CSEP 517
Natural Language Processing

Frame Semantics

Luke Zettlemoyer

Slides adapted from Yejin Choi, Martha Palmer, Chris Manning,

Ray Mooney, Lluis Marquez, Luheng He

Frames

“Case for Case”

- Theory:
  - Frame Semantics (Fillmore 1968)
- Resources:
  - VerbNet(Kipper et al., 2000)
  - FrameNet (Fillmore et al., 2004)
  - PropBank (Palmer et al., 2005)
  - NomBank
- Statistical Models:
  - Task: Semantic Role Labeling (SRL)
  - Deep SRL

reddit AMA (ask me anything): Michael Jordan

(Sep 2014)

- [–]CyberByte
- If you got a billion dollars to spend on a huge research project that you get to lead, what would you like to do?

- [–]michaelijordan
- I'd use the billion dollars to build a **NASA-size program** focusing on natural language processing (NLP), in all of its glory (**semantics**, **pragmatics**, etc).
- Intellectually I think that NLP is fascinating, allowing us to focus on **highly-structured inference problems**, on issues that go to the core of "**what is thought**" but remain eminently practical, and on a technology that surely would make the world a better place.

reddit

AMA (ask me anything): Michael Jordan

(Sep 2014)

- Although current deep learning research tends to claim to encompass NLP, I'm (1) much less convinced about the strength of the results, compared to the results in, say, vision; (2) much less convinced in the case of NLP than, say, vision, the way to go is to couple huge amounts of data with black-box learning architectures.

- I'd invest in some of the human-intensive labeling processes that one sees in projects like **FrameNet** and (gasp) projects like Cyc. I'd do so in the context of a full merger of "data" and "knowledge", where the representations used by the humans can be connected to data and the representations used by the learning systems are directly tied to linguistic structure. I'd do so in the context of clear concern with the usage of language (e.g., causal reasoning).

Frames

- Theory:
  - Frame Semantics (Fillmore 1968)
- Resources:
  - VerbNet(Kipper et al., 2000)
  - FrameNet (Fillmore et al., 2004)
  - PropBank (Palmer et al., 2005)
  - NomBank
- Statistical Models:
  - Task: Semantic Role Labeling (SRL)
  - Deep SRL

“Case for Case”

Frame Semantics

- Frame: Semantic frames are schematic representations of situations involving various participants, propositions, and other conceptual roles.
- Frame Elements (FEs) include events, states, relations and entities.

✓ **Frame**: *"The case for case"* (Fillmore 1968)

- 8k citations in Google Scholar.

✓ **Script:** knowledge about situations like eating in a restaurant.

- "*Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures*" (Schank & Abelson 1977)

✓ **Political Framings**: George Lakoff's recent writings on the framing of political discourse.

Capturing Generalizations over Related Predicates & Arguments

verb	BUYER	GOODS	SELLER	MONEY	PLACE
Buy	subject	object	from	for	at
Sell	to	object	subject	for	at
Cost	Ind. object	subject	--	object	at
Spend	subject	on	--	object	at

Case Grammar -> Frames

- Valency: Predicates have arguments (optional & required)
  - Example: “give” requires 3 arguments:
    - Agent (A), Object (O), and Beneficiary (B)
    - Jones (A) gave money (O) to the school (B)
- Frames:
  - commercial transaction frame: Buy/Sell/Pay/Spend
  - Save <good thing> from <bad situation>
  - Risk <valued object> for <situation>|<purpose>|<beneficiary>|<motivation>
- Collocations & Typical predicate argument relations
  - Save whales from extinction (not vice versa)
  - Ready to risk everything for what he believes
- Representation Challenges: What matters for practical NLP?

Slide from Ken Church (at Fillmore tribute workshop)

Thematic (Semantic) Roles

- AGENT - the volitional causer of an event
  - The waiter spilled the soup
- EXPERIENCER - the experiencer of an event
  - John has a headache
- FORCE - the non-volitional causer of an event
  - The wind blows debris from the mall into our yards.
- THEME - the participant most directly affected by an event
  - Only after Benjamin Franklin broke the ice ...
- RESULT - the end product of an event
  - The French government has built a regulation-size baseball diamond ...

Thematic (Semantic) Roles

- INSTRUMENT - an instrument used in an event
  - He turned to poaching catfish, stunning them with a shocking device ...
- BENEFICIARY - the beneficiary of an event
  - Whenever Ann makes hotel reservations for her boss ...
- SOURCE - the origin of the object of a transfer event
  - I flew in from Boston
- GOAL - the destination of an object of a transfer event
  - I drove to Portland

- Can we read semantic roles off from PCFG or dependency parse trees?

Semantic roles ≠ Grammatical roles

- Agent – the volitional causer of an event
  - usually “subject”, sometimes “prepositional argument”, ...
- Theme – the participant directly affected by an event
  - usually “object”, sometimes “subject”, ...
- Instrument – an instrument (method) used in an event
  - usually prepositional phrase, but can also be a “subject”

- John broke the window.
- John broke the window with a rock.
- The rock broke the window.
- The window broke.
- The window was broken by John.

Ergative Verbs

- **Ergative verbs**
  - **subject when intransitive** = **direct object when transitive**.
  - "it broke the window" (transitive)
  - "the window broke" (intransitive).
- Most verbs in English are *not* ergative (the subject role does not change whether transitive or not)
  - "He ate the soup" (transitive)
  - "He ate" (intransitive)
- Ergative verbs generally describe some sort of “changes” of states:
  - Verbs suggesting a change of state — *break, burst, form, heal, melt, tear, transform*
  - Verbs of cooking — *bake, boil, cook, fry*
  - Verbs of movement — *move, shake, sweep, turn, walk*
  - Verbs involving vehicles — *drive, fly, reverse, run, sail*

FrameNet

Frames

“Case for Case”

- Theory:
  - Frame Semantics (Fillmore 1968)
- Resources:
  - VerbNet(Kipper et al., 2000)
  - FrameNet (Fillmore et al., 2004)
  - PropBank (Palmer et al., 2005)
  - NomBank
- Statistical Models:
  - Task: Semantic Role Labeling (SRL)

Words in “change_position_on _a_scale” frame:

VERBS:	dwindle	move	soar	escalation	shift
advance	edge	mushroom	swell	explosion	tumble
climb	explode	plummet	swing	fall	
decline	fall	reach	triple	fluctuation	**ADVERBS:**
decrease	fluctuate	rise	tumble	gain	increasingly
diminish	gain	rocket		growth	
dip	grow	shift	**NOUNS:**	hike	
double	increase	skyrocket	decline	increase	
drop	jump	slide	decrease	rise	

- Frame := the set of words sharing a similar predicate-argument relations
- Predicate can be a verb, noun, adjective, adverb
- The same word with multiple senses can belong to multiple frames

Roles in "change_position_on _a_scale" frame

Core Roles	
ATTRIBUTE	The ATTRIBUTE is a scalar property that the ITEM possesses.
DIFFERENCE	The distance by which an ITEM changes its position on the scale.
FINAL_STATE	A description that presents the ITEM's state after the change in the ATTRIBUTE's value as an independent predication.
FINAL_VALUE	The position on the scale where the ITEM ends up.
INITIAL_STATE	A description that presents the ITEM's state before the change in the ATTRIBUTE's value as an independent predication.
INITIAL_VALUE	The initial position on the scale from which the ITEM moves away.
ITEM	The entity that has a position on the scale.
VALUE_RANGE	A portion of the scale, typically identified by its end points, along which the values of the ATTRIBUTE fluctuate.
Some Non-Core Roles	
DURATION	The length of time over which the change takes place.
SPEED	The rate of change of the VALUE.
GROUP	The GROUP in which an ITEM changes the value of an ATTRIBUTE in a specified way.

