Exploring Controllable Text Generation Techniques

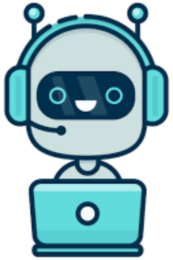
Shrimai Prabhumoye, Alan W Black,
Ruslan Salakhutdinov

Carnegie Mellon University
Language Technologies Institute

Controllable Text Generation allows us to add “knobs” to control the attributes of the text to be generated

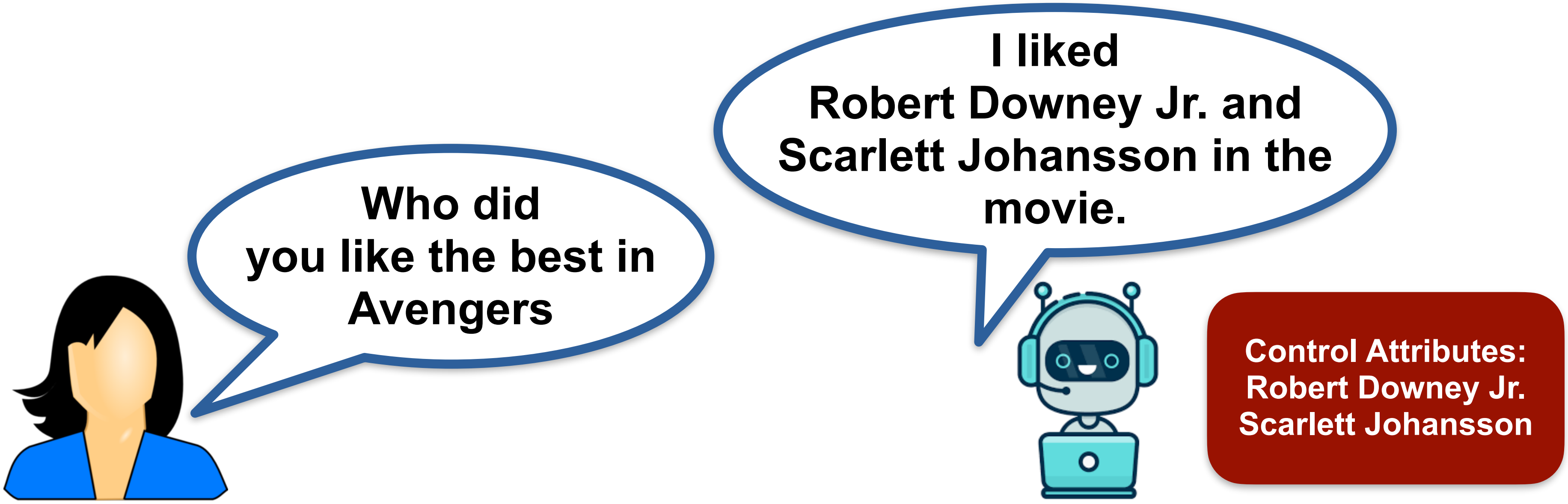
Controllable Text Generation allows us to add "knobs" to control the attributes of the text to be generated

Controllable Text Generation allows us to add "knobs" to control the attributes of the text to be generated

