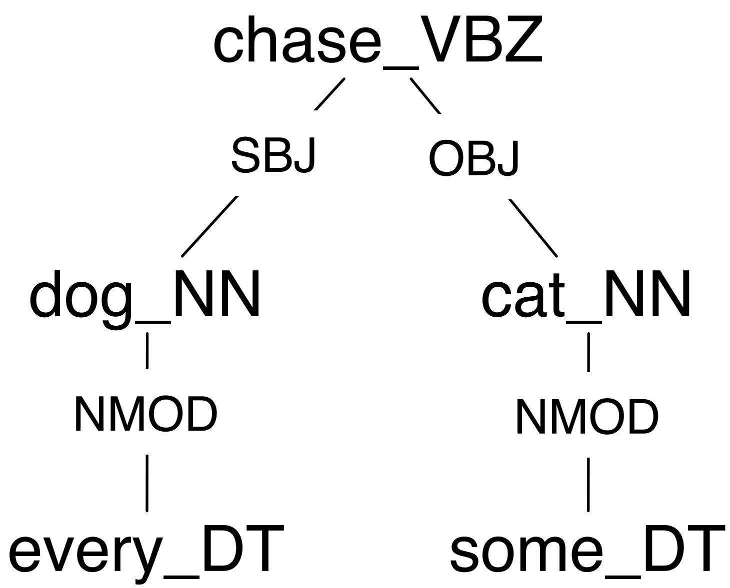
Sentence Realization with Unlexicalized Tree Linearization Grammars

Rui Wang
(Joint work with Yi Zhang)
DFKI GmbH, Germany

The Task

- Parsing
  - (Ordered) text → structure: tokenization, POS tagging, constituent/dependency parsing, …
- Generation
  - (Unordered) structure → Text: content planning, lexical choices, surface realization, …
- Syntactic dependency tree → Linearization of tokens
  - Generation Challenge 2011 Surface Realization Shared Task (Belz et al., 2011)

An Example



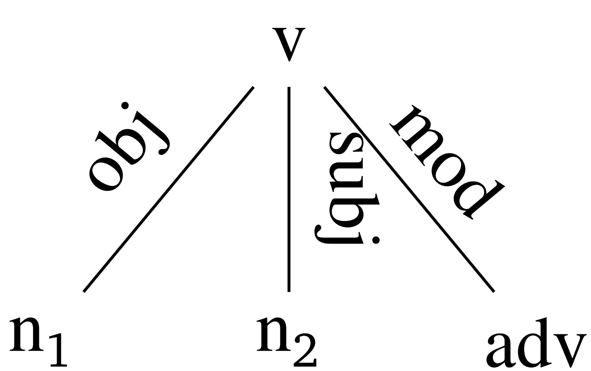
- every dog chase some cat
- Every dog chases some cat .

Outline

- The basic model
- Problems
- Extensions
- Related/Future Work

The Basic Model

- Unlexicalized Tree Linearization Grammar
  - A set of linearization rules
  - Rule: (Local) configuration → Linear order

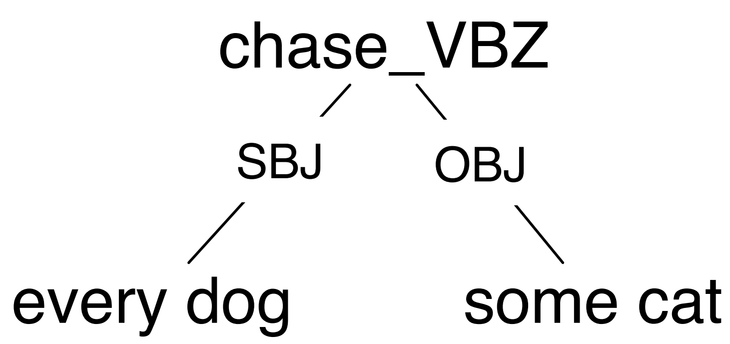


$$\Rightarrow \langle n_2, adv, v, n_1 \rangle$$
$$\Rightarrow \langle n_2, v, n_1, adv \rangle$$

