Resolving Relative Clause Attachment Ambiguities
using Machine Learning Techniques and WordNet Hierarchies

Advaith Siddharthan

Natural Language and Information Processing Group

Computer Laboratory, Cambridge

as372@cl.cam.ac.uk

Structure of Talk

- Background
- Problem Definition
- Methodology
- Results
- Work in Progress

Relative Clause Attachment

A Parsing Problem?

Non-restrictive relative clauses are increasingly being treated by parsers as sentential adjuncts, leaving the attachment decisions to anaphora resolution algorithms.

An Example

The board is dominated by *the heirs of **the late John T. Dorrance Jr.***, who controlled about 58% of Campbell's stock.

```
(T/leta_s
 (S/np_vp (NP/det_n The_AT (N1/n board_NN1))
  (V/be_ppart/- be+s_VBZ (V/pp dominate+ed_VVN
   (PP/p1 (P1/p_np by_II
    (NP/n2_name (NP/det_n the_AT (N1/n_of heir+s_NN2
     (PP/p1 (P1/p_np of_IO
      (NP/det_n the_AT (N1/ap_n1/- (AP/a1 (A1/a late_JJ))
       (N1/n John_NP1))))))) (NP/n1_posttit
        (N1/name+ T._NP1 (N1/n Dorrance_NP1)) Jr._NNSA1)))))))
 (Tacl/comma-e ,_, (S/whnp_vp who_PNQS
  (V/pp_pp control+ed_VVD
   (PP/p1 (P1/p_np about_II
    (NP/plu3 (N1/num2_nms (NP/num (N1/n 58_MC))
     (N1/n_of %_NNU (PP/p1 (P1/p_np of_IO
      (NP/n2_poss (NP/n1_name/- (N1/n Campbell_NP1)) 's+_POSS)
       (N1/n stock_NN1))))))))))
```

Relative Clause Attachment

An Anaphora Resolution Problem?

On the other hand, anaphora resolution algorithms based on:

- discourse oriented approaches (Hobbs, 1986; Reichman, 1978; Grosz, 1978)
- global focus (Grosz and Sidner, 1986)
- local focus (Carter, 1987; Webber, 1978; Sidner, 1981)

do not deal with relative clause attachment directly.

Motivation

Important for dis-embedding relative clauses, an important aspect of text simplification, an NLP task that:

- Restructures sentences, making text easier to read (or process)
- Preserves meaning and information content.
- Reduces grammatical (or lexical) complexity

Applications of Text Simplification:

- People with Language Disabilities like Aphasia (Carroll et al., 1998; Carroll et al., 1999)
- Preprocessing before parsing (Chandrasekar et al., 1996; Chandrasekar and Srinivas, 1997)
- Displaying text on Limited Channel Devices

Dis-Embedding Clauses

Previously published work on dis-embedding relative clauses make use of simplification rules that act on:

- Linear text (Chandrasekar et al., 1996)
- Some form of parse tree (Chandrasekar and Srinivas, 1997; Carroll et al., 1998)

A Hand-Crafted Rule:

$\texttt{W X:NP, Y:Rel\_Pr Z.} \longrightarrow \texttt{W X. X Z.}$

'[The pace] of [life] was slower in [those days],' says [51-year-old Cathy Tinsall] from [South London], *who had [five children]*.

$$\downarrow$$

'[The pace] of [life] was slower in [those days],' says [51-year-old Cathy Tinsall] from [South London]. [51-year-old Cathy Tinsall] from [South London] had [five children].

Defining the Problem

I focus on deciding local vs wide attachment when the noun phrase preceding the relative clause has the structure:

`NP1 Prep NP2`

‘[The pace] of [life] was slower in [those days],’ says [51-year-old Cathy Tinsall] from [South London], *who had [five children]*.

[The suicide note] included [lurid references] to [the economy] run under [the influence] of [Herr Pohl], *who might stop [a British government] from running [its own economic policy]*.

How Important is this Problem?

In the Penn Wall Street Journal Treebank (Marcus et al., 1993):

- 19% of *who* relative clauses
- 24% of *which* relative clauses

are preceded by complex noun phrases having the structure: `NP1 Prep NP2`

Lexicalization over Prepositions

Prep (Freq)	Prob	Prep (Freq)	Prob
under (1)	0	across (1)	0
outside (1)	0	around (1)	0
in (17)	0	at (8)	0.25
with (11)	0.55	on (9)	0.56
of (115)	0.67	from (14)	0.71
for (22)	0.83	to (15)	0.88
by (13)	1	against (5)	1
among (5)	1	like (2)	1
about (7)	1	but (1)	1

Table 1: Lexicalization over prepositions: Probability of local attachment for *who* clauses. Derived from the Penn WSJ Treebank.

