Unsupervised machine learning as acquisition modeling

Cory Shain, Dept of Linguistics, The Ohio State University

Collaborators: William Bryce, Finale Doshi-Velez, Micha Elsner, Lifeng Jin, Victoria Krakovna, Timoth Miller, William Schuler, Lane Schwartz

16 Aug. 2017, MIT

Introduction

Unsupervised machine learning as acquisition modeling

- Both humans and computers can learn (aspects of) language
- Human language acquisition is not well understood
- Computer "language acquisition" is well understood
- Can we use machine learning to shed light on human learning?

Unsupervised machine learning as acquisition modeling

- Both humans and computers can learn (aspects of) language
- Human language acquisition is not well understood
- Computer "language acquisition" is well understood
- Can we use machine learning to shed light on human learning?

Unsupervised machine learning as acquisition modeling

- Both humans and computers can learn (aspects of) language
- Human language acquisition is not well understood
- Computer "language acquisition" is well understood
- Can we use machine learning to shed light on human learning?

Unsupervised machine learning as acquisition modeling

- Both humans and computers can learn (aspects of) language
- Human language acquisition is not well understood
- Computer “language acquisition” is well understood
- Can we use machine learning to shed light on human learning?

Roadmap for this talk

- Motivation
- Modeling lexical acquisition with unsupervised speech segmentation
- Modeling grammar acquisition with unsupervised PCFG induction
- Discussion and future directions

Roadmap for this talk

- Motivation
- Modeling lexical acquisition with unsupervised speech segmentation
- Modeling grammar acquisition with unsupervised PCFG induction
- Discussion and future directions

Roadmap for this talk

- Motivation
- Modeling lexical acquisition with unsupervised speech segmentation
- Modeling grammar acquisition with unsupervised PCFG induction
- Discussion and future directions

Roadmap for this talk

- Motivation
- Modeling lexical acquisition with unsupervised speech segmentation
- Modeling grammar acquisition with unsupervised PCFG induction
- Discussion and future directions

Motivation

Why model language acquisition computationally?

Of interest to both science and engineering

Why model language acquisition computationally?

- Science:
  - Test predictions of hypotheses about language acquisition
  - Dissect the language learning problem
  - Explore learnability of linguistic phenomena

Why model language acquisition computationally?

- Science:
  - Test predictions of hypotheses about language acquisition
  - Dissect the language learning problem
  - Explore learnability of linguistic phenomena

Why model language acquisition computationally?

- Science:
  - Test predictions of hypotheses about language acquisition
  - Dissect the language learning problem
  - Explore learnability of linguistic phenomena

Why model language acquisition computationally?

- Engineering:
  - Humans are better than computers at learning and using language
  - We learn from cheap and abundant sources of data
  - Low-resource NLP
  - Study and preservation of endangered languages

Why model language acquisition computationally?

- Engineering:
  - Humans are better than computers at learning and using language
  - We learn from cheap and abundant sources of data
  - Low-resource NLP
  - Study and preservation of endangered languages

Why model language acquisition computationally?

- Engineering:
  - Humans are better than computers at learning and using language
  - We learn from cheap and abundant sources of data
  - Low-resource NLP
  - Study and preservation of endangered languages

Why model language acquisition computationally?

- Engineering:
  - Humans are better than computers at learning and using language
  - We learn from cheap and abundant sources of data
  - Low-resource NLP
  - Study and preservation of endangered languages

Why model language acquisition computationally?

- Engineering:
  - Humans are better than computers at learning and using language
  - We learn from cheap and abundant sources of data
  - Low-resource NLP
  - Study and preservation of endangered languages

But...

NLP is (usually) cognitively implausible

- Normally requires lots of annotated data
  - Microsoft's Bing used 2100 hours of transcribed speech to train its speech recognizer (Dahl et al. 2011)
  - Humans don't have direct access to the right answers
- Unrealistically large memory capacity
- Non-incremental processing

NLP is (usually) cognitively implausible

- Normally requires lots of annotated data
  - Microsoft's Bing used 2100 hours of transcribed speech to train its speech recognizer (Dahl et al. 2011)
  - Humans don't have direct access to the right answers
- Unrealistically large memory capacity
  - Human working memory constraints are severe compared to those of computers (Miller 1956; Cowan 2001; McElree 2001)
- Non-incremental processing

NLP is (usually) cognitively implausible

- Normally requires lots of annotated data
  - Microsoft's Bing used 2100 hours of transcribed speech to train its speech recognizer (Dahl et al. 2011)
  - Humans don't have direct access to the right answers
- Unrealistically large memory capacity
  - Human working memory constraints are severe compared to those of computers (Miller 1956; Cowan 2001; McElree 2001)
- Non-incremental processing

NLP is (usually) cognitively implausible

- Normally requires lots of annotated data
  - Microsoft's Bing used 2100 hours of transcribed speech to train its speech recognizer (Dahl et al. 2011)
  - Humans don't have direct access to the right answers
- Unrealistically large memory capacity
  - Human working memory constraints are severe compared to those of computers (Miller 1956; Cowan 2001; McElree 2001)
- Non-incremental processing
  - Humans process language incrementally (Marslen-Wilson 1975; Tanenhaus et al. 1995)

NLP is (usually) cognitively implausible

- Normally requires lots of annotated data
  - Microsoft's Bing used 2100 hours of transcribed speech to train its speech recognizer (Dahl et al. 2011)
  - Humans don't have direct access to the right answers
- Unrealistically large memory capacity
  - Human working memory constraints are severe compared to those of computers (Miller 1956; Cowan 2001; McElree 2001)
- Non-incremental processing
  - Humans process language incrementally (Marslen-Wilson 1975; Tanenhaus et al. 1995)

NLP is (usually) cognitively implausible

- Normally requires lots of annotated data
  - Microsoft's Bing used 2100 hours of transcribed speech to train its speech recognizer (Dahl et al. 2011)
  - Humans don't have direct access to the right answers
- Unrealistically large memory capacity
  - Human working memory constraints are severe compared to those of computers (Miller 1956; Cowan 2001; McElree 2001)
- Non-incremental processing
  - Humans process language incrementally (Marslen-Wilson 1975; Tanenhaus et al. 1995)

NLP is (usually) cognitively implausible

- Normally requires lots of annotated data
  - Microsoft's Bing used 2100 hours of transcribed speech to train its speech recognizer (Dahl et al. 2011)
  - Humans don't have direct access to the right answers
- Unrealistically large memory capacity
  - Human working memory constraints are severe compared to those of computers (Miller 1956; Cowan 2001; McElree 2001)
- Non-incremental processing
  - Humans process language incrementally (Marslen-Wilson 1975; Tanenhaus et al. 1995)

Unsupervised NLP uses a more plausible training objective

- Unsupervised = no direct access to the right answers
- More similar training feedback to that received by humans
- Can still be cognitively implausible in other ways
- Cognitively-constrained unsupervised NLP can shed light on language acquisition

Unsupervised NLP uses a more plausible training objective

- Unsupervised = no direct access to the right answers
- More similar training feedback to that received by humans
- Can still be cognitively implausible in other ways
- Cognitively-constrained unsupervised NLP can shed light on language acquisition

Unsupervised NLP uses a more plausible training objective

- Unsupervised = no direct access to the right answers
- More similar training feedback to that received by humans
- Can still be cognitively implausible in other ways
- Cognitively-constrained unsupervised NLP can shed light on language acquisition

Unsupervised NLP uses a more plausible training objective

- Unsupervised = no direct access to the right answers
- More similar training feedback to that received by humans
- Can still be cognitively implausible in other ways
- Cognitively-constrained unsupervised NLP can shed light on language acquisition

This work: Modeling lexical and grammar acquisition

- Models of two related acquisition tasks:
  - **Lexical acquisition:** Learning to segment the speech signal
  - **Grammar acquisition:** Learning to parse

This work: Modeling lexical and grammar acquisition

- Models of two related acquisition tasks:
  - **Lexical acquisition:** Learning to segment the speech signal
  - **Grammar acquisition:** Learning to parse

This work: Modeling lexical and grammar acquisition

- Models of two related acquisition tasks:
  - **Lexical acquisition:** Learning to segment the speech signal
  - **Grammar acquisition:** Learning to parse

This work: Modeling lexical and grammar acquisition

- Both of these models accept arbitrary naturally-occurring training data in any language
- C.f. computational models that use “toy” input (e.g. Elman 1991; Briscoe 2000; Fodor and Sakas 2004)
- Our approach more accurately represents input to human learners

This work: Modeling lexical and grammar acquisition

- Both of these models accept arbitrary naturally-occurring training data in any language
- C.f. computational models that use “toy” input (e.g. Elman 1991; Briscoe 2000; Fodor and Sakas 2004)
- Our approach more accurately represents input to human learners

This work: Modeling lexical and grammar acquisition

- Both of these models accept arbitrary naturally-occurring training data in any language
- C.f. computational models that use "toy" input (e.g. Elman 1991; Briscoe 2000; Fodor and Sakas 2004)
- Our approach more accurately represents input to human learners

Modeling lexical acquisition with unsupervised speech segmentation

Speech segmentation: Cognitive background

- Phonological memory limits may encourage sparse encodings (Baddeley and Hitch 1974)
- Thought to affect learning as well as processing (Baddeley, Gathercole, and Papagno 1998)
- We model this learning pressure by seeking compressible segmentations

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Cognitive background

- Phonological memory limits may encourage sparse encodings (Baddeley and Hitch 1974)
- Thought to affect learning as well as processing (Baddeley, Gathercole, and Papagno 1998)
- We model this learning pressure by seeking compressible segmentations

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Cognitive background

- Phonological memory limits may encourage sparse encodings (Baddeley and Hitch 1974)
- Thought to affect learning as well as processing (Baddeley, Gathercole, and Papagno 1998)
- We model this learning pressure by seeking compressible segmentations

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Model overview

- Two RNN's:
  - Auto-encoder (AE) network: Reconstructs its input
  - Proposal network: Predicts segmentation points

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Model overview

- Two RNN's:
  - Auto-encoder (AE) network: Reconstructs its input
  - Proposal network: Predicts segmentation points

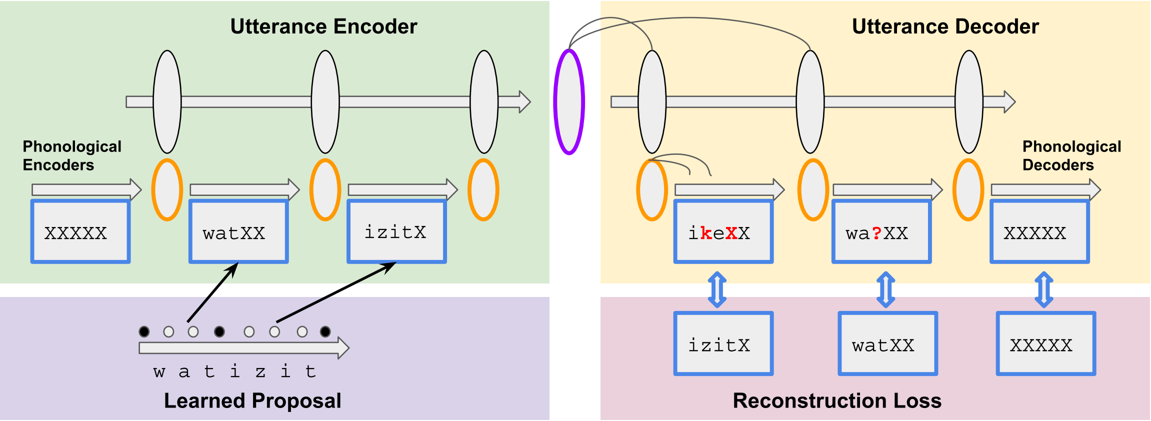
Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Model overview

- Two RNN's:
  - Auto-encoder (AE) network: Reconstructs its input
  - Proposal network: Predicts segmentation points

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Auto-encoder network architecture



Micha Elsner and Cory Shain (to appear). “Speech segmentation with a neural encoder model of working memory”. In: *EMNLP 2017*

Speech segmentation: Segmenter network architecture

- LSTM trained to predict segmentation probability at each time step

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Model overview

- Overall segmentation loss is non-differentiable (segmentation decisions are hard)
- Estimated via importance sampling (e.g. Mnih et al. 2014; Xu et al. 2015), using reconstruction loss for scoring

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Model overview

- Overall segmentation loss is non-differentiable (segmentation decisions are hard)
- Estimated via importance sampling (e.g. Mnih et al. 2014; Xu et al. 2015), using reconstruction loss for scoring

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Model overview

- Memory limits simulated using:
  - Dropout
    - Simulates forgetting
    - Phonemes drop at rate D_p, words drop at rate D_w
  - LSTM hidden state size

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Model overview

- Memory limits simulated using:
  - Dropout
    - Simulates forgetting
    - Phonemes drop at rate D_p, words drop at rate D_u
  - LSTM hidden state size

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Model overview

- Memory limits simulated using:
  - Dropout
    - Simulates forgetting
    - Phonemes drop at rate D_p, words drop at rate D_u
  - LSTM hidden state size
    - Simulates capacity limitations
    - Number of hidden units in phonological (H_p) and utterance (H_u) auto-encoders
    - Too little memory might cause total reconstruction failure
    - Too much memory might not encourage efficient segmentations

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Model overview

- Memory limits simulated using:
  - Dropout
    - Simulates forgetting
    - Phonemes drop at rate D_p, words drop at rate D_u
  - LSTM hidden state size
    - Simulates capacity limitations
    - Number of hidden units in phonological (H_p) and utterance (H_u) autoencoders
    - Too little memory might cause total reconstruction failure
    - Too much memory might not encourage efficient segmentations

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Model overview

- Memory limits simulated using:
  - Dropout
    - Simulates forgetting
    - Phonemes drop at rate D_p, words drop at rate D_u
  - LSTM hidden state size
    - Simulates capacity limitations
    - Number of hidden units in phonological (H_p) and utterance (H_u) auto-encoders
    - Too little memory might cause total reconstruction failure
    - Too much memory might not encourage efficient segmentations

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Model overview

- Memory limits simulated using:
  - Dropout
    - Simulates forgetting
    - Phonemes drop at rate D_p, words drop at rate D_u
  - LSTM hidden state size
    - Simulates capacity limitations
    - Number of hidden units in phonological (H_p) and utterance (H_u) auto-encoders
    - Too little memory might cause total reconstruction failure
    - Too much memory might not encourage efficient segmentations

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Model overview

- Memory limits simulated using:
  - Dropout
    - Simulates forgetting
    - Phonemes drop at rate D_p, words drop at rate D_u
  - LSTM hidden state size
    - Simulates capacity limitations
    - Number of hidden units in phonological (H_p) and utterance (H_u) auto-encoders
    - Too little memory might cause total reconstruction failure
    - Too much memory might not encourage efficient segmentations

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Model overview

- Memory limits simulated using:
  - Dropout
    - Simulates forgetting
    - Phonemes drop at rate D_p, words drop at rate D_u
  - LSTM hidden state size
    - Simulates capacity limitations
    - Number of hidden units in phonological (H_p) and utterance (H_u) auto-encoders
    - Too little memory might cause total reconstruction failure
    - Too much memory might not encourage efficient segmentations

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Model overview

- Memory limits simulated using:
  - Dropout
    - Simulates forgetting
    - Phonemes drop at rate D_p, words drop at rate D_u
  - LSTM hidden state size
    - Simulates capacity limitations
    - Number of hidden units in phonological (H_p) and utterance (H_u) auto-encoders
    - Too little memory might cause total reconstruction failure
    - Too much memory might not encourage efficient segmentations

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Input

- Architecture is very flexible
- Can accept any vectorial representation of the input sequence
- Characters (1-hot)
- Acoustic features (MFCC)
- First system to perform unsupervised segmentation of either text or acoustics using same code base

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Input

- Architecture is very flexible
- Can accept any vectorial representation of the input sequence
- Characters (1-hot)
- Acoustic features (MFCC)
- First system to perform unsupervised segmentation of either text or acoustics using same code base

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Input

- Architecture is very flexible
- Can accept any vectorial representation of the input sequence
- Characters (1-hot)
- Acoustic features (MFCC)
- First system to perform unsupervised segmentation of either text or acoustics using same code base

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Input

- Architecture is very flexible
- Can accept any vectorial representation of the input sequence
- Characters (1-hot)
- Acoustic features (MFCC)
- First system to perform unsupervised segmentation of either text or acoustics using same code base

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Input

+ Architecture is very flexible
+ Can accept any vectorial representation of the input sequence
+ Characters (1-hot)
+ Acoustic features (MFCC)
+ First system to perform unsupervised segmentation of either text or acoustics using same code base

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Experiments

- Text: Brent corpus (Brent 1999)
- Acoustics: Zerospeech '15 English (Versteegh et al. 2015)

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Experiments

- Text: Brent corpus (Brent 1999)
- Acoustics: Zerospeech '15 English (Versteegh et al. 2015)

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Results (Brent)

	Bd P	Bd R	Bd F	Wd F
Our system	81	85	83	72

- Examples:
  - yu wanttu si D6bUk
    You wantto see thebook?
  - oke yusIt D* &nd 9l pUty) Suz b&kan
    Okay, yousit there and I'll putyour shoes backon
  - &nd lUk&t WAt D6kItiz pleIN wIT
    And lookat what thekitty's playing with
  - dId yu kQnt Ol6v DEm
    Did you count allof them?
  - wan6 lUk&t 6nADR bUk
    Wanna lookat another book?
  - If pApi b9ts pApi gEtss p&Nkt
    If puppy bites, puppy getss panked.

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Results (Brent)

	Bd P	Bd R	Bd F	Wd F
Our system	81	85	83	72

- Examples:
  - yu wanttu si D6bUk
    You wantto see thebook?
  - oke yusIt D* &nd 9l pUty) Suz b&kan
    Okay, yousit there and I'll putyour shoes backon
  - &nd lUk&t WAt D6kItiz pleIN wIT
    And lookat what thekitty's playing with
  - dId yu kQnt Ol6v DEm
    Did you count allof them?
  - wan6 lUk&t 6nADR bUk
    Wanna lookat another book?
  - If pApi b9ts pApi gEtss p&Nkt
    If puppy bites, puppy getss panked.

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Results (Brent)

	Bd P	Bd R	Bd F	Wd F
Our system	81	85	83	72

- Examples:
  - yu wanttu si D6bUk
    You wantto see thebook?
  - oke yusIt D* &nd 9I pUty) Suz b&kan
    Okay, yousit there and I'll putyour shoes backon
  - &nd lUk&t WAt D6kItiz pleIN wIT
    And lookat what thekitty's playing with
  - dId yu kQnt Ol6v DEm
    Did you count allof them?
  - wan6 lUk&t 6nADR bUk
    Wanna lookat another book?
  - If pApi b9ts pApi gEtss p&Nkt
    If puppy bites, puppy getss panked.

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Results (Brent)

	Bd P	Bd R	Bd F	Wd F
Our system	81	85	83	72

- Examples:
  - yu wanttu si D6bUk
    You wantto see thebook?
  - oke yusIt D* &nd 9l pUty) Suz b&kan
    Okay, yousit there and I'll putyour shoes backon
  - &nd lUk&t WAt D6kItiz pleIN wIT
    And lookat what thekitty's playing with
  - dId yu kQnt Ol6v DEm
    Did you count allof them?
  - wan6 lUk&t 6nADR bUk
    Wanna lookat another book?
  - If pApi b9ts pApi gEtss p&Nkt
    If puppy bites, puppy getss panked.

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Results (Brent)

	Bd P	Bd R	Bd F	Wd F
Our system	81	85	83	72

- Examples:
  - yu wanttu si D6bUk
    You wantto see thebook?
  - oke yusIt D* &nd 9l pUty) Suz b&kan
    Okay, yousit there and I'll putyour shoes backon
  - &nd lUk&t WAt D6kItiz pleIN wIT
    And lookat what thekitty's playing with
  - dId yu kQnt Ol6v DEm
    Did you count allof them?
  - wan6 lUk&t 6nADR bUk
    Wanna lookat another book?
  - If pApi b9ts pApi gEtss p&Nkt
    If puppy bites, puppy getss panked.

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Results (Brent)

	Bd P	Bd R	Bd F	Wd F
Our system	81	85	83	72

- Examples:
  - yu wanttu si D6bUk
    You wantto see thebook?
  - oke yusIt D* &nd 9I pUty) Suz b&kan
    Okay, yousit there and I'll putyour shoes backon
  - &nd lUk&t WAt D6kItiz pleIN wIT
    And lookat what thekitty's playing with
  - dId yu kQnt Ol6v DEm
    Did you count allof them?
  - wan6 lUk&t 6nADR bUk
    Wanna lookat another book?
  - If pApi b9ts pApi gEtss p&Nkt
    If puppy bites, puppy getss panked.

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Results (Brent)

	Bd P	Bd R	Bd F	Wd F
Our system	81	85	83	72

- Examples:
  - yu wanttu si D6bUk
    You wantto see thebook?
  - oke yusIt D* &nd 9l pUty) Suz b&kan
    Okay, yousit there and I'll putyour shoes backon
  - &nd lUk&t WAt D6kItiz pleIN wIT
    And lookat what thekitty's playing with
  - dId yu kQnt Ol6v DEm
    Did you count allof them?
  - wan6 lUk&t 6nADR bUk
    Wanna lookat another book?
  - If pApi b9ts pApi gEtss p&Nkt
    If puppy bites, puppy getss panked.