Linearization

- For each configuration
  – Apply the linearization rule

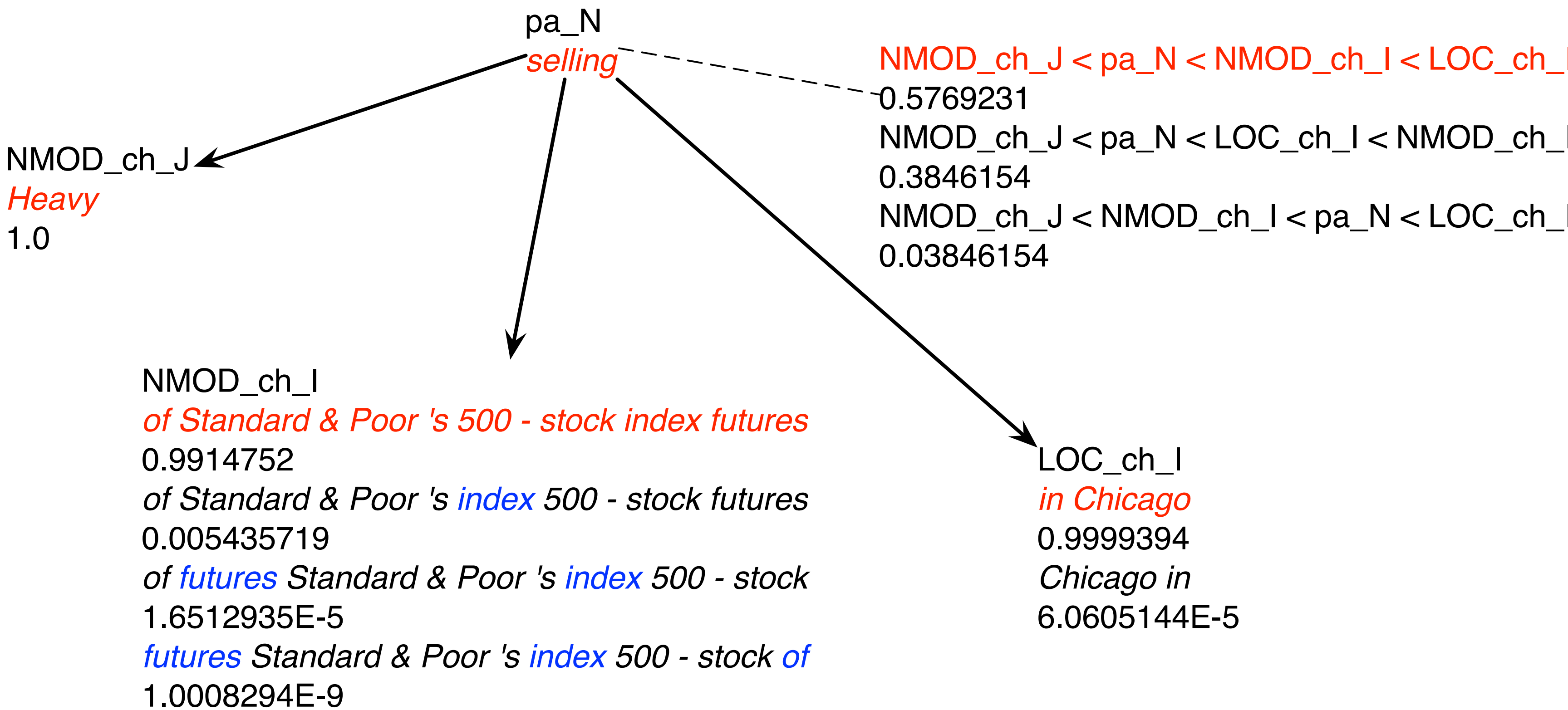
- For each subtree
  – The linearization is a continuous string



deepandance

N-Best Linearization

- *Heavy selling of Standard & Poor 's 500 - stock index futures in Chicago relentlessly beat stocks downward .*