Example

ATTRIBUTE
DIFFERENCE
FINAL_STATE
FINAL_VALUE
INITIAL_STATE
INITIAL_VALUE
ITEM
VALUE_RANGE
DURATION
SPEED
GROUP

- [Oil] rose [in price] [by 2%].
- [It] has increased [to having them 1 day a month].
- [Microsoft shares] fell [to 7 5/8].
- [cancer incidence] fell [by 50%] [among men].
- a steady increase [from 9.5] [to 14.3] [in dividends].
- a [5%] [dividend] increase…

Find "Item" roles?

ATTRIBUTE
DIFFERENCE
FINAL_STATE
FINAL_VALUE
INITIAL_STATE
INITIAL_VALUE
ITEM
VALUE_RANGE
DURATION
SPEED
GROUP

- [Oil] rose [in price] [by 2%].
- [It] has increased [to having them] [1 day a month].
- [Microsoft shares] fell [to 7 5/8].
- [cancer incidence] fell [by 50%] [among men].
- a steady increase [from 9.5] [to 14.3] [in dividends].
- a [5%] [dividend] increase…

Find "Difference" & "Final_Value" roles?

ATTRIBUTE DIFFERENCE FINAL_STATE FINAL_VALUE INITIAL_STATE INITIAL_VALUE ITEM VALUE_RANGE
DURATION SPEED GROUP

- [Oil] rose [in price] [by 2%].
- [It] has increased [to having them] [1 day a month].
- [Microsoft shares] fell [to 7 5/8].
- [cancer incidence] fell [by 50%] [among men].
- a steady increase [from 9.5] [to 14.3] [in dividends].
- a [5%] [dividend] increase…

FrameNet (2004)

- Project at UC Berkeley led by Chuck Fillmore for developing a database of frames, general semantic concepts with an associated set of roles.
- Roles are specific to frames, which are “invoked” by the predicate, which can be a verb, noun, adjective, adverb
  - JUDGEMENT frame
    - Invoked by: V: blame, praise, admire; N: fault, admiration
    - Roles: JUDGE, EVALUEE, and REASON
- Specific frames chosen, and then sentences that employed these frames selected from the British National Corpus and annotated by linguists for semantic roles.
- Initial version: 67 frames, 49,013 sentences, 99,232 role fillers

PropBank
(proposition bank)

PropBank := proposition bank (2005)

- Project at Colorado led by Martha Palmer to add semantic roles to the Penn treebank.
- Proposition := verb + a set of roles
- Annotated over 1M words of Wall Street Journal text with existing gold-standard parse trees.
- Statistics:
  - 43,594 sentences 99,265 propositions
  - 3,324 unique verbs 262,281 role assignments

PropBank argument numbering

- Numbered roles, rather than named roles.
  - Arg0, Arg1, Arg2, Arg3, …
- Different numbering scheme for each **verb sense**.
- The general pattern of numbering is as follows.

- **Arg0** = "Proto-Agent" (agent)
- **Arg1** = "Proto-Patient" (direct object / theme / patient)
- Arg2 = indirect object (benefactive / instrument / attribute / end state)
- Arg3 = start point (benefactive / instrument / attribute)
- Arg4 = end point

Different “frameset” for each verb sense

- Mary left the room.
- Mary left her daughter-in-law her pearls in her will.

Frameset **leave.01** "move away from":
Arg0: entity leaving
Arg1: place left

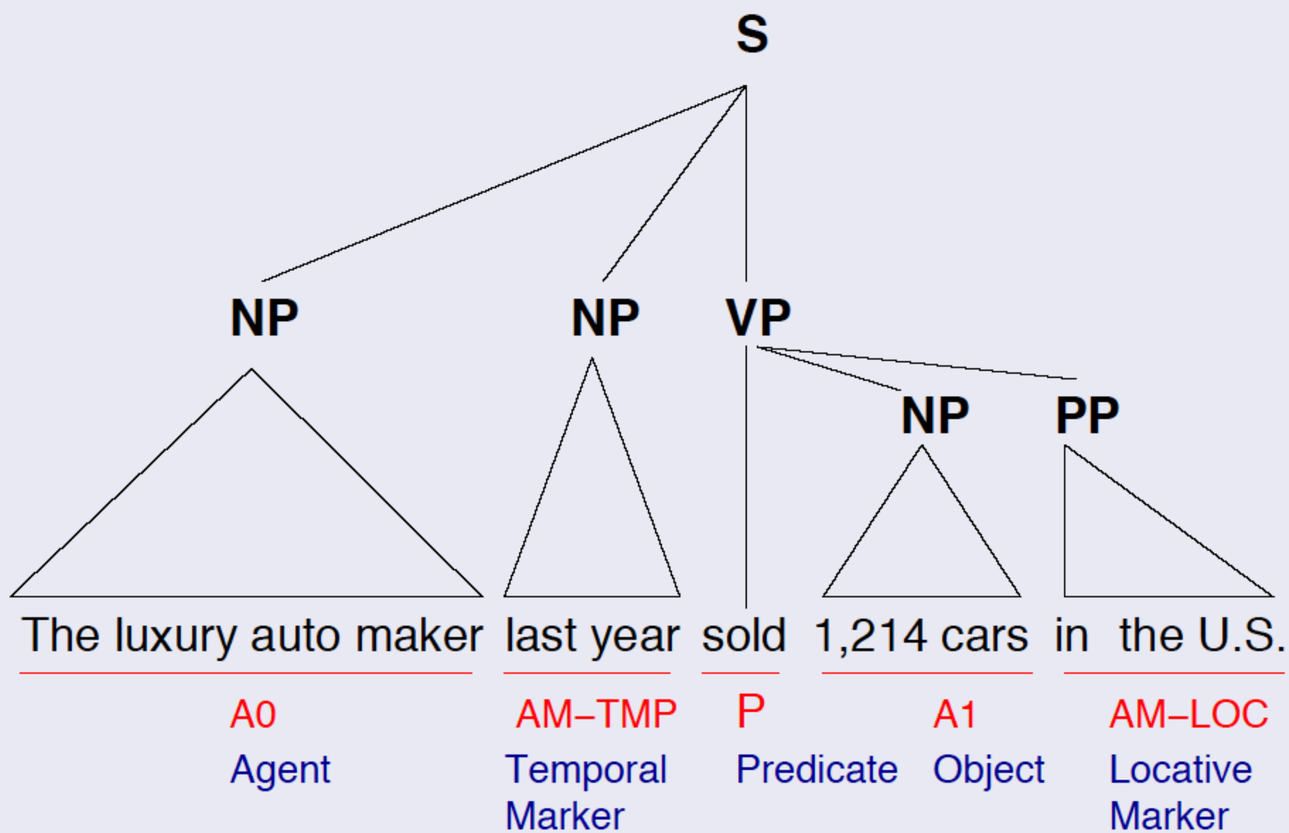
Frameset **leave.02** "give":
Arg0: giver
Arg1: thing given
Arg2: beneficiary

Semantic Role Labeling

Semantic Role Labeling (Task)