Lexicalization over Prepositions

Prep (Freq)	Prob	Prep (Freq)	Prob
between (1)	0	near (1)	0
before (1)	0	in (70)	0.49
into (2)	0.50	on (25)	0.60
about (3)	0.67	of (216)	0.67
with (11)	0.73	as (4)	0.75
by (13)	0.76	for (47)	0.87
at (16)	0.93	from (22)	0.94
to (26)	0.96	over (3)	1
among (1)	1	against (2)	1
toward (1)	1	off (1)	1

Table 2: Lexicalization over prepositions: Probability of local attachment for *which* clauses. Derived from the Penn WSJ Treebank.

WordNet Classes

- It is useful to differentiate between *who* and *which* clauses
- Quirk (Quirk et al., 1985):

  The relative pronoun *who* is used to refer to something with personality and *which* to something without.

- In terms of the WordNet hierarchy (Miller et al., 1993), *who* can only refer to hyponyms of
  - *humans*
  - *groups (organizations)*
  - *animals*
- *which* cannot refer to *humans*

Machine Learning

0: Target (wide attachment)
1: Target (local attachment)
2: Restrictive Clause (defined by absence of comma)
3: `NP1` is a *person*
4: `NP1` is a *group*
5: `NP1` is an *animal*
6: `NP1` is a *possession*
7: `NP1` is an *entity*
8: `NP1` is an *act*
9: `NP1` is an *abstraction*
10: `NP1` has no WordNet class
11: `NP1` is a proper noun
12: `NP1` is a definite NP (presence of definite determiner)
13: `NP1` has no determiner
14 -17: Presence of top 4 prepositions
18: `Prep` favours local attachment
19: `Prep` favours wide attachment

Machine Learning

20: `NP2` is a *person*

21: `NP2` is a *group*

22: `NP2` is an *animal*

23: `NP2` is a *possession*

24: `NP2` is an *entity*

25: `NP2` is an *act*

26: `NP2` is an *abstraction*

27: `NP2` has no WordNet class

28: `NP2` is a proper noun

29: `NP2` is a definite NP (presence definite determiner)

30: `NP2` has no determiner

31: `Verb` selects for singular subject

32: `Verb` selects for plural subject

33: `NP1` is singular

34: `NP2` is singular

Machine Learning

Examples
An example is a list of the indexes of the features that are present in any particular sentence.

> **Ex:** 0,2,3,4,7,13,33,19,20,21,24,29,34:

Algorithm
SNoW machine learning package (Carlson et al., 1999) using the WINNOW algorithm.

Data
Parse trees from the Penn Wall Street Journal Treebank (Marcus et al., 1993)

Results (*who* clauses)

Data Set	Size	Baseline*	Winnow
Training Set	~200	66.5%	91.6%
Test Set	~50	66.5%	86.2%

* Baseline: Always attach locally.

Another baseline:

I converted the first 100 of these sentences to plain text and parsed them with the ANLT chart parser (Carroll, 1993; Briscoe and Carroll, 1995). An analysis of the parse trees gave:

- Recall = 62%
- Precision = 69.35%
- Local Attachment Baseline: Precision = 68%

Error Analysis

39% of the errors came from noun phrases like:

1. {hundreds|thousands|dozens} of investors
2. a {number|lot} of people
3. the percentage of Americans
4. the kind of guy

where `NP1` does not have a WordNet semantic class.

Example

> A Fannie Mae seminar this week promises to attract hundreds of investors, who can be expected to channel tens of billions of dollars into the market.

can be simplified to either of:

> A Fannie Mae seminar this week promises to attract hundreds of investors. These investors can be expected to channel tens of billions of dollars into the market.

and:

> A Fannie Mae seminar this week promises to attract hundreds of investors. These hundreds of investors can be expected to channel tens of billions of dollars into the market.

Error Analysis

21% of errors arose because the network didn't learn the rule that a restrictive clause cannot attach to a proper noun (in instances like *A former backup singer for Ms Midler who had...*). If this is enforced as a hard rule, the precision figures go up by almost 2%. It is possible that this rule would have been learnt by the network had it been presented with more training data.