The Probabilistic Model

- For each LHS

$$\mathbf{Pr}: \mathscr{L} \rightarrow [0,1] s.t. \forall \mathscr{C} \in \mathscr{C}, \sum_{\forall \mathscr{L} \in \mathscr{L}, LHS(\mathscr{L}) = \mathscr{C}} \Pr(\mathscr{L}) = 1$$

$$\Pr(\mathscr{L}) = \frac{Freq(\mathscr{L})}{Freq(LHS(\mathscr{L}))}$$

- For the subtree

$$P(\mathscr{L}_n) = \Pr(rule(\mathscr{L}_n)) * \prod_{d \in dependents(n)} P(\mathscr{L}_d)$$

deependance

Assumptions

- Connected
- Single-headed
- Projective

deependance

Evaluation

- Data: dependency treebanks from the CoNLL-shared task 2009 (Hajic et al., 2009)
- Evaluation metrics: BLEU (Papineni et al., 2002)

Sentence Coverage	451 / 1334 (33.8%)
Configuration Coverage	15843 / 17282 (91.7%)
1-best	92.65
Upper bound (1000)	96.31

Problems

- Out of grammar
  - Coarse-grained rules
  - Backup strategy
- In-grammar performance
  - N-gram-based Smoothing

deependance

Out of Grammar

- Coarse-grained rules (POS → CPOS)
- Backup models (Pair-wised ranking)

Models		POS	CPOS
Coverage	Sent. (1334)	451 (33.8%)	711 (53.3%)
	Conf. (17282)	15843 (91.7%)	16423 (95.0%)
Covered	1-best	**92.65**	90.64
	upper (1000)	**96.31**	95.31
Overall	1-best	81.63	**83.28**
	upper (1000)	84.08	**87.13**

Examples

- Gold: ***[“ The market is overvalued , not cheap , ” says]*** *Alan Gaines of the New York money - management firm Gaines Berland .*
- System: *Alan Gaines of the New York money - management firm Gaines Berland* ***[says , “ The market is overvalued , not cheap . ”]***

- Gold: *... than many taxpayers working at the same kinds of jobs and* ***[perhaps]*** *supporting families .*
- System: *... than many taxpayers* ***[perhaps]*** *working at the same kinds of jobs and supporting families .*

Examples (cont.)

- Gold: *... to set **[aside]** provisions covering all its C$ 1.17 billion in non - Mexican LDC debt .*
- System: *... to set provisions covering all C$ its 1.17 billion in non - Mexican LDC debt **[aside]** .*

- Gold: *Good service programs require recruitment , screening , training and supervision – **[all of high quality]** .*
- System: ***[all of high quality]** – Good service programs require recruitment , screening , training and supervision .*