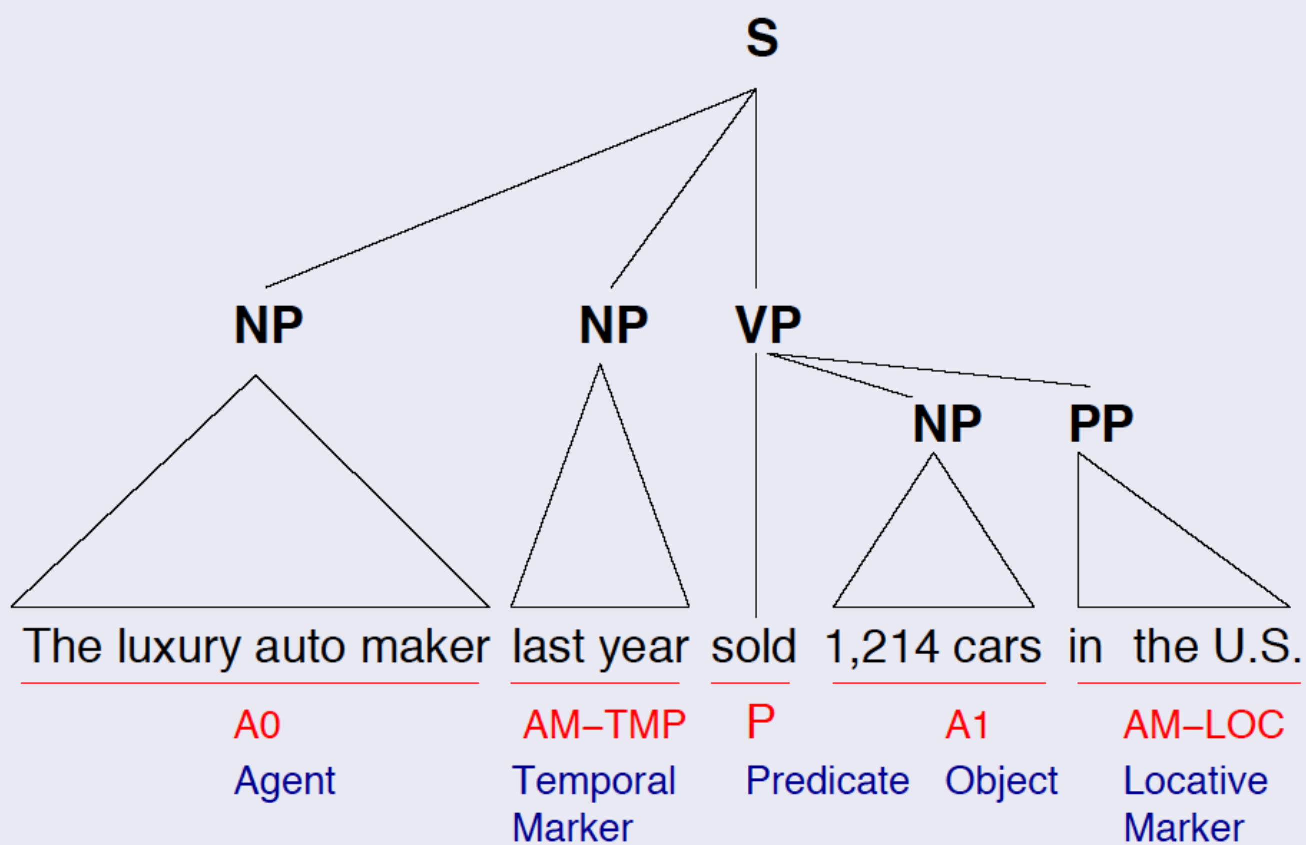
- Shallow meaning representation beyond syntactic parse trees
- Question Answering
  - “Who” questions usually use Agents
  - “What” question usually use Patients
  - “How” and “with what” questions usually use Instruments
  - “Where” questions frequently use Sources and Destinations.
  - “For whom” questions usually use Beneficiaries
  - “To whom” questions usually use Destinations
- Machine Translation Generation
  - Semantic roles are usually expressed using particular, distinct syntactic constructions in different languages.
- Summarization, Information Extraction

SRL $\stackrel{def}{=}$ detecting basic event structures such as *who* did *what* to *whom*, *when* and *where* [**IE** point of view]



Example from Lluis Marquez

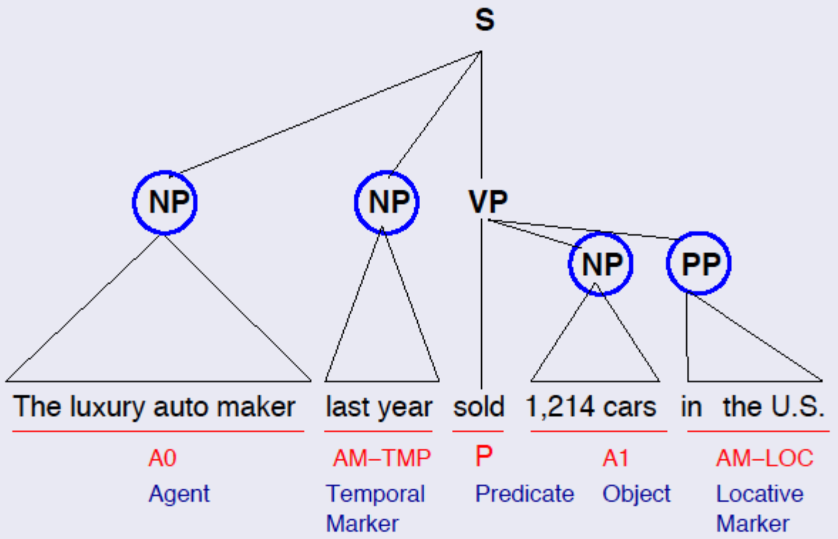
SRL $\stackrel{def}{=}$ identify the *arguments* of a given verb and assign them semantic labels describing the *roles* they play in the predicate (i.e., identify predicate argument structures) [**CL** point of view]



Example from Lluis Marquez

Linguistic nature of the problem

- Argument identification is strongly related to syntax



- Role labeling is a semantic task
  - e.g., selectional preferences should play an important role

Example from Lluis Marquez

SRL as Parse Node Classification

- Assume that a syntactic parse is available
- Treat problem as classifying parse-tree nodes.
- Can use any machine-learning classification method.
- Critical issue is engineering the right set of features for the classifier to use.

Color Code:

not-a-role

agent

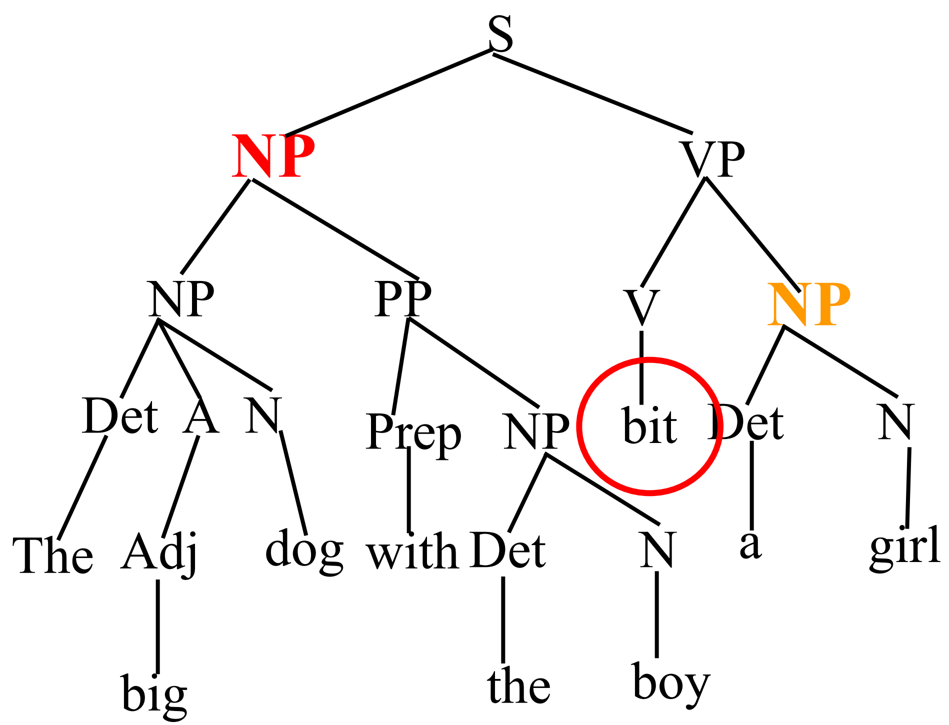
patient

source

destination

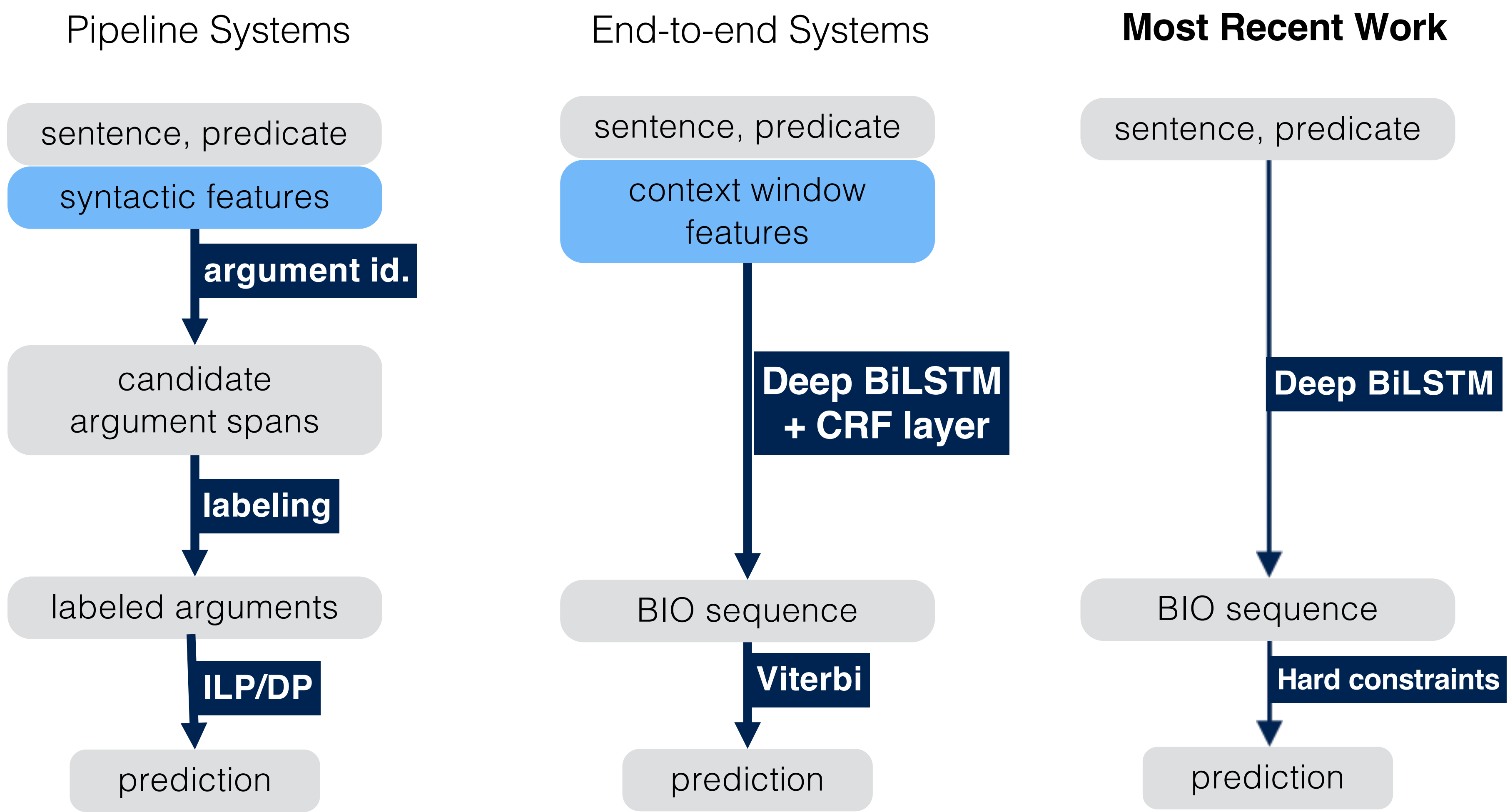
instrument

beneficiary



Deep Semantic Role Labeling

SRL Systems

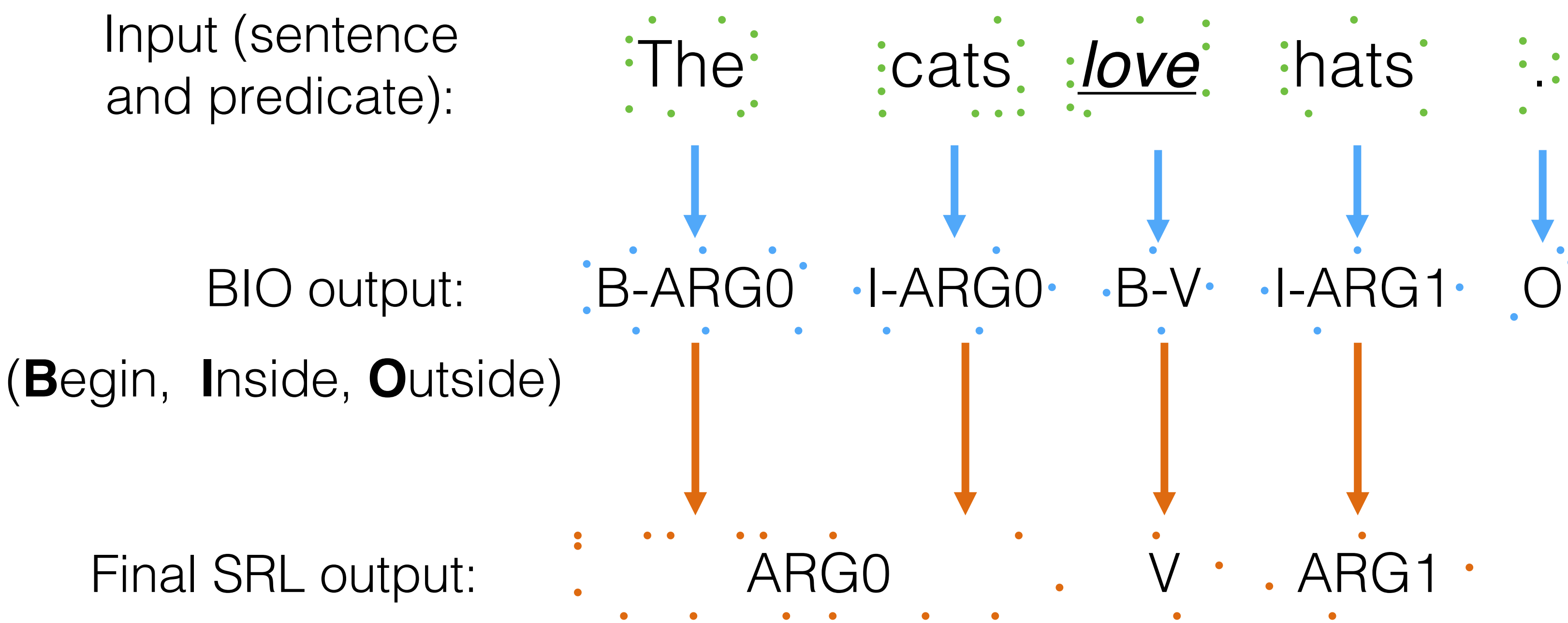


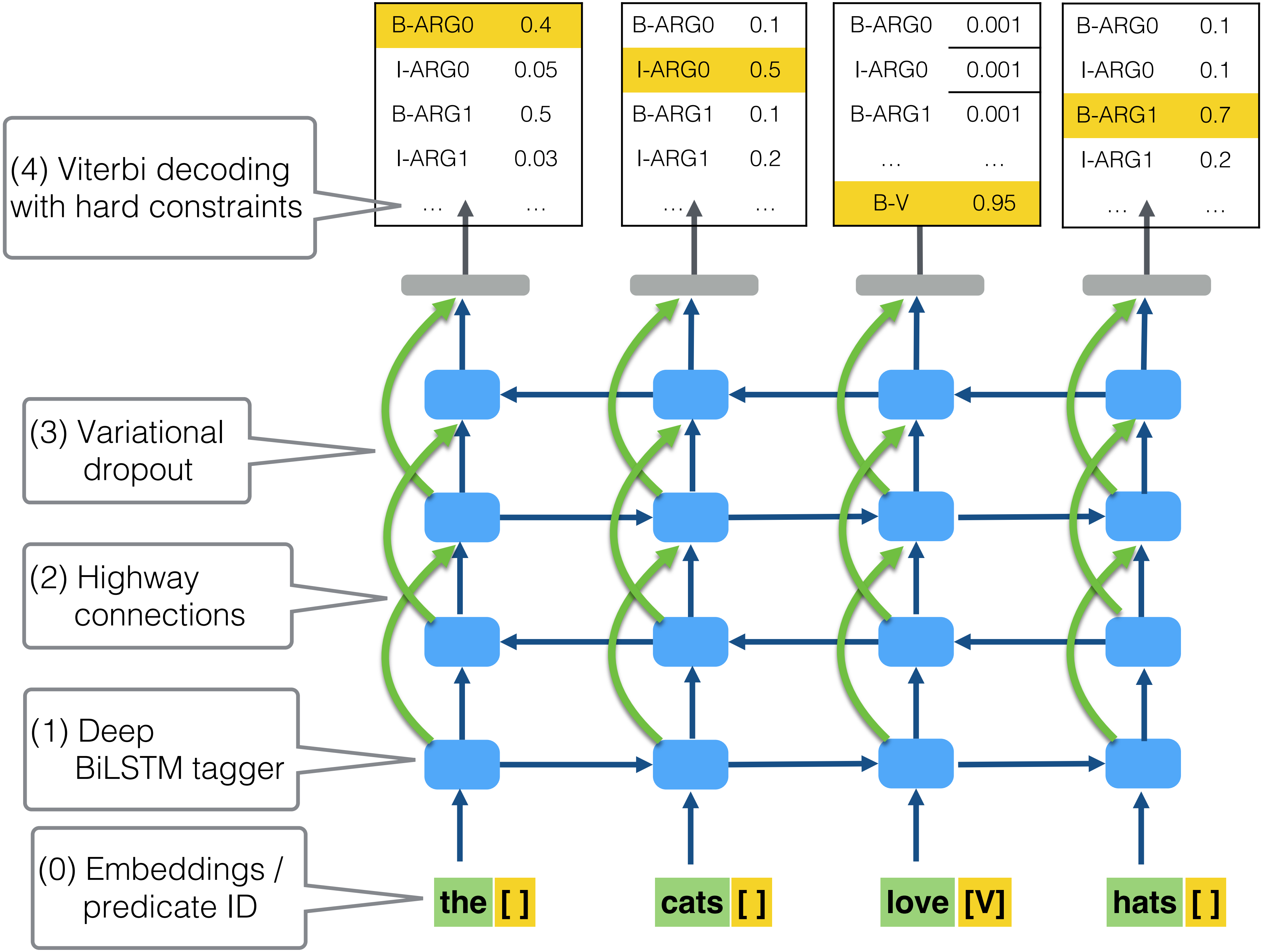
Punyakanok et al., 2008
Täckström et al., 2015
FitzGerald et al., 2015

Collobert et al., 2011
Zhou and Xu, 2015
Wang et. al, 2015

He et al. 2017, 2018

SRL as BIO Tagging Problem





B-ARG0 0.4
I-ARG0 0.05
B-ARG1 0.5
I-ARG1 0.03
