The CoNLL-2008 Shared Task on Joint Parsing of Syntactic and Semantic Dependencies

http://www.yr-bcn.es/conll2008

Mihai Surdeanu[†,⋆], Richard Johansson[‡], Adam Meyers[⋄], Lluís Màrquez[††], and Joakim Nivre[‡‡,⋆⋆]

†: Barcelona Media Innovation Center, ⋆: Yahoo! Research Barcelona, ‡: Lund University, ⋄: New York University, ††: Technical University of Catalonia, ‡‡: Växjö University, ⋆⋆: Uppsala University

August 13, 2008

Outline

Introduction

Task Definition

Data

Submissions and Results

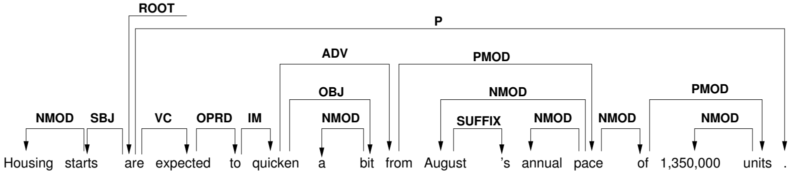
Approaches and Analysis

Concluding Remarks

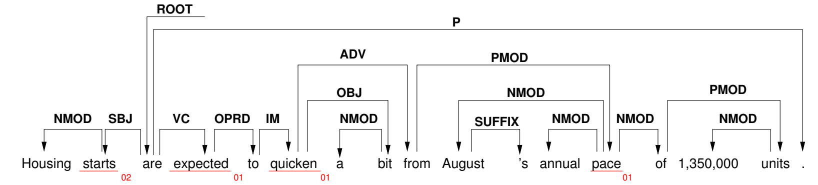
Example

Housing starts are expected to quicken a bit from August ’s annual pace of 1,350,000 units .

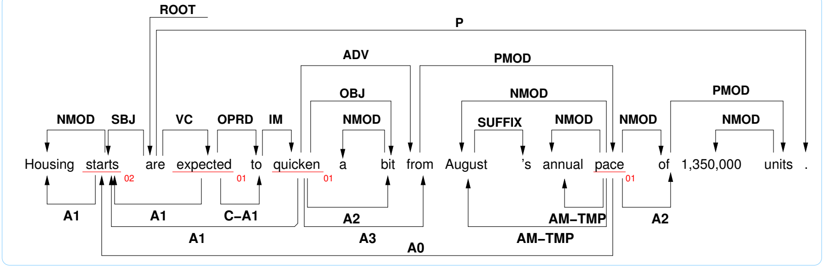
Example


ROOT
P
ADV
PMOD
OBJ
NMOD
PMOD
NMOD
SBJ
VC
OPRD
IM
NMOD
SUFFIX
NMOD
NMOD
NMOD
Housing starts are expected to quicken a bit from August 's annual pace of 1,350,000 units .

Example


ROOT
P
ADV
PMOD
OBJ
NMOD
PMOD
NMOD
SBJ
VC
OPRD
IM
NMOD
SUFFIX
NMOD
NMOD
NMOD
Housing
starts
02
are
expected
01
to
quicken
01
a
bit
from
August
's
annual
pace
01
of
1,350,000
units
.

Example


ROOT
P
ADV
PMOD
OBJ
NMOD
PMOD
NMOD
SBJ
VC
OPRD
IM
NMOD
SUFFIX
NMOD
NMOD
NMOD
Housing starts are expected to quicken a bit from August 's annual pace of 1,350,000 units .
02
01
01
01
A1
A1
C–A1
A2
AM–TMP
A2
A1
A3
AM–TMP
A0

Objectives

Novelties

- SRL using a dependency-based representation
- SRL for both verbal and nominal predicates
- More complex syntactic dependencies
- Merged representation for syntax and semantics

Why?