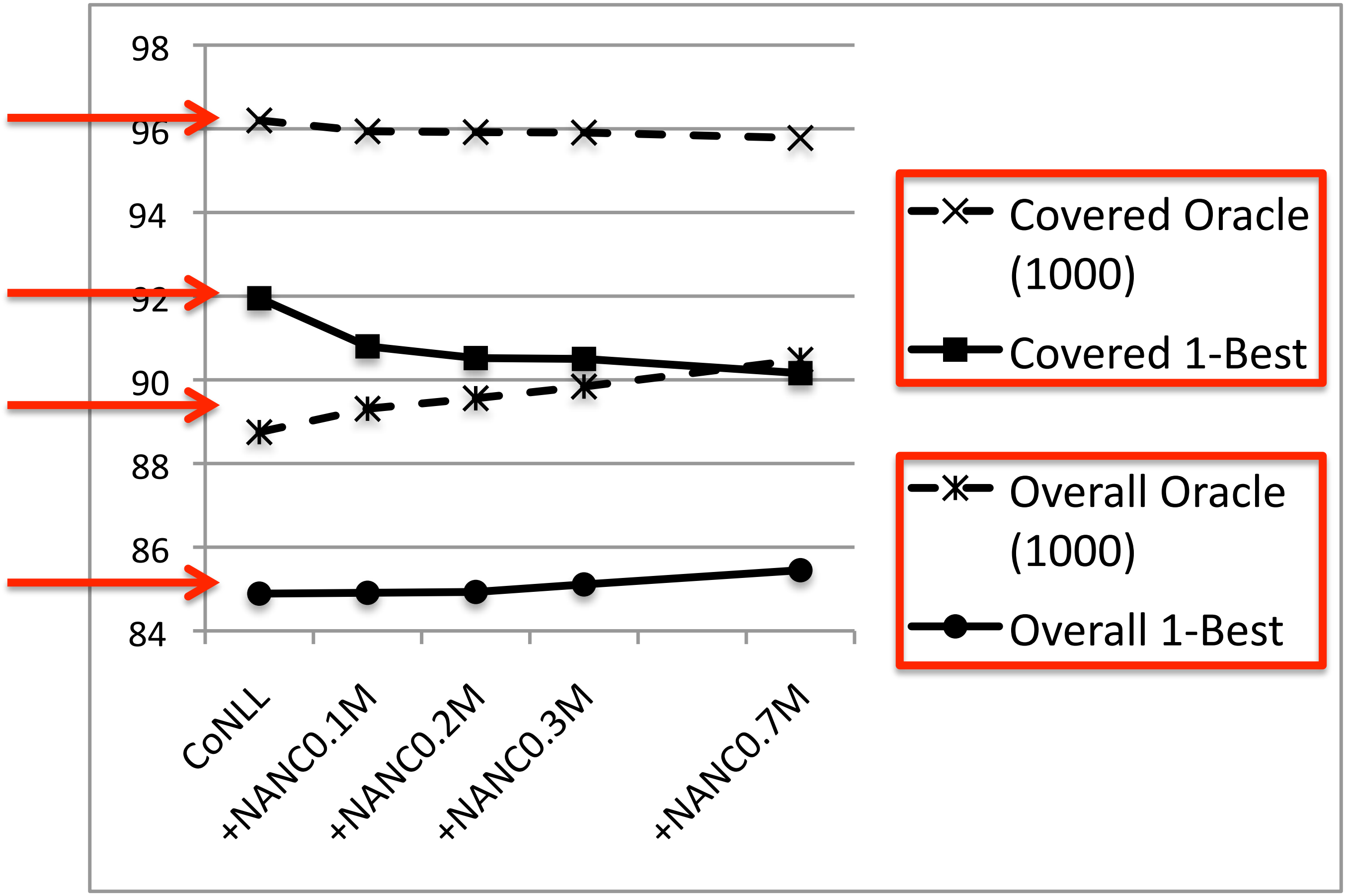
Related Work

- Generation Challenge 2011 Surface Realization Shared Task (Belz et al., 2011)
- Filippova and Strube (2009) (and their previous paper)'s evaluation is at the clause level instead of full sentences
- Bohnet et al. (2010) relied on discriminative modeling for the selection of the realization
- Guo et al. (2011)'s dependency-based N-gram approach

Extensions

- Does the size of training data matter?
- Are automatic evaluation metrics sufficient?
- Can we apply it to other languages than English?
- Can we break the projectivity assumption?

Additional Training Data



Manual Evaluation

- Comprehensiveness
  - 2 The meaning is the same as the gold standard.
  - 1 The meaning changed slightly from the gold standard, but comprehensible.
  - 0 The meaning is unclear, or totally different from the gold standard.
- Grammaticality
  - 2 The sentence is grammatical and fluent.
  - 1 The sentence is grammatical, but not natural or fluent (including the punctuation errors).
  - 0 The sentence is ungrammatical.

