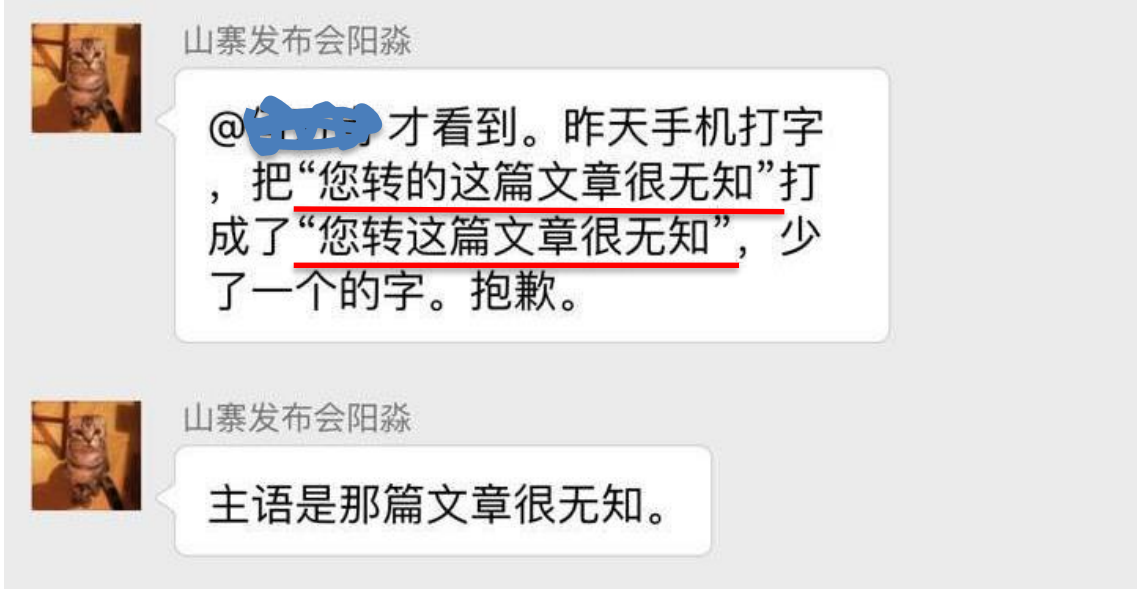
Recent Advances in Parsing

Wanxiang Che

Harbin Institute of Technology

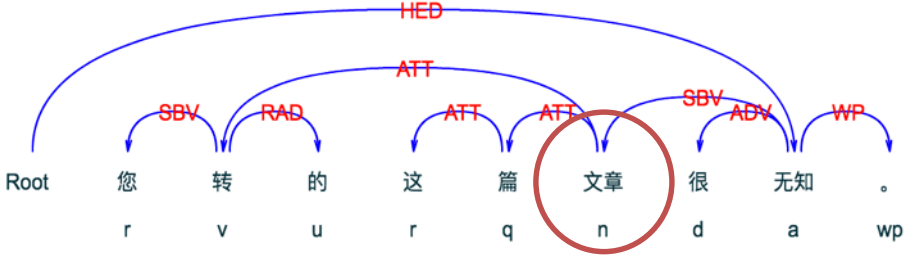
2016-10

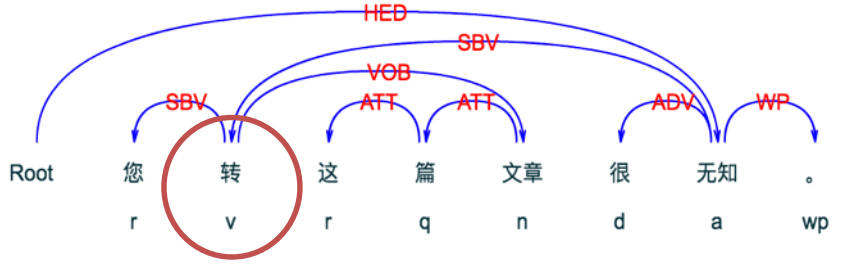
An Example


山寨发布会阳淼
@[redacted]才看到。昨天手机打字，把“您转的这篇文章很无知”打成了“您转这篇文章很无知”，少了一个的字。抱歉。
山寨发布会阳淼
主语是那篇文章很无知。

Syntactic Parsing

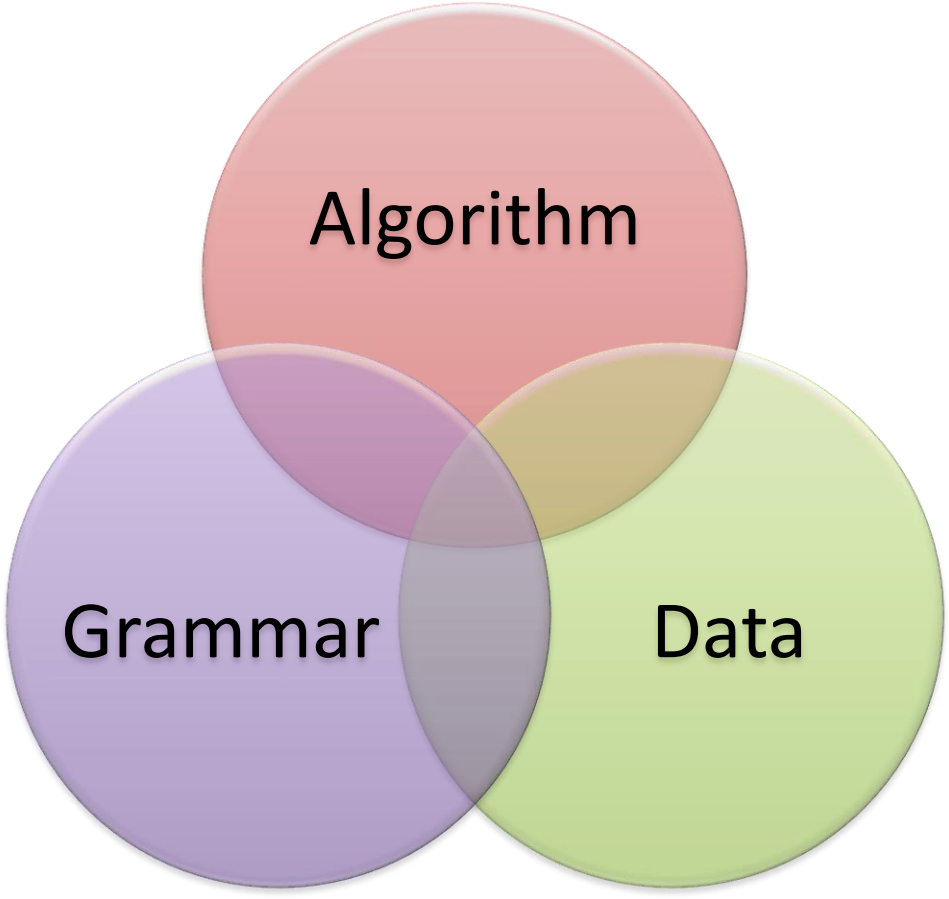
- Analyzing a natural language string conforming to the rules of a formal grammar, emphasizing subject, predicate, object, etc.
- A traditional and core NLP task