… …
B-ARG0 0.1
I-ARG0 0.5
B-ARG1 0.1
I-ARG1 0.2
… …
B-ARG0 0.001
I-ARG0 0.001
B-ARG1 0.001
… …
B-V 0.95
B-ARG0 0.1
I-ARG0 0.1
B-ARG1 0.7
I-ARG1 0.2
… …
(4) Viterbi decoding with hard constraints
(3) Variational dropout
(2) Highway connections
(1) Deep BiLSTM tagger
(0) Embeddings / predicate ID
the [ ]
cats [ ]
love [V]
hats [ ]

Model - (2) Highway Connections

Trend: Deeper models for higher accuracy

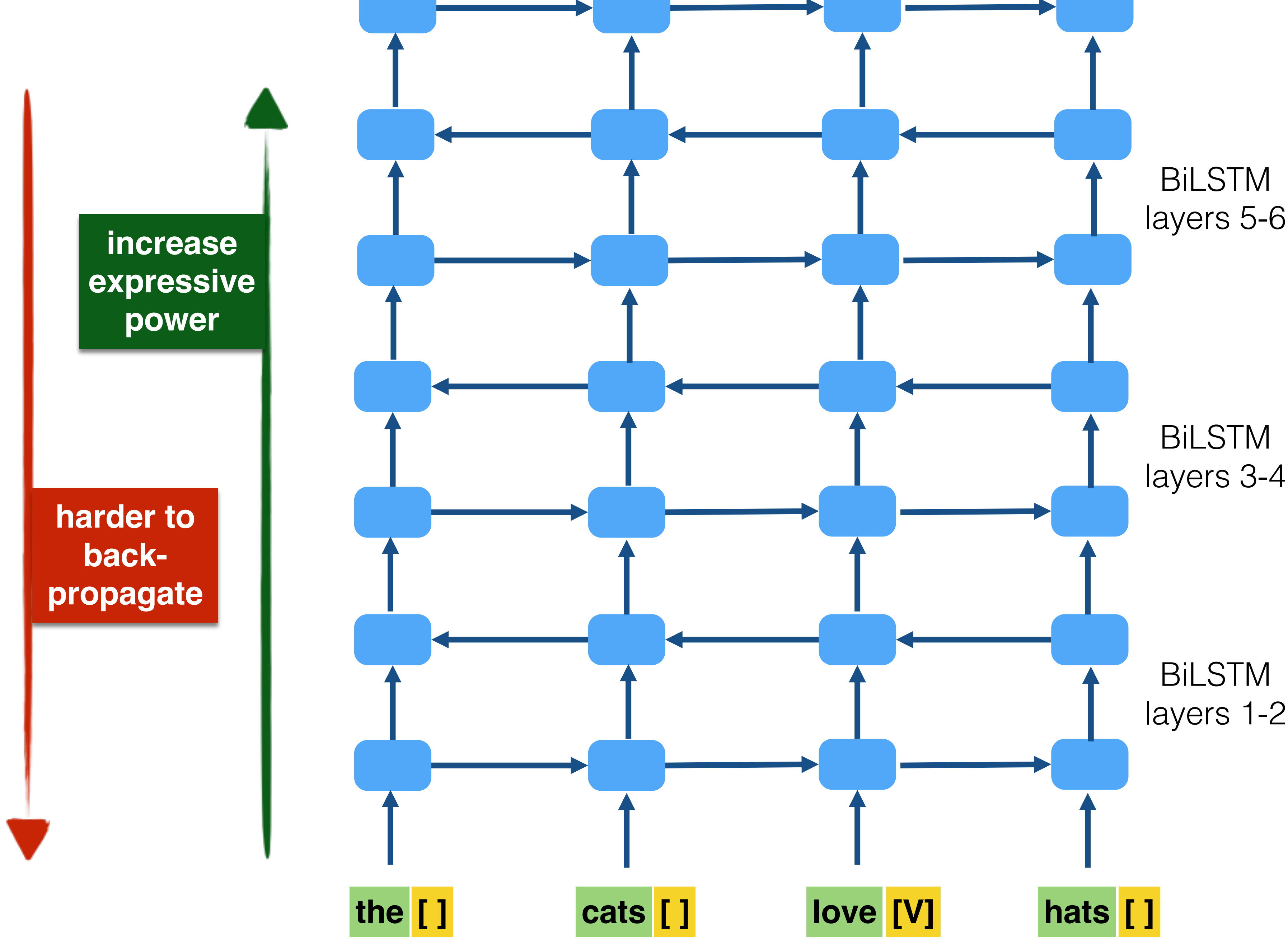
Grammar as a Foreign Language (Vinyals et al., 2014): **3** layers