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Results (Brent)

System	Bd P	Bd R	Bd F	Wd F
Goldwater 09	90	74	87	74
Johnson 09	-	-	-	88
Berg-Kirkpatrick 10	-	-	-	88
Fleck 08	95	74	83	71
Our system	81	85	83	72

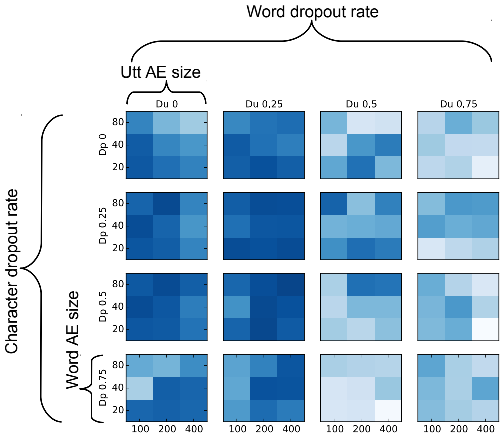
Micha Elsner and Cory Shain (to appear). “Speech segmentation with a neural encoder model of working memory”. In: *EMNLP 2017*

Speech segmentation: Results (Zerospeech '15)

System	Bd P	Bd R	Bd F	Wd F
Lyzinski 15	18.8	64.0	29.0	2.4
Räsänen 15	75.7	33.7	46.7	9.6
Räsänen new	61.1	50.1	55.2	12.4
Kamper 16	66.5	58.8	62.4	20.6
Ours	62.4	43.2	51.1	9.3

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

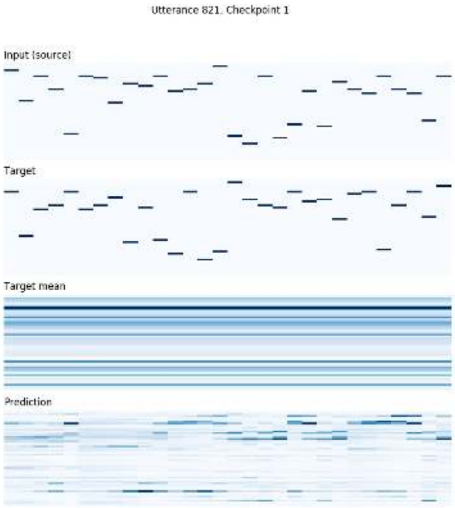
Speech segmentation: Dropout



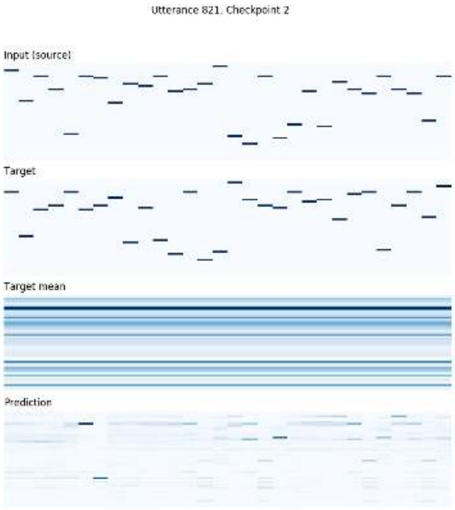
Dropout and memory limits encourage better segmentations

