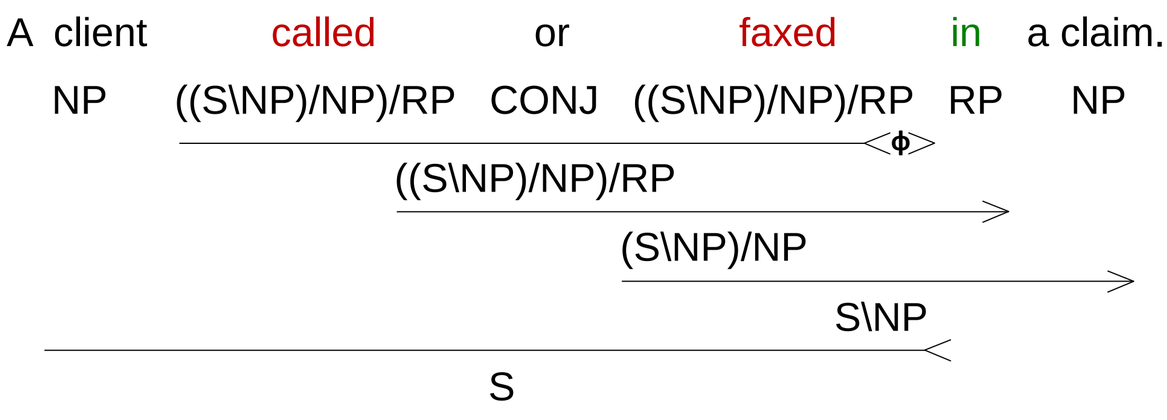
The Split Configuration:

This configuration is accommodated using a combination of type-raising and backward crossed composition.

Verb Coordination across VPCs:

A relatively rare construction, our system gracefully handles verb coordination using CCG's simplified coordination combinator.





6. Results

After modifying the parser to include support for the new categories, it was retrained and tested on the modified corpus. The table below shows the results (C&C + VPC), along with those from the unmodified parser (C&C).

Model	LP	LR	LF	LF (POS)	Sent. Acc.	UP	UR	UF	Cat. Acc.	Cov.
C&C	88.06	86.43	87.24	85.25	35.67	93.88	92.13	93.00	94.16	99.06
C&C + VPC	87.90	86.34	87.11	85.11	35.73	93.80	92.13	92.96	94.06	99.06

There has been a very slight drop in performance, however this is a small penalty given the increased linguistic fidelity of the resulting parse, and the fact that the task has been made more difficult by the addition of new categories.

7. References

1. Stephen Clark and James R. Curran. Wide-Coverage Efficient Statistical Parsing with CCG and Log-Linear Models. *Computational Linguistics*, 33(4):493-552, 2007.
2. Julia Hockenmaier and Mark Steedman. CCGbank: A Corpus of CCG Derivations and Dependency Structures Extracted from the Penn Treebank. *Computational Linguistics*, 33(3):355-396, 2007.
3. Paul Kingsbury and Martha Palmer. PropBank: The Next Level of Treebank. In *Proceedings of Treebanks and Lexical Theories*, 2003.
4. Mark Steedman. *The Syntactic Process*. Massachusetts Institute of Technology, USA, 2000.

The University of Sydney