- Research questions: Is the dependency-based representation better for SRL than the constituent-based formalism? Is the merged representation more helpful than the individual ones?
- Ease adoption of this NLP technology: linear time processing possible, better fit for many applications

Outline

Introduction

Task Definition

Data

Submissions and Results

Approaches and Analysis

Concluding Remarks

Two Challenges

- Closed challenge – systems have to be built strictly with information contained in the given training corpus and the given PropBank and NomBank lexical frames
  - Fair environment to compare participating systems

- Open Challenge – systems can be developed making use of any kind of external tools and resources. The output of several state-of-the-art processors were provided by the organizers
  - Does other semantic information help?
  - The output of a parser was provided → groups could participate only in SRL

Data Format: General Rules

- The files contain sentences separated by a blank line.
- A sentence consists of one or more tokens and the information for each token is represented on a separate line.
- A token consists of at least 11 fields. The fields are separated by one or more whitespace characters (spaces or tabs). Whitespace characters are not allowed within fields.

Data Format: Closed Challenge

Number	Name	Description
1	ID	Token counter, starting at 1 for each new sentence.
2	FORM	Unsplit word form or punctuation symbol.
3	LEMMA	Predicted lemma of FORM.
4	GPOS	Gold part-of-speech tag from the Treebank (empty at test time).
5	PPOS	Predicted POS tag.
6	SPLIT_FORM	Tokens split at hyphens and slashes.
7	SPLIT_LEMMA	Predicted lemma of SPLIT_FORM.
8	PPOSS	Predicted POS tags of the split forms.
9	HEAD	Syntactic head of the current token, which is either a value of ID or zero (0).
10	DEPREL	Syntactic dependency relation to the HEAD.
11	PRED	Rolesets of the semantic predicates in this sentence.
12+	ARG	Columns with argument labels for each semantic predicate following textual order.

Data Format: Open Challenge

Extra information provided:

Number	Name	Description
1	CONLL2003	Named entity labels using the tag set from the CoNLL-2003 shared task.
2	BBN	NE labels using the tag set from the BBN Wall Street Journal Entity Corpus.
3	WNSS	WordNet super senses.
4	MALT_HEAD	Head of the syntactic dependencies generated by MaltParser.
5	MALT_DEPREL	Label of syntactic dependencies generated by MaltParser.

Official Evaluation Measures

- Syntactic dependencies – Labeled Attachment Score (LAS): percentage of tokens with the correct HEAD and DEPREL values
- Semantic dependencies – Labeled F_1
  - One dependency from every predicate to each of its arguments, labeled with the argument label
  - One dependency from each predicate to a virtual ROOT node, labeled with the predicate sense

Correct	`verb.01:   ARG0, ARG1, ARGM-TMP`
Predicted	`verb.02:   ARG0, ARGM-LOC`

$LP_{sem} = 1/3$, $LR_{sem} = 1/4$

- Global measure – macro average between the two tasks:

$$LMP = W_{sem} * LP_{sem} + (1 - W_{sem}) * LAS$$

$$LMR = W_{sem} * LR_{sem} + (1 - W_{sem}) * LAS$$

Additional Evaluation Measures

- ExactMatch – percentage of sentences that are completely correct
  - Should award systems that performed joint learning or optimization for all subtasks
- Perfect Proposition F_1 – harmonic mean of precision and recall for complete semantic frames, or propositions
  - Measures the capacity to recognize entire frames rather than individual semantic dependencies
- Ratio between labeled F_1 for semantic dependencies and LAS
  - Estimates the performance on the semantic subtask independent of the syntactic parser

Outline

Introduction

Task Definition

Data

Submissions and Results

Approaches and Analysis

Concluding Remarks

Input Corpora

- **Penn Treebank 3** – hand-coded parses of Wall Street Journal and Brown corpora
- **BBN Pronoun Coreference and Entity Corpus** – NE annotations of the Wall Street Journal; extended by us to include a subset of the Brown corpus
  - We only use NE boundaries to derive `NAME` dependencies.
- **Proposition Bank I** – semantic arguments of the main Treebank verbs, other than *be*. We started from the version used for CoNLL-2005:
  - Added the concept of continuation arguments, e.g.: [*This sentence*]$_{\mathrm{A1}}$, *Mary claims,* [*is self-referential*]$_{\mathrm{C\text{-}A1}}$
  - Empty fillers are not annotated.
- **NomBank** – semantic arguments for nominal predicates in Treebank
  - Has support chains for long-distance dependencies, e.g., *took dozens of* in [*Mary*]$_{\mathrm{A1}}$ *took dozens of* [*walks*]$_{\mathrm{PRED}}$