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

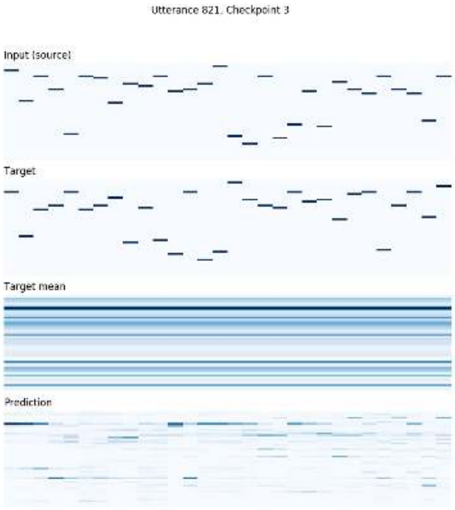
Example: Reconstruction learning



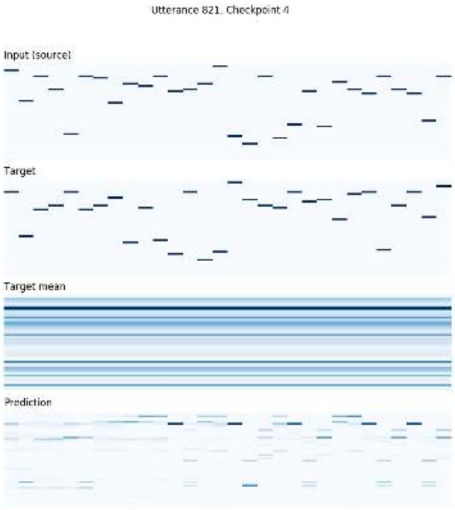
Example: Reconstruction learning



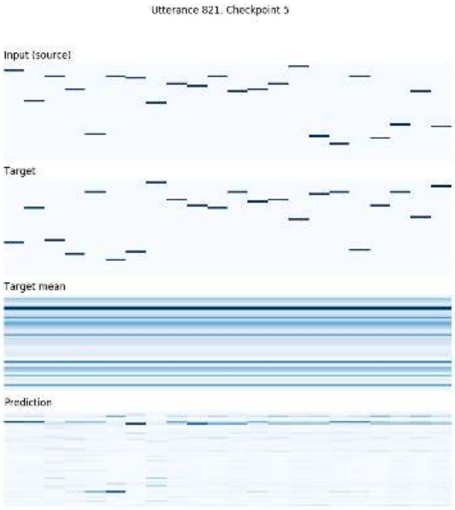
Example: Reconstruction learning



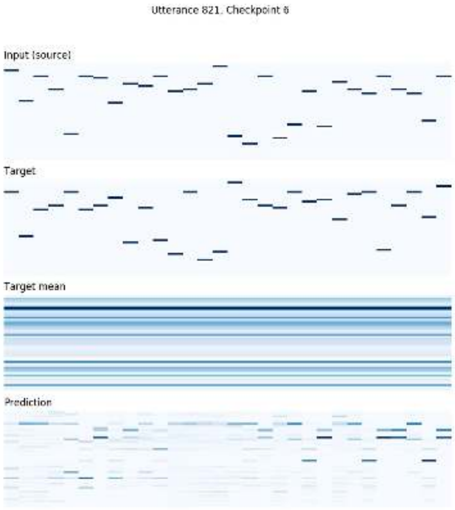
Example: Reconstruction learning



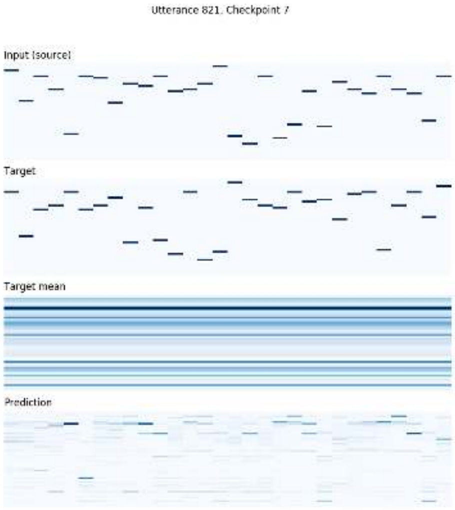
Example: Reconstruction learning



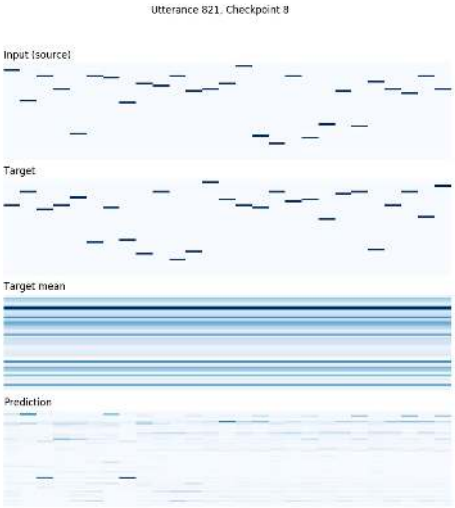
Example: Reconstruction learning



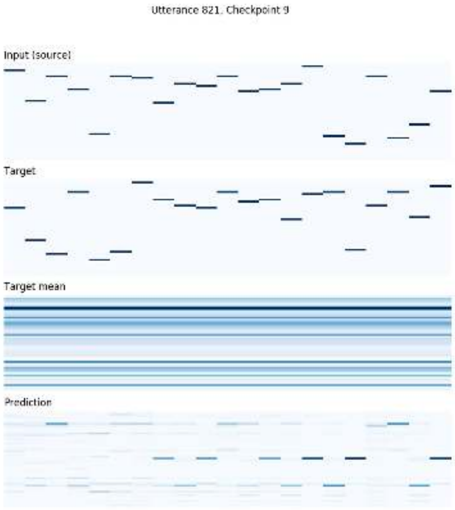
Example: Reconstruction learning



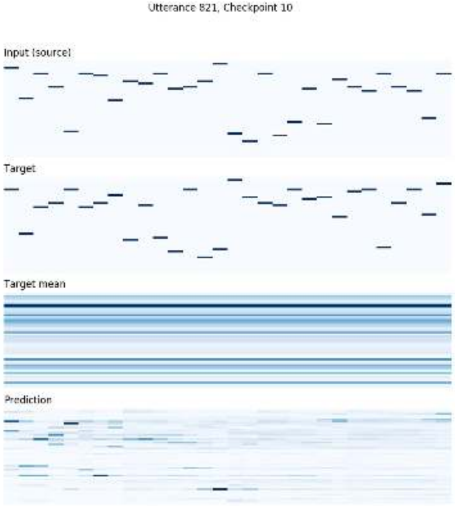
Example: Reconstruction learning



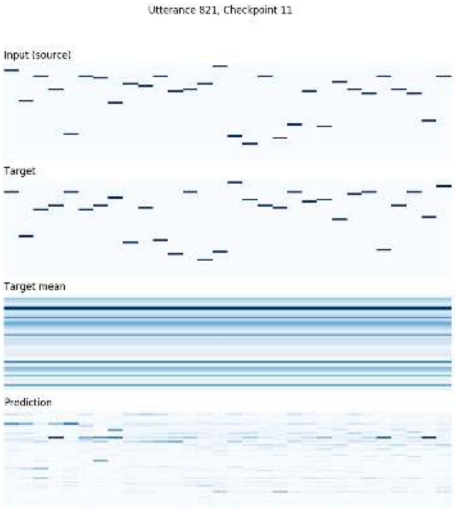
Example: Reconstruction learning



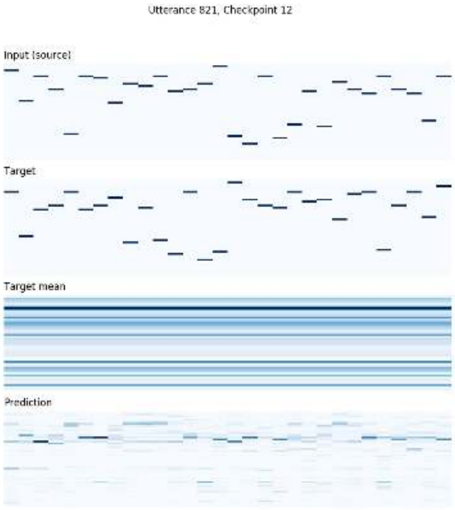
Example: Reconstruction learning



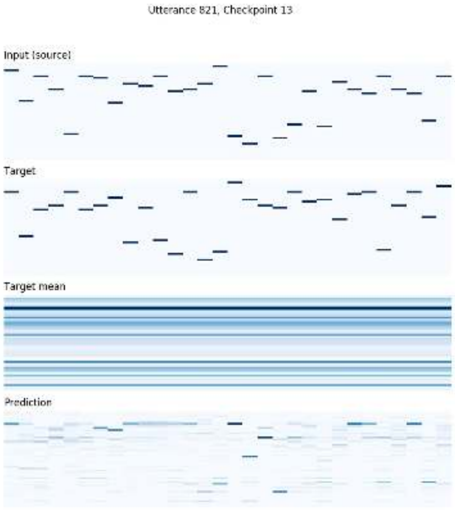
Example: Reconstruction learning



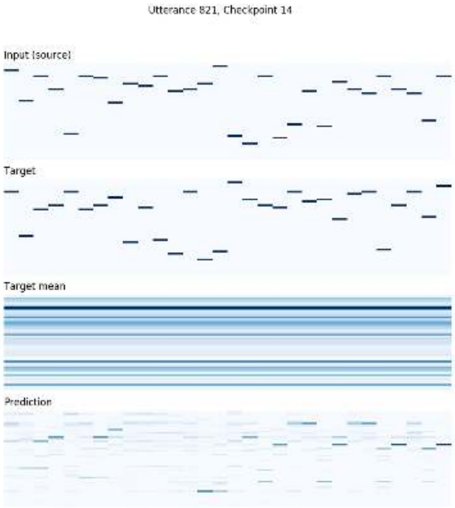
Example: Reconstruction learning



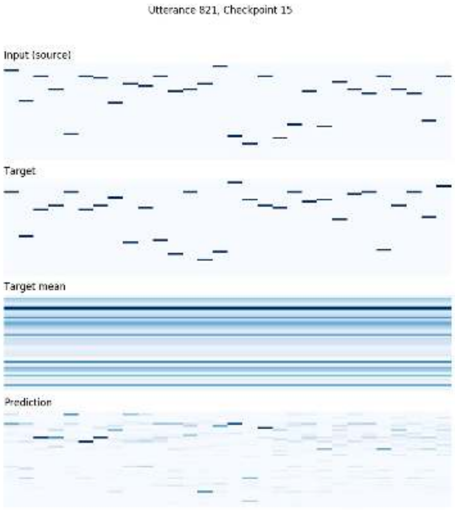
Example: Reconstruction learning



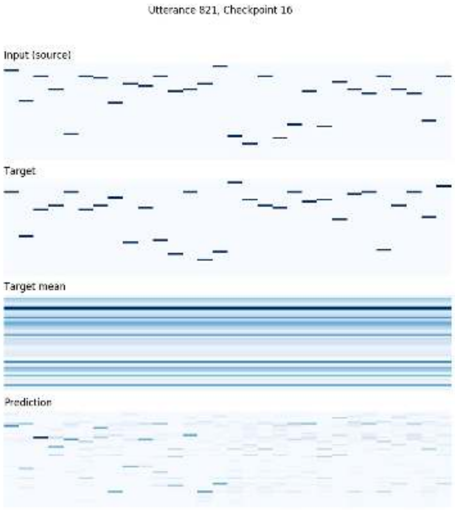
Example: Reconstruction learning



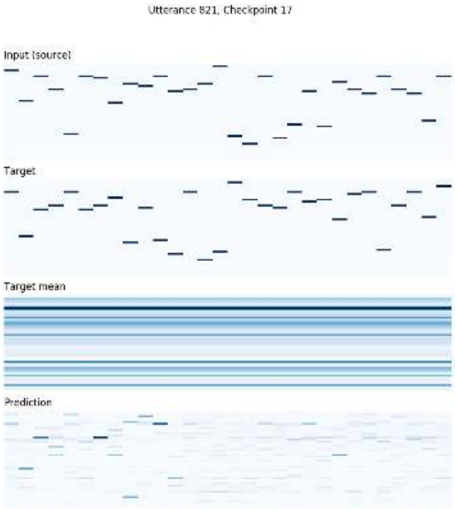
Example: Reconstruction learning



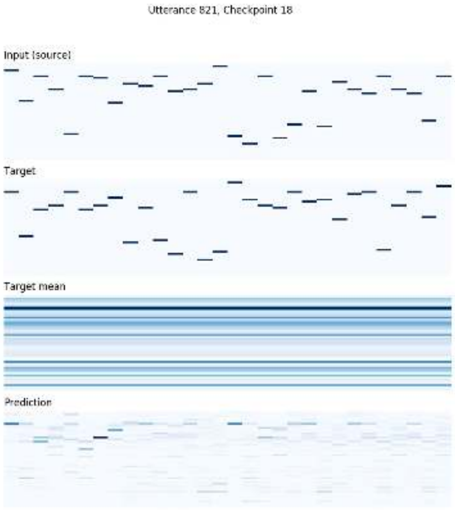
Example: Reconstruction learning



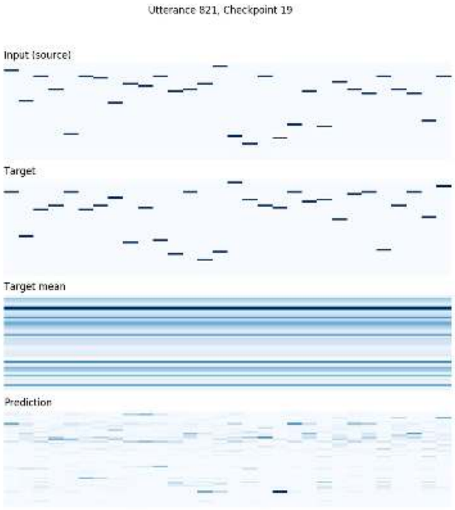
Example: Reconstruction learning



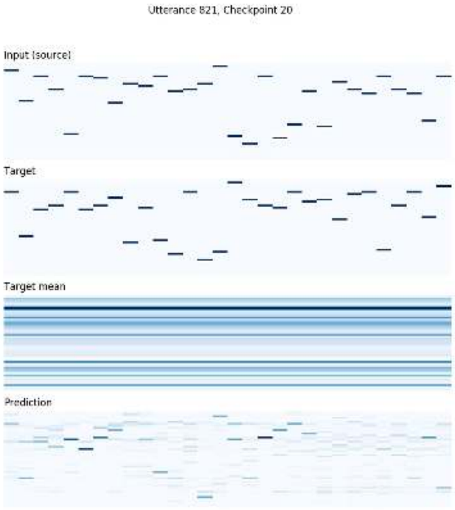
Example: Reconstruction learning



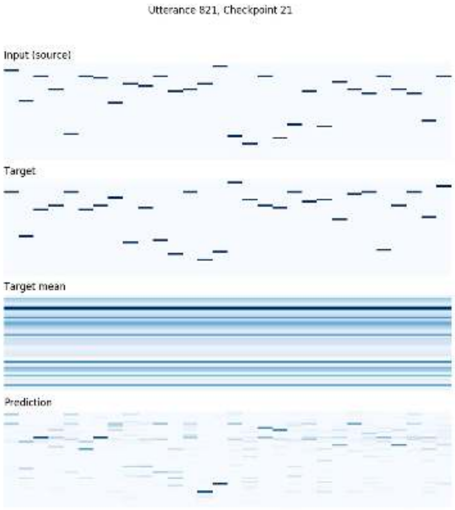
Example: Reconstruction learning



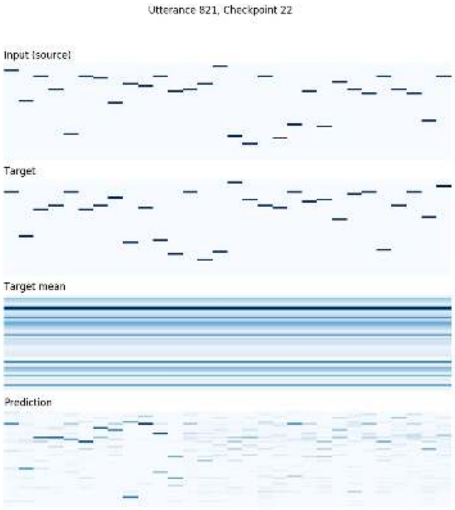
Example: Reconstruction learning



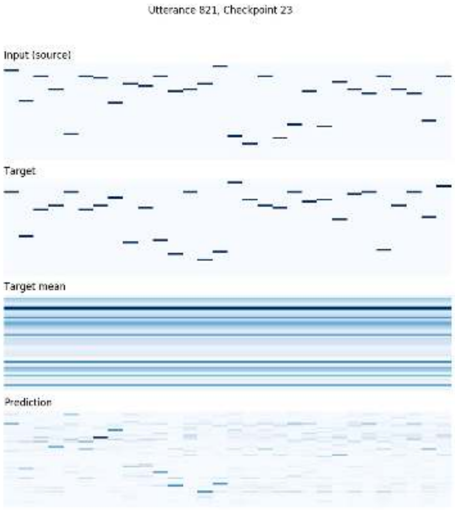
Example: Reconstruction learning



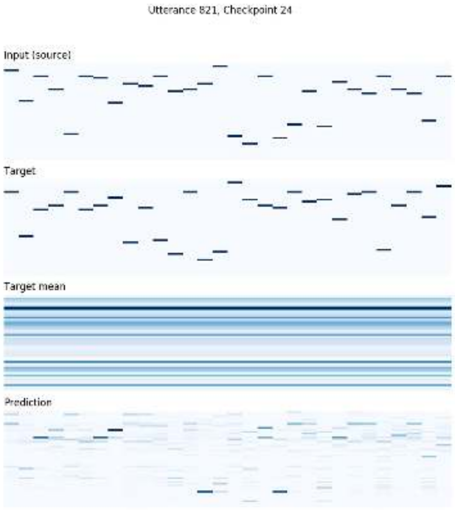
Example: Reconstruction learning



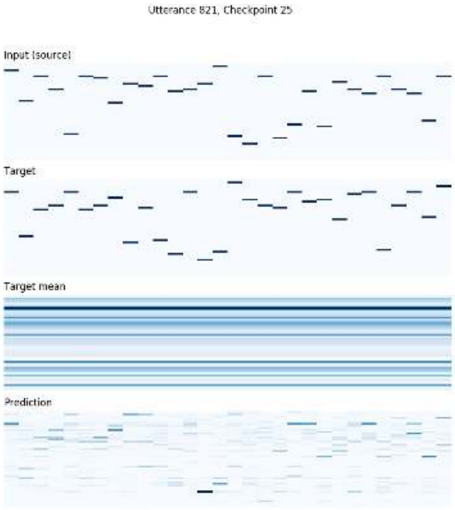
Example: Reconstruction learning



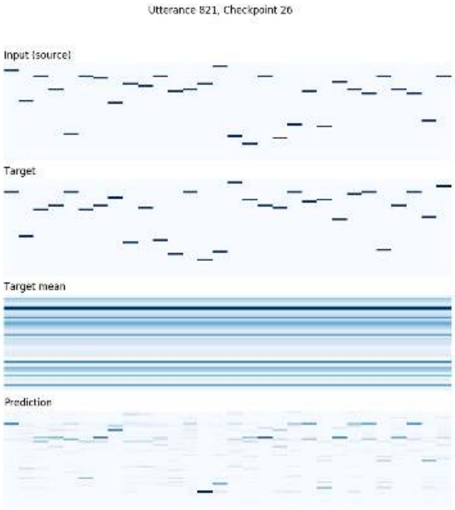
Example: Reconstruction learning



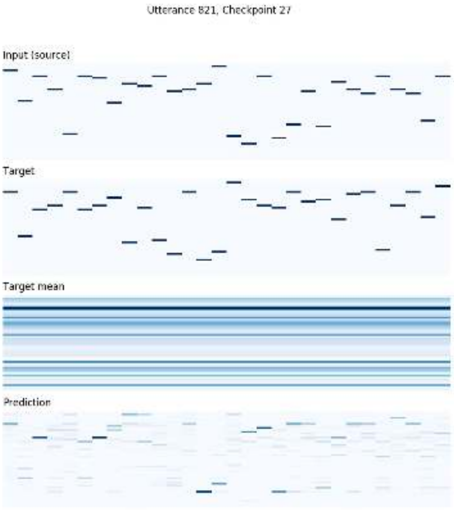
Example: Reconstruction learning



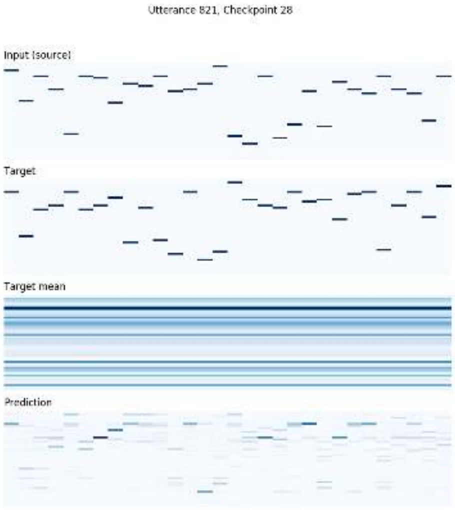
Example: Reconstruction learning



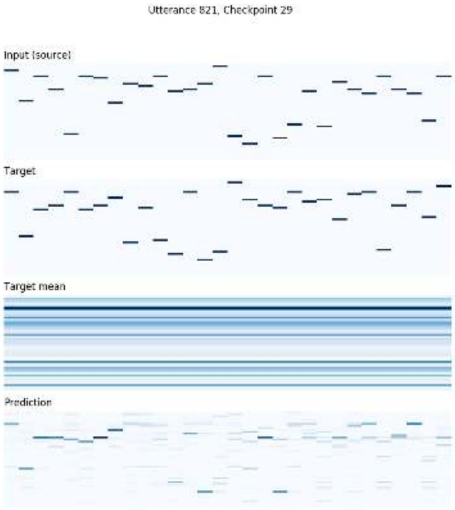
Example: Reconstruction learning



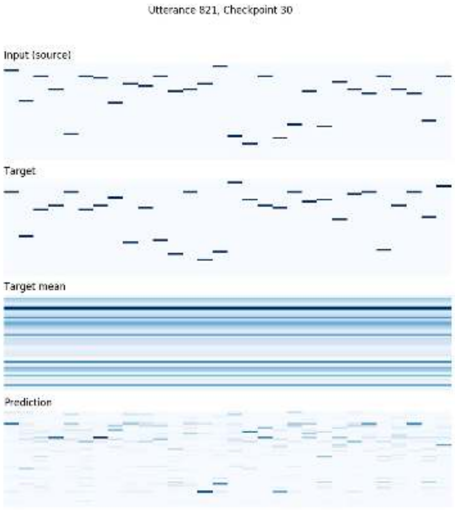
Example: Reconstruction learning



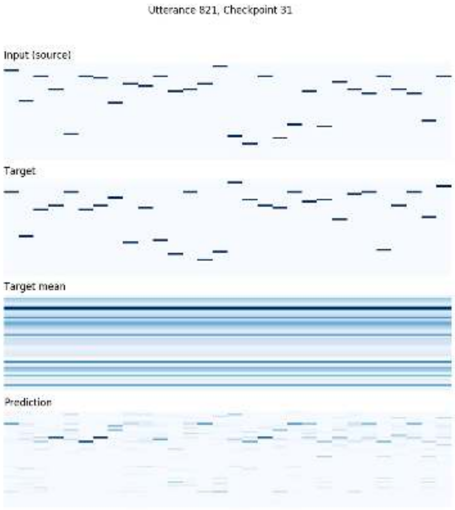
Example: Reconstruction learning



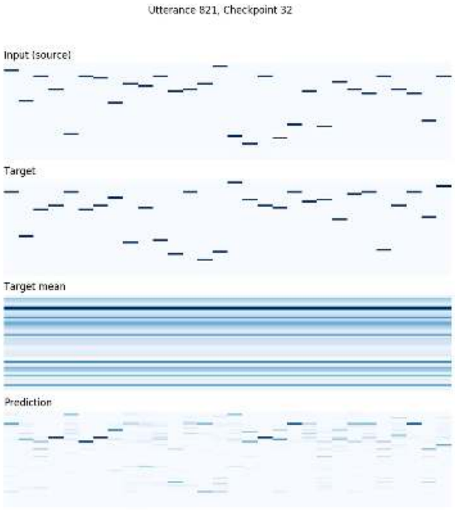
Example: Reconstruction learning



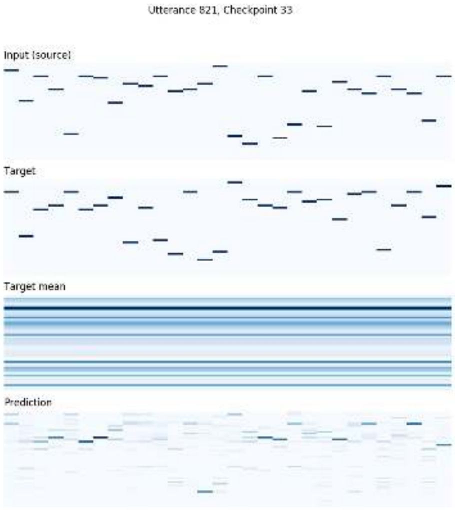
Example: Reconstruction learning



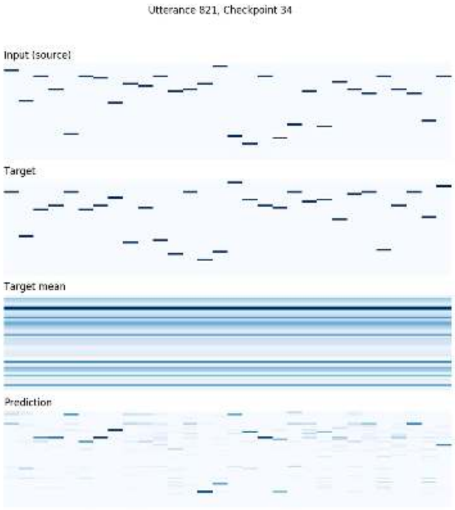
Example: Reconstruction learning



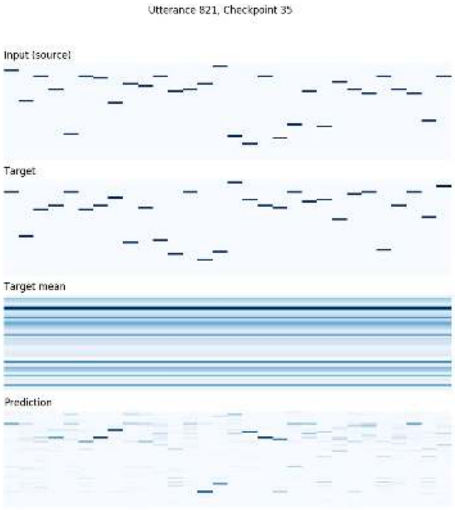
Example: Reconstruction learning



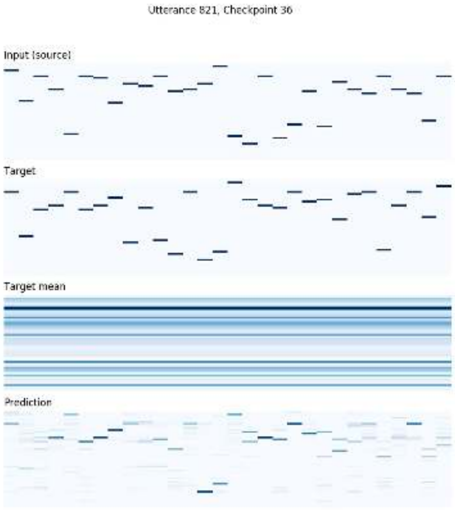
Example: Reconstruction learning



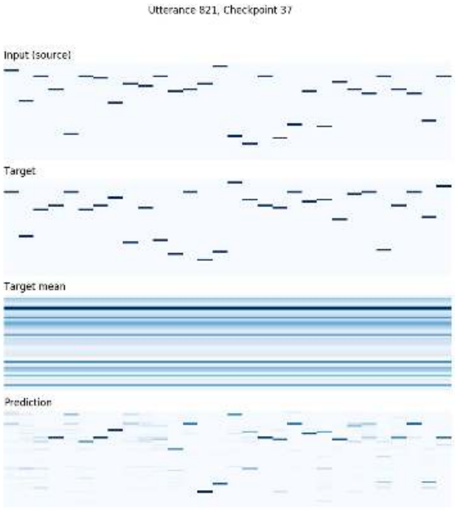
Example: Reconstruction learning



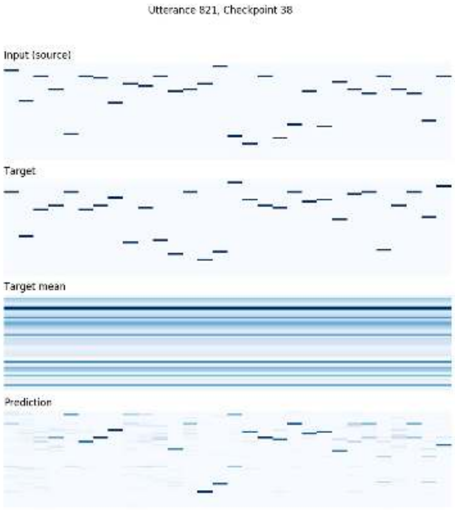
Example: Reconstruction learning



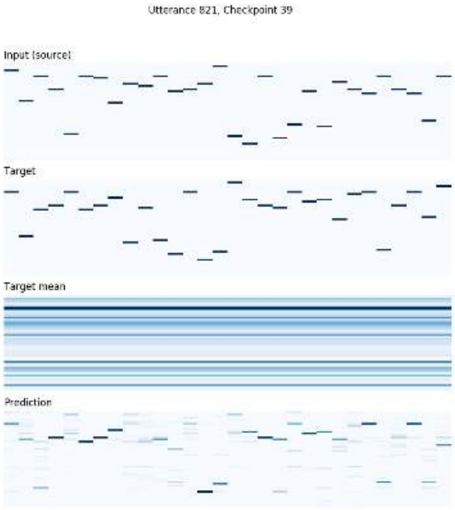
Example: Reconstruction learning



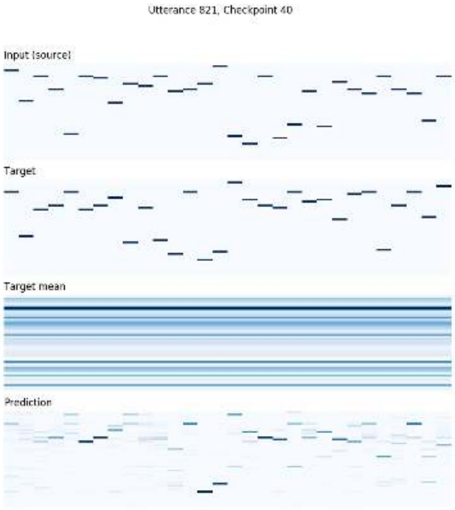
Example: Reconstruction learning



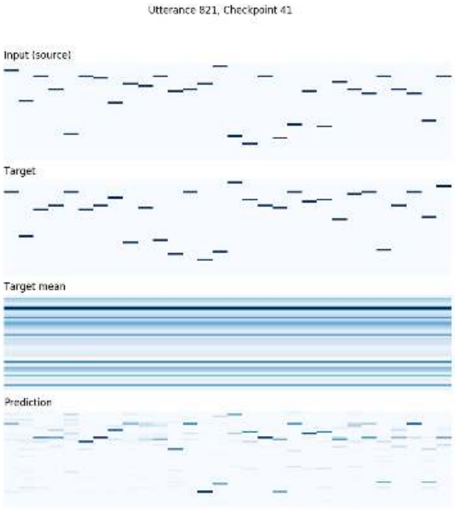
Example: Reconstruction learning



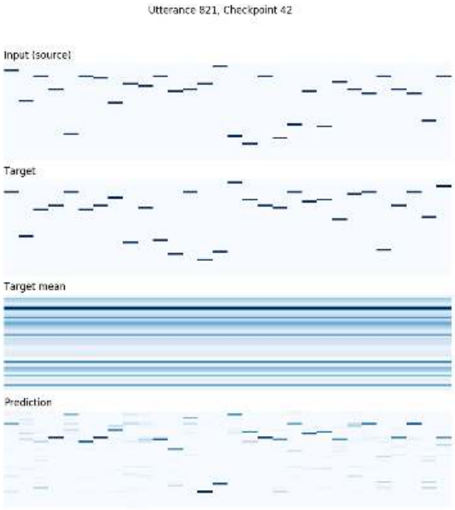
Example: Reconstruction learning



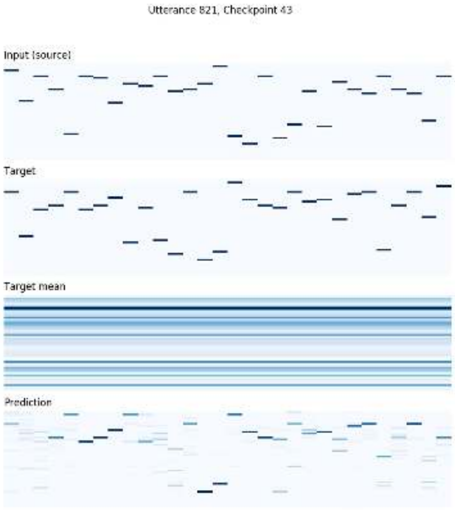
Example: Reconstruction learning



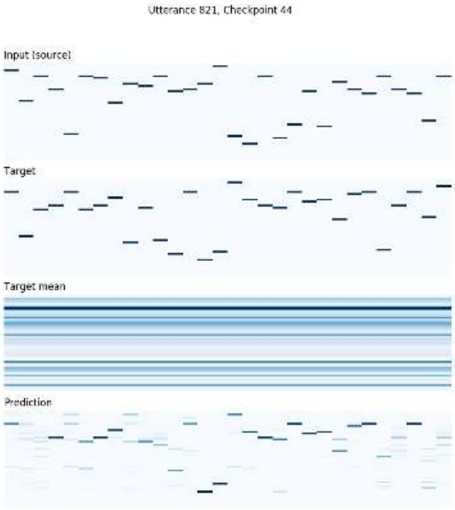
Example: Reconstruction learning



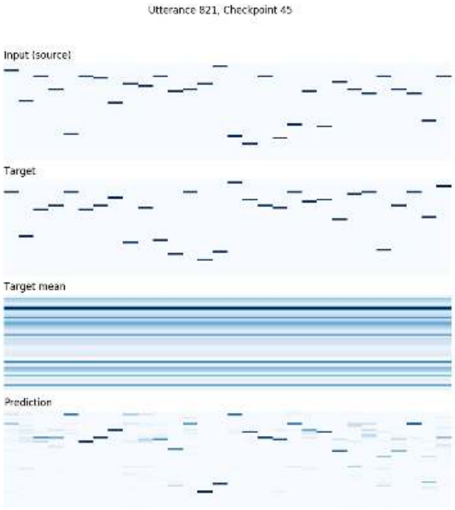
Example: Reconstruction learning



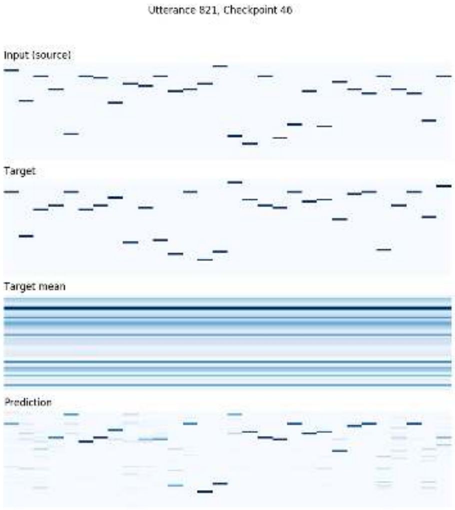
Example: Reconstruction learning



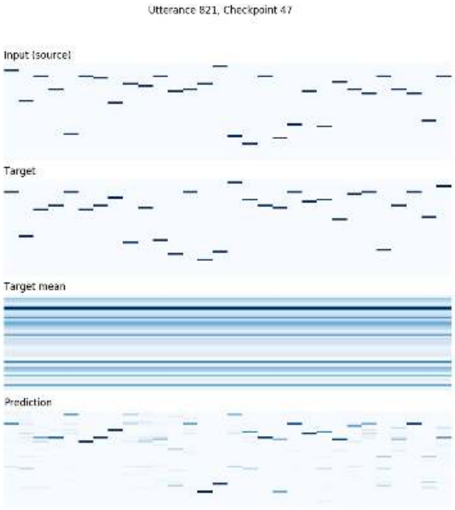
Example: Reconstruction learning



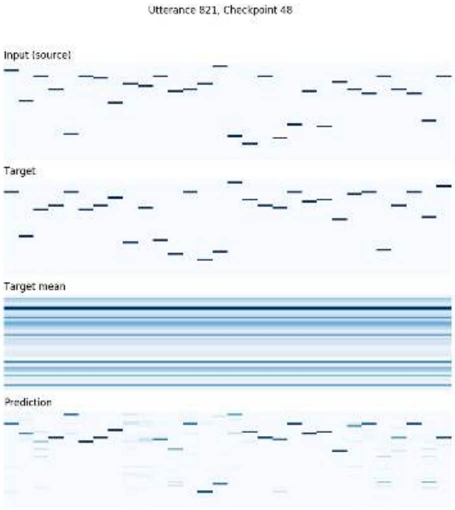
Example: Reconstruction learning



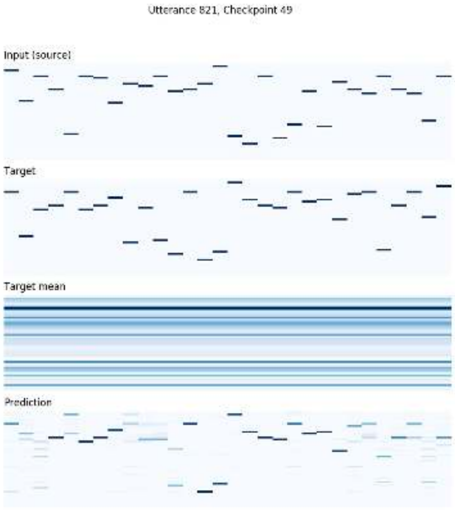
Example: Reconstruction learning



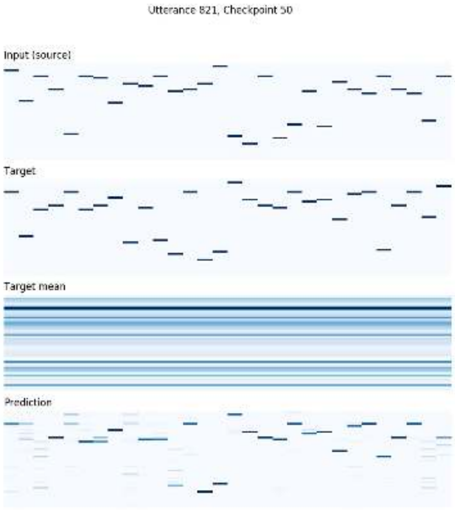
Example: Reconstruction learning



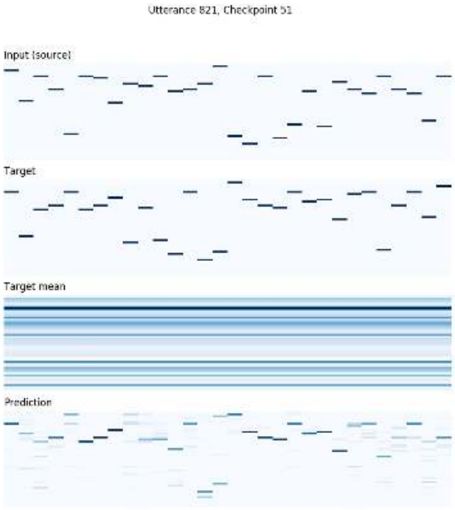
Example: Reconstruction learning



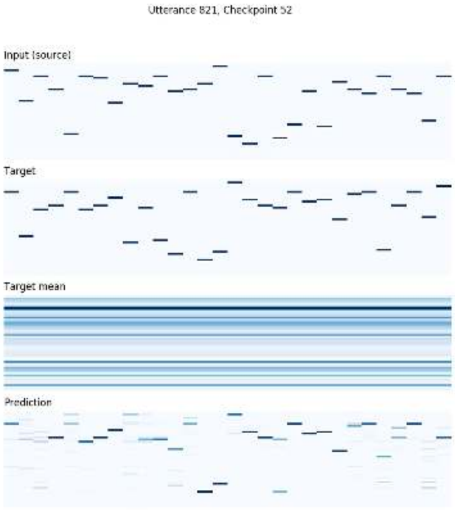
Example: Reconstruction learning



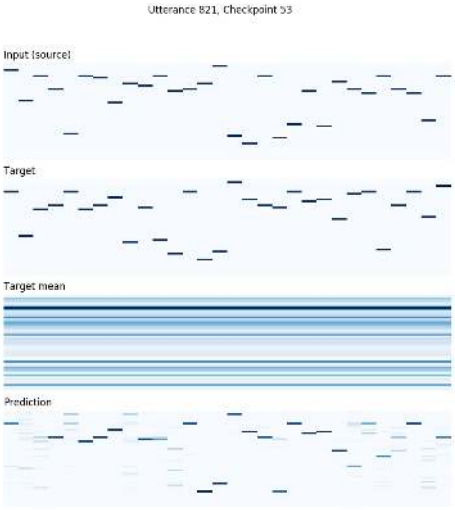
Example: Reconstruction learning



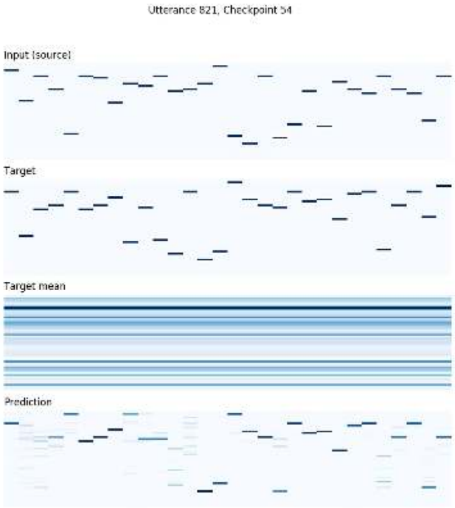
Example: Reconstruction learning



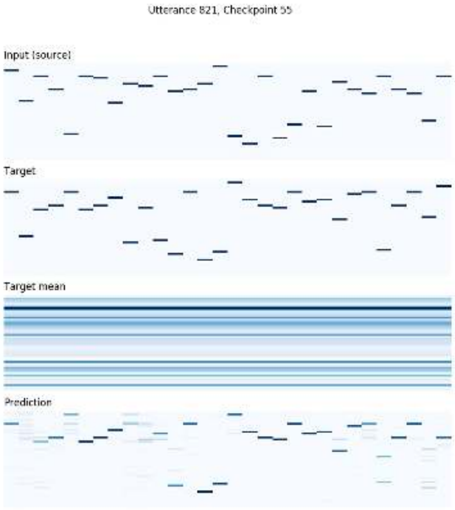
Example: Reconstruction learning



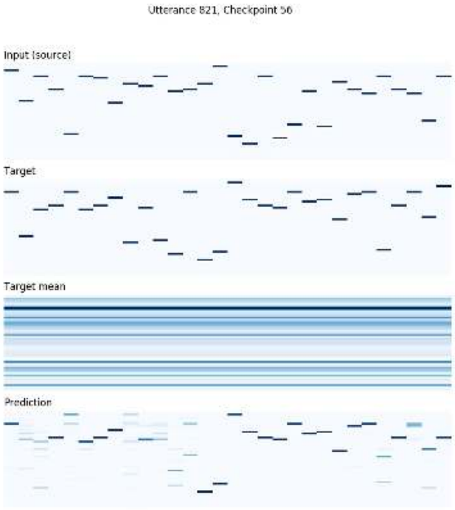
Example: Reconstruction learning



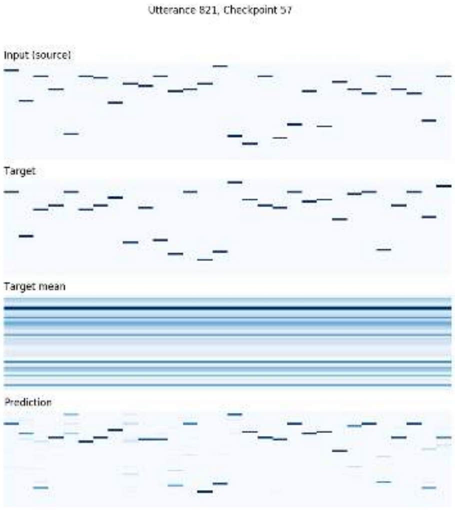
Example: Reconstruction learning



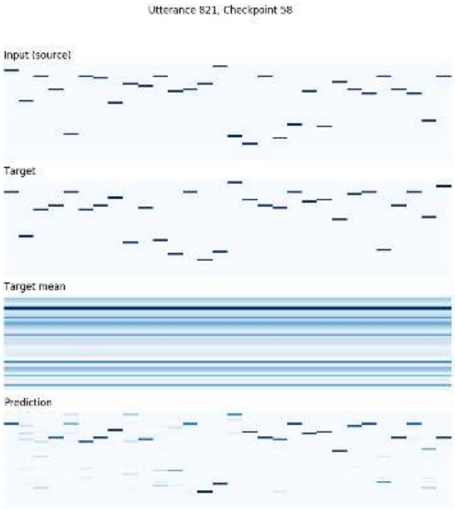
Example: Reconstruction learning



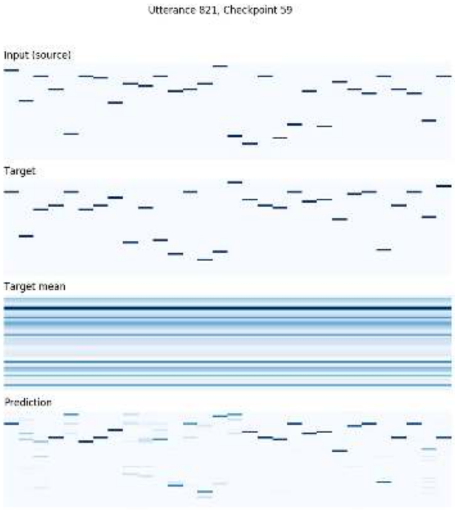
Example: Reconstruction learning



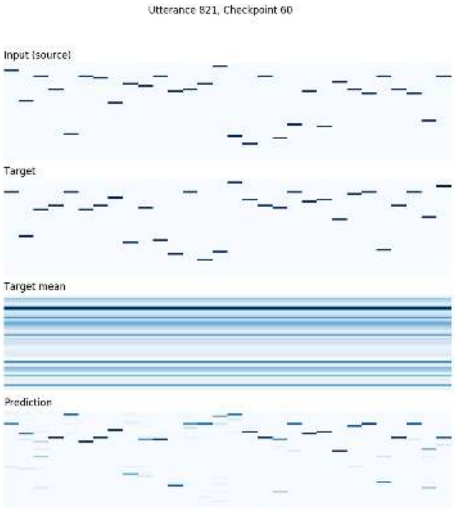
Example: Reconstruction learning



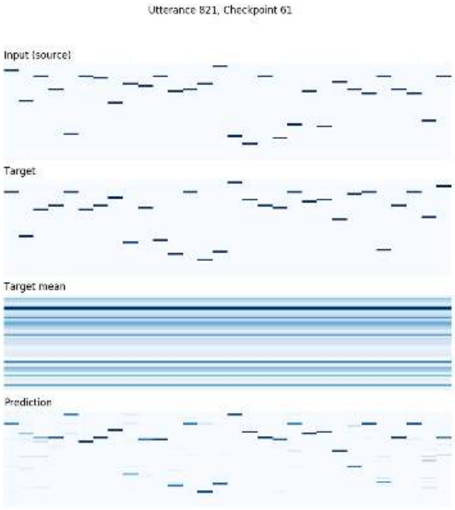
Example: Reconstruction learning