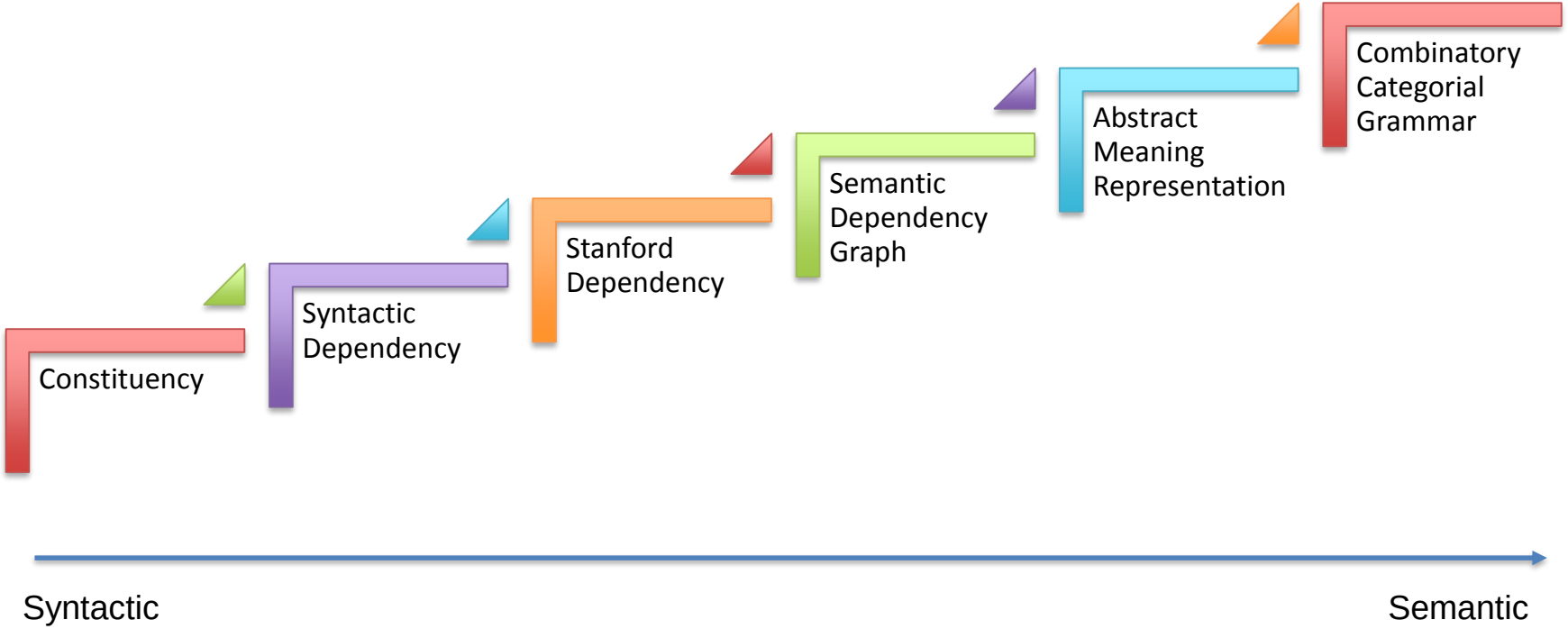


http://www.ltp-cloud.com/demo/

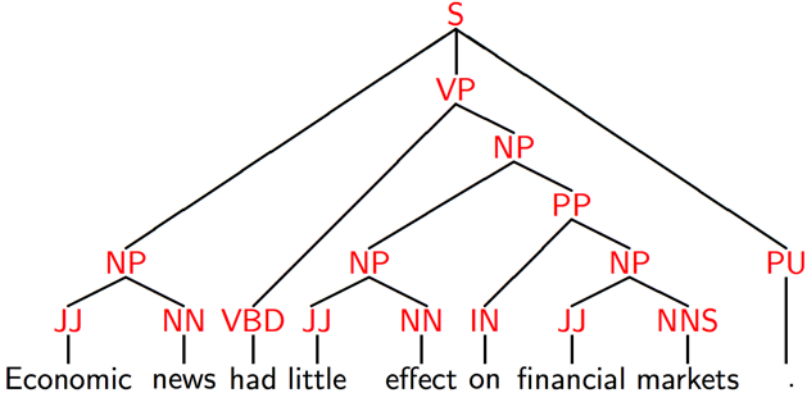
Elements of Parsing

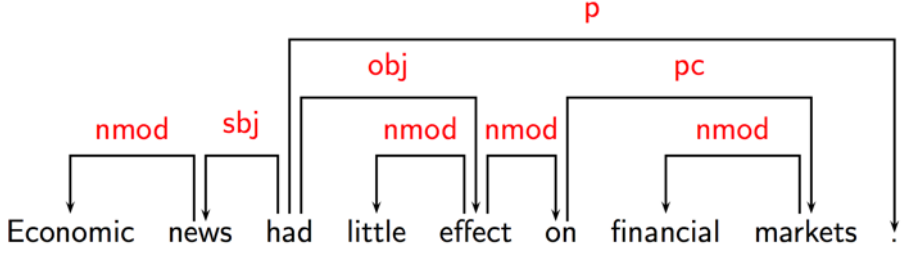


Grammar

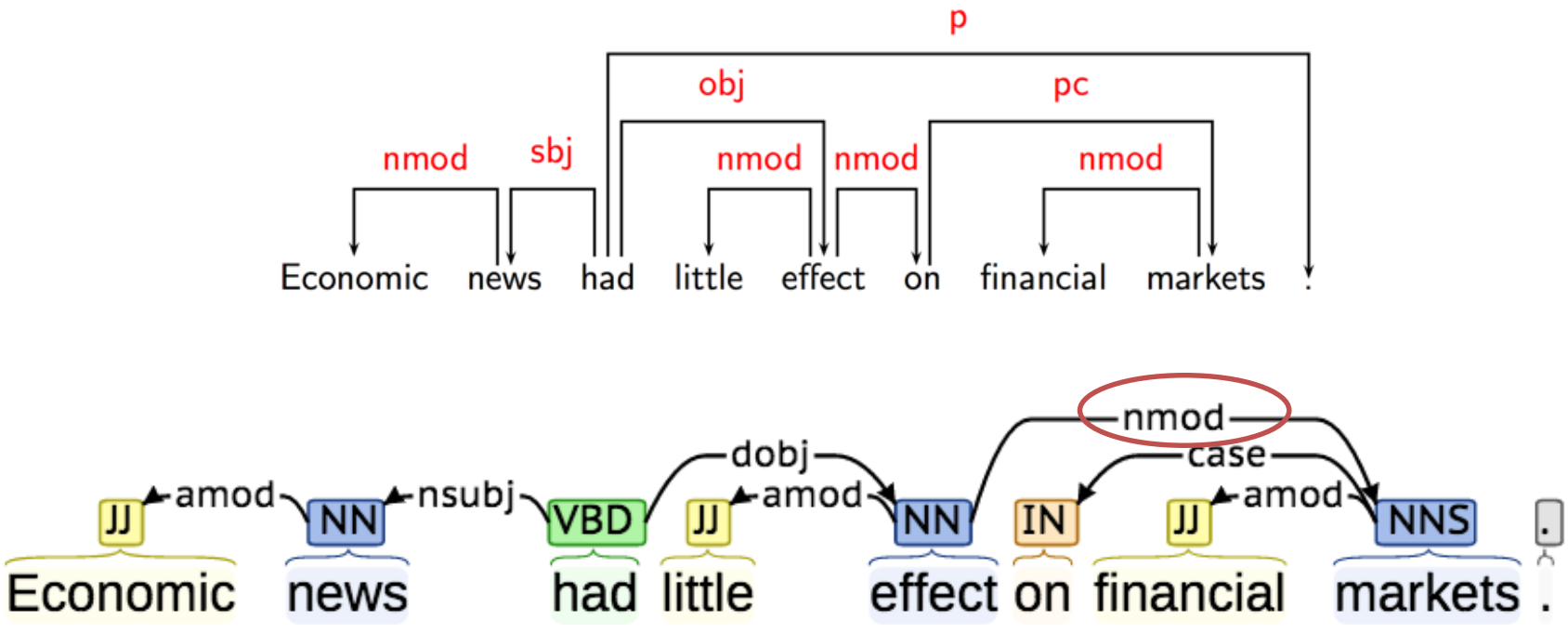

Constituency
Syntactic Dependency
Stanford Dependency
Semantic Dependency Graph
Abstract Meaning Representation
Combinatory Categorial Grammar