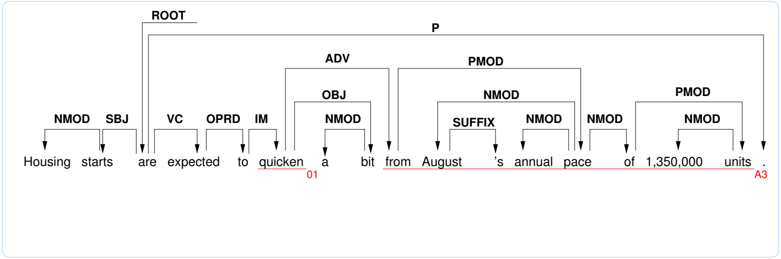
Conversion to Dependencies: Syntactic Dependencies

- Assign a syntactic *head* to every constituent
- Links from traces in the Treebank may result in nonprojective dependencies
- Grammatical functions (SBJ, LOC, TMP, . . .) from the Treebank
- Rules to assign grammatical functions to remaining dependencies

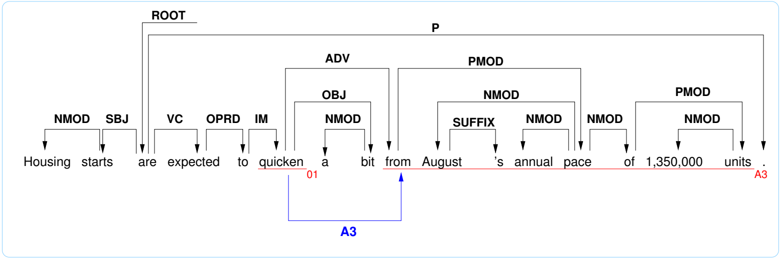
Conversion to Dependencies: Semantic Dependencies

- Necessary for argument constituents
- Input:
  - Argument boundaries (from PropBank and NomBank)
  - Syntactic dependencies (from the previous process)
- Conversion heuristic:
  - The head of a semantic argument is assigned to the token inside the argument boundaries whose head is a token outside the argument boundaries
  - Handles over 99% of the argument constituents

Conversion to Dependencies: Semantic Dependencies – Example

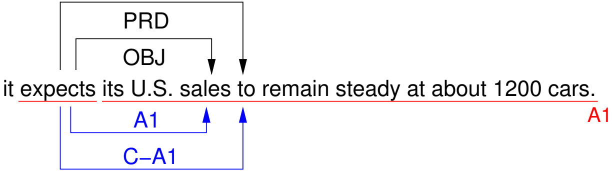

ROOT
P
ADV
PMOD
OBJ
NMOD
PMOD
NMOD
SBJ
VC
OPRD
IM
NMOD
SUFFIX
NMOD
NMOD
NMOD
Housing starts are expected to quicken a bit from August 's annual pace of 1,350,000 units .
01
A3

Conversion to Dependencies: Semantic Dependencies – Example



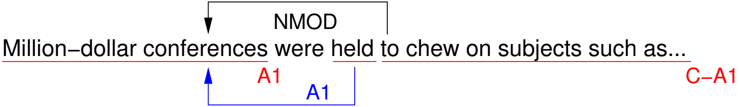
Conversion to Dependencies: Semantic Dependencies – Exceptions

- Arguments with several syntactic heads



it expects its U.S. sales to remain steady at about 1200 cars.

- Merging discontinuous arguments



Million–dollar conferences were held to chew on subjects such as...

- Empty categories, annotation disagreements between Treebank and Nombank, support chains

Outline

Introduction

Task Definition

Data

Submissions and Results

Approaches and Analysis

Concluding Remarks

Participants

- 55 groups signed up for the task: 23 Europe, 17 Asia, 15 North America
- 22 actually submitted results – 40% completion...
- 5 groups submitted post-evaluation improvements (posted on the website)