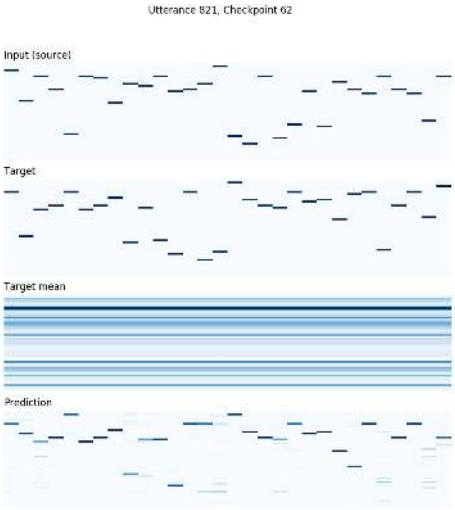
Example: Reconstruction learning



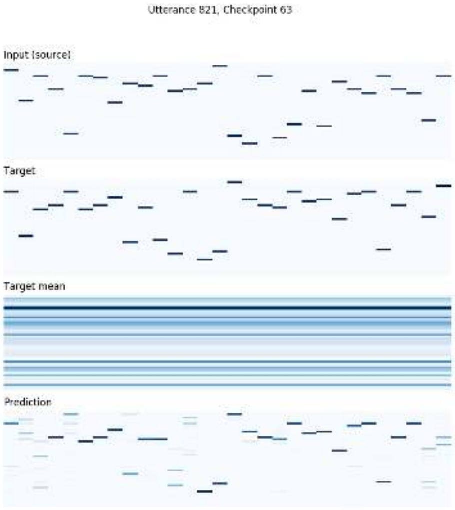
Example: Reconstruction learning



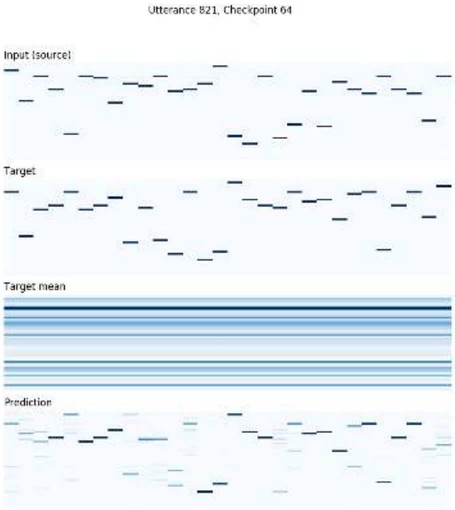
Example: Reconstruction learning



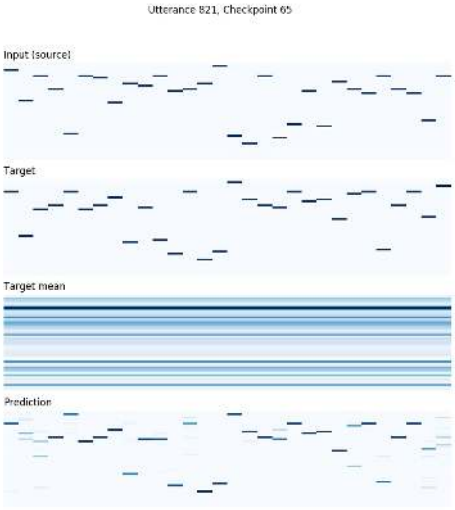
Example: Reconstruction learning



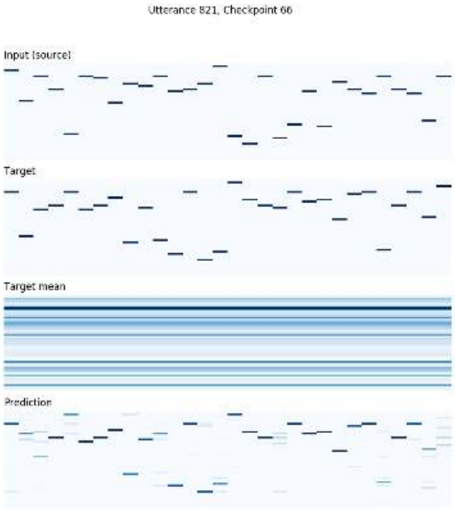
Example: Reconstruction learning



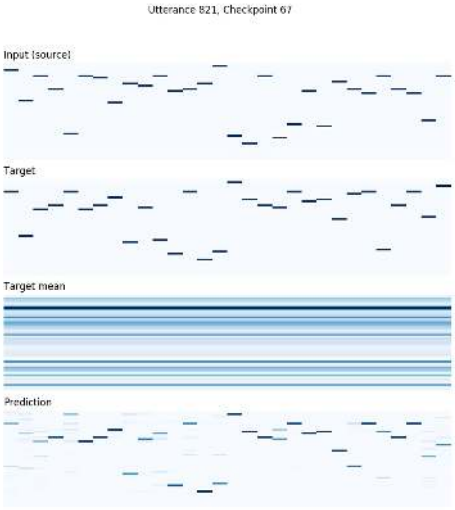
Example: Reconstruction learning



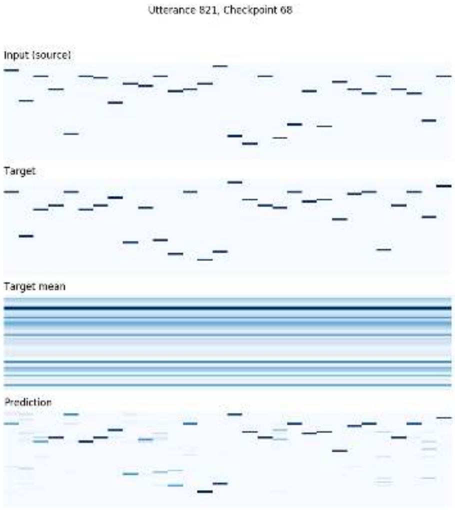
Example: Reconstruction learning



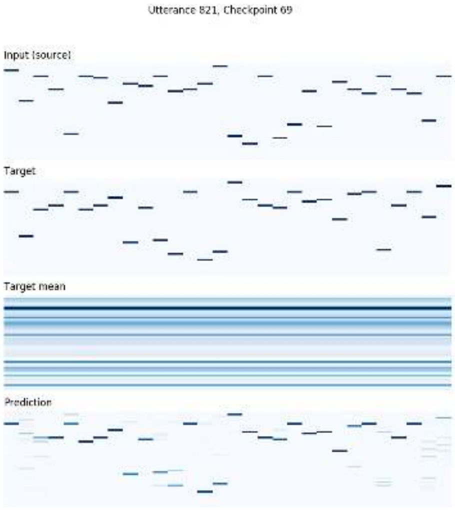
Example: Reconstruction learning



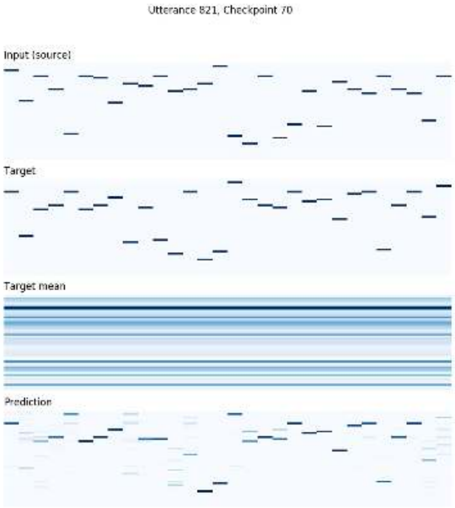
Example: Reconstruction learning



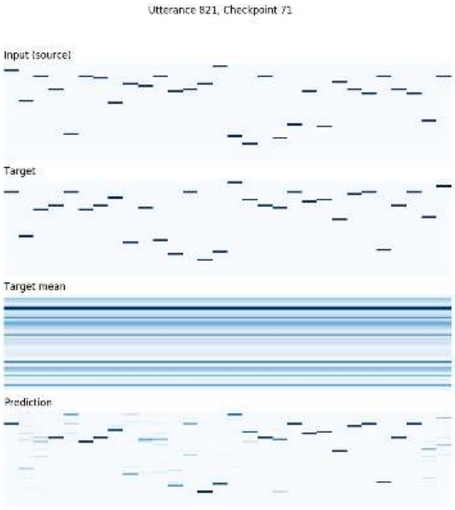
Example: Reconstruction learning



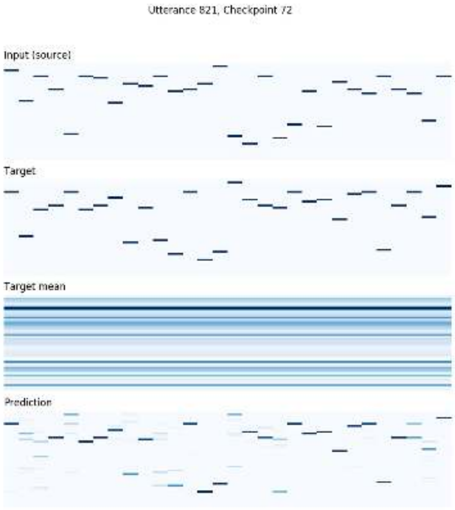
Example: Reconstruction learning



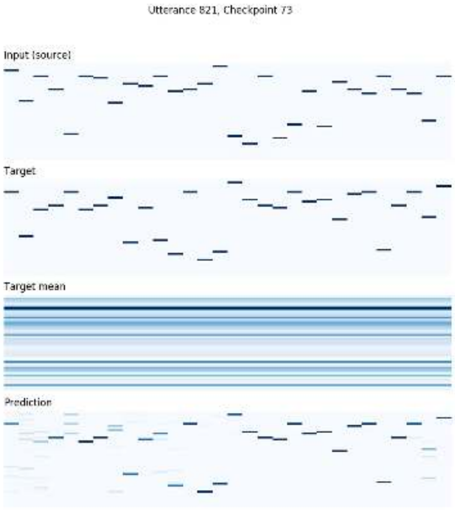
Example: Reconstruction learning



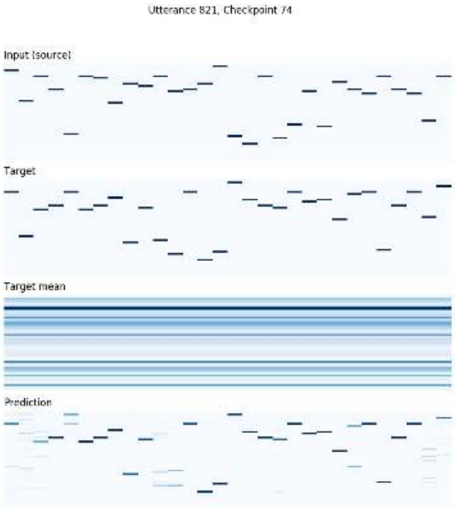
Example: Reconstruction learning



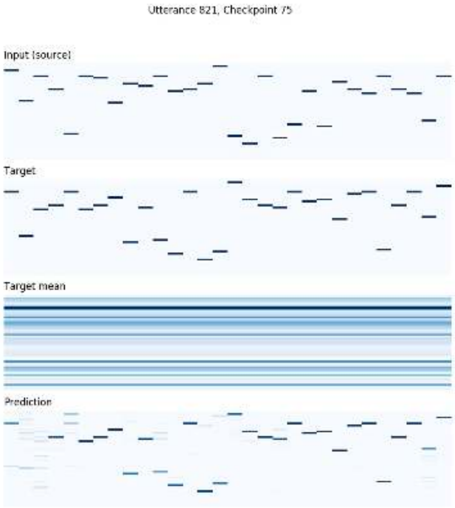
Example: Reconstruction learning



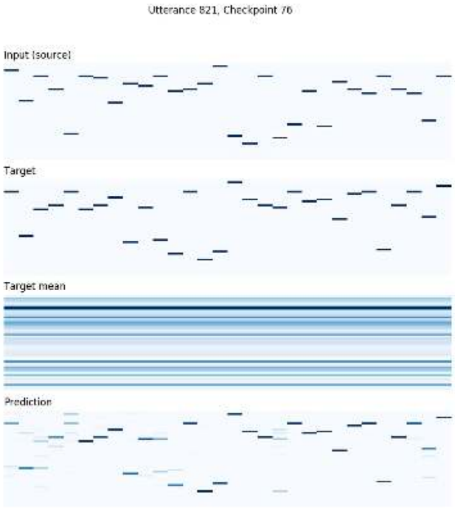
Example: Reconstruction learning



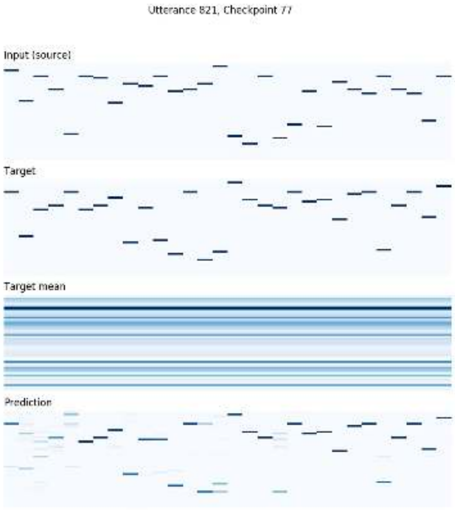
Example: Reconstruction learning



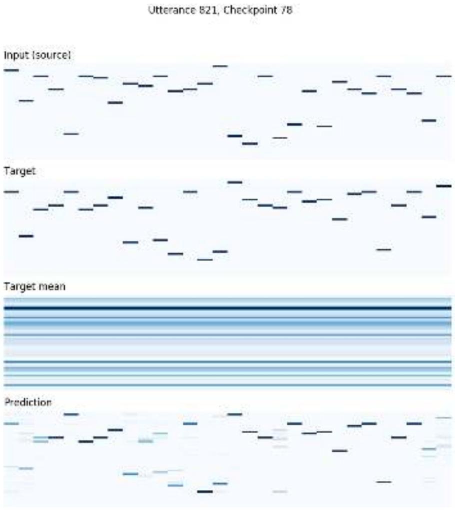
Example: Reconstruction learning



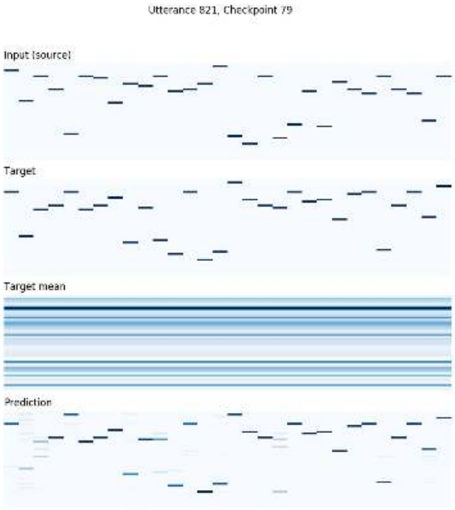
Example: Reconstruction learning



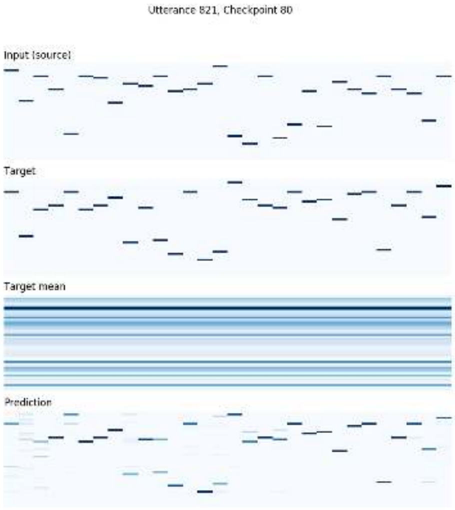
Example: Reconstruction learning



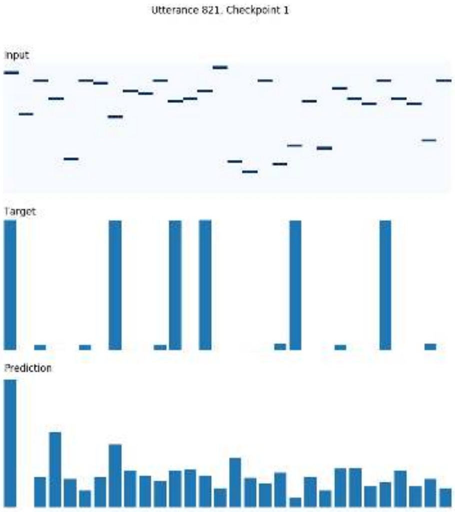
Example: Reconstruction learning



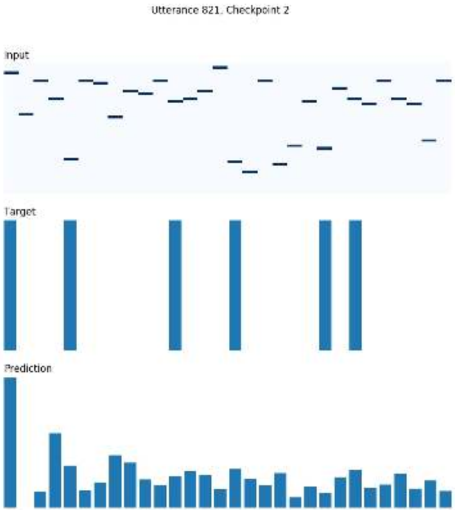
Example: Reconstruction learning



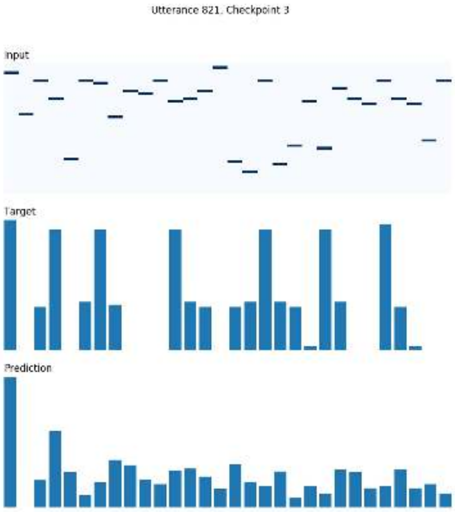
Example: Segmentation learning



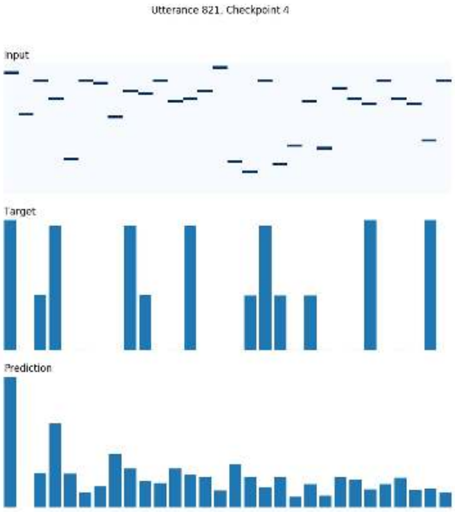
Example: Segmentation learning



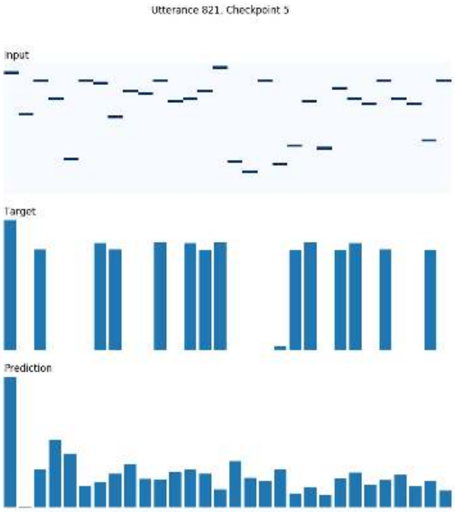
Example: Segmentation learning



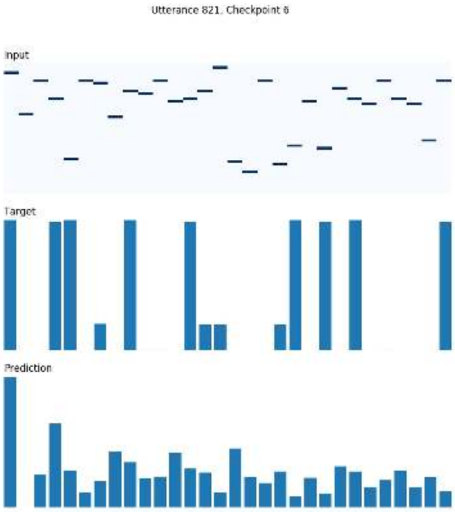
Example: Segmentation learning



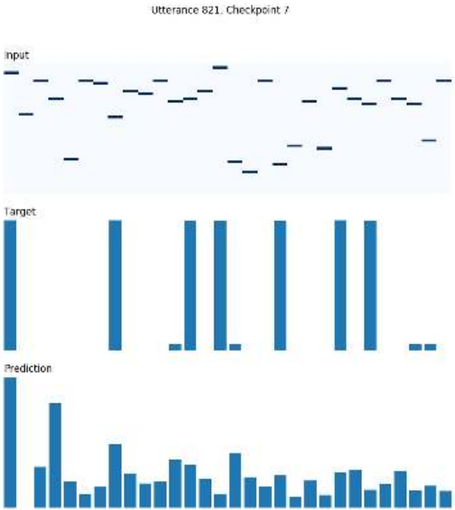
Example: Segmentation learning



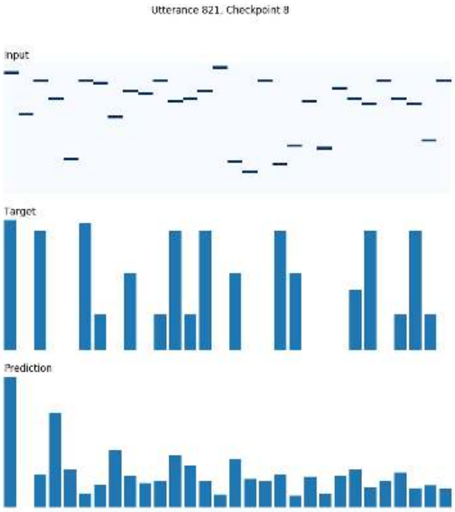
Example: Segmentation learning



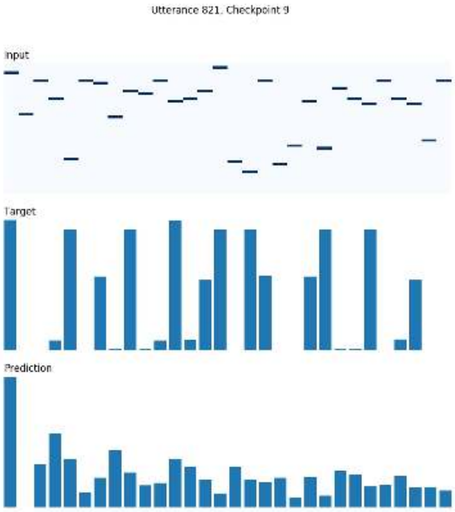
Example: Segmentation learning



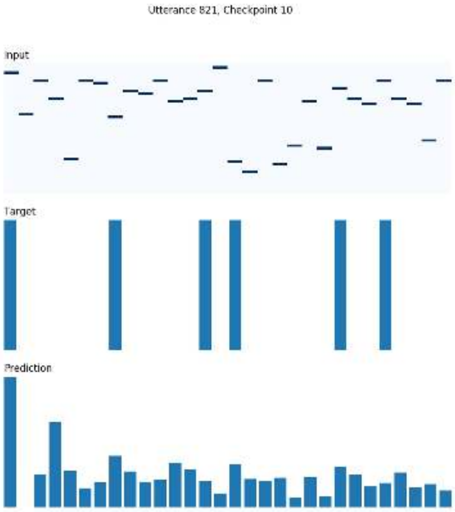
Example: Segmentation learning



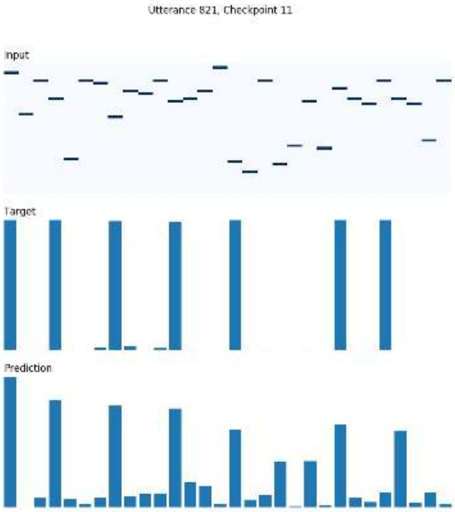
Example: Segmentation learning



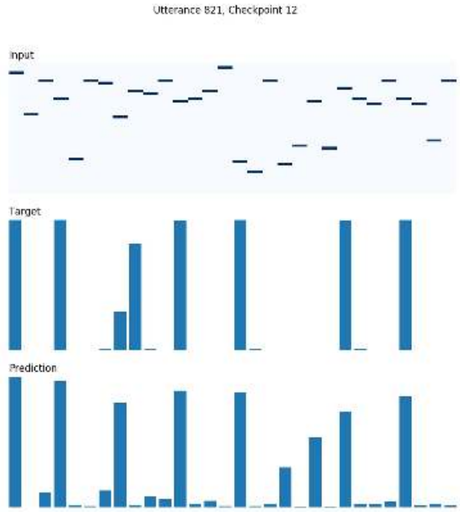
Example: Segmentation learning



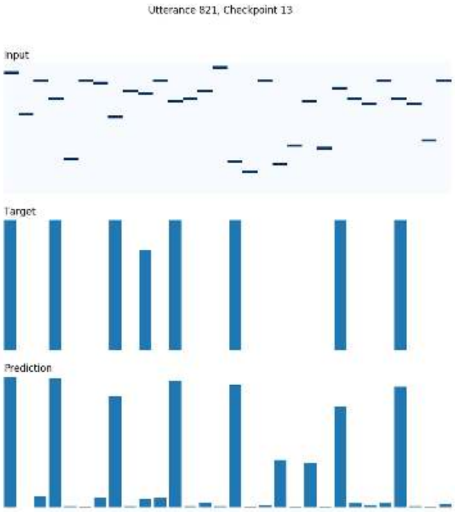
Example: Segmentation learning



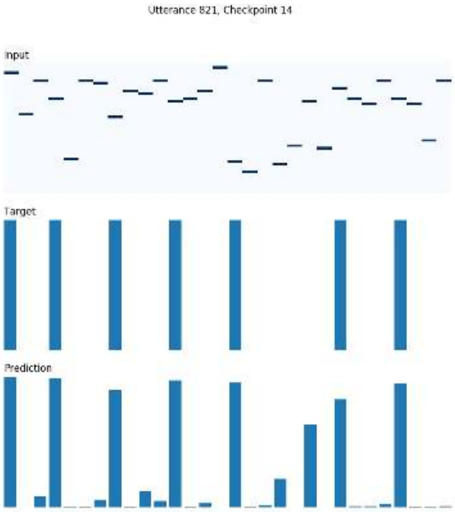
Example: Segmentation learning



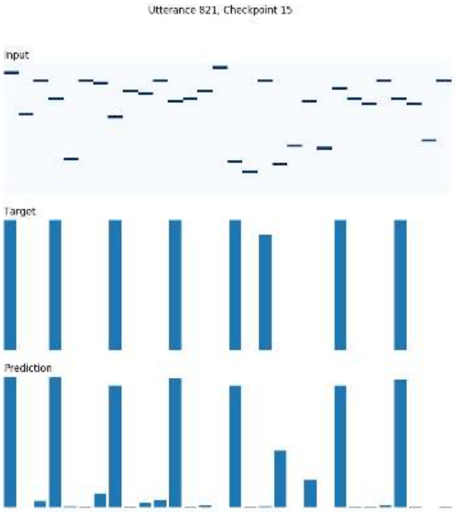
Example: Segmentation learning



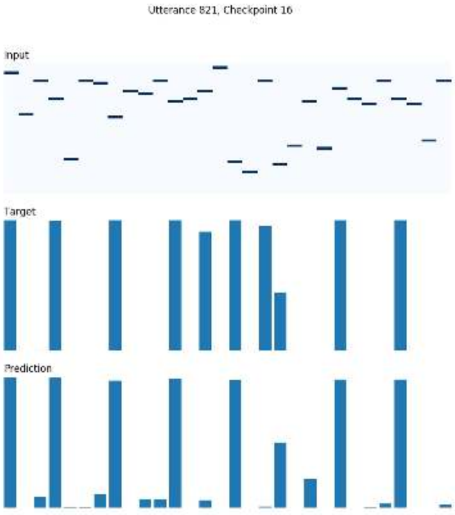
Example: Segmentation learning



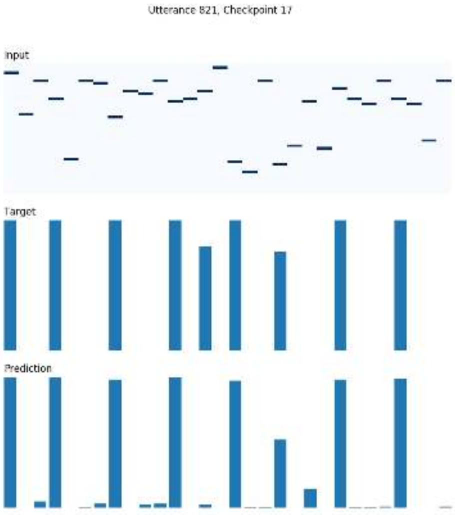
Example: Segmentation learning



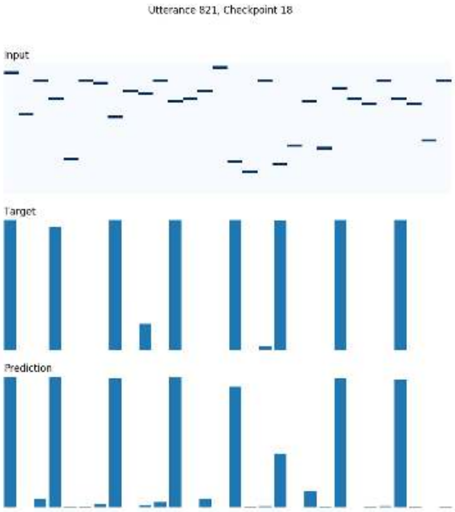
Example: Segmentation learning



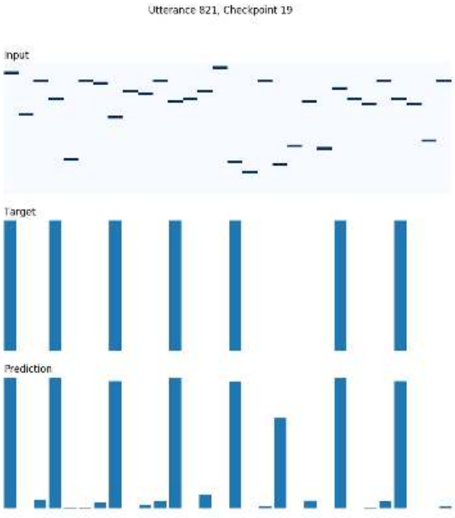
Example: Segmentation learning



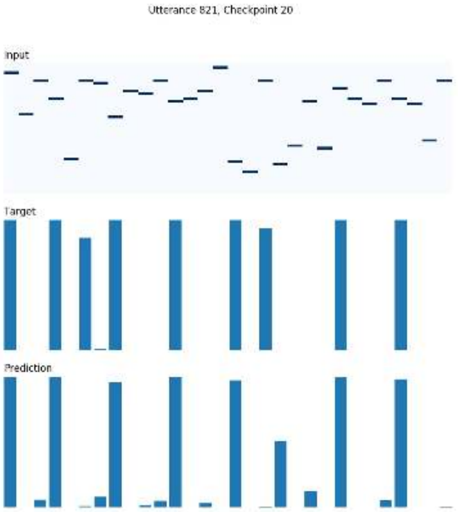
Example: Segmentation learning



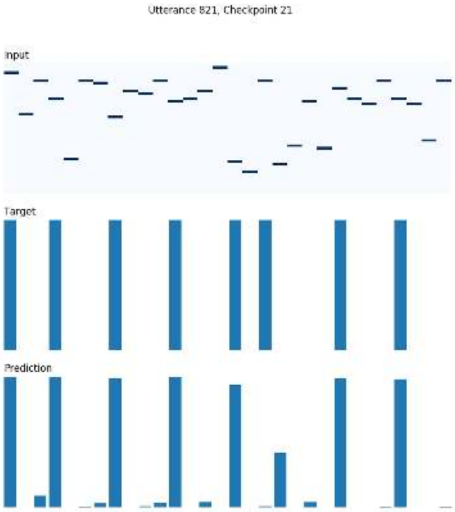
Example: Segmentation learning



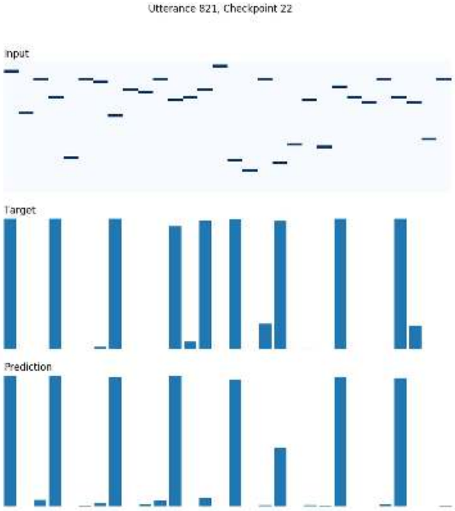
Example: Segmentation learning



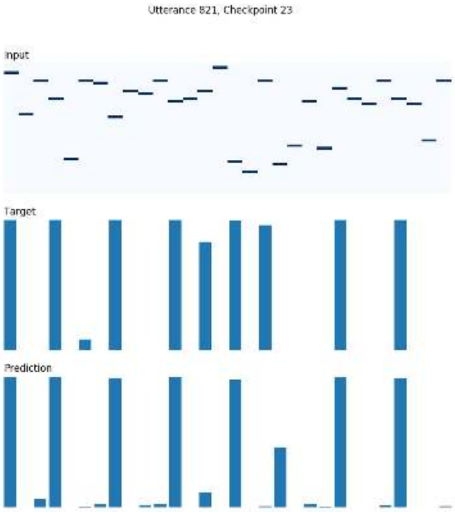
Example: Segmentation learning



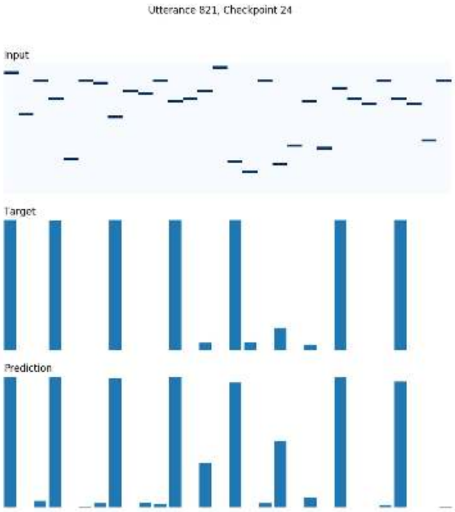
Example: Segmentation learning



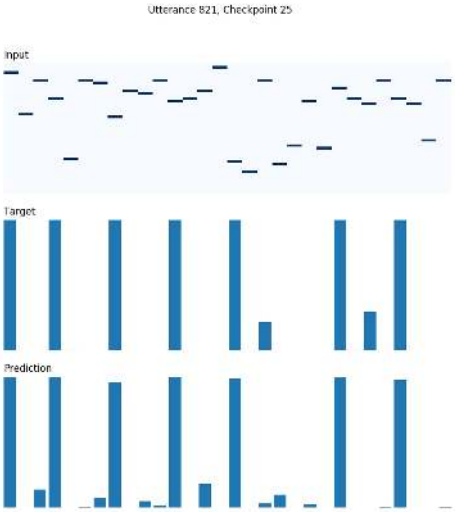
Example: Segmentation learning



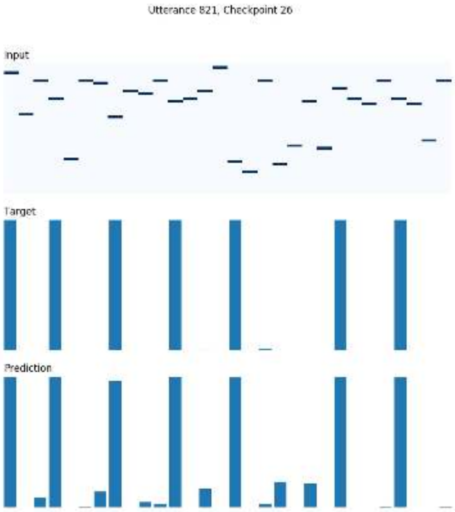
Example: Segmentation learning



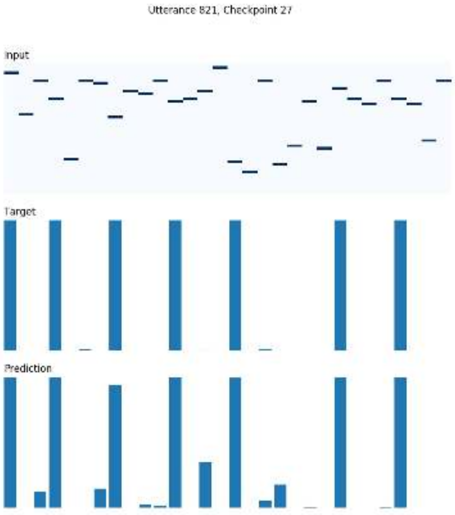
Example: Segmentation learning



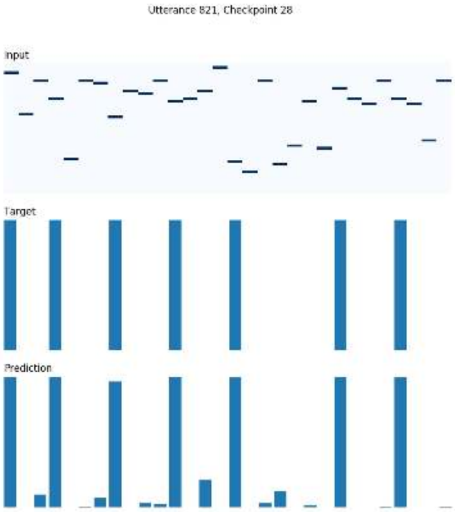
Example: Segmentation learning



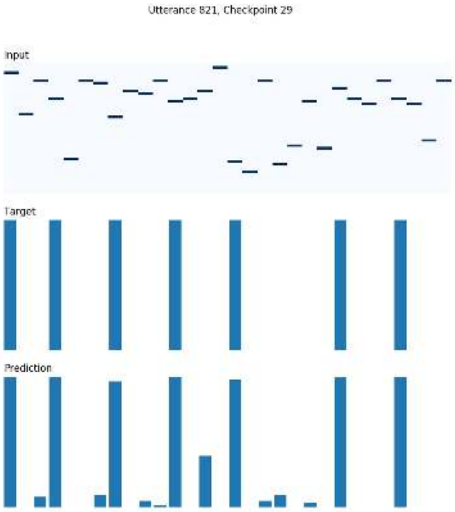
Example: Segmentation learning



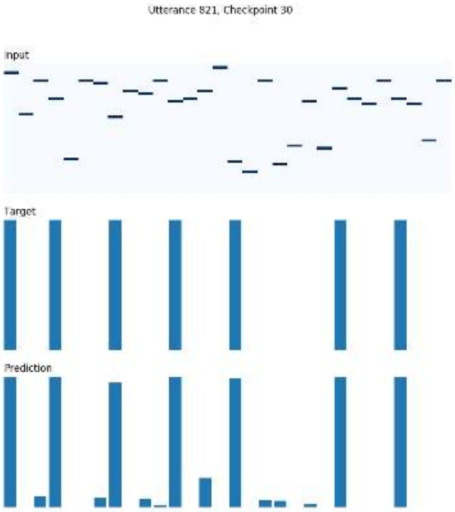
Example: Segmentation learning



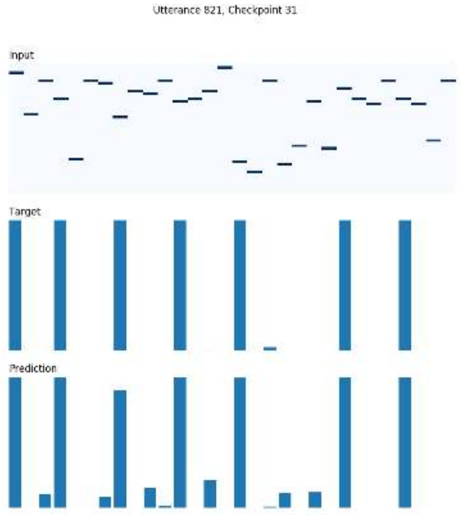
Example: Segmentation learning



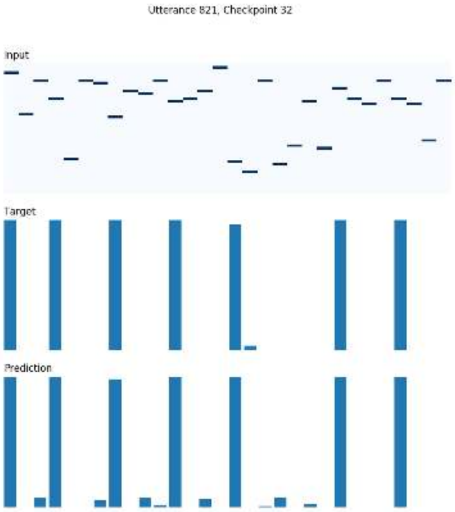
Example: Segmentation learning



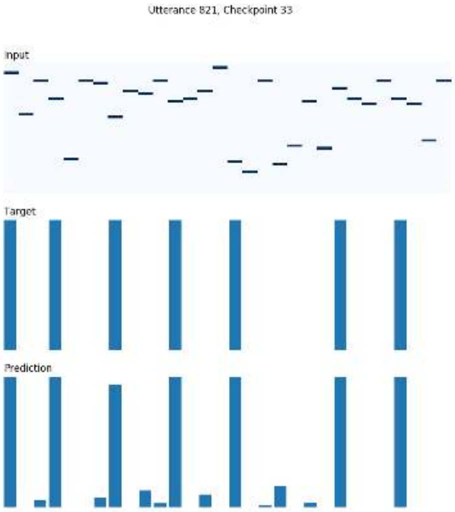
Example: Segmentation learning



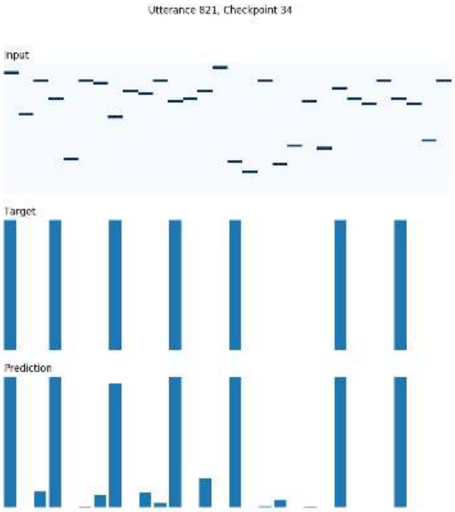
Example: Segmentation learning



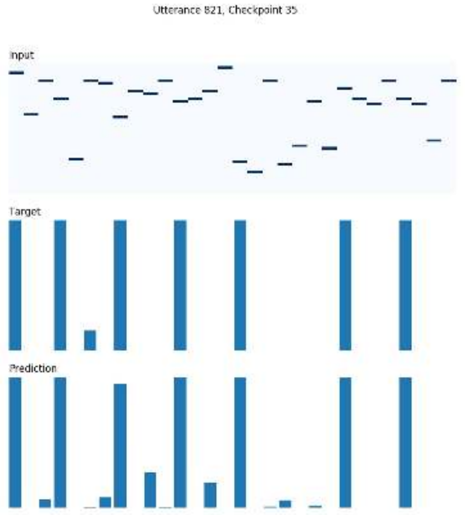
Example: Segmentation learning



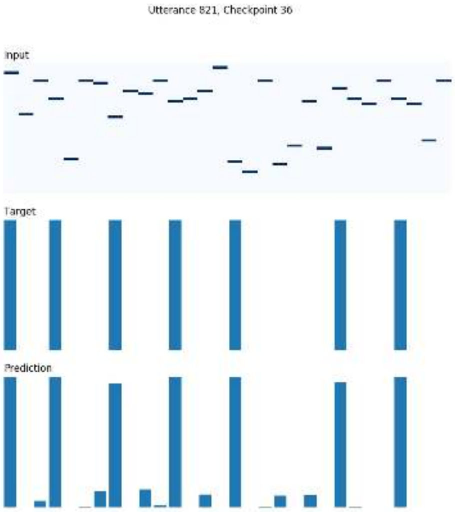
Example: Segmentation learning



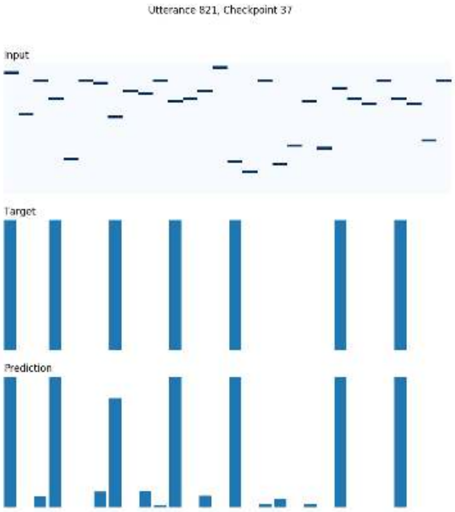
Example: Segmentation learning



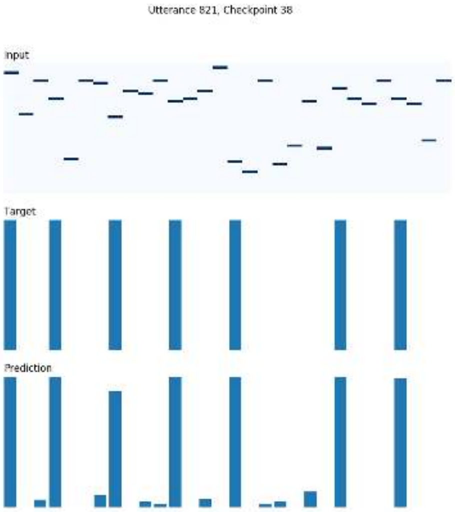
Example: Segmentation learning



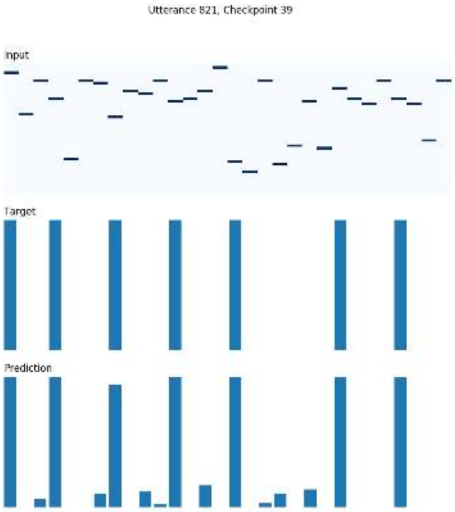
Example: Segmentation learning



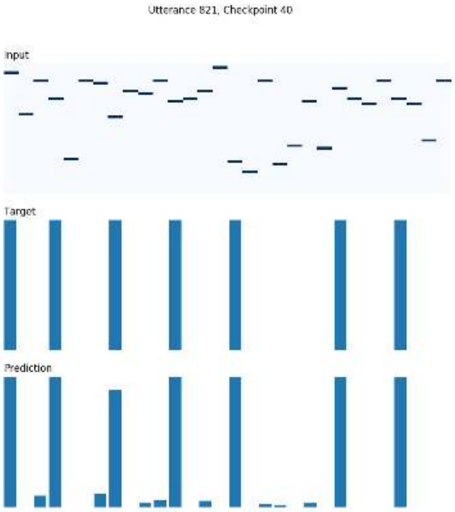
Example: Segmentation learning



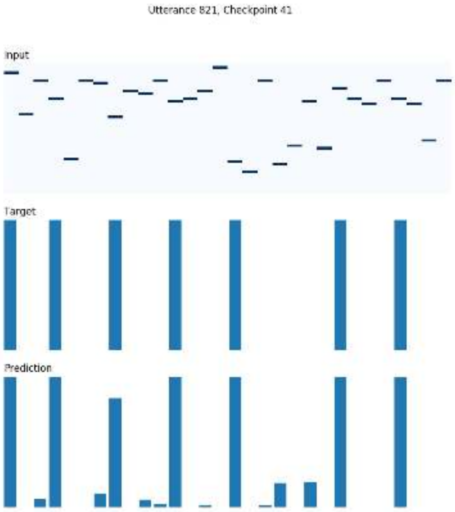
Example: Segmentation learning



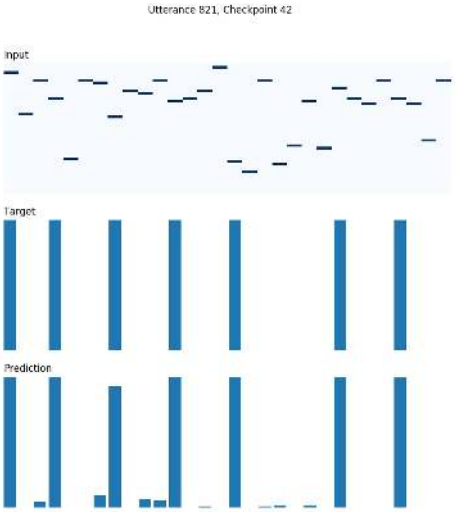
Example: Segmentation learning