Syntactic
Semantic

Constituency vs. Dependency



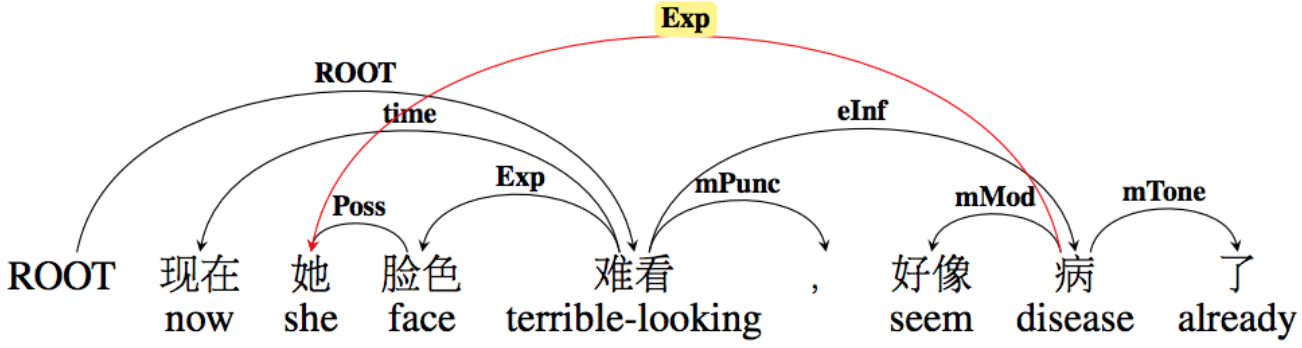


Syntactic vs. Stanford Dependency



http://nlp.stanford.edu/software/stanford-dependencies.shtml

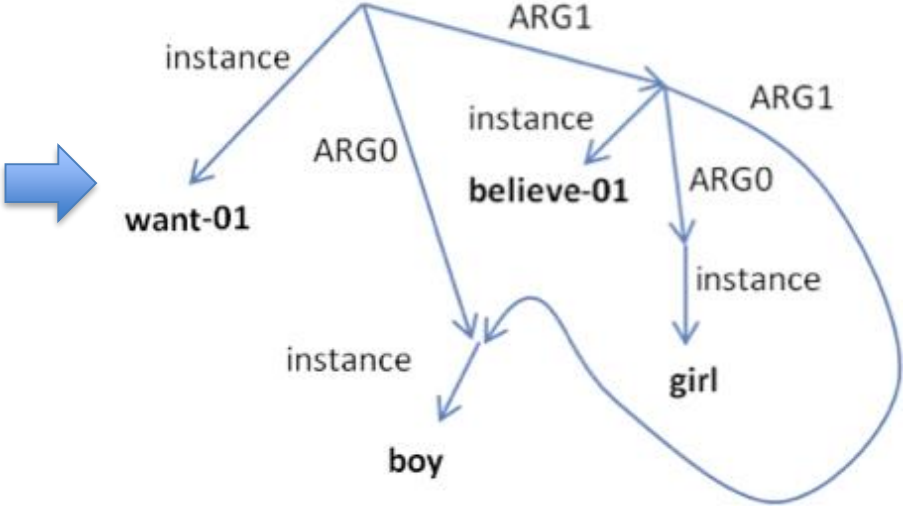
Semantic Dependency Graph



http://www.ltp-cloud.com/intro/#sdp_how

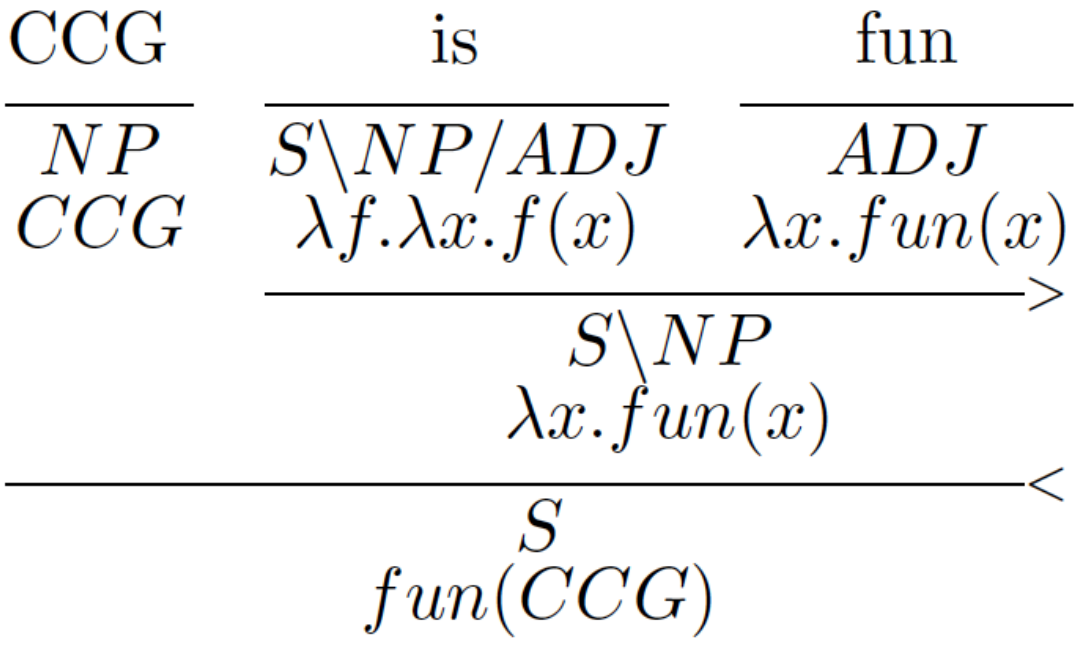
Abstract Meaning Representation (AMR)