Closed Challenge: Complete Task

	Labeled Macro F_1 (complete task)		
	WSJ+Brown	WSJ	Brown
johansson	84.86 (1)	85.95	75.95
che	82.66 (2)	83.78	73.57
ciaramita	82.06 (3)	83.25	72.46
zhao	81.44 (4)	82.62	71.78
yuret	79.84 (5)	80.97	70.55
samuelsson	79.79 (6)	80.92	70.49
zhang	79.32 (7)	80.41	70.48
henderson	79.11 (8)	80.19	70.34
watanabe	79.10 (9)	80.30	69.29
morante	78.43 (10)	79.52	69.55
li	78.35 (11)	79.38	70.01
baldridge	77.49 (12)	78.57	68.53
chen	77.00 (13)	77.95	69.23
lee	76.90 (14)	77.96	68.34
sun	76.28 (15)	77.10	69.58
choi	71.23 (16)	72.22	63.44
trandabat	63.45 (17)	64.21	57.41
lluis	63.29 (18)	63.74	59.65
neumann	19.93 (19)	20.13	18.14

Closed Challenge: the Two Subtasks

	Labeled Attachment Score (syntactic dependencies)			Labeled F_1 (semantic dependencies)		
	WSJ+Brown	WSJ	Brown	WSJ+Brown	WSJ	Brown
johansson	89.32 (1)	90.13	82.81	80.37 (1)	81.75	69.06
che	86.75 (5)	87.51	80.73	78.52 (2)	80.00	66.37
ciaramita	86.60 (11)	87.47	79.67	77.50 (3)	79.00	65.24
zhao	86.66 (8)	87.52	79.83	76.16 (4)	77.67	63.69
yuret	86.62 (10)	87.39	80.46	73.06 (5)	74.54	60.62
samuelsson	86.63 (9)	87.36	80.77	72.94 (6)	74.47	60.18
zhang	87.32 (2)	88.14	80.80	71.31 (7)	72.67	60.16
henderson	86.91 (4)	87.78	80.01	70.97 (8)	72.26	60.38
watanabe	87.18 (3)	88.06	80.17	70.84 (9)	72.37	58.21
morante	86.07 (12)	86.88	79.58	70.51 (10)	71.88	59.23
li	86.69 (6)	87.42	80.80	69.95 (11)	71.27	59.17
baldridge	86.67 (7)	87.42	80.64	67.92 (14)	69.35	55.95
chen	84.47 (16)	85.20	78.58	69.45 (12)	70.62	59.81
lee	84.82 (15)	85.69	77.83	68.71 (13)	69.95	58.63
sun	85.75 (13)	86.37	80.75	66.61 (15)	67.62	58.26
choi	77.56 (17)	78.58	69.46	64.78 (16)	65.72	57.4
trandabat	85.21 (14)	85.96	79.24	40.63 (17)	41.36	34.75
lluis	71.95 (18)	72.30	69.14	54.52 (18)	55.09	49.95
neumann	16.25 (19)	16.22	16.47	22.36 (19)	22.86	17.94

Open Challenge: Complete Task

	Labeled Macro F_1 (complete task)		
	WSJ+Brown	WSJ	Brown
zhang	79.61 (1)	80.61	71.45
li	77.84 (2)	78.87	69.51
wang	76.19 (3)	78.39	59.89
vickrey	–	–	–
riedel	–	–	–

Open Challenge: the Two Subtasks

	Labeled Attachment Score (syntactic dependencies)			Labeled F_1 (semantic dependencies)		
	WSJ+Brown	WSJ	Brown	WSJ+Brown	WSJ	Brown
vickrey	–	–	–	76.17 (1)	77.38	66.23
riedel	–	–	–	74.59 (2)	75.72	65.38
zhang	87.32 (1)	88.14	80.80	71.89 (3)	73.08	62.11
li	86.69 (2)	87.42	80.80	68.99 (4)	70.32	58.22
wang	84.56 (3)	85.50	77.06	67.12 (5)	70.41	42.67

Outline

Introduction

Task Definition

Data

Submissions and Results

Approaches and Analysis

Concluding Remarks

Summary of System Architectures

- Overall architectures:
  - Mostly pipeline
  - Only five systems combined the syntactic and semantic subtasks: Johansson and Nugues, Henderson et al., Samuelsson et al., Lluís and Màrquez, Sun et al.
- Parsing approaches:
  - Most transition-based + greedy inference, or graph-based + MST inference
  - Strategies to mitigate errors: voting (2), stacking (2), meta-learning (1), second order model (1)
- SRL approaches:
  - Most token-by-token classification + greedy inference. Exceptions: Riedel and Meza-Ruiz + most joint systems.
  - Strategies to mitigate errors: voting (4)