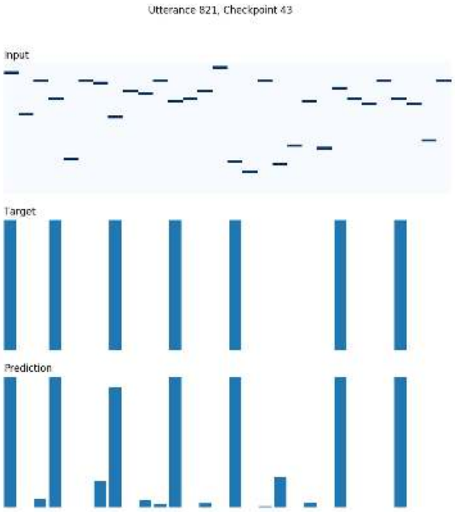
Example: Segmentation learning



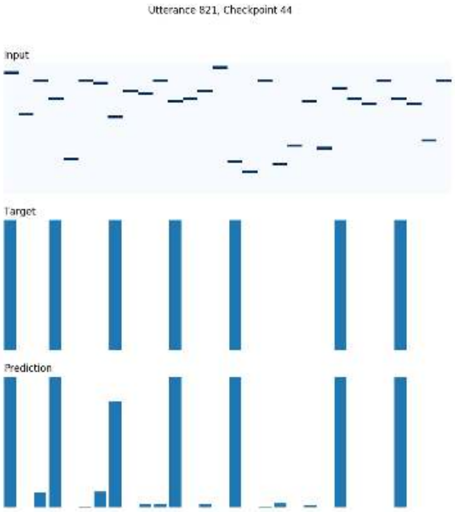
Example: Segmentation learning



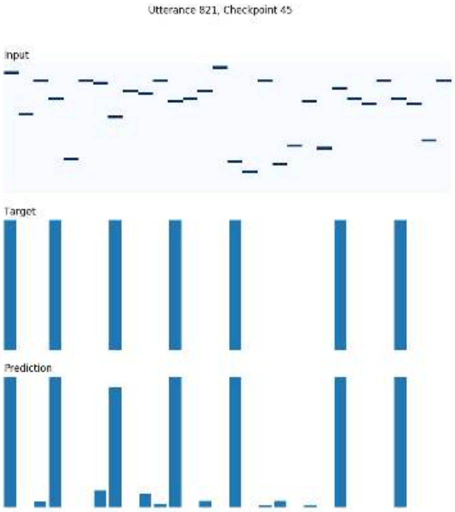
Example: Segmentation learning



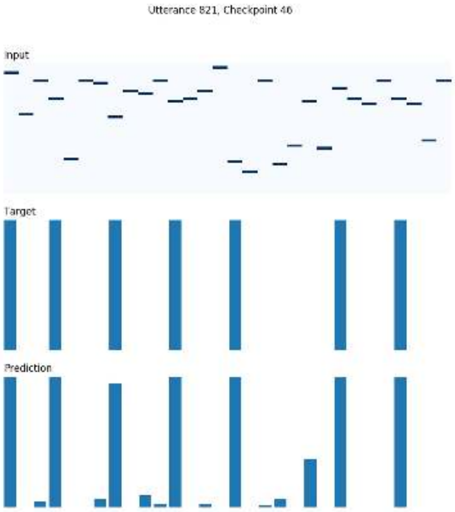
Example: Segmentation learning



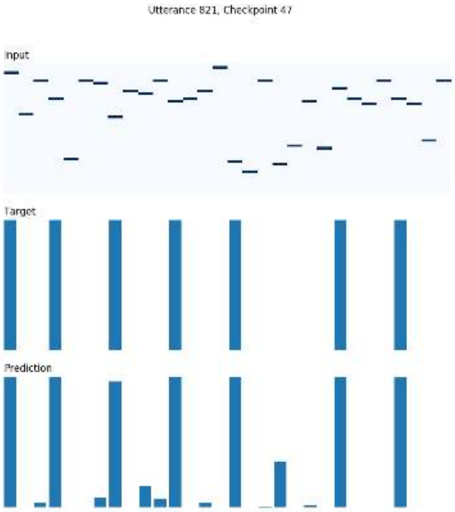
Example: Segmentation learning



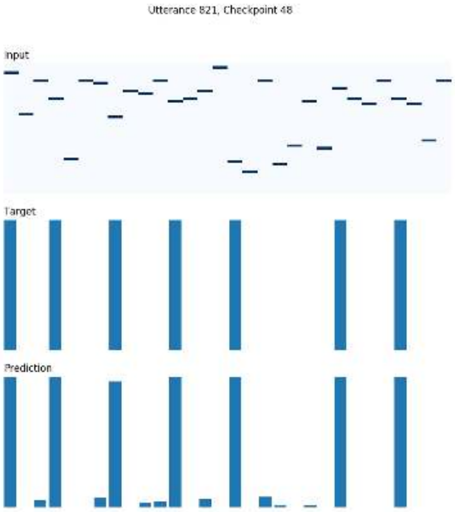
Example: Segmentation learning



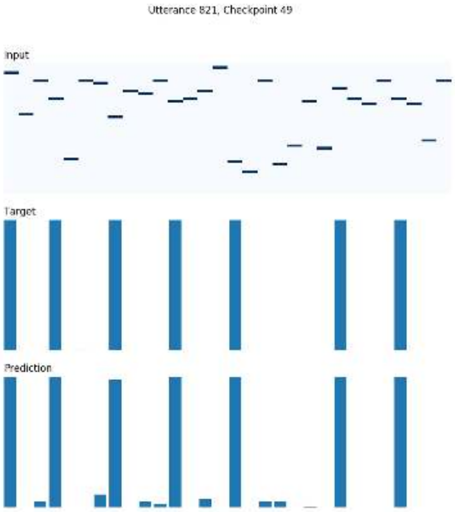
Example: Segmentation learning



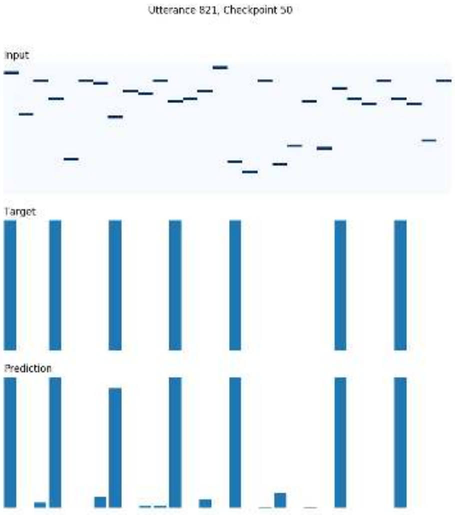
Example: Segmentation learning



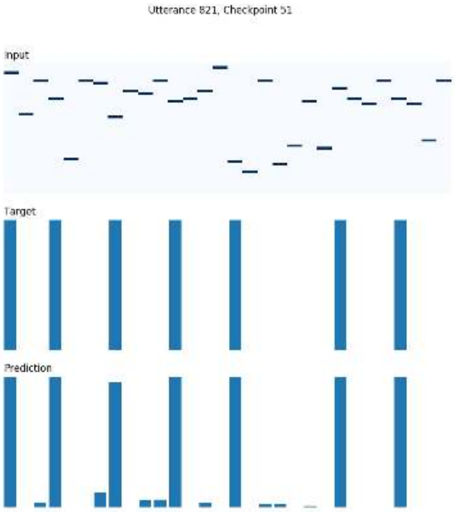
Example: Segmentation learning



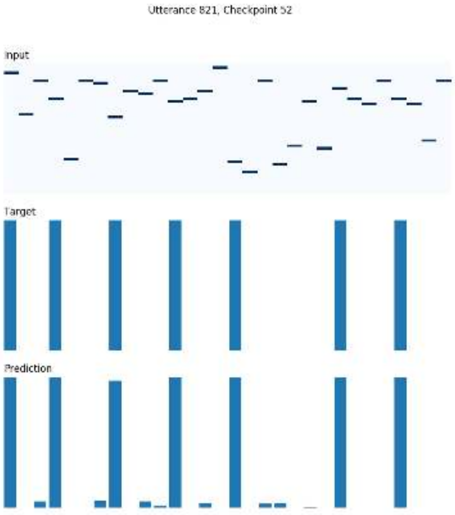
Example: Segmentation learning



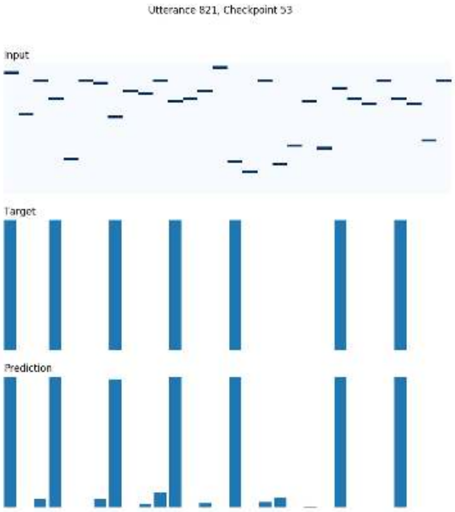
Example: Segmentation learning



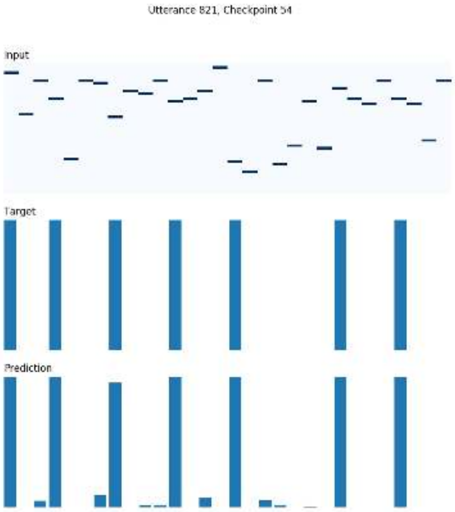
Example: Segmentation learning



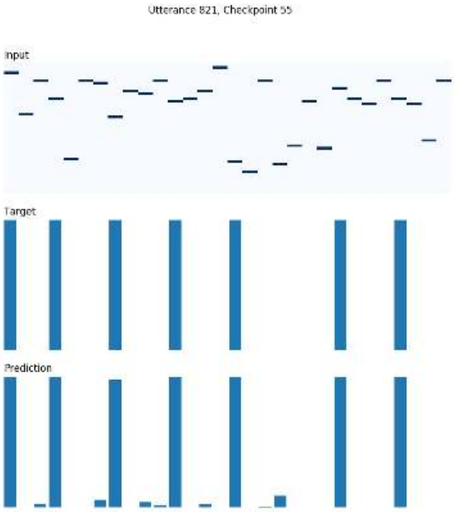
Example: Segmentation learning



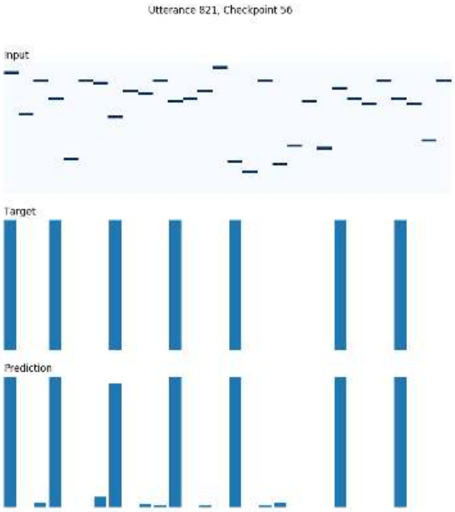
Example: Segmentation learning



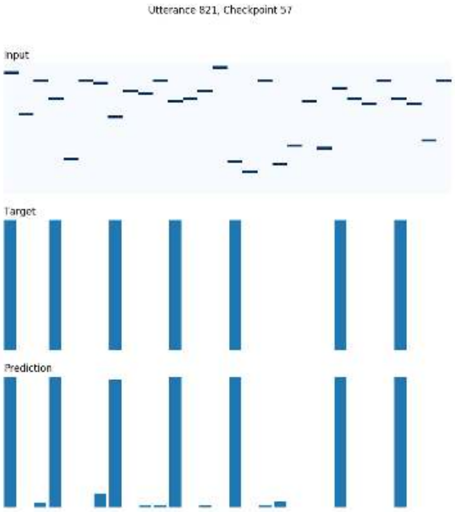
Example: Segmentation learning



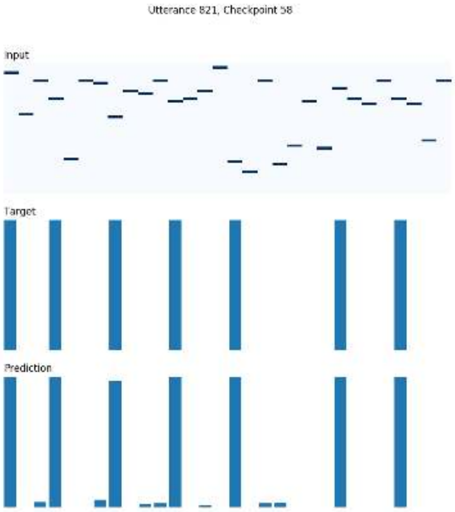
Example: Segmentation learning



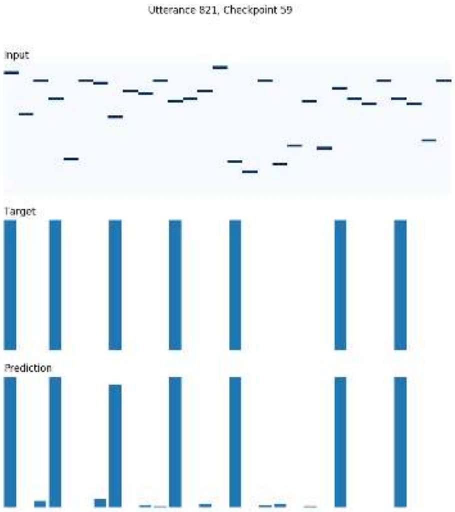
Example: Segmentation learning



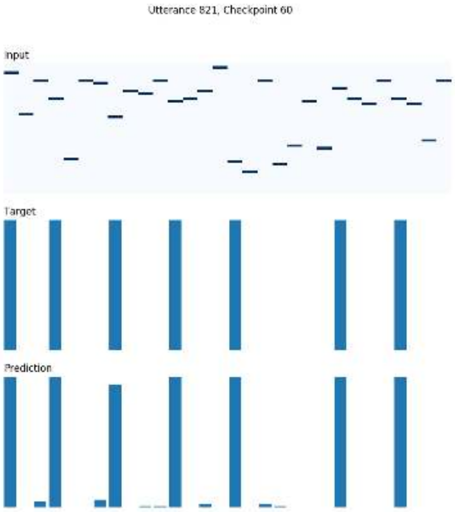
Example: Segmentation learning



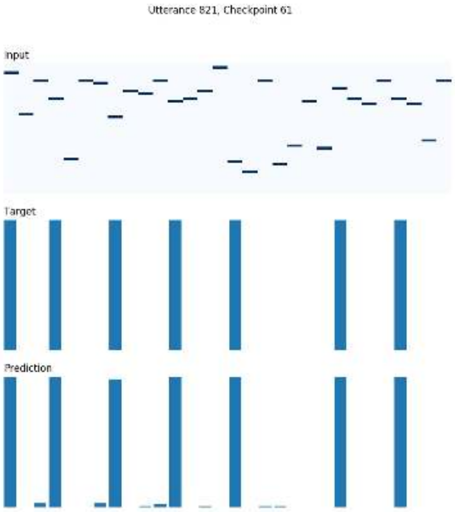
Example: Segmentation learning



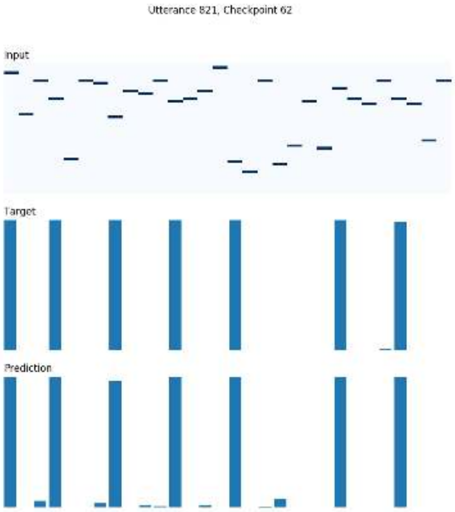
Example: Segmentation learning



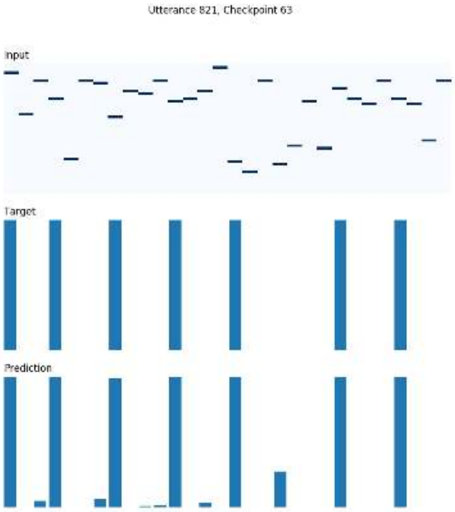
Example: Segmentation learning



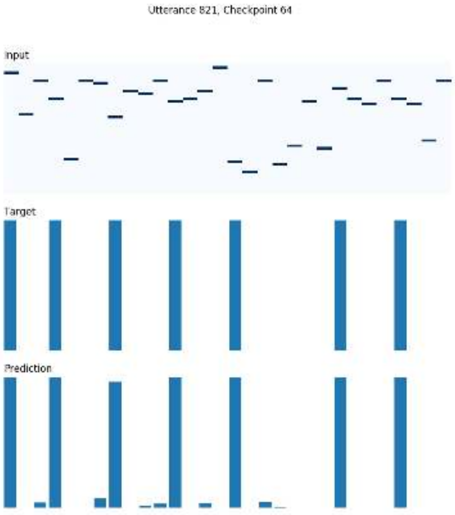
Example: Segmentation learning



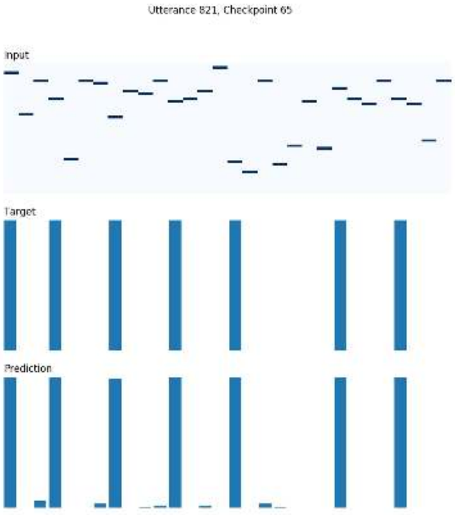
Example: Segmentation learning



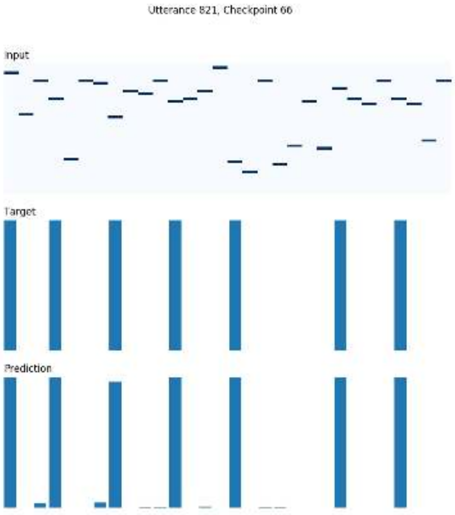
Example: Segmentation learning



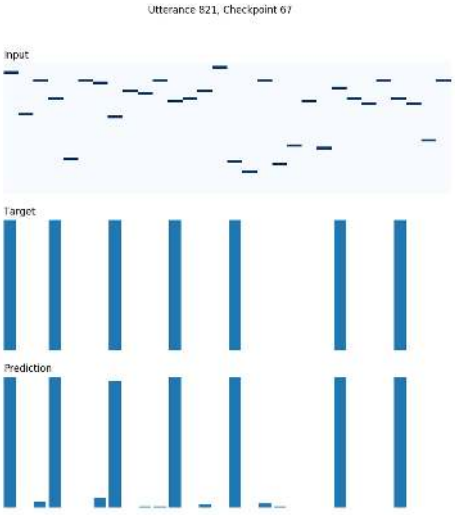
Example: Segmentation learning



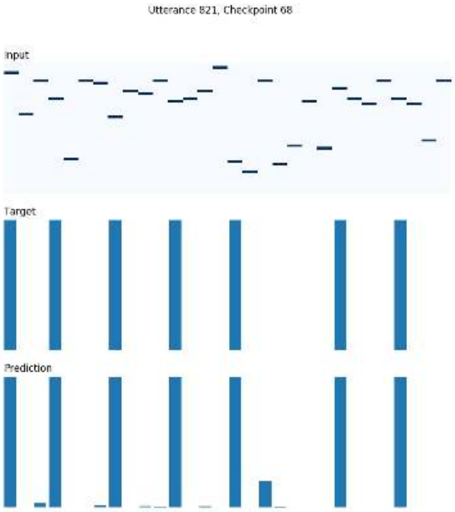
Example: Segmentation learning



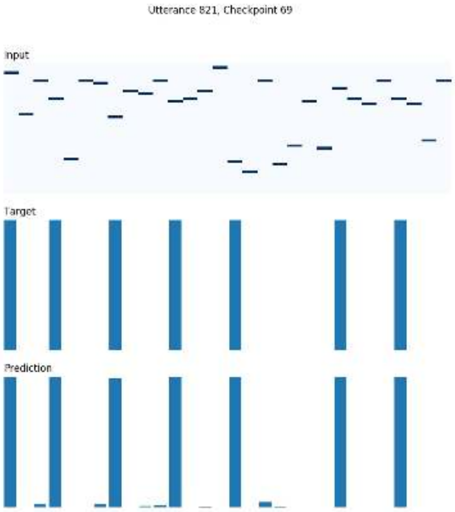
Example: Segmentation learning



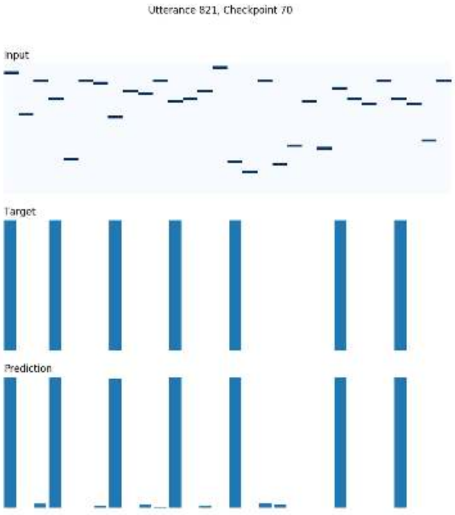
Example: Segmentation learning



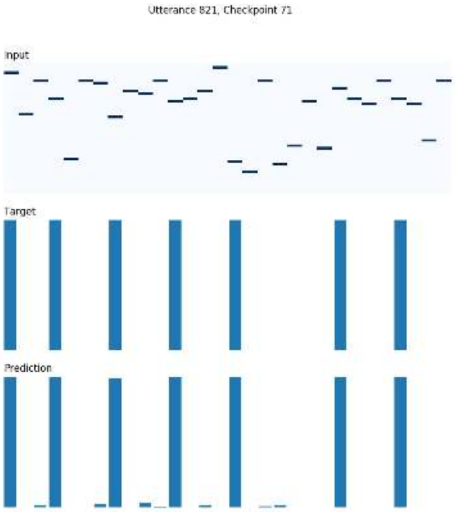
Example: Segmentation learning



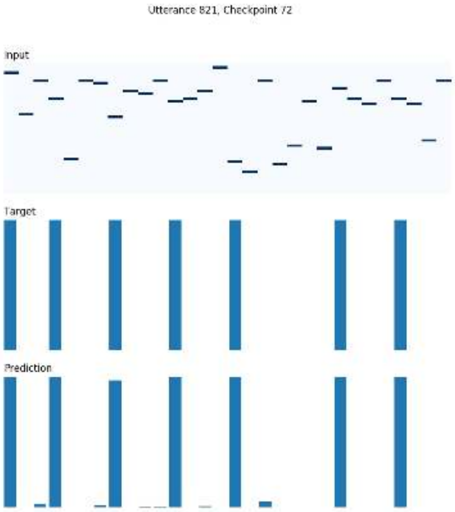
Example: Segmentation learning



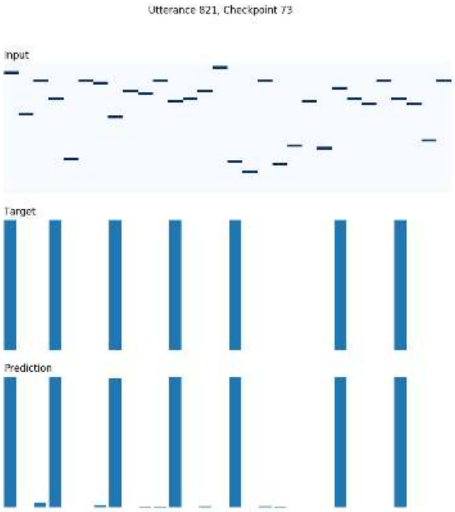
Example: Segmentation learning



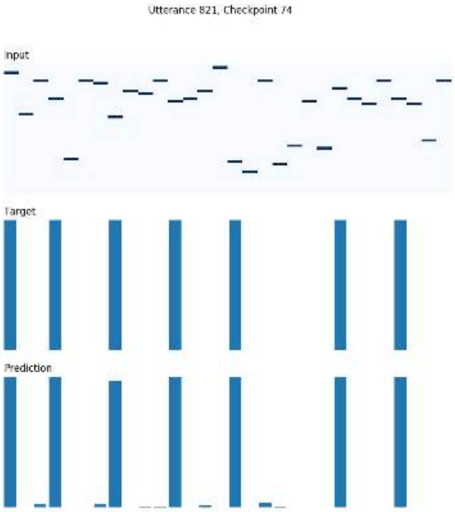
Example: Segmentation learning



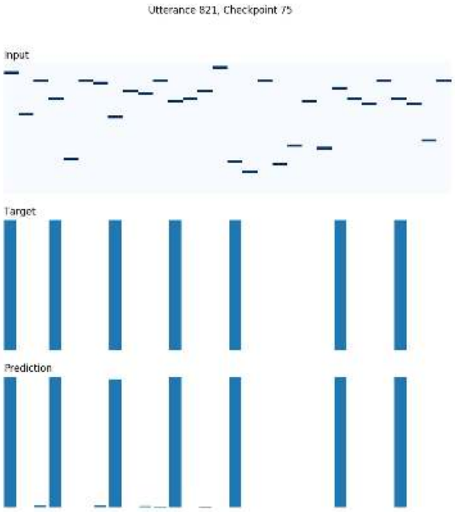
Example: Segmentation learning



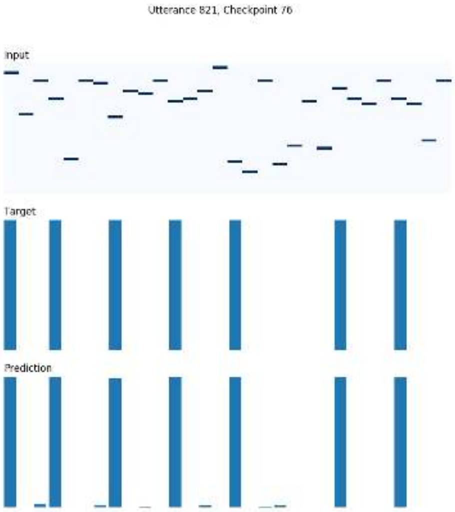
Example: Segmentation learning



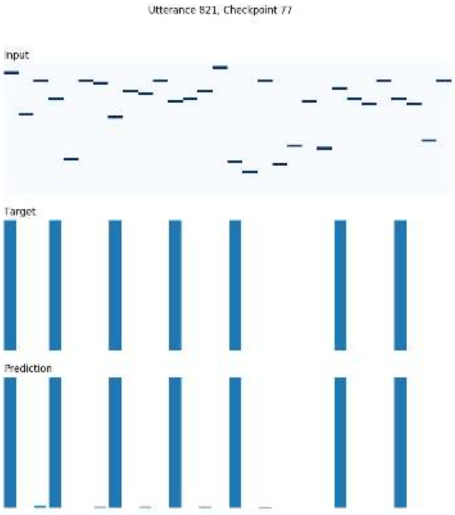
Example: Segmentation learning



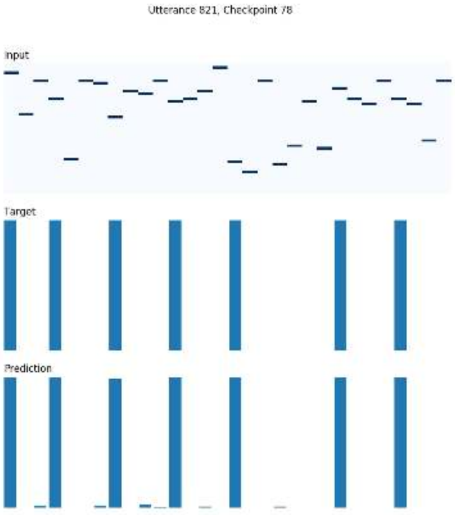
Example: Segmentation learning



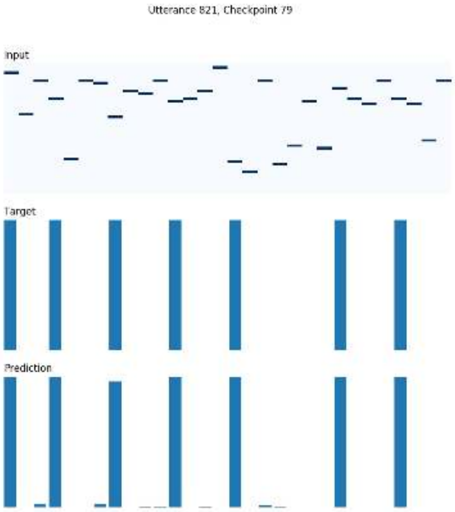
Example: Segmentation learning



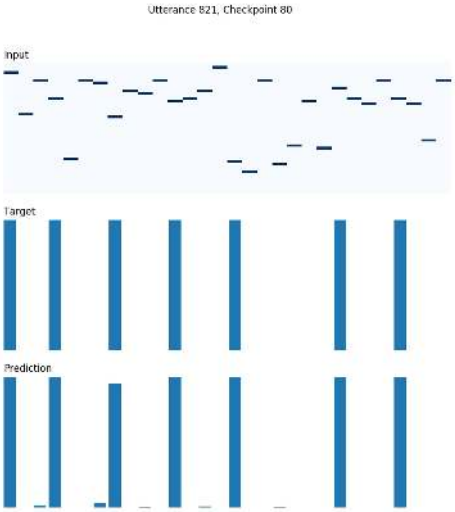
Example: Segmentation learning



Example: Segmentation learning



Example: Segmentation learning



Speech segmentation: Conclusion

Our results support the hypothesis that limited phonological memory facilitates lexical acquisition by encouraging efficient segmentation

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Modeling grammar acquisition with unsupervised PCFG induction

Grammar induction: Cognitive background

- Humans have been shown to use distributional statistics in language acquisition (Saffran et al. 1999)
- Cognitively-constrained grammar induction allows us to study:
    - Utility of word distributions to syntax acquisition
    - Advantages/disadvantages of cognitive constraints

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Cognitive background

- Humans have been shown to use distributional statistics in language acquisition (Saffran et al. 1999)
- Cognitively-constrained grammar induction allows us to study:
  - Utility of word distributions to syntax acquisition
  - Advantages/disadvantages of cognitive constraints

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Cognitive background

- Humans have been shown to use distributional statistics in language acquisition (Saffran et al. 1999)
- Cognitively-constrained grammar induction allows us to study:
  - Utility of word distributions to syntax acquisition
  - Advantages/disadvantages of cognitive constraints

Cory Shain et al. (2016). “Memory-bounded left-corner unsupervised grammar induction on child-directed input”. In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Cognitive background

- Humans have been shown to use distributional statistics in language acquisition (Saffran et al. 1999)
- Cognitively-constrained grammar induction allows us to study:
  - Utility of word distributions to syntax acquisition
