Chart Mining-based Lexical Acquisition with Precision Grammars

Yi Zhang,[♠] Timothy Baldwin,[♡♢] Valia Kordoni,[♠] David Martinez[♢]
Jeremy Nicholson[♡♢]
♠ DFKI GmbH and Saarland University
♡ University of Melbourne
♢ NICTA Victoria Research Laboratory

Introduction: Parsability

- Parsing with precision grammars has made great strides in terms of scalability and coverage, but still room for improvement, esp. with coverage

  precision grammar = grammar which has been engineered to model grammaticality (avoid overgeneration)

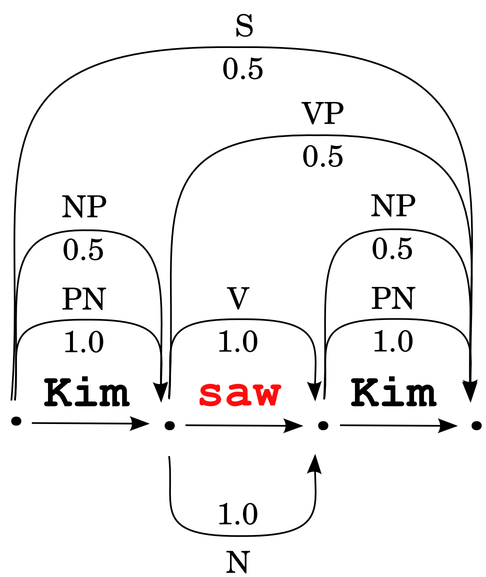
- Our approach to improving coverage = (off-line) lexical acquisition based on **chart mining**

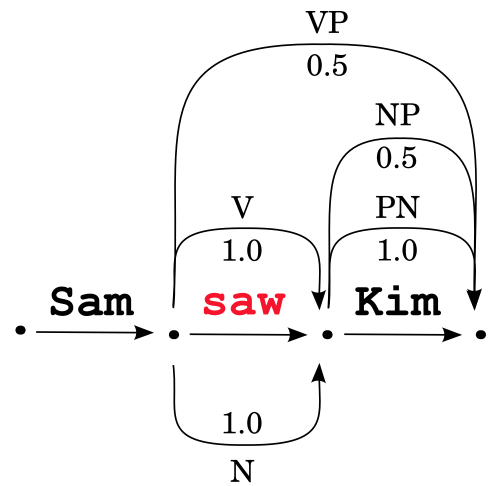
  ⋆ relative "lifetime" and probability of different analyses provide valuable insights into their plausibility

Illustration of Chart Mining

$$\Pr(\texttt{S} \rightarrow \texttt{NP VP}) = 1.0$$
$$\Pr(\texttt{VP} \rightarrow \texttt{V NP}) = 1.0$$
$$\Pr(\texttt{NP} \rightarrow \texttt{PN}) = 0.5$$
$$\Pr(\texttt{NP} \rightarrow \texttt{'the' N}) = 0.5$$

$$\Pr(\texttt{V} \rightarrow \texttt{'saw'}) = 1.0$$
$$\Pr(\texttt{PN} \rightarrow \texttt{'Kim'}) = 1.0$$
$$\Pr(\texttt{N} \rightarrow \texttt{'saw'}) = 1.0$$





Introduction to Chart Parsing

- A chart is used to record the partial analysis during parsing
- Together with its variants, chart parser can be used for a variety of grammar formalisms (CFG, TAG, LFG, HPSG, . . .)
- We use the agenda-driven bottom-up search strategy
- Constituent-based chart parser records potential constituents as passive edges
- The size of the parsing chart can be reduced by local ambiguity packing (based on certain "equivalence classes")

Methodology: General Approach

- Populate the chart with bottom-up search strategy
- Mine relevant features from the densely populated chart, even if a full parse is not available
- Use customised set of chart-mined features as appropriate for task