Exact Match and Perfect Propositions
Closed Challenge

	Exact Match (complete task)			Perfect Proposition F_1 (semantic dependencies)		
closed	WSJ+Brown	WSJ	Brown	WSJ+Brown	WSJ	Brown
johansson	12.46 (1)	12.46	12.68	54.12 (1)	56.12	36.90
che	10.37 (2)	10.21	11.50	48.05 (2)	50.15	30.90
ciaramita	9.27 (3)	9.04	10.80	46.05 (3)	48.05	28.61
zhao	9.20 (4)	9.00	10.56	43.19 (4)	45.23	26.14
henderson	8.11 (5)	7.75	10.33	39.24 (5)	40.64	27.51
watanabe	7.79 (6)	7.54	9.39	36.44 (6)	38.09	22.72
yuret	7.65 (7)	7.33	9.62	34.61 (9)	36.13	21.78
zhang	7.40 (8)	7.46	7.28	34.96 (8)	36.25	24.22
li	7.12 (9)	6.71	9.62	32.08 (10)	33.45	20.62
samuelsson	6.94 (10)	6.62	8.92	35.20 (7)	36.96	20.22
chen	6.83 (11)	6.46	9.15	31.02 (12)	32.08	22.14
lee	6.69 (12)	6.29	9.15	31.40 (11)	32.52	22.18
morante	6.44 (13)	6.04	8.92	30.41 (14)	31.97	17.49
sun	5.38 (14)	4.96	7.98	30.43 (13)	31.51	21.40
baldridge	5.24 (15)	4.92	7.28	25.35 (15)	26.57	15.26
choi	3.33 (16)	3.50	2.58	24.77 (16)	25.71	17.37
trandabat	3.26 (17)	3.08	4.46	6.59 (18)	6.81	4.76
lluis	2.55 (18)	1.96	6.10	16.07 (17)	16.46	13.00
neumann	0.11 (19)	0.12	0.23	0.30 (19)	0.31	0.20

Exact Match and Perfect Propositions Open Challenge

	Exact Match (complete task)			Perfect Proposition F_1 (semantic dependencies)		
open	WSJ+Brown	WSJ	Brown	WSJ+Brown	WSJ	Brown
vickrey	–	–	–	44.94 (1)	46.68	30.28
riedel	–	–	–	42.77 (2)	44.18	31.15
zhang	8.14 (1)	8.04	8.92	35.46 (3)	36.74	24.84
li	6.90 (2)	6.46	9.62	29.91 (4)	31.30	18.41
wang	5.17 (3)	5.12	5.63	18.63 (5)	20.31	7.09

Nonprojectivity

System	All	*wh* Movement	Split Clauses	Split NPs
lee	46.26	50.30	64.84	20.69
nugues	46.15	58.96	59.26	11.32
titov	42.32	50.56	48.71	0
choi	25.43	49.49	45.47	8.72
samuelsson	24.47	38.15	0	9.83
zhang	13.39	5.71	12.33	7.3