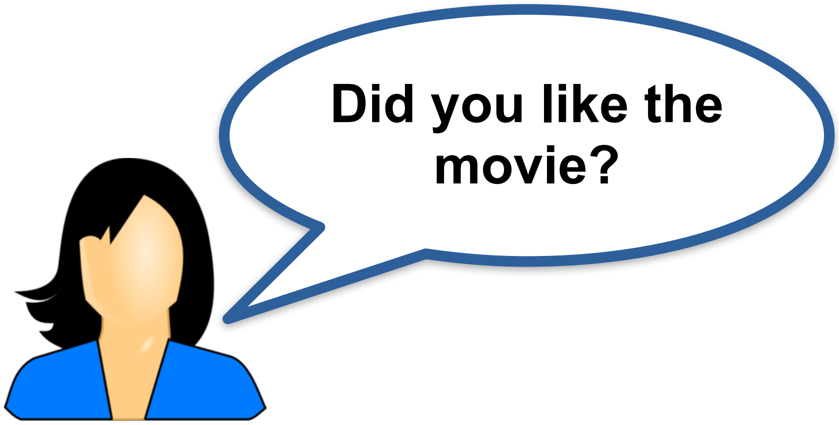


Control Attributes:
Robert Downey Jr.
Scarlett Johansson

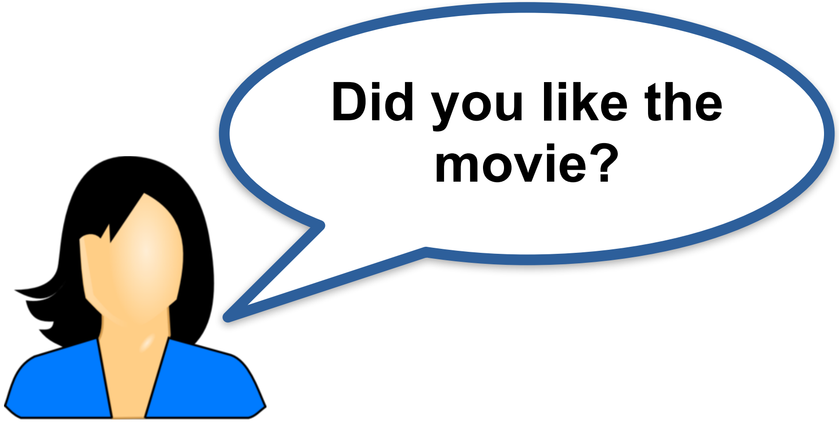
Controllable Text Generation allows us to add "knobs" to control the attributes of the text to be generated

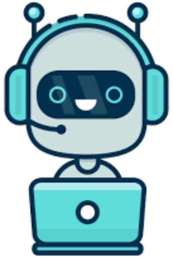


Controllable Text Generation allows us to add “knobs” to control the attributes of the text to be generated



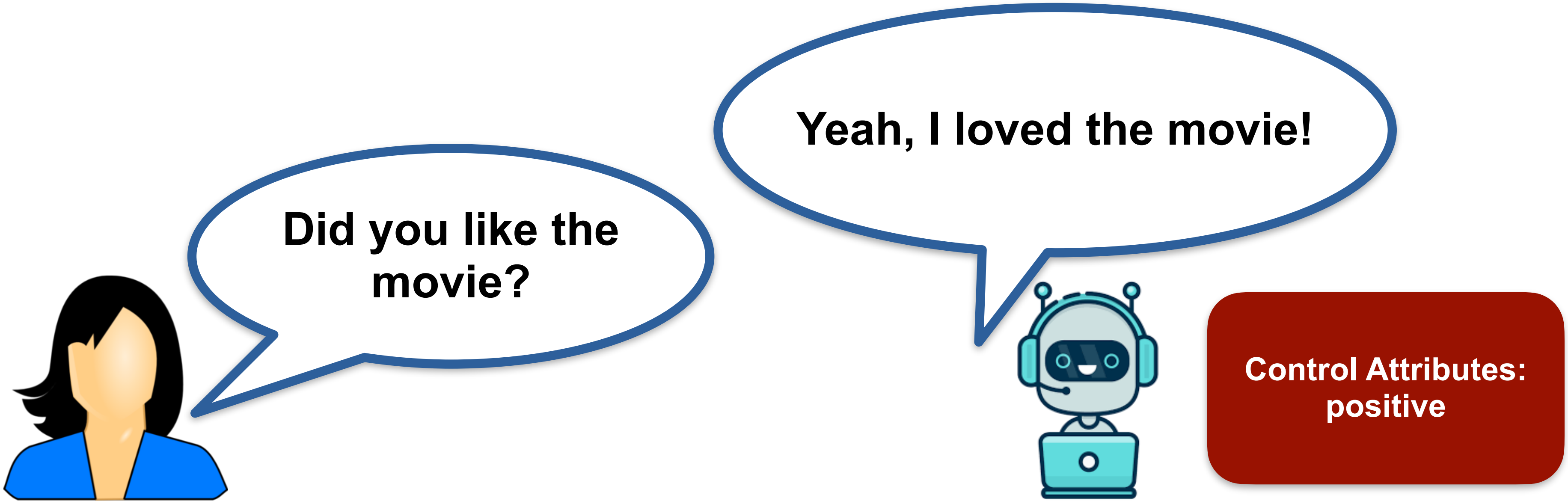
Controllable Text Generation allows us to add “knobs” to control the attributes of the text to be generated