End-to-end Semantic Role Labeling (Zhou and Xu, 2015): **8** layers

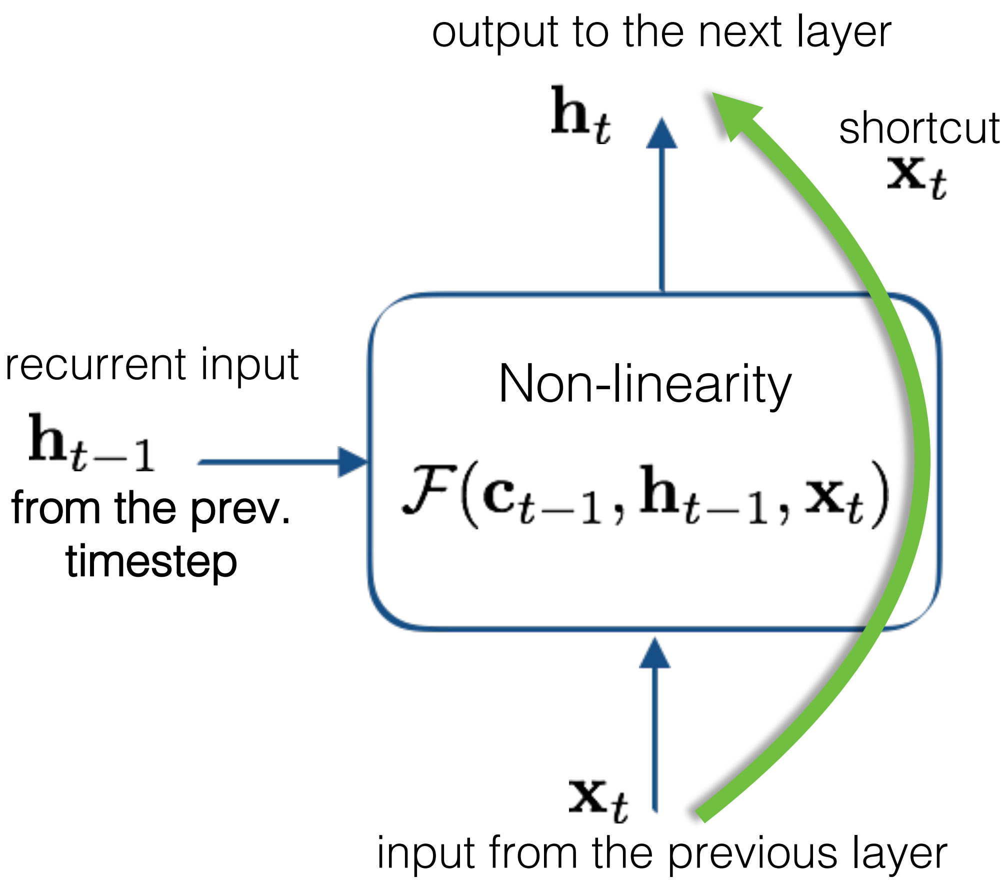
Google's Neural Machine Translation (GNMT, Wu et al., 2016): **8** layers

Deep Semantic Role Labeling (He et al 2017): **8** layers

Deep Residual Learning for Image Recognition (He et al, 2016): **152** layers


increase expressive power
harder to back-propagate
BiLSTM layers 5-6
BiLSTM layers 3-4
BiLSTM layers 1-2
the [ ]
cats [ ]
love [V]
hats [ ]

Model - (2) Highway Connections



new output:

residual net $\mathbf{h}_t + \mathbf{x}_t$

gated highway network:

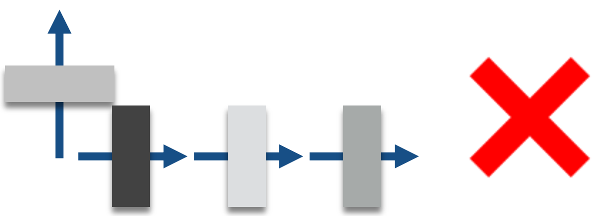
$$\mathbf{r}_t \circ \mathbf{h}_t + (1 - \mathbf{r}_t) \circ \mathbf{x}_t$$

$$\mathbf{r}_t = \sigma(f(\mathbf{h}_{t-1}, \mathbf{x}_t))$$

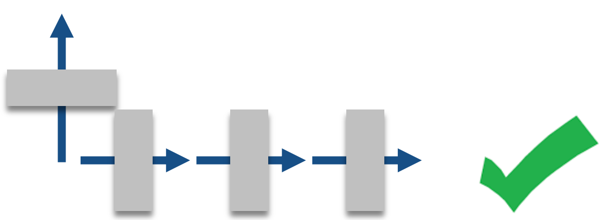
References:
Deep Residual Networks, Kaiming He, ICML 2016 Tutorial
Training Very Deep Networks, Srivastava et al., 2015

Model - (3) Variational Dropout

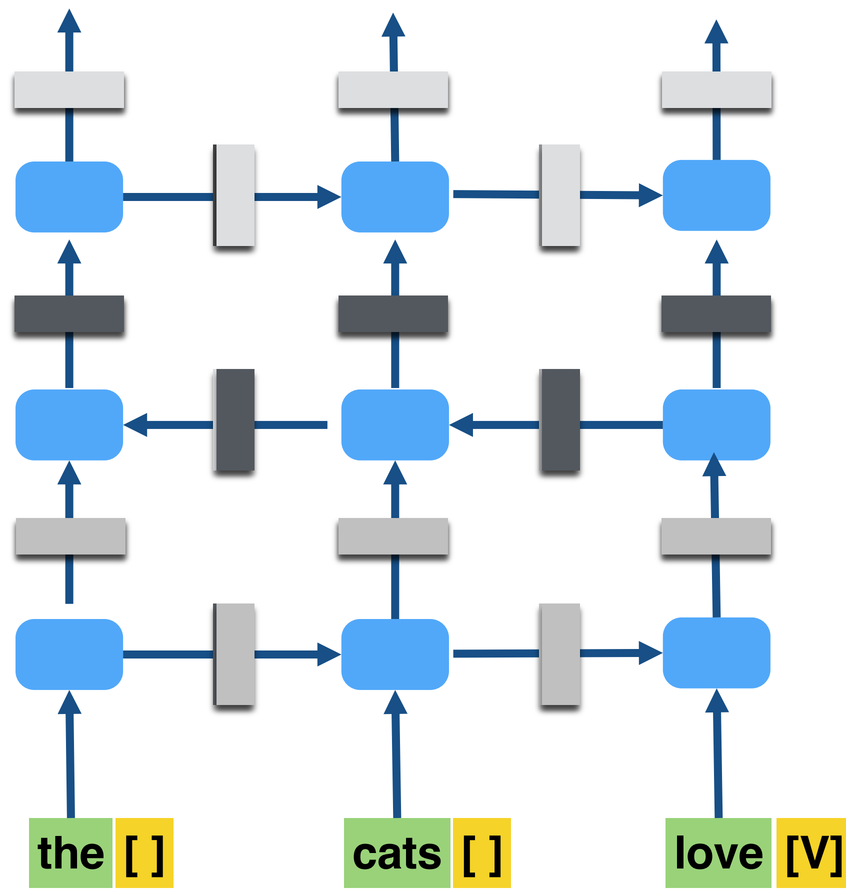
Traditionally, dropout masks are only applied to vertical connections.