PropBank versus NomBank
Closed Challenge

	Labeled F_1 (verbal predicates)			Labeled F_1 (nominal predicates)		
closed	WSJ+Brown	WSJ	Brown	WSJ+Brown	WSJ	Brown
johansson	84.45 (1)	86.37	71.87	74.32 (2)	75.42	60.13
che	80.46 (2)	82.17	69.33	75.18 (1)	76.64	56.87
ciaramita	80.15 (3)	82.09	67.62	73.17 (4)	74.42	57.69
zhao	77.67 (4)	79.40	66.38	73.28 (3)	74.69	54.81
samuelsson	76.17 (5)	78.03	64.00	68.13 (7)	69.58	49.24
yuret	75.91 (6)	77.88	63.02	68.81 (5)	69.98	53.58
zhang	74.82 (7)	76.62	63.15	65.61 (11)	66.82	50.18
li	74.36 (8)	76.14	62.92	62.61 (14)	63.76	47.09
henderson	73.80 (9)	75.40	63.36	66.26 (10)	67.44	50.73
watanabe	73.06 (10)	75.02	60.34	67.15 (8)	68.37	50.92
sun	72.97 (11)	74.45	63.50	58.68 (15)	59.73	45.75
morante	72.81 (12)	74.36	62.72	66.50 (9)	67.92	47.97
lee	72.34 (13)	74.15	60.49	62.83 (13)	63.66	52.18
chen	72.02 (14)	73.49	62.46	65.02 (12)	66.14	50.48
choi	70.00 (15)	71.28	61.71	56.16 (16)	57.19	44.05
baldridge	67.02 (16)	68.64	56.50	68.57 (6)	69.78	52.96
lluis	62.42 (17)	63.49	55.49	42.15 (17)	42.81	34.22
trandabat	42.88 (18)	43.79	37.06	37.14 (18)	37.89	27.50
neumann	22.87 (19)	23.53	18.24	21.7 (19)	22.04	17.14

PropBank versus NomBank
Open Challenge

	Labeled F_1 (verbal predicates)			Labeled F_1 (nominal predicates)		
open	WSJ+Brown	WSJ	Brown	WSJ+Brown	WSJ	Brown
vickrey	78.41 (1)	79.75	69.57	71.86 (1)	73.29	53.25
riedel	77.13 (2)	78.72	66.75	70.25 (2)	71.03	60.17
zhang	75.00 (3)	76.62	64.44	66.76 (3)	67.79	53.76
li	73.74 (4)	75.57	62.05	61.24 (5)	62.38	46.36
wang	67.50 (5)	70.34	49.72	66.53 (4)	69.83	28.96

Predicate Identification and Classification Closed Challenge

	Labeled F_1		
	WSJ+Brown	WSJ	Brown
johansson	85.40 (1)	86.75	74.19
che	85.31 (2)	86.82	73.00
ciaramita	83.46 (5)	84.86	71.98
zhao	78.26 (12)	79.76	65.72
yuret	83.20 (6)	84.87	69.14
samuelsson	81.28 (8)	82.89	67.48
zhang	82.65 (7)	84.19	69.83
henderson	79.60 (10)	81.14	66.69
watanabe	77.19 (14)	79.02	62.10
morante	77.21 (13)	78.28	68.34
li	83.80 (4)	85.26	71.67
baldridge	84.32 (3)	85.94	70.96
chen	78.45 (11)	79.65	68.41
lee	80.12 (9)	81.51	68.69
sun	74.53 (16)	75.42	67.05
choi	76.35 (15)	77.31	68.77
trandabat	66.33 (17)	67.59	55.95
lluis	70.60 (18)	71.43	64.09
neumann	55.30 (19)	56.65	43.69

Predicate Identification and Classification
Open Challenge

	Labeled F_1		
	WSJ+Brown	WSJ	Brown
vickrey	80.81 (4)	82.15	69.51
riedel	82.12 (3)	83.22	73.03
zhang	83.24 (2)	84.56	72.33
li	83.80 (1)	85.26	71.67

Outline

Introduction

Task Definition

Data

Submissions and Results

Approaches and Analysis

Concluding Remarks

Concluding Remarks

- Shared task dedicated to the joint parsing of syntactic and semantic dependencies
- Largest initial interest of all shared tasks (55 groups) → interesting and important problem
- One of the lowest completion rates (40%) → complex problem
- Proposal for future shared tasks:
  - Multiple languages
  - Larger out-of-domain corpora
  - How to minimize startup effort?

Acknowledgments: Many Thanks!

- Helped with data generation: Jesús Giménez (POS tags), Massimiliano Ciaramita (open challenge data)
- Helped the organization: Paola Merlo and James Henderson (ExactMatch score), Sebastian Riedel (visualization software), Hai Zhao (F_1 ratio score), Carlos Castillo (website)
- Organizers of the previous four shared tasks: Sabine Buchholz, Xavier Carreras, Ryan McDonald, Amit Dubey, Johan Hall, Yuval Krymolowski, Sandra Kübler, Erwin Marsi, Jens Nilsson, Sebastian Riedel, and Deniz Yuret

Thank you! Questions, feedback?

Reminder: please attend the poster session after the oral presentations!