The boy wants the girl to believe him.
The boy wants to be believed by the girl.
The boy has a desire to be believed by the girl.
The boy's desire is for the girl to believe him.
The boy is desirous of the girl believing him.



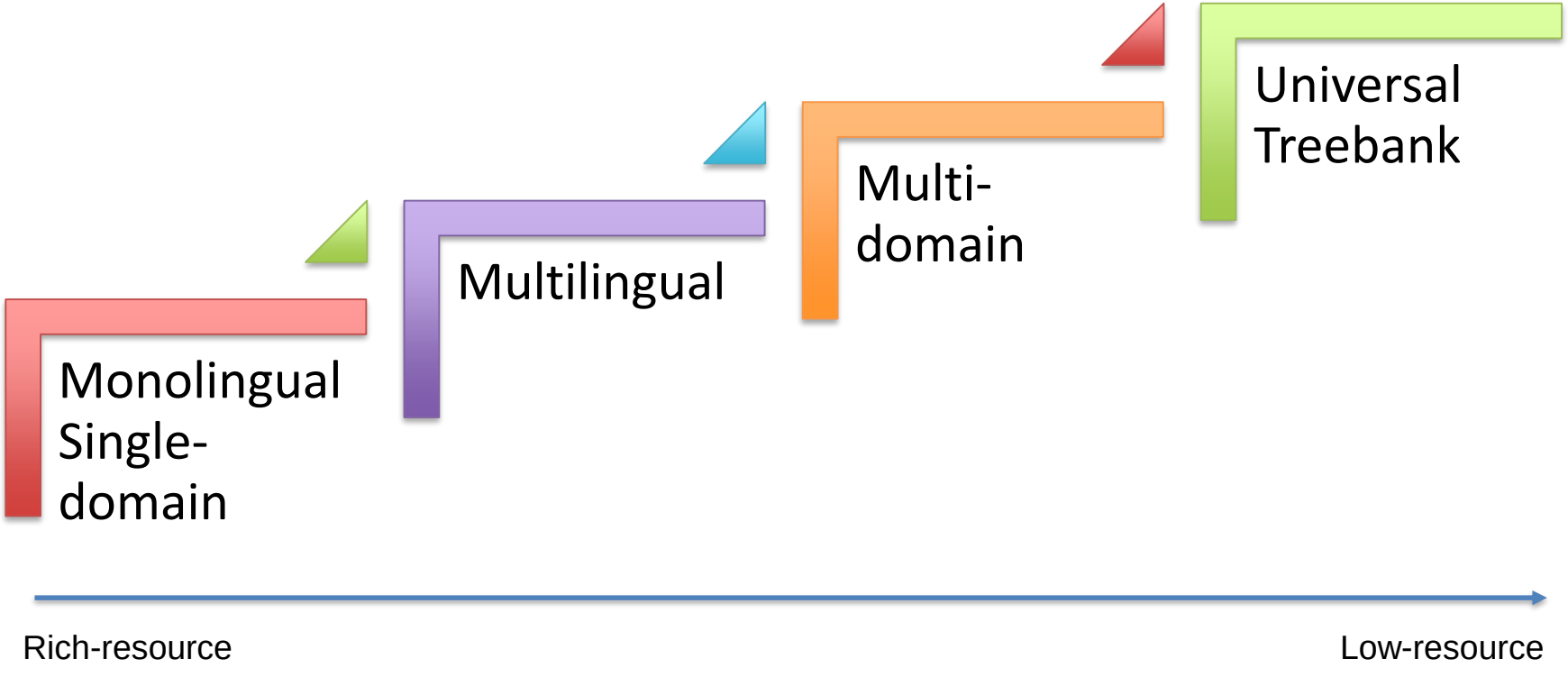
http://www.isi.edu/~ulf/amr/help/amr-guidelines.pdf

Combinatory Categorial Grammars (CCG)



http://groups.inf.ed.ac.uk/ccg/

Data


Monolingual
Single-
domain
Multilingual
Multi-
domain
Universal
Treebank
Rich-resource
Low-resource

Multilingual

- CoNLL 2007 Shared Tasks
  - Multi-lingual Dependency Parsing
  - 12 languages
  - http://ilk.uvt.nl/conll/
- CoNLL 2009 Shared Tasks
  - Syntactic and Semantic Dependencies in Multiple Languages
  - 7 languages
  - http://ufal.mff.cuni.cz/conll2009-st/

Multi-domain

- Syntactic Analysis of Non-Canonical Language (SANCL) 2012 Shared Task
  - https://sites.google.com/site/sancl2012/
  - Organized by Google
- Corpus
  - Google Web Treebank
  - Three web domains: CQA, Newsgroup, Online Review

Universal Treebank

- http://universaldependencies.org/
  - 40+ languages with **universal** dependencies and POS tags
- For multi-lingual and cross-lingual research
  - "Many Languages, One Parser" – CMU
- CoNLL 2017 Shared Task
  - Multilingual Parsing from Raw Text to Universal Dependencies

