Recognising Clauses

Benjamin Hachey
Department of Theoretical & Applied Linguistics
University of Edinburgh

Supervised by *Mirella Lapata* and *Alex Lascarides*

- Introduction
- Previous work
- Performance bounds
- SNoW predictor
- Future work

Introduction: What is a clause?

- Example (from Penn Treebank):

  (`M`
      (`M` *Mr. Stone thought*
          (`S` *the discipline was unfair* `S)M)`
      ;
      (`M` *he believed*
          (`S` *that*
              (`S` *his manager wanted to*
              get rid of him for personal
              reasons `S)S)M)`
  `. M)`

- Embedded structure!
  - Main Clause: `(M ... M)`
    - possibly coordinated
      `(M(M ... M) ... (M ... M)M)`
  - Subordinate Clause: `(S ... S)`
    - embedded w/in other clause
      `(M ... (S ... (S ... S)S)M)`

Introduction: What good are clauses?

- Phrasing in TTS synthesis (Ejerhed, 1988)
- Text alignment for machine translation (Papageorgiou, 1997)
- An intermediate step in full parsing (Ejerhed, 1988; Sang, 2002)

Previous work: Symbolic approaches

- Ejerhed (1988)
  - Regular expression grammar for clauses
  - “error rate:” 13%
- Leffa (1998)
  - Three phases
    1. identify clause initiators
    2. identify clause terminators
    3. identify full clauses
  - “clauses correctly classified:” 95%

Previous work: CoNLL-2001

- Re-caste the problem as one of tagging

- Three phases as in Leffa (1998)

Phase 1	Consumers NNS B-NP S may MD B-VP X want VB I-VP X to TO I-VP X move VB I-VP X their PRP$ B-NP X telephones NNS I-NP X . . O X
Phase 2	Consumers NNS B-NP S X may MD B-VP X X want VB I-VP X X to TO I-VP X X move VB I-VP X X their PRP$ B-NP X X telephones NNS I-NP X X . . O X E
Phase 3	Consumers NNS B-NP S X (S* may MD B-VP X X * want VB I-VP X X * to TO I-VP X X * move VB I-VP X X * their PRP$ B-NP X X * telephones NNS I-NP X X * . . O X E *S)

Previous work: CoNLL-2001 Results

$$F_{\beta=1} = (\beta^2+1)*precision*recall/(\beta^2*precision+recall)$$

where

$$precision = correctpredictions/selectedlabels$$
$$recall = correctpredictions/targetlabels$$

SYSTEM	STARTS	ENDS	FULL
Adaboost (Carreras & Már.)	91.72	89.22	78.63
HMMs (Molina & Pla)	86.48	78.38	66.79
MBL (Tjong Kim Sang)	87.49	82.28	66.67
ALLiS (Déjean)	87.43	65.47	62.77
Connectionist (Hammerton)	-	-	50.42
Adaboost (Patrick & Goyal)	50.34	60.26	18.49
baseline	53.34	65.34	47.71

Previous work: CoNLL-2001 disagreement

	TESTB1	TESTB2	TESTB3
Agree	89.30%	92.24%	78.48%
on incorrect label	0.74%	0.49%	1.59%
Disagree	10.70%	7.76%	21.52%
Carreras & Már. correct	89.17%	82.72%	77.16%
Molina & Pla correct	80.91%	60.09%	64.85%
Tjong Kim Sang correct	83.67%	68.55%	67.78%
Déjean correct	80.91%	50.02%	66.47%
Hammerton correct	–	–	56.55%
Patrick & Goyal correct	80.14%	65.53%	66.90%

Previous work: Shortcomings of CoNLL-2001 shared task

- ML methods are only motivated if they have some advantage over symbolic methods:
  - Better performance
  - Less labor intensive
  - Quicker to produce
  - No need to generalise outside of data domain
  - Data available
  - Complexity (memory and run time) not an issue
- Baseline doesn't say much about relative capability/merits of symbolic methods
- Does not distinguish "main" and "subordinate"

Previous work: Why not full parsing?

- Each tag a classification problem
- Less time consuming to design and run
- Portable to other languages

Data preparation

- Corpus: newswire (subset of the Penn Treebank)
- Encoding
  - 1 line per word:
    WORD POS CHUNK (S) (E) (F)
  - chunklink script (Buccholz)
- Main vs. subordinate
  - Main
    * All top-level full strings
    * Clauses coordinated under top-level
  - All else subordinate
- Checked with random sample of 500 sentences

Performance bounds

- New symbolic baseline
  - Following Leffa (1998)
- Upper bounds
  - Full parse
  - Human performance

Performance bounds: Baseline version 1

- Does not distinguish “main” and “subordinate”