  - Advantages/disadvantages of cognitive constraints

Cory Shain et al. (2016). “Memory-bounded left-corner unsupervised grammar induction on child-directed input”. In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Previous work

- Several raw-text constituency parsers exist (e.g. Seginer 2007; Ponvert, Baldridge, and Erik 2011)
- No system besides ours is
  - Depth-bounded (memory-limited)
  - Incremental
- Typically constituents are not labeled

Cory Shain et al. (2016). “Memory-bounded left-corner unsupervised grammar induction on child-directed input”. In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Previous work

- Several raw-text constituency parsers exist (e.g. Seginer 2007; Ponvert, Baldridge, and Erik 2011)
- No system besides ours is
  - Depth-bounded (memory-limited)
  - Incremental
- Typically constituents are not labeled

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Previous work

- Several raw-text constituency parsers exist (e.g. Seginer 2007; Ponvert, Baldridge, and Erik 2011)
- No system besides ours is
  - Depth-bounded (memory-limited)
  - Incremental
- Typically constituents are not labeled

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Previous work

- Several raw-text constituency parsers exist (e.g. Seginer 2007; Ponvert, Baldridge, and Erik 2011)
- No system besides ours is
  - Depth-bounded (memory-limited)
  - Incremental
- Typically constituents are not labeled

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Previous work

- Several raw-text constituency parsers exist (e.g. Seginer 2007; Ponvert, Baldridge, and Erik 2011)
- No system besides ours is
  - Depth-bounded (memory-limited)
  - Incremental
- Typically constituents are not labeled

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Model overview

- Bayesian depth-bounded incremental left-corner PCFG induction system
- Parses with depth-bounded hierarchical hidden Markov model (Schuler et al. 2010)
- Trained using block Gibbs sampling
- Produces a full labeled tree structure and PCFG model

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Model overview

- Bayesian depth-bounded incremental left-corner PCFG induction system
- Parses with depth-bounded hierarchical hidden Markov model (Schuler et al. 2010)
- Trained using block Gibbs sampling
- Produces a full labeled tree structure and PCFG model

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Model overview

- Bayesian depth-bounded incremental left-corner PCFG induction system
- Parses with depth-bounded hierarchical hidden Markov model (Schuler et al. 2010)
- Trained using block Gibbs sampling
- Produces a full labeled tree structure and PCFG model

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Model overview

- Bayesian depth-bounded incremental left-corner PCFG induction system
- Parses with depth-bounded hierarchical hidden Markov model (Schuler et al. 2010)
- Trained using block Gibbs sampling
- Produces a full labeled tree structure and PCFG model

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Experiment

- Experimental conditions designed to mimic conditions of early language learning:
  - **Child-directed input:** Child-directed utterances from the Eve corpus of Brown (1973), distributed with CHILDES (MacWhinney 2000)
  - **Limited depth:** Depth was limited to 2
    - Children have more severe working memory limits than adults (Gathercole 1998)
    - Greater depths rarely needed for child-directed utterances
  - **Small hypothesis space (Newport 1990):** 4 left child categories, 4 right child categories, 8 parts of speech

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Experiment

- Experimental conditions designed to mimic conditions of early language learning:
  - **Child-directed input:** Child-directed utterances from the Eve corpus of Brown (1973), distributed with CHILDES (MacWhinney 2000)
  - **Limited depth:** Depth was limited to 2
    - Children have more severe working memory limits than adults (Gathercole 1998)
    - Greater depths rarely needed for child-directed utterances
  - **Small hypothesis space (Newport 1990):** 4 left child categories, 4 right child categories, 8 parts of speech

Cory Shain et al. (2016). “Memory-bounded left-corner unsupervised grammar induction on child-directed input”. In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Experiment

- Experimental conditions designed to mimic conditions of early language learning:
  - **Child-directed input:** Child-directed utterances from the Eve corpus of Brown (1973), distributed with CHILDES (MacWhinney 2000)