Algorithm

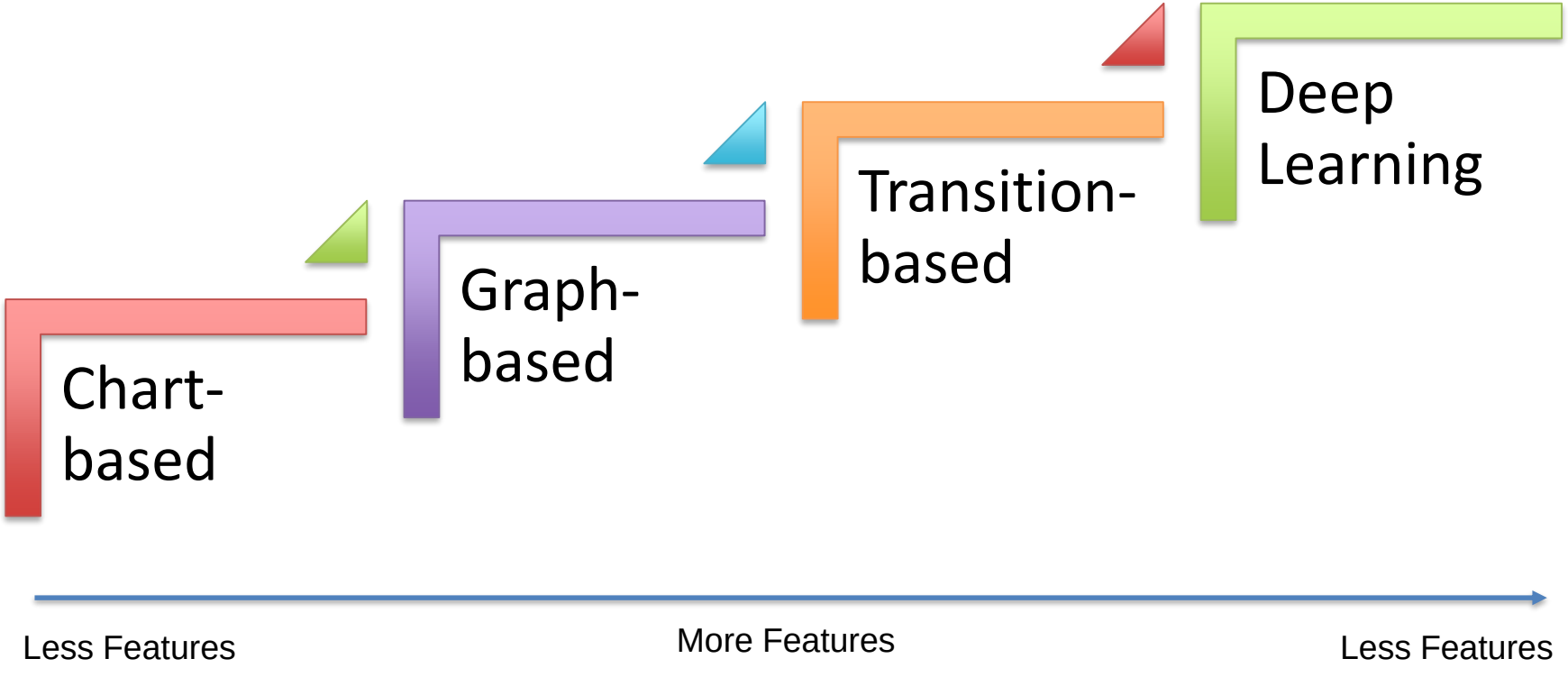
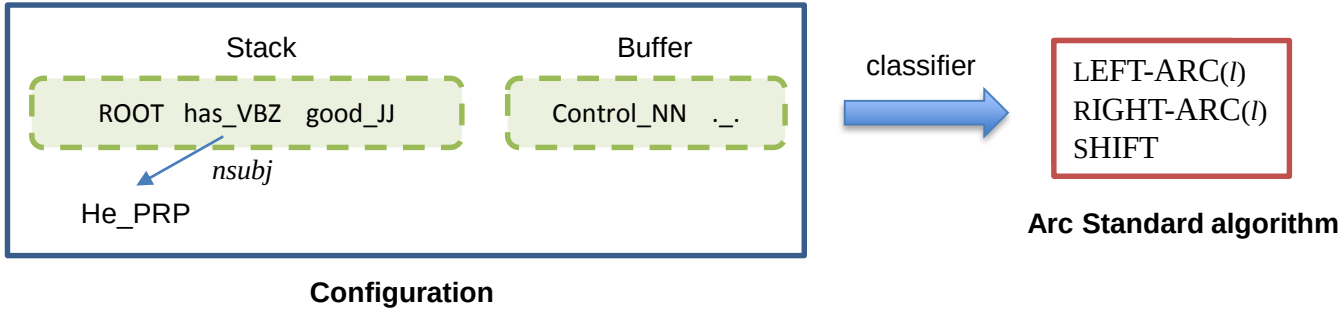


Chart-based
Graph-based
Transition-based
Deep Learning
Less Features
More Features
Less Features

Transition-based Dependency Parsing

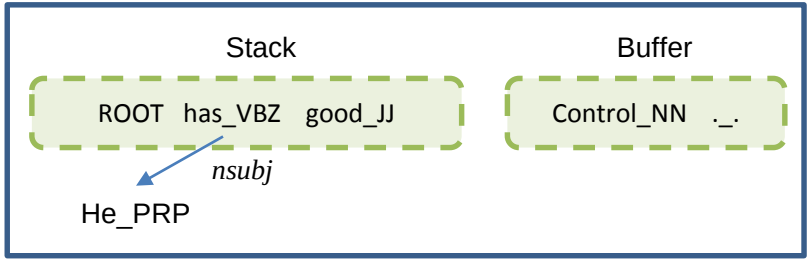
- Greedily predict a transition sequence from an initial parser state to some terminal states
- State (configuration)

  = Stack + Buffer + Dependency Arcs



Traditional Features

Configuration



Feature Vector:
- Binary
- Sparse
- High-dimensional

Feature templates: a combination of elements from the configuration.
- For example: (Zhang and Nivre, 2011): 72 feature templates

from single words
S_0wp; S_0w; S_0p; N_0wp; N_0w; N_0p; N_1wp; N_1w; N_1p; N_2wp; N_2w; N_2p;
from word pairs
S_0wpN_0wp; S_0wpN_0w; S_0wN_0wp; S_0wpN_0p; S_0pN_0wp; S_0wN_0w; S_0pN_0p N_0pN_1p
from three words
$N_0pN_1pN_2p$; $S_0pN_0pN_1p$; $S_{0h}pS_0pN_0p$; $S_0pS_{0l}pN_0p$; $S_0pS_{0r}pN_0p$; $S_0pN_0pN_{0l}p$

Table 1: Baseline feature templates.
w – word; p – POS-tag.

distance
S_0wd; S_0pd; N_0wd; N_0pd; S_0wN_0wd; S_0pN_0pd;
valency
S_0wv_r; S_0pv_r; S_0wv_l; S_0pv_l; N_0wv_l; N_0pv_l;
unigrams
$S_{0h}w$; $S_{0h}p$; S_0l; $S_{0l}w$; $S_{0l}p$; $S_{0l}l$; $S_{0r}w$; $S_{0r}p$; $S_{0r}l$;$N_{0l}w$; $N_{0l}p$; $N_{0l}l$;
third-order
$S_{0h2}w$; $S_{0h2}p$; $S_{0h}l$; $S_{0l2}w$; $S_{0l2}p$; $S_{0l2}l$; $S_{0r2}w$; $S_{0r2}p$; $S_{0r2}l$; $N_{0l2}w$; $N_{0l2}p$; $N_{0l2}l$; $S_0pS_{0l}pS_{0l2}p$; $S_0pS_{0r}pS_{0r2}p$; $S_0pS_{0h}pS_{0h2}p$; $N_0pN_{0l}pN_{0l2}p$;
label set
S_0ws_r; S_0ps_r; S_0ws_l; S_0ps_l; N_0ws_l; N_0ps_l;

Table 2: New feature templates.
w – word; p – POS-tag; v_l, v_r – valency; l – dependency label, s_l, s_r – labelset.