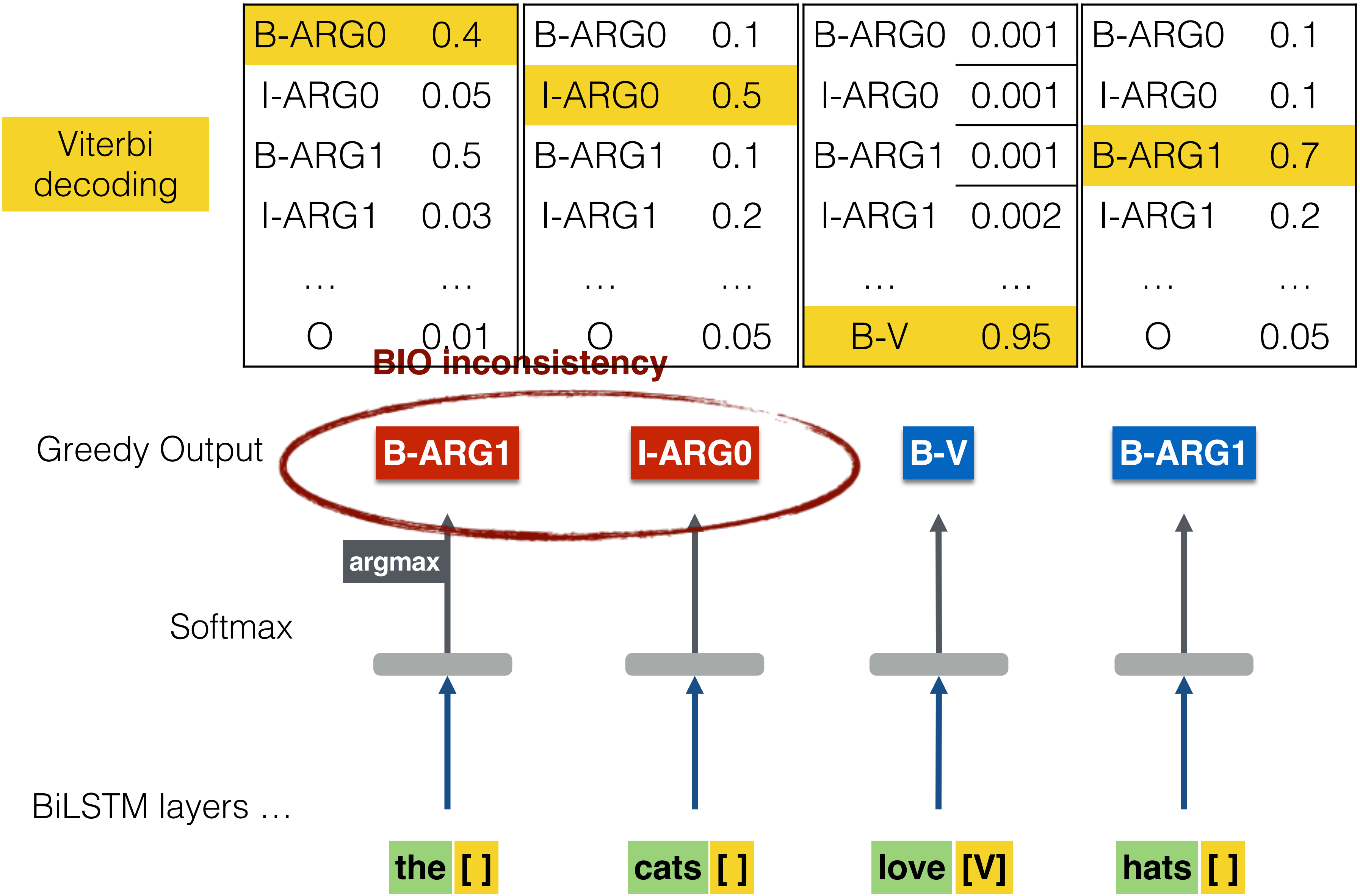
Applying dropout to recurrent connections causes too much noise amplification.



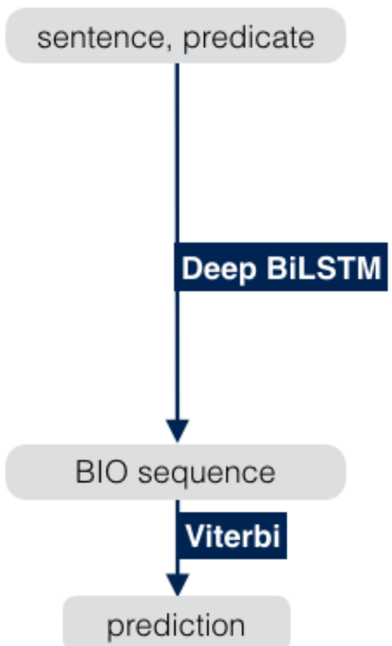
Variational dropout: Reuse the same dropout mask for each timestep.
Gal and Ghahramani, 2016