Subsumption-based Packing and Selective Unpacking

- Packing under subsumption allows efficient storage of local ambiguities
- Selective unpacking to mine relevant features
- Probabilities on each selectively-unpacked edge from discriminative parse selection model (Toutanova et al., 2005)
- Dynamic programming used to decode the N-best (partial) readings from packed parse forest

Verb Particle Constructions

- **Verb Particle Construction:**

  English Verb Particle Constructions (VPCs) consist of a head verb and one or more obligatory (prepositional) particles

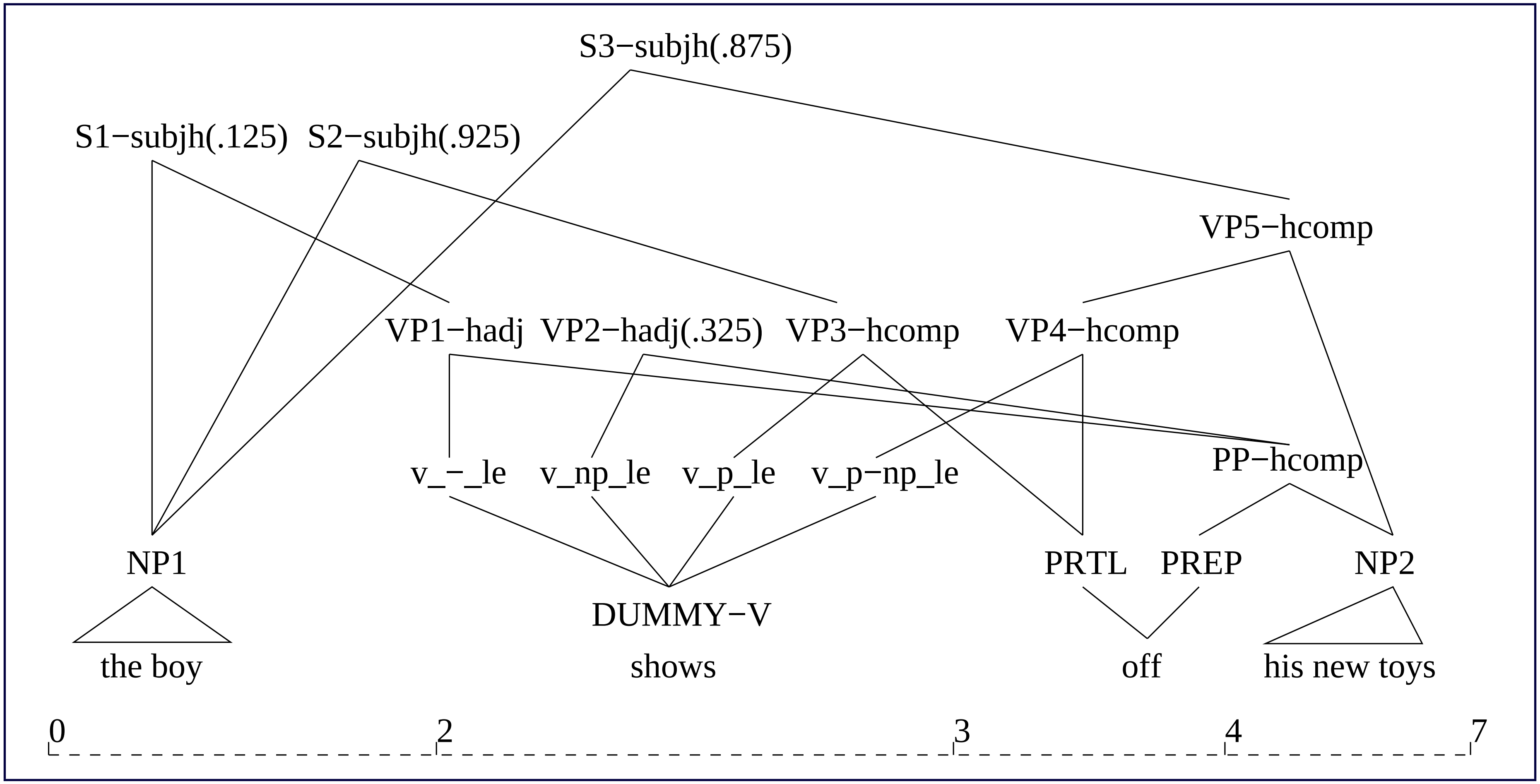
- We are interested in extracting:
  - ⋆ **non-compositional** VPCs: *look up* vs. *battle on*
  - ⋆ with **valence**: *hand in* vs. *back off*