Chen and Manning NN Parser

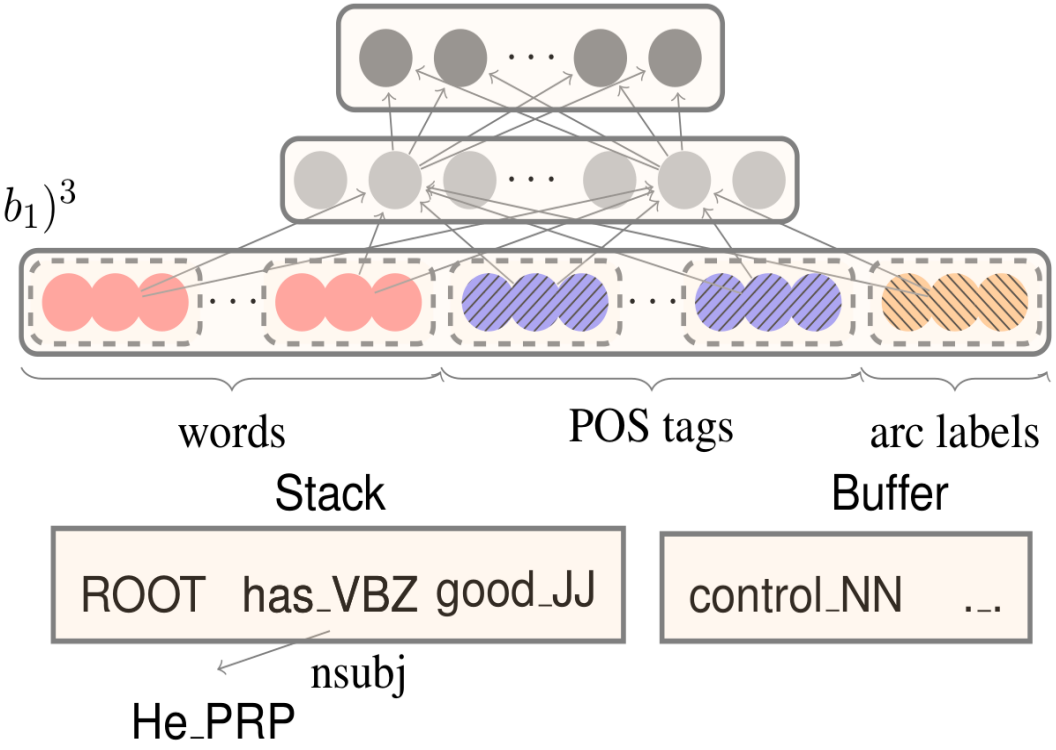
Softmax layer:

$$p = \texttt{softmax}(W_2 h)$$

Hidden layer:

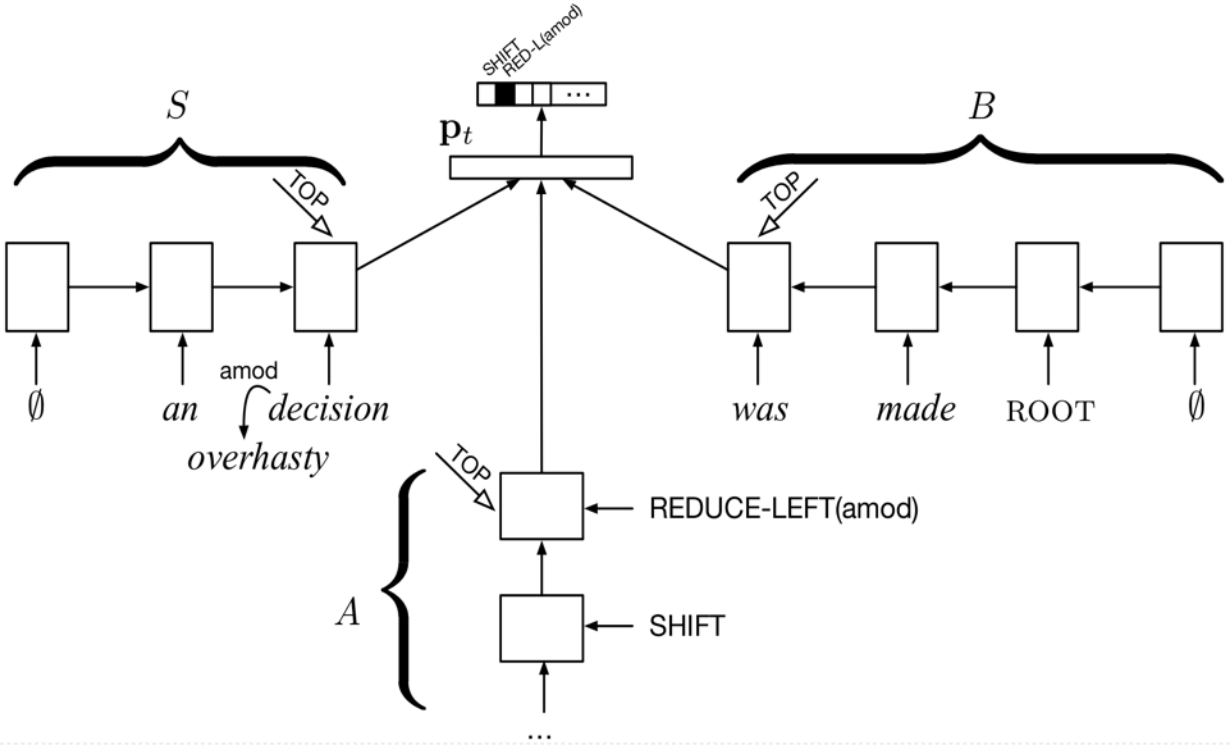
$$h = (W_1^w x^w + W_1^t x^t + W_1^l x^l + b_1)^3$$

Input layer: $[x^w, x^t, x^l]$



Chen, D., & Manning, C. D. (2014). A Fast and Accurate Dependency Parser using Neural Network. ACL.

Stack LSTM Parser



Dyer, C., Ballesteros, M., Ling, W., Matthews, A., & Smith, N. A. (2015). Transition-Based Dependency Parsing with Stack Long Short-Term Memory. ACL.