- Algorithm:

  for each word
    if the word is a (clause-level) conjunction
      or subordinator
      then mark as clause initiator
    if the word is a subject-less verb
      then mark as clause initiator
    if the word is the subject of a verb
      then mark as clause initiator
  for each word
    if the word is followed by a clause initiator and
      (the phrase preceding it contains a verb)
      then mark as clause terminator
  for each clause initiator
    if verb phrase and clause terminator follow
      then mark as a clause

- Employed to find best clues for identifying clauses

Performance bounds: Baseline version 1 scores

CLUES	ACCURACY	PRECISION	RECALL	$F_{\beta=1}$
STARTS				
Baseline	92.81	98.44	36.58	53.34
Relativiser	93.57	97.20	43.98	60.56
After relat'r	93.66	97.76	44.59	61.24
B-SBAR	93.45	97.76	42.70	59.43
B-VP				
VBG	92.95	92.92	40.27	56.19
TO	93.27	95.00	42.27	58.51
B-NP...B-VP	91.23	60.37	63.91	62.09
All above	93.29	64.91	87.57	74.56
ENDS				
Baseline	95.64	98.44	48.90	65.34
Prec Clause	92.25	53.95	52.73	53.34
End-1	95.92	77.99	71.73	74.73
FULL				
Baseline		98.44	31.48	47.71
New S & Es	86.07	48.35	55.10	51.51

Performance bounds: Baseline version 2

- *Does* distinguish "main" and "subordinate"
  - Clues are those found in version 1
  - Whole sentences given "main" label
  - Everything else given "subordinate" label
    - does not recover coordinated main clauses
- Performance

CLUES	ACCURACY	PRECISION	RECALL	$F_{\beta=1}$
Full				
M v. S	79.24	48.35	59.29	53.27

Performance Bounds: Upper

- Voting
  - Combine CoNLL-2001 system results
  - Must have balanced brackets
  - Problems correlated
    - predicting multiple starts is difficult
  - Results

SYSTEM	FULL
Carreras & Már.	78.63
Molina & Pla	66.79
Tjong Kim Sang	66.67
Déjean	62.77
Hammerton	50.42
Patrick & Goyal	18.49
baseline	47.71
vote	67.50

Performance bounds: Upper

- Clause identication by state of the art full parser (Collins, 1999):

TYPE	PRECISION	RECALL	$F_{\beta=1}$
S	90.96	91.21	91.08
SBAR	88.87	87.81	88.34
SINV	88.46	83.64	85.98
SQ	66.67	53.33	58.26
SBARQ	88.89	66.67	76.19
All	90.37	90.16	90.26

- Human performance?
  - Give definition of a clause to study with practice sentences.
  - Evaluate performance on a sample of the data.

Sparse Network of Winnows

(Carlson et al., 1999)

- Advantages
  - Winnow successful in CoNLL-2000 chunking task (Zhang et al., 2001)
  - Good for problems with many irrelevant attributes!
- Limits
  - Regularised version of Winnow used in CoNLL-2000

SNoW predictor: Experiment

- Experiment 1
  - For all combinations of available features (eg. w, p, c, wp, wc, pc, wpc)
  - For window sizes 0..4
- Results ($F_{\beta=1}$ score):

FEATS	0	1	2	3	4
w	0.00	65.69	70.36	71.30	70.14
p	0.00	64.78	65.38	63.35	64.14
c	0.00	65.06	61.54	63.87	63.05
wp	0.00	73.96	76.03	74.54	75.37
wc	0.00	70.45	71.79	73.81	74.12
pc	0.00	73.67	74.53	73.05	74.16
wpc	0.00	76.94	77.96	78.17	76.84

SNoW predictor: More sophisticated features

(Carreras & Márquez)

- Sentence patterns: occurence of following items to both sides of current position
  - {“ ” ( ) , . : CC}
  - WORD=that + its POS
  - relative pronouns
  - VP chunks
- Sentence features: Count and existence variables for the following items to both sides of the current position
  - verb phrase, and wh-pronoun chunks
  - puncutation parts of speech
  - WORD=that
  - S- and E- points when available

Future Work

- Maximum Entropy
- Support Vector Machines
- Improve symbolic approach
- Estimate human performance

Maximum entropy

(Berger et al., 1996)

- Has shown good results for full parsing (Ratnaparkhi, 1999)

- Makes no independence assumptions

- Can use any features considered relevant

  – Allows linguistic knowledge to be incorporated

Support Vector Machines

(Vapnik, 1998)

- SVMs successful in CoNLL-2000 chunking task (Kudoh & Matsumoto, 2001)
- Finds optimal hyperplane that separates data into two classes
- Robust to non-linearly-separable data
  - Allows some misclassifications
- Can handle non-linear classification

Other Future Work

- Estimate human performance
- Improve symbolic approach
  - not tag fragments
  - better end point identification
  - better full clause identification
  - multiple starts