  - **Limited depth:** Depth was limited to 2
    - Children have more severe working memory limits than adults (Gathercole 1998)
    - Greater depths rarely needed for child-directed utterances
  - **Small hypothesis space (Newport 1990):** 4 left child categories, 4 right child categories, 8 parts of speech

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Experiment

- Experimental conditions designed to mimic conditions of early language learning:
  - **Child-directed input:** Child-directed utterances from the Eve corpus of Brown (1973), distributed with CHILDES (MacWhinney 2000)
  - **Limited depth:** Depth was limited to 2
    - Children have more severe working memory limits than adults (Gathercole 1998)
    - Greater depths rarely needed for child-directed utterances
  - **Small hypothesis space (Newport 1990):** 4 left child categories, 4 right child categories, 8 parts of speech

Cory Shain et al. (2016). “Memory-bounded left-corner unsupervised grammar induction on child-directed input”. In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Experiment

- Experimental conditions designed to mimic conditions of early language learning:
  - **Child-directed input:** Child-directed utterances from the Eve corpus of Brown (1973), distributed with CHILDES (MacWhinney 2000)
  - **Limited depth:** Depth was limited to 2
    - Children have more severe working memory limits than adults (Gathercole 1998)
    - Greater depths rarely needed for child-directed utterances
  - **Small hypothesis space (Newport 1990):** 4 left child categories, 4 right child categories, 8 parts of speech

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Experiment

- Experimental conditions designed to mimic conditions of early language learning:
  - **Child-directed input:** Child-directed utterances from the Eve corpus of Brown (1973), distributed with CHILDES (MacWhinney 2000)
  - **Limited depth:** Depth was limited to 2
    - Children have more severe working memory limits than adults (Gathercole 1998)
    - Greater depths rarely needed for child-directed utterances
  - **Small hypothesis space (Newport 1990):** 4 left child categories, 4 right child categories, 8 parts of speech

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Evaluation

- **Gold standard:** Hand-corrected PTB-style trees for Eve (Pearl and Sprouse 2013)
- **Competitors:**
  - CCL (Seginer 2007)
  - UPPARSE (Ponvert, Baldridge, and Erk 2011)
  - BMMM+DMV (Christodoulopoulos, Goldwater, and Steedman 2012)

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Evaluation

- **Gold standard:** Hand-corrected PTB-style trees for Eve (Pearl and Sprouse 2013)
- **Competitors:**
  - CCL (Seginer 2007)
  - UPPARSE (Ponvert, Baldridge, and Erik 2011)
  - BMMM+DMV (Christodoulopoulos, Goldwater, and Steedman 2012)

Cory Shain et al. (2016). “Memory-bounded left-corner unsupervised grammar induction on child-directed input”. In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Evaluation

- **Gold standard:** Hand-corrected PTB-style trees for Eve (Pearl and Sprouse 2013)
- **Competitors:**
  - CCL (Seginer 2007)
  - UPPARSE (Ponvert, Baldridge, and Erik 2011)
  - BMMM+DMV (Christodoulopoulos, Goldwater, and Steedman 2012)

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Evaluation

- **Gold standard:** Hand-corrected PTB-style trees for Eve (Pearl and Sprouse 2013)
- **Competitors:**
  - CCL (Seginer 2007)
  - UPPARSE (Ponvert, Baldridge, and Erik 2011)
  - BMMM+DMV (Christodoulopoulos, Goldwater, and Steedman 2012)

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Evaluation

- **Gold standard:** Hand-corrected PTB-style trees for Eve (Pearl and Sprouse 2013)
- **Competitors:**
  - CCL (Seginer 2007)
  - UPPARSE (Ponvert, Baldridge, and Erik 2011)
  - BMMM+DMV (Christodoulopoulos, Goldwater, and Steedman 2012)

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: COLING results

	P	R	F_1
Our system	68.83	57.18	62.47
Random baseline (Ours 1st iter)	51.69	38.75	44.30

Unlabeled bracketing accuracy on Eve

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

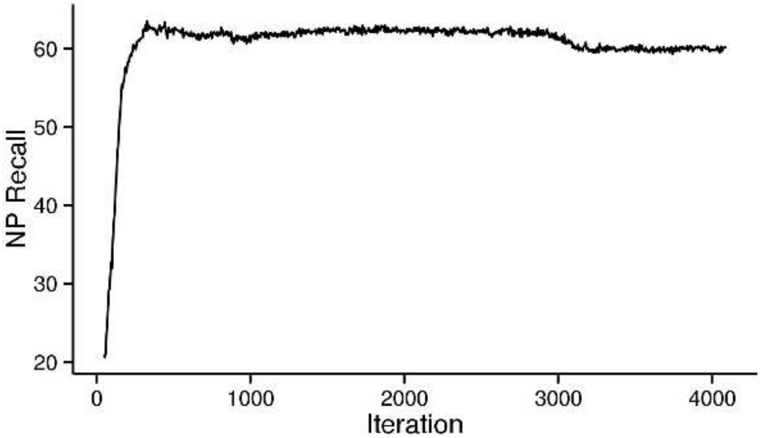
Grammar induction: COLING results

	P	R	F_1
UPPARSE	60.50	51.96	55.90
CCL	64.70	53.47	58.55
BMMM+DMV	63.63	**64.02**	**63.82**
Our system	**68.83**	57.18	62.47
Random baseline (Ours 1st iter)	51.69	38.75	44.30

Unlabeled bracketing accuracy on Eve

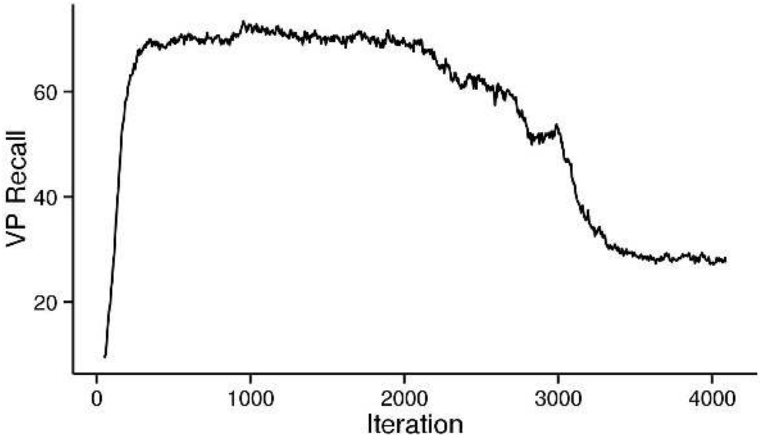
Cory Shain et al. (2016). “Memory-bounded left-corner unsupervised grammar induction on child-directed input”. In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Error analysis