- Dataset from LREC-2008-MWE shared task (Baldwin 2008)
  - ⋆ 4,090 candidate VPC triples (verb, particle, valence)
  - ⋆ up to 50 sentences containing the given VPC from BNC

VPC Feature Engineering

Feature	Description	Examples
LE:MaxCons	A lexical entry together with the maximal constituent constructed from it	*v_-_le:subjh*, *v_np_le:hadj*, ...
LE:MaxSpan	A lexical entry together with the length of the span of the maximal constituent constructed from the LE	*v_-_le:7*, *v_np_le:5*, ...
LE:MaxLevel	A lexical entry together with the levels of projections before it reaches its maximal constituent	*v_-_le:2*, *v_np_le:1*, ...
LE:MaxCRank	A lexical entry together with the relative disambiguation score ranking of its maximal constituent among all MaxCons from different LEs	*v_-_le:4*, *v_np_le:3*, ...
PARTICLE	The stem of the particle in the candidate VPC	*off*

Putting It All Together



Three VPC Tasks

Task	Description
GOLD VPC	Determine the valence for a verb–preposition combination which is known to occur as a non-compositional VPC (i.e. known VPC, with unknown valence(s))
FULL	Determine whether each verb–preposition combination is a VPC or not, and further predict its valence(s) (i.e. unknown if VPC, and unknown valence(s))
VPC	Determine whether each verb–preposition combination is a VPC or not *ignoring valence* (i.e. unknown if VPC, and don't care about valence)

Experimental Details

- PET parser (Callmeier 2001)
- English Resource Grammar (Flickinger 2002), version nov-06
- Unknown word handling with lexical type prediction model trained on LOGON
- 4 dummy lexical entries:

  $v\_\text{-}\_le$, v_np_le, v_p_le, $v\_p\text{-}np\_le$

- Features are mined from the parsing chart

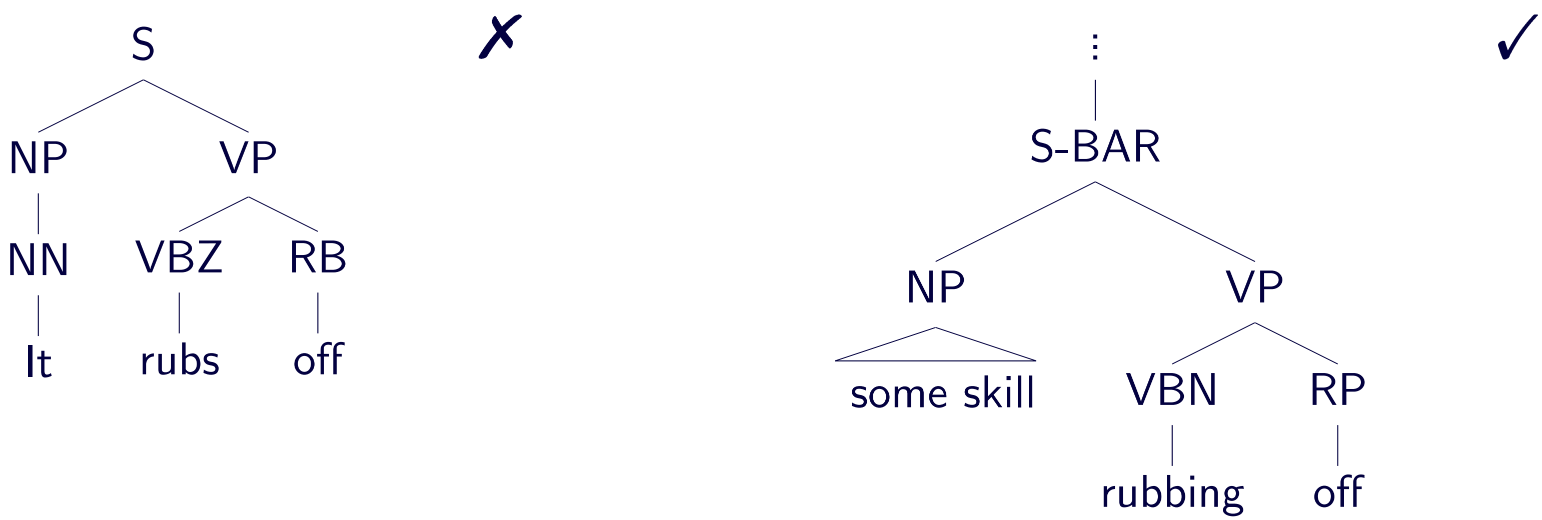
Experimental Details

- Probabilistic baseline:

$$\tilde{P}(s|v,p) = P(s|v) \cdot P(s|p) \text{ for } s \in \{intrans, trans, null\}$$

- Benchmark: Charniak parser

majority vote over `RB/IN/TO` vs. `RP` for each valence



Experimental Details

- Probabilistic baseline:

  $\tilde{P}(s|v,p) = P(s|v) \cdot P(s|p)$ for $s \in \{intrans, trans, null\}$

- Benchmark: Charniak parser

  majority vote over `RB/IN/TO` vs. `RP` for each valence

- Remove VPCs which are attested in WSJ Sections 1–21 from test data on each iteration, for comparability with Charniak parser

- 5-fold cross-validation

Results: GOLD

VPC Type	Naïve Baseline			Charniak Parser			Chart-Mining		
	P	R	F	P	R	F	P	R	F
Intrans-VPC	.300	.018	.034	.549	.753	.635	.845	.621	.716
Trans-VPC	.676	.348	.459	.829	.648	.728	.877	.956	.915
All	.576	.236	.335	.691	.686	.688	.875	.859	**.867**

Results: FULL/VPC

VPC Type	Naïve Baseline			Charniak Parser			Chart-Mining		
	P	R	F	P	R	F	P	R	F
Intrans-VPC	.060	.018	.028	.102	.593	.174	.153	.155	.154
Trans-VPC	.083	.348	.134	.179	.448	.256	.179	.362	.240
All	.080	.236	.119	.136	.500	.213	.171	.298	**.218**
VPC	.123	.348	.182	.173	.782	.284	.259	.332	**.291**

Findings

- Chart mining superior to Charniak parser overall

  Charniak parser much better over VPCs lexicalised in the training data (unsurprisingly!) → potential for our method to similarly benefit from lexicalisation

- FULL harder than due to 7/8 of candidates not in fact being VPCs

- Intransitive VPCs harder to extract than transitive

Discussion

- Considerable scope for extra experimentation over other tasks (MWEs and non-MWEs) and languages
- Grammar-based nature means particularly well suited to lexical acquisition tasks over discontinuous lexemes/non-configurational languages
- Unlexicalised nature, non-requirement of spanning parse means suited to lexical acquisition over low-density languages/under-developed grammars
- Applications beyond lexical acquisition (e.g. partial parsing)

Conclusion

- Precision grammar-based chart mining method proposed
- Highly encouraging results achieved over VPC lexical acquisition task
- Lots of scope for follow-up experimentation/applications beyond lexical acquisition