	Comprehensiveness	Grammaticality	Perfect
Base	84.1%	**77.1%**	28.8%
LM-Rerank	**90.1%**	73.2%	**36.7%**

Multilinguality

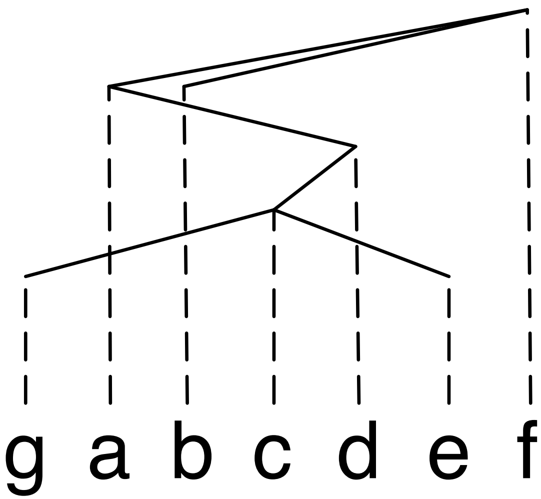
Languages		CA	CN	CZ	EN	DE	ES
No. of CPOS Tag		12	13	12	24	10	12
Avg. Token / Sent.		31.0	30.0	16.8	25.0	16.0	30.4
Grammar							
Avg. Config. / Sent.		13.1	14.0	8.3	12.4	6.0	13.2
Coverage	Sent.	578 / 1724 (33.5%)	790 / 1762 (44.8%)	498 / 5228 (9.5%)	724 / 1334 (54.3%)	1512 / 2000 (**75.6%**)	650 / 1655 (39.3%)
	Config.	22526 / 24546 (91.8%)	24749 / 26250 (94.3%)	43552 / 49751 (87.5%)	16536 / 17369 (95.2%)	11925 / 12503 (**95.4%**)	21920 / 23511 (93.2%)
BLEU							
Covered	1-best	84.51	88.67	82.00	91.95	78.52	79.93
	upper bound (1000)	91.77	94.49	93.60	96.20	88.01	89.78
Overall	1-best	75.79	**81.48**	66.59	**84.89**	73.85	73.10
	upper bound (1000)	80.61	86.52	76.85	88.75	82.09	79.75
Ulam's distance							
Covered	1-best	0.890	0.946	0.867	0.950	0.857	0.871
	upper bound (1000)	0.949	0.973	0.965	0.978	0.934	0.941
Overall	1-best	0.838	**0.891**	0.771	**0.911**	0.829	0.820
	upper bound (1000)	0.875	0.914	0.856	0.934	0.897	0.869

Non-Projective Trees

- Symbols
  - _x: the gap
  - |: the split

- Rules
  - f->a,b => a b _a f
  - a->d => a |2 d
  - d->c => _c |2 c d _c
  - c->g,e => g |2 c |1 e

- Application
  - g |2 c d e
  - g a |1 c d e
  - g a b c d e f



Evaluation

- Non-projective only

	German		Czech	
	Old	New	Old	New
1-best	**59.3**	59.0	55.1	**56.4**
upper bound (1000)	67.8	**70.8**	63.2	**70.1**

- Overall

	German		Czech	
	Old	New	Old	New
1-best	72.0	**72.1**	66.0	**66.1**
upper bound (1000)	81.0	**81.9**	77.6	**78.8**

Conclusion

- Rule-based and treebank-induced
- Generative model: n-best
- Language-independent

Future Work

- Generation from semantic representation, i.e., (D)MRS
- Lexical selection, morphological generation
- Interoperation with deep generation based on DELPH-IN grammars
- Better evaluation methodology

thank_VB | thank you | viel_PIAT | viel dank
OBJ | | NK |
you_PRP | Thank you! | dank_NN | Vielen Dank!

(see one application scenario in the next presentation)