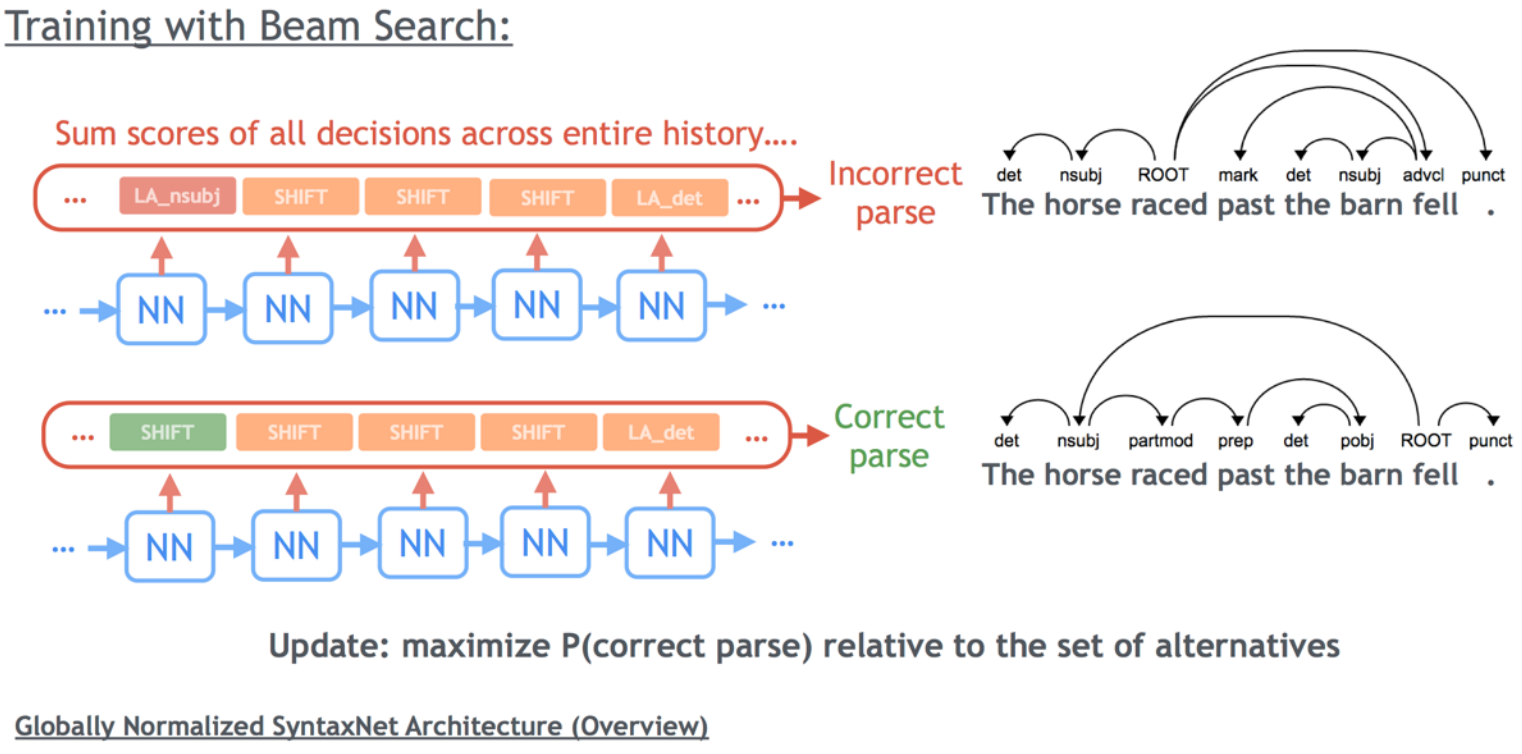
Sentence-level Log Likelihood

$$p(y_i \mid x, \theta) = \frac{e^{f(x,\,\theta)_i}}{\sum_{y_j \in \mathbf{GEN}(x)} e^{f(x,\,\theta)_j}}$$

$$f(x,\,\theta)_i = \sum_{a_k \in y_i} o(x,\,y_i,\,k,\,a_k)$$

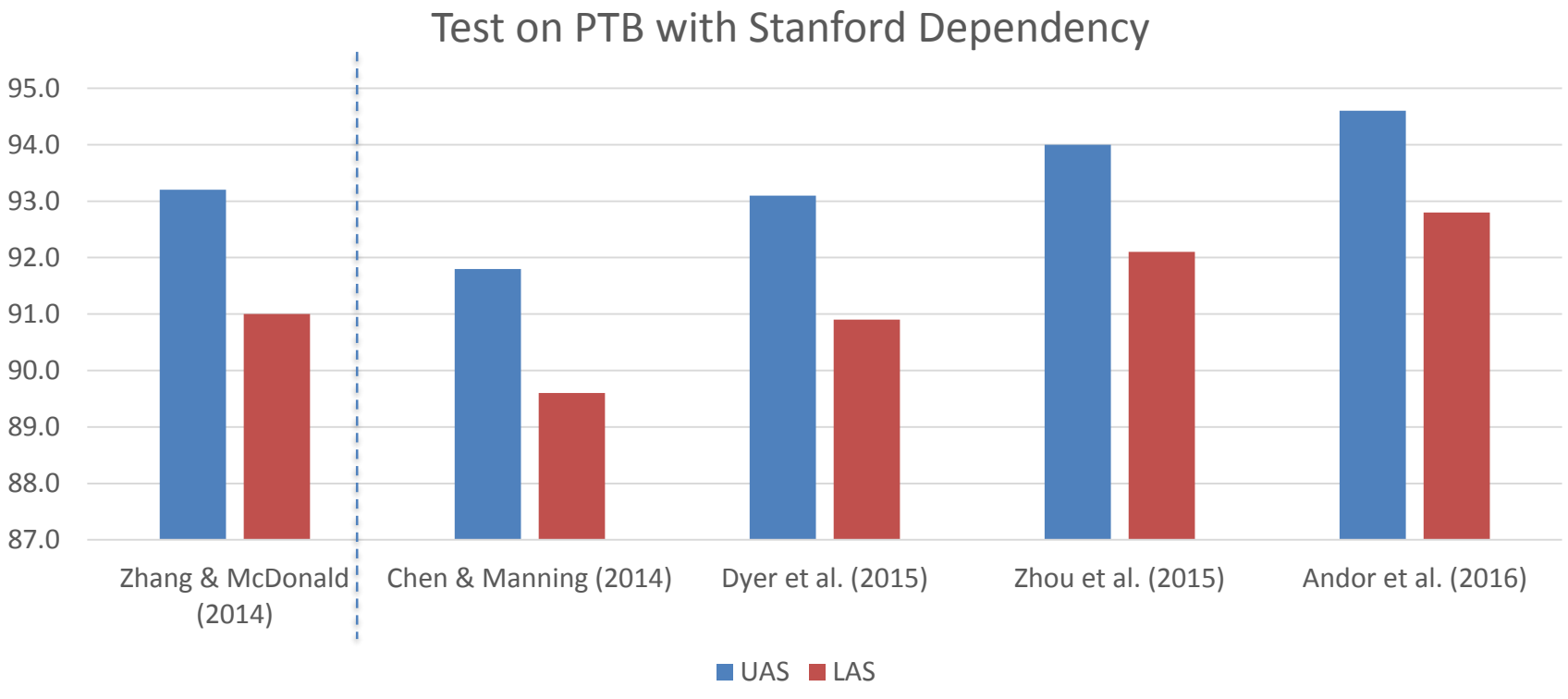
Zhou, H., Zhang, Y., Huang, S., & Chen, J. (2015). A Neural Probabilistic Structured-Prediction Model for Transition-Based Dependency Parsing. ACL.