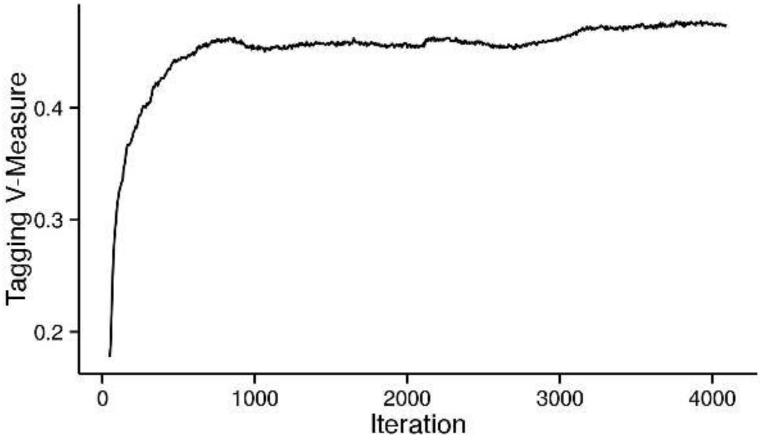
Percent gold noun phrases (NPs) discovered

Grammar induction: Error analysis



Percent gold verb phrases (VPs) discovered

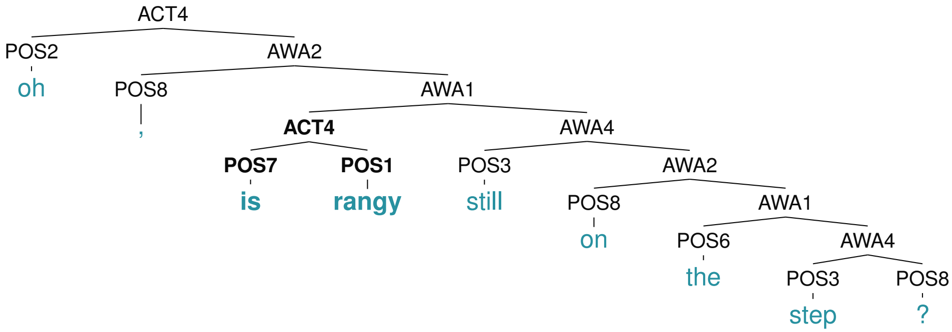
Grammar induction: Error analysis



Part-of-speech tagging (V-Measure)

Grammar induction: Constructions of interest

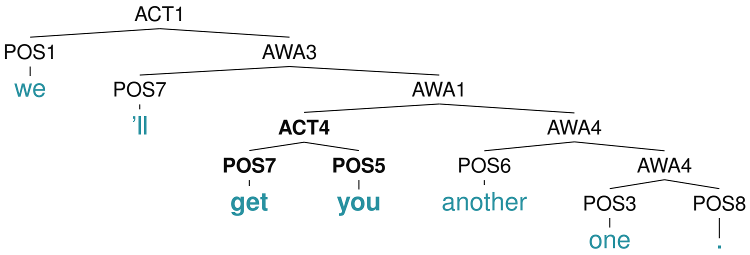
Subject-auxiliary inversion: (c.f. Chomsky 1968)



Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Constructions of interest

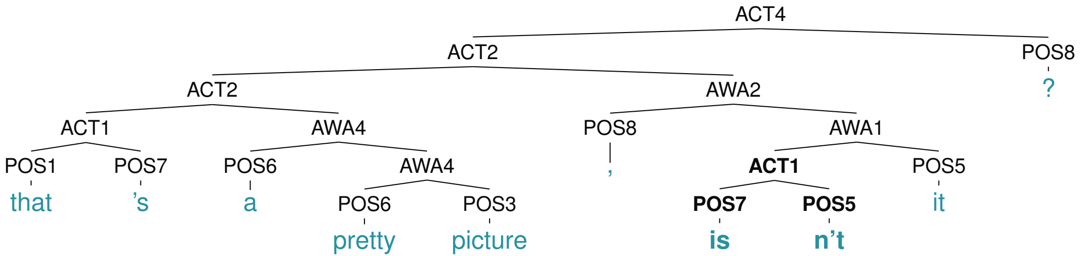
Ditransitive:



Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Constructions of interest

Contraction:



Cory Shain et al. (2016). “Memory-bounded left-corner unsupervised grammar induction on child-directed input”. In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Hot off the press

- Since COLING:
  - Merged left, right, and PoS category spaces
  - Depth=1 run on Eve got $F_1 = 71$
- Additional constraints on search space facilitate learning

Grammar induction: Hot off the press

- Since COLING:
  - Merged left, right, and PoS category spaces
  - Depth=1 run on Eve got $F_1 = 71$
- Additional constraints on search space facilitate learning

Grammar induction: Hot off the press

- Since COLING:
  - Merged left, right, and PoS category spaces
  - Depth=1 run on Eve got $F_1 = 71$
- Additional constraints on search space facilitate learning

Grammar induction: Hot off the press

- Since COLING:
  - Merged left, right, and PoS category spaces
  - Depth=1 run on Eve got $F_1 = 71$
- Additional constraints on search space facilitate learning

Grammar induction: Conclusions

- Word distributions contain a substantial amount of information about English syntax
- This information is detectible by a cognitively-constrained learner
- There is still much room for improvement
  - Some residue may be unlearnable without additional cues (e.g. vision) or innate bias
  - Some residue may be captured by improved induction techniques

Grammar induction: Conclusions

- Word distributions contain a substantial amount of information about English syntax
- This information is detectible by a cognitively-constrained learner
- There is still much room for improvement
  - Some residue may be unlearnable without additional cues (e.g., vision) or innate bias
  - Some residue may be captured by improved induction techniques

Grammar induction: Conclusions

- Word distributions contain a substantial amount of information about English syntax
- This information is detectible by a cognitively-constrained learner
- There is still much room for improvement
  - Some residue may be unlearnable without additional cues (e.g. vision) or innate bias
  - Some residue may be captured by improved induction techniques

Grammar induction: Conclusions

- Word distributions contain a substantial amount of information about English syntax
- This information is detectible by a cognitively-constrained learner
- There is still much room for improvement
  - Some residue may be unlearnable without additional cues (e.g. vision) or innate bias
  - Some residue may be captured by improved induction techniques

Grammar induction: Conclusions

- Word distributions contain a substantial amount of information about English syntax
- This information is detectible by a cognitively-constrained learner
- There is still much room for improvement
  - Some residue may be unlearnable without additional cues (e.g. vision) or innate bias
  - Some residue may be captured by improved induction techniques

Conclusion

Conclusion

- Unsupervised NLP approaches to speech segmentation and parsing can shed light on language acquisition
- Speech segmenter results show that memory pressures encourage learning efficient representations
- Grammar induction results show that much syntax can be acquired from word distributions alone

Conclusion

- Unsupervised NLP approaches to speech segmentation and parsing can shed light on language acquisition
- Speech segmenter results show that memory pressures encourage learning efficient representations
- Grammar induction results show that much syntax can be acquired from word distributions alone

Conclusion

- Unsupervised NLP approaches to speech segmentation and parsing can shed light on language acquisition
- Speech segmenter results show that memory pressures encourage learning efficient representations
- Grammar induction results show that much syntax can be acquired from word distributions alone

Conclusion

- Together, these systems might take us closer to a full computational model of language acquisition
- Dense word embeddings can be obtained from raw speech
- (Soon:) PCFG can be trained from dense word representations
- If pipelined, these approaches could go from acoustics to syntax trees, completely unsupervised

Conclusion

- Together, these systems might take us closer to a full computational model of language acquisition
- Dense word embeddings can be obtained from raw speech
- (Soon:) PCFG can be trained from dense word representations
- If pipelined, these approaches could go from acoustics to syntax trees, completely unsupervised

Conclusion

- Together, these systems might take us closer to a full computational model of language acquisition
- Dense word embeddings can be obtained from raw speech
- (Soon:) PCFG can be trained from dense word representations
- If pipelined, these approaches could go from acoustics to syntax trees, completely unsupervised

Conclusion

- Together, these systems might take us closer to a full computational model of language acquisition
- Dense word embeddings can be obtained from raw speech
- (Soon:) PCFG can be trained from dense word representations
- If pipelined, these approaches could go from acoustics to syntax trees, completely unsupervised

Thank you!

To you, co-authors, anonymous reviewers of submitted papers, and members of various discussion groups who gave feedback.

Computations for this project were run on a Titan-X GPU donated by the NVIDIA Hardware Grant program and on the Ohio Supercomputer (1987). Funding was provided by NSF #1422987.

This project was sponsored by the Defense Advanced Research Projects Agency award #HR0011-15-2-0022. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

Segmenter Github:

`https://github.com/melsner/neural-segmentation`

Parser Github:

`https://github.com/tmills/uhhmm/`

References I

Baddeley, Alan, Susan Gathercole, and Costanza Papagno (1998). "The Phonological Loop as a Language Learning Device". In: *Psychological Review* 105.1, pp. 158–173.

Baddeley, Alan D. and Graham Hitch (1974). *Working Memory*. Stirling, Scotland: University of Stirling.

Brent, Michael R. (1999). "An efficient, probabilistically sound algorithm for segmentation and word discovery". In: *Machine Learning* 34, pp. 71–105.

Briscoe, Ted (2000). "Grammatical Acquisition: Inductive Bias and Coevolution of Language and the Language Acquisition Device". In: *Language* 76.2, pp. 245–296. ISSN: 00978507, 15350665. URL: http://www.jstor.org/stable/417657.

Brown, R. (1973). *A First Language*. Cambridge, MA: Harvard University Press.

Center, Ohio Supercomputer (1987). *Ohio Supercomputer Center*. http://osc.edu/ark:/19495/f5s1ph73.

Chomsky, Noam (1968). *Language and Mind*. New York: Harcourt, Brace & World.

References II

Christodoulopoulos, Christos, Sharon Goldwater, and Mark Steedman (2012). "Turning the pipeline into a loop: Iterated unsupervised dependency parsing and PoS induction". In: *NAACL-HLT Workshop on the Induction of Linguistic Structure*. Montreal, Canada, pp. 96–99.

Cowan, Nelson (2001). "The magical number 4 in short-term memory: A reconsideration of mental storage capacity". In: *Behavioral and Brain Sciences* 24. A frequent error owed to a misprint in Cowan 2001, which see., pp. 87–185.

Dahl, George et al. (2011). "Large Vocabulary Continuous Speech Recognition With Context-Dependent DBN-HMMS". In: *Proc. ICASSP, Prague*. IEEE. URL: https://www.microsoft.com/en-us/research/publication/large-vocabulary-continuous-speech-recognition-with-context-dependent-dbn-hmms/.

Elman, Jeffrey L. (1991). "Distributed representations, simple recurrent networks, and grammatical structure". In: *Machine Learning* 7, pp. 195–225.

Elsner, Micha and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*.

References III

Fodor, Janet Dean and William Gregory Sakas (2004). "Evaluating models of parameter setting". In: *BUCLD 28*. Ed. by A. Brugos, L. Micciulla, and C. E. Smith. Somerville: Cascadilla Press.

Gathercole, Susan E. (1998). "The development of memory". In: *Journal of Child Psychology and Psychiatry* 39 (1), pp. 3–27.

Johnson-Laird, Philip N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA, USA: Harvard University Press. ISBN: 0-674-56882-6.

MacWhinney, Brian (2000). *The CHILDES project: Tools for analyzing talk*. Third. Mahwah, NJ: Lawrence Elrbaum Associates.

Marslen-Wilson, William D. (1975). "Sentence Perception as an Interactive Parallel Process". In: *Science* 189.4198, pp. 226–228.

McElree, Brian (2001). "Working Memory and Focal Attention". In: *Journal of Experimental Psychology, Learning Memory and Cognition* 27.3, pp. 817–835.

References IV

Miller, George A. (1956). “The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information”. In: *Psychological Review* 63, pp. 81–97.

Mnih, Volodymyr et al. (2014). “Recurrent models of visual attention”. In: *Advances in Neural Information Processing Systems*, pp. 2204–2212.

Newport, Elissa (1990). “Maturational constraints on language learning”. In: *Cognitive Science* 14, pp. 11–28.

Pearl, Lisa and Jon Sprouse (2013). “Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem”. In: *Language Acquisition* 20, pp. 23–68.

Ponvert, Elias, Jason Baldridge, and Katrin Erik (2011). “Simple unsupervised grammar induction from raw text with cascaded finite state models”. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*. Portland, Oregon, pp. 1077–1086.

Saffran, Jenny R et al. (1999). “Statistical learning of tone sequences by human infants and adults”. In: *Cognition* 70.1, pp. 27–52.

References V

Schijndel, Marten van, Andy Exley, and William Schuler (2013). "A model of language processing as hierarchic sequential prediction". In: *Topics in Cognitive Science* 5.3. Ed. by John Hale and David Reitter, pp. 522–540.

Schuler, William et al. (2010). "Broad-Coverage Incremental Parsing using Human-Like Memory Constraints". In: *Computational Linguistics* 36.1, pp. 1–30.

Seginer, Yoav (2007). "Fast Unsupervised Incremental Parsing". In: *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pp. 384–391.

Shain, Cory et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975.

Tanenhaus, Michael K. et al. (1995). "Integration of visual and linguistic information in spoken language comprehension". In: *Science* 268, pp. 1632–1634.

Versteegh, Maarten et al. (2015). "The Zero Resource Speech Challenge 2015". In: *Proceedings of Interspeech 2015*.

References VI

Xu, Kelvin et al. (2015). “Show, Attend and Tell: Neural Image Caption Generation with Visual Attention”. In: *Proceedings of ICML*. Lille, France: JMLR.

Appendix

Speech segmentation: Algorithm

1. For each training epoch:
    - 1.1 For each batch of n utterances in the training data
        - 1.1.1 Generate a proposal distribution (segmenter network output)
        - 1.1.2 Sample m segmentations from proposal distribution
        - 1.1.3 Compute new proposal distribution as normalized sum of segmentations weighted by reconstruction loss
        - 1.1.4 Train segmenter network on new proposal distribution

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Algorithm

1. For each training epoch:
   1.1 For each batch of *n* utterances in the training data
      1.1.1 Generate a proposal distribution (segmenter network output)
      1.1.2 Sample *m* segmentations from proposal distribution
      1.1.3 Compute new proposal distribution as normalized sum of segmentations weighted by reconstruction loss
      1.1.4 Train segmenter network on new proposal distribution

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Algorithm

1. For each training epoch:
   1.1 For each batch of *n* utterances in the training data
      1.1.1 Generate a proposal distribution (segmenter network output)
      1.1.2 Sample *m* segmentations from proposal distribution
      1.1.3 Compute new proposal distribution as normalized sum of segmentations weighted by reconstruction loss
      1.1.4 Train segmenter network on new proposal distribution

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Algorithm

1. For each training epoch:
   1.1 For each batch of *n* utterances in the training data
      1.1.1 Generate a proposal distribution (segmenter network output)
      1.1.2 Sample *m* segmentations from proposal distribution
      1.1.3 Compute new proposal distribution as normalized sum of segmentations weighted by reconstruction loss
      1.1.4 Train segmenter network on new proposal distribution

Micha Elsner and Cory Shain (to appear). “Speech segmentation with a neural encoder model of working memory”. In: *EMNLP 2017*

Speech segmentation: Algorithm

1. For each training epoch:
   1.1 For each batch of *n* utterances in the training data
      1.1.1 Generate a proposal distribution (segmenter network output)
      1.1.2 Sample *m* segmentations from proposal distribution
      1.1.3 Compute new proposal distribution as normalized sum of segmentations weighted by reconstruction loss
      1.1.4 Train segmenter network on new proposal distribution

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Algorithm

1. For each training epoch:
   1.1 For each batch of *n* utterances in the training data
      1.1.1 Generate a proposal distribution (segmenter network output)
      1.1.2 Sample *m* segmentations from proposal distribution
      1.1.3 Compute new proposal distribution as normalized sum of segmentations weighted by reconstruction loss
      1.1.4 Train segmenter network on new proposal distribution

Micha Elsner and Cory Shain (to appear). "Speech segmentation with a neural encoder model of working memory". In: *EMNLP 2017*

Speech segmentation: Sampling procedure

Given a set of m sampled boundary sequences $B_1..B_m$ with associated reconstruction losses $L_1...L_m$:

$$P(x|B_i) = \frac{P(B_i|x)P(B_i)}{P(x)} \approx \frac{\exp(L_i)}{\sum_j \exp(L_j)} \tag{1}$$

$$w_i^t = \frac{P(x|B_i)}{P_{seg}^t(B_i^t)} \tag{2}$$

$$\mathbb{E}[B(t)] \approx \frac{1}{\sum_i w_i^t} \sum_i w_i^t B_i^t \tag{3}$$

Speech segmentation: Tweaks for acoustics

- Importance sampling caused oversegmentation
- We suspect that this is due to non-independence between samples, exaggerated by longer sequences
- Acoustic results were obtained via 1-best sampling

Speech segmentation: Experiment parameters

- **Brent:**
  - Max characters per utterance: 30
  - Max words per utterance: 10
  - Max characters per word: 7
  - Phonological AE hidden units: 80
  - Utterance AE hidden units: 400
  - Segmenter hidden units: 100
  - Phonological AE dropout probability: 0.5
  - Utterance AE dropout probability: 0.25

Speech segmentation: Experiment parameters

- **Zerospeech:**
  - Max frames per utterance: 400
  - Max words per utterance: 16
  - Max frames per word: 100
  - Phonological AE hidden units: 20
  - Utterance AE hidden units: 400
  - Segmenter hidden units: 1500
  - Phonological AE dropout probability: 0
  - Utterance AE dropout probability: 0.25

Grammar induction: Algorithm

1. **Initalization:** Randomly sample HHMM parameters
2. For each training iteration:
   2.1 **Parsing:** For each sentence in input:

   Update HHMM parameters from sampled trees

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Algorithm

1. **Initalization:** Randomly sample HHMM parameters
2. For each training iteration:
   2.1 **Parsing:** For each sentence in input:
       2.1.1 **Forward pass:** Compute posterior over HHMM states left to right
       2.1.2 **Backward pass:** Sample states right to left
   2.2 Update HHMM parameters from sampled counts

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Algorithm

1. **Initalization:** Randomly sample HHMM parameters
2. For each training iteration:
   2.1 **Parsing:** For each sentence in input:
      2.1.1 **Forward pass:** Compute posterior over HHMM states left to right
      2.1.2 **Backward pass:** Sample states right to left

   2.2 Update HHMM parameters from sampled counts

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Algorithm

1. **Initalization:** Randomly sample HHMM parameters
2. For each training iteration:
   2.1 **Parsing:** For each sentence in input:
      2.1.1 **Forward pass:** Compute posterior over HHMM states left to right
      2.1.2 **Backward pass:** Sample states right to left
   2.2 Update HHMM parameters from sampled counts

Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Algorithm

1. **Initalization:** Randomly sample HHMM parameters
2. For each training iteration:
    2.1 **Parsing:** For each sentence in input:
        2.1.1 **Forward pass:** Compute posterior over HHMM states left to right
        2.1.2 **Backward pass:** Sample states right to left
    2.2 Update HHMM parameters from sampled counts

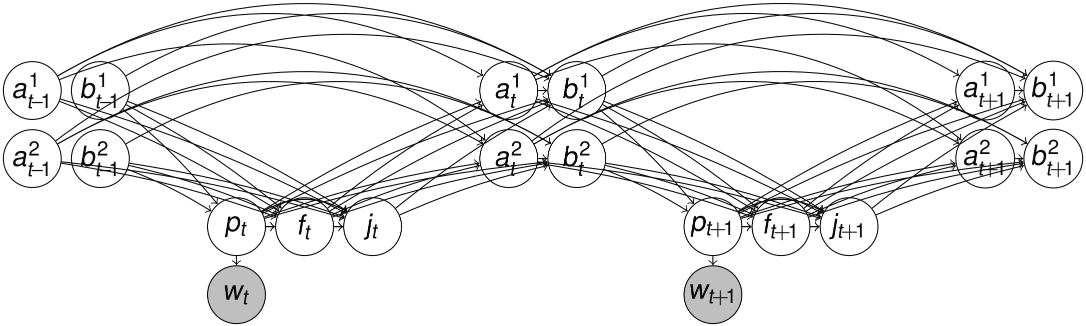
Cory Shain et al. (2016). “Memory-bounded left-corner unsupervised grammar induction on child-directed input”. In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: Algorithm

1. **Initalization:** Randomly sample HHMM parameters
2. For each training iteration:
   2.1 **Parsing:** For each sentence in input:
      2.1.1 **Forward pass:** Compute posterior over HHMM states left to right
      2.1.2 **Backward pass:** Sample states right to left
   2.2 Update HHMM parameters from sampled counts

Cory Shain et al. (2016). “Memory-bounded left-corner unsupervised grammar induction on child-directed input”. In: *Proceedings of The 26th International Conference on Computational Linguistics*. Osaka, pp. 964–975

Grammar induction: HHMM Graphical model



Grammar induction: Punctuation

- Punctuation poses a problem — keep or remove?
  - **Remove:** Doesn't exist in input to human learners.
  - **Keep:** Might be proxy for intonational phrasal cues.
- Punctuation was kept in training data in main result presented above.
- We did an additional UHHMM run trained on data with punctuation removed (2000 iterations).

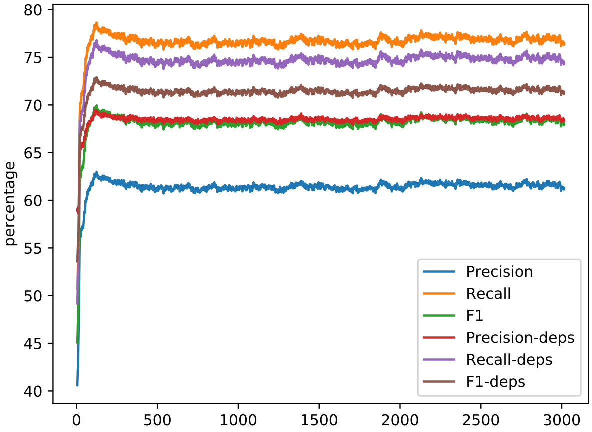
Grammar induction: Full COLING Results

	With punc			No punc		
	P	R	F1	P	R	F1
UPPARSE	60.50	51.96	55.90	38.17	48.38	42.67
CCL	64.70	53.47	58.55	56.87	47.69	51.88
BMMM+DMV (directed)	62.08	62.51	62.30	61.01	59.24	60.14
BMMM+DMV (undirected)	63.63	**64.02**	**63.82**	61.34	**59.33**	**60.32**
UHHMM-4000, binary	46.68	58.28	51.84	37.62	46.97	41.78
UHHMM-4000, flattened	**68.83**	57.18	62.47	**61.78**	45.52	52.42
Right-branching	68.73	**85.81**	**76.33**	**68.73**	**85.81**	**76.33**

Table 1: Parsing accuracy on Eve with and without punctuation (phrasal cues) in the input. The UHHMM systems were given 8 PoS categories while the BMMM+DMV systems were given 45. UPPARSE and CCL do not learn PoS tags. Only the UHHMM systems model limited working memory capacity or incremental left-corner parsing.

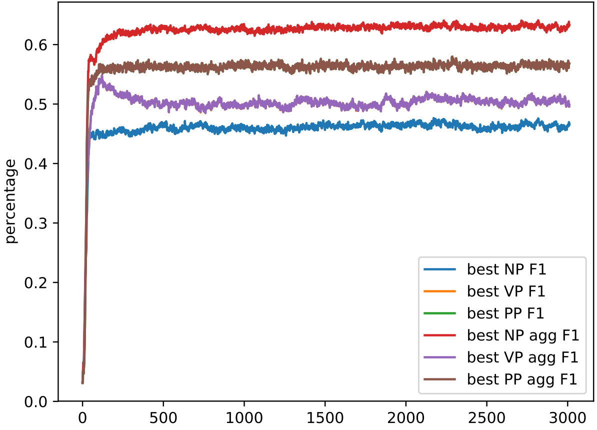
Cory Shain et al. (2016). "Memory-bounded left-corner unsupervised grammar induction on child-directed input". In: *Proceedings of The 26th International Conference on Computational*

Grammar induction: Newer results



Learning curves on Eve

Grammar induction: Newer results



Category learning on Eve