Many of the other errors arose because the network had genuinely little information to go on; for example:

- *Some 3.8 million of the 5 million who will...*
- *A major piece of Hollywood manpower who has...*

Results (*which* clauses)

Data Set	Size	Baseline*	Winnow
Training Set	~400	69.7%	79.6%
Test Set	~50	69.7%	76.5%

* Baseline: Always attach locally.

Hand-Crafted Rules Again...

`V W:NP, X:Rel_Pr Y, Z.` $\longrightarrow$ `W Y. V W Z.`

John, who loves Mary, is sad.

$\downarrow$

John loves Mary. John is sad.

`V W:NP, X:REL_PRON Y.` $\longrightarrow$ `V W. W Y.`

I met John, who loves Mary.

$\downarrow$

I met John. John loves Mary.

Problems with these Rules

I tested these rules on 75 occurrences of non-restrictive *who* clauses in a corpus of featured articles from the Guardian Newspaper (January, 1991).

1. Wrong NP: 15 occurrences

   One man who is likely to reap the benefits is Vaino Heikkinen, aged 67, a farmer in Lieksa, 10km from the Soviet border, *who claims a Finnish record for shooting 36 bears since 1948*.

2. Wrong Determiner: 5 occurrences

   A former Ceremonial Officer, *who was at the heart of Whitehall's patronage machinery from 1966-77*, says there is a general review of the state of the honours list every five years or so.

3. Wrong Clause Boundary: 12 occurrences

   ...especially Debra Winger, *who plays a young American woman named Kit, based on his dead wife, Jane*.

References

Ted Briscoe and John Carroll. 1995. Developing and evaluating a probabilistic LR parser of part-of-speech and punctuation labels. In *Proceedings of the ACL/SIGPARSE 4th International Workshop on Parsing Technologies, Prague / Karlovy Vary, Czech Republic*, pages 48–58.

Andrew J. Carlson, Chad M. Cumby, Jeff L. Rosen, and Dan Roth. 1999. The SNoW learning architecture. Technical report, Tech. Report UIUCDCS-R-99-2101, UIUC Computer Science Department, May.

John Carroll, Guido Minnen, Yvonne Canning, Siobhan Devlin, and John Tait. 1998. Practical simplication of English newspaper text to assist aphasic readers. In *Proceedings of AAAI98 Workshop on Integrating Artificial Intelligence and Assistive Technology, Madison, Wisconsin*.

John Carroll, Guido Minnen, Darren Pearce, Yvonne Canning, Siobhan Devlin, and John Tait. 1999. Simplifying English text for language impaired readers. In *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics (EACL), Bergen, Norway.*

John Carroll. 1993. Practical unification-based parsing of natural language. Technical report, PhD thesis, University of Cambridge, Cambridge, UK, October 1993. Computer Laboratory Technical Report 314.

David Carter. 1987. *Interpreting Anaphors in Natural Language Texts*. Ellis Horwood, Chichester.

Raman Chandrasekar and Bangalore Srinivas. 1997. Automatic induction of rules for text simplification. *Knowledge-Based Systems*, 10:183–190.

Raman Chandrasekar, Christine Doran, and Bangalore Srinivas. 1996. Motivations and methods for text simplification. In *Proceedings of the Sixteenth International Conference on Computational Linguistics (COLING '96)*, Copenhagen, Denmark.

Barbara J. Grosz and Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.

Barbara J. Grosz. 1978. Discourse analysis. In D.E. Walker, editor, *Understanding Spoken Language*, pages 235–268. North-Holland, New York.

Jerry R. Hobbs. 1986. Resolving pronoun references. In Barbara J. Grosz, Karen Sparck-Jones, and Bonnie L. Webber, editors, *Readings in Natural Language Processing*, pages 339–352. Morgan Kaufmann, Los Altos, California.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large natural language corpus of English: The Penn Treebank. *Computational Linguistics*, 19:313–330.

George A. Miller, Richard Beckwith, Christiane D. Fellbaum, Derek Gross, and Katherine Miller. 1993. Five Papers on WordNet. Technical report, Princeton University, Princeton, N.J.

Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman.

Rachel Reichman. 1978. Conversational coherency. *Cognitive Science*, 2:283–327.

Candace L. Sidner. 1981. Focusing for interpretation of pronouns. *American Journal of Computational Linguistics*, 7(4):217–231.

Bonnie L. Webber. 1978. A formal approach to discourse anaphora. Technical Report 3761, Cambridge, Massachusetts.