SyntaxNet: Google

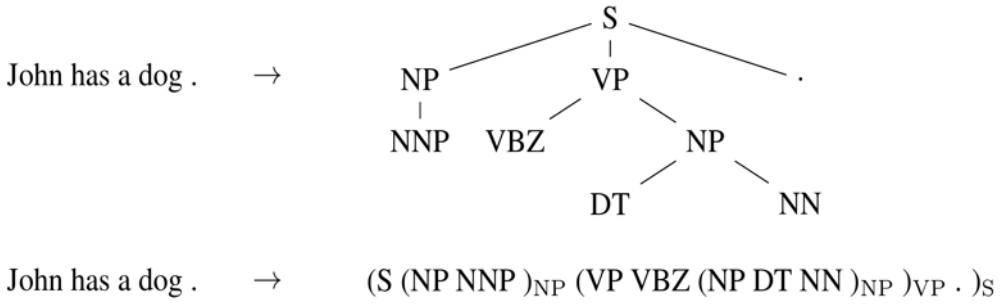


Andor, D., Alberti, Chris., Weiss, D., Severyn, A., Presta, A., Ganchev, K., Petrov, S., & Collins, M. (2016). Globally Normalized Transition-Based Neural Networks. ACL.

Changes of Performance



Encoder-decoder Parsing



Vinyals, O., Kaiser, L., Koo, T., Petrov, S., & Sutskever, I. (2015). Grammar as a Foreign Language. ICLR.

Encoding: *But it was the Quotron problems that ...*

Decoding: *But it was @L SBJ @L DEP the Quotron problems @L NMOD @L NMOD that ...*

Wiseman, S., & Rush, A. (2016). Sequence-to-Sequence Learning as Beam-Search Optimization. arxiv.

Some Open Questions

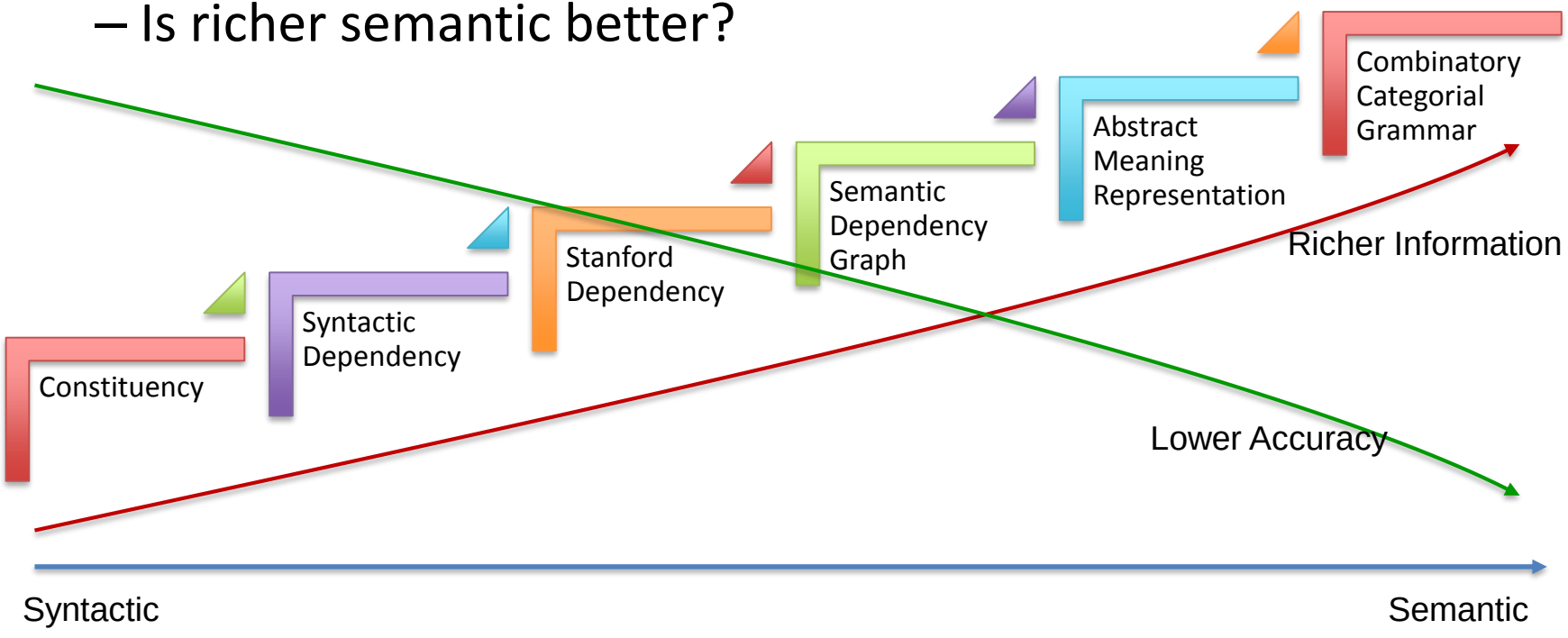
- Is Parsing Necessary?

	Bi-LSTM	Tree-LSTM
Stanford Sentiment TreeBank	49.8 / 50.7 (Segment)	50.4
Binary Sentiment Classification	79.0	77.4
Question-Answer Matching	56.4	55.8
Semantic Relationship Classification	**75.2**	**76.7**
Discourse Parsing	57.5	56.4

Jiwei Li, Minh-Thang Luong, Dan Jurafsky and Eduard Hovy. When Are Tree Structures Necessary for Deep Learning of Representations? EMNLP, 2015.

More Questions

- Grammar
  - Is richer semantic better?



More Questions

- Grammar
  - Is richer semantic better?
- Data
  - How to obtain large (pseudo) annotation data?
- Algorithm
  - Can we utilize other supervision?

Future Trends

- Grammar
  - Trade-off between rich information and performance
- Data
  - Exploiting multiple treebanks
- Algorithm
  - Deep reinforcement learning

Thanks!