Model - (4) Viterbi Decoding with Hard Constraints

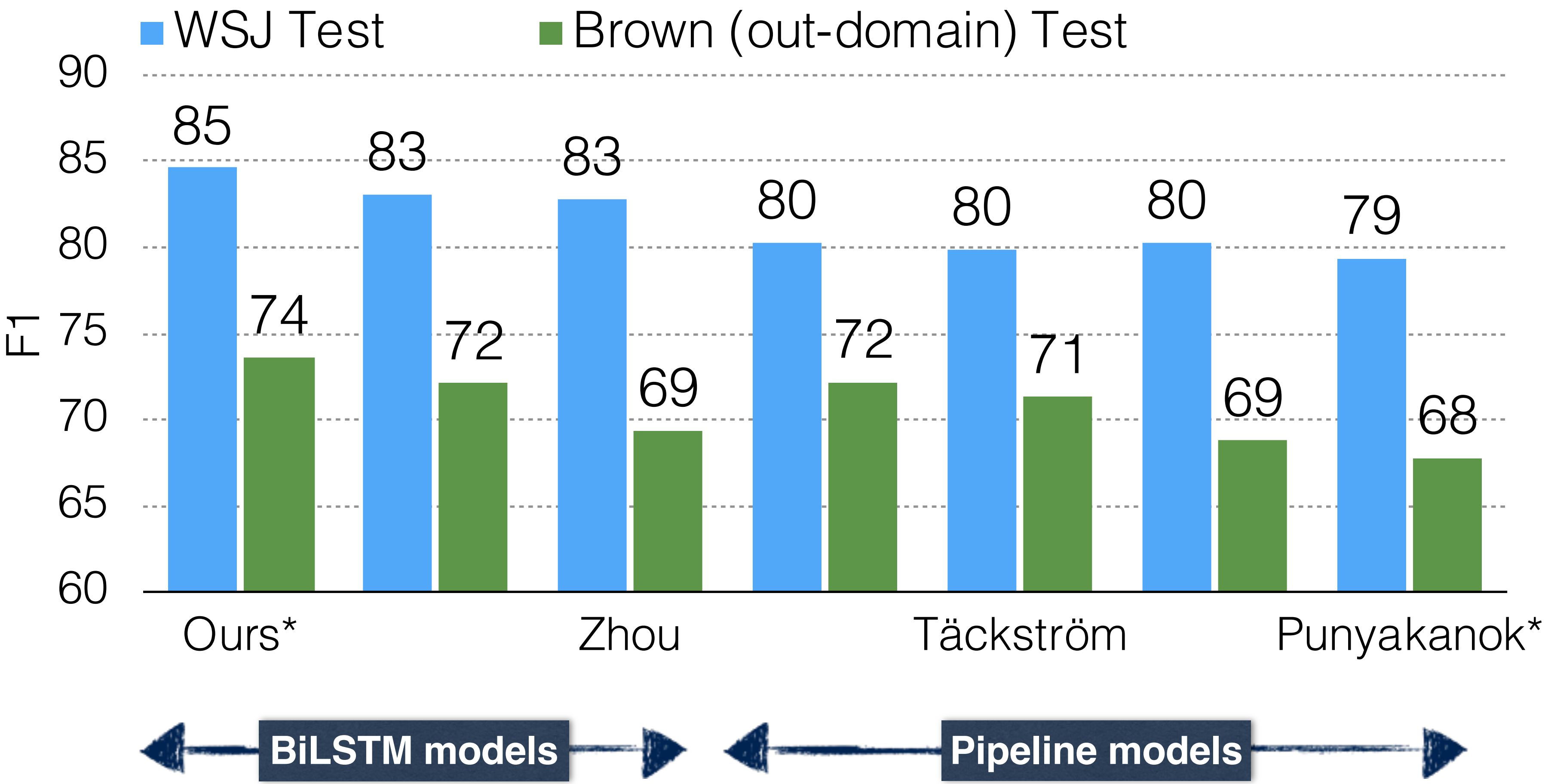


Other Implementation Details …



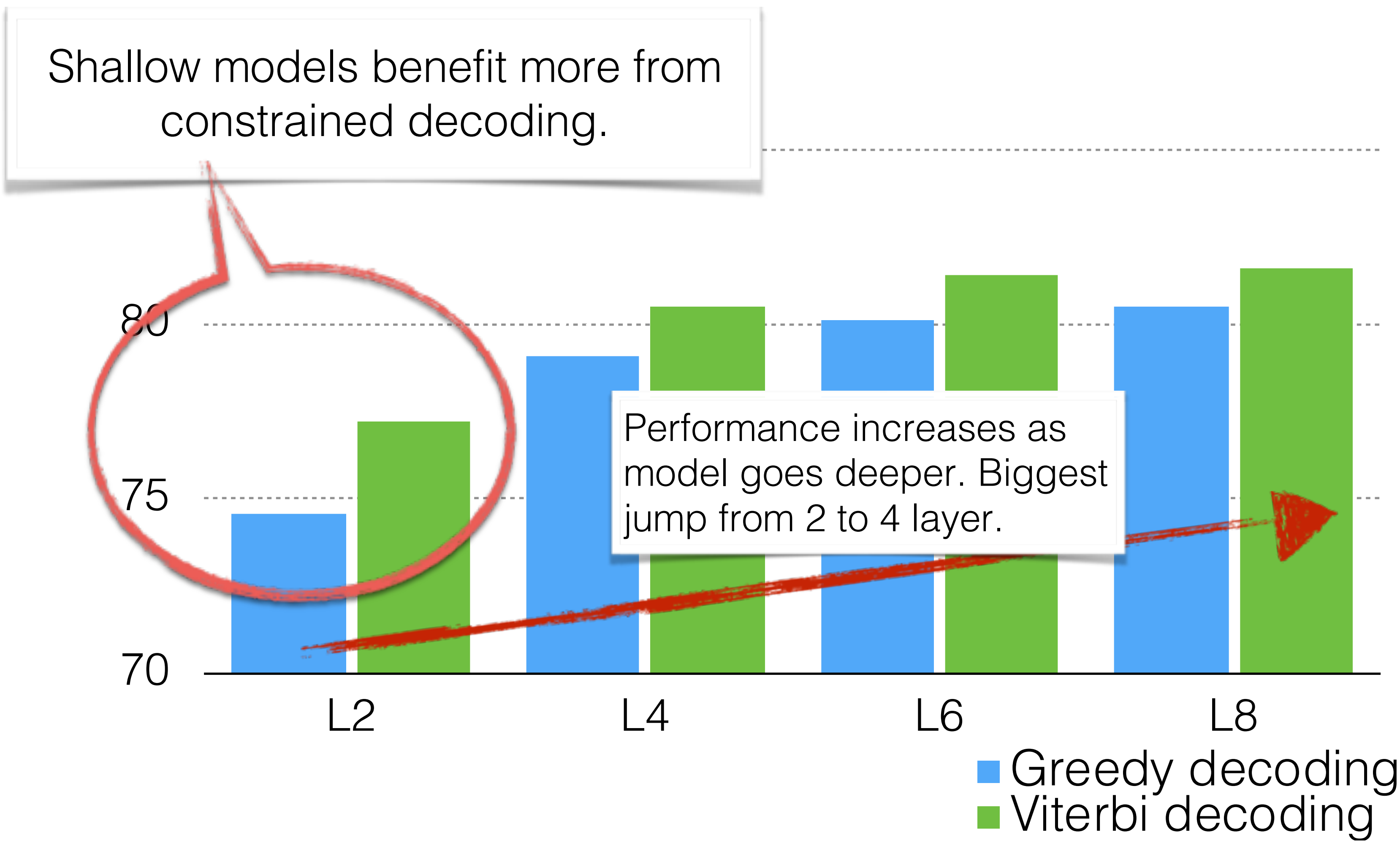
- 8 layer BiLSTMs with 300D hidden layers.
- 100D GloVe embeddings, updated during training.
- **Orthonormal initialization** for LSTM weight matrices (Saxe et al., 2013)
- 5 model ensemble with **product-of-experts** (Hinton 2002)
- Trained for 500 epochs.

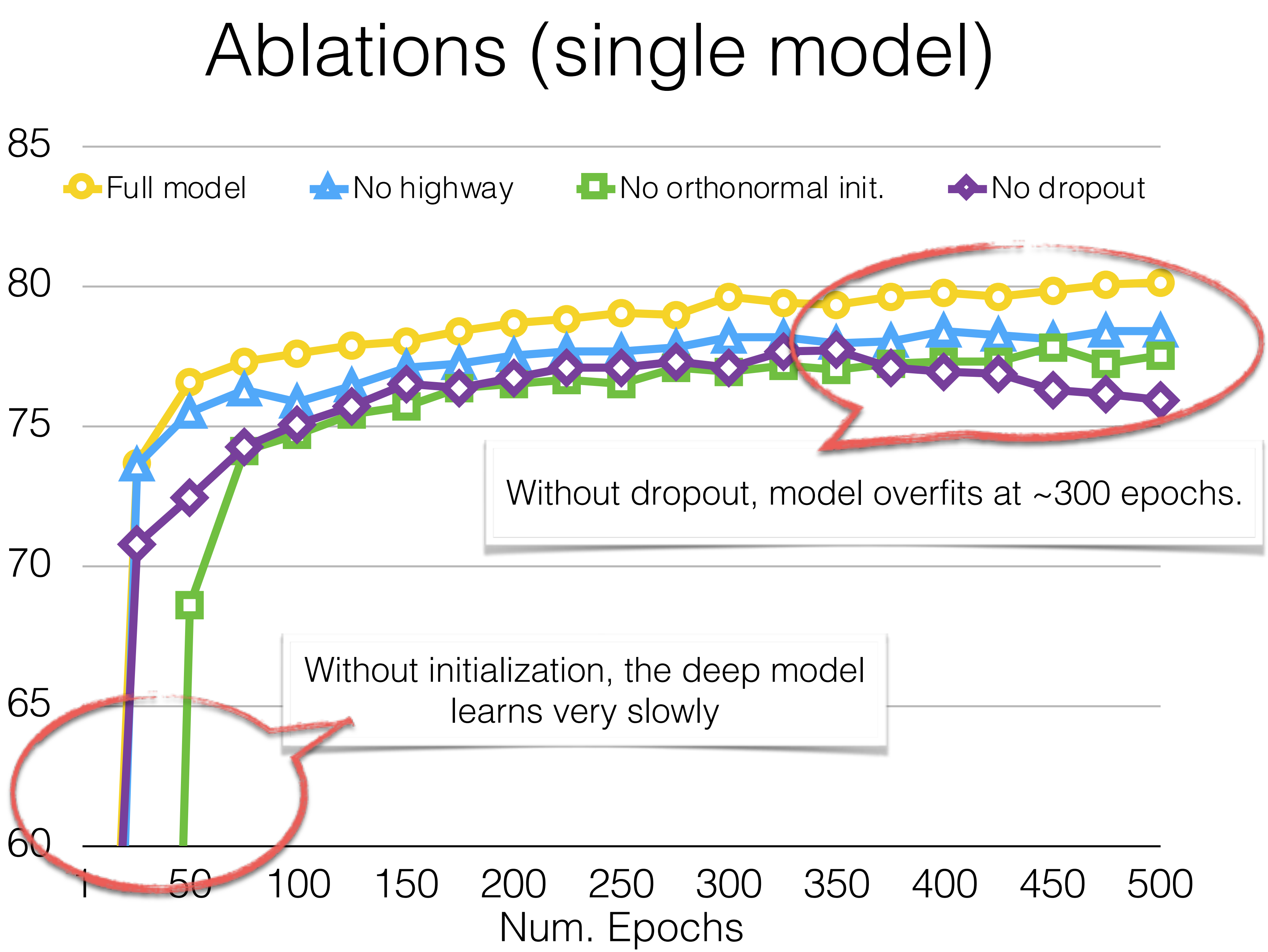
CoNLL 2005 Results



*:Ensemble models

Ablations on Number of Layers (2,4,6 and 8)





Ablations (single model)
85
80
75
70
65
60
Full model
No highway
No orthonormal init.
No dropout
Without dropout, model overfits at ~300 epochs.
Without initialization, the deep model learns very slowly
1
50
100
150
200
250
300
350
400
450
500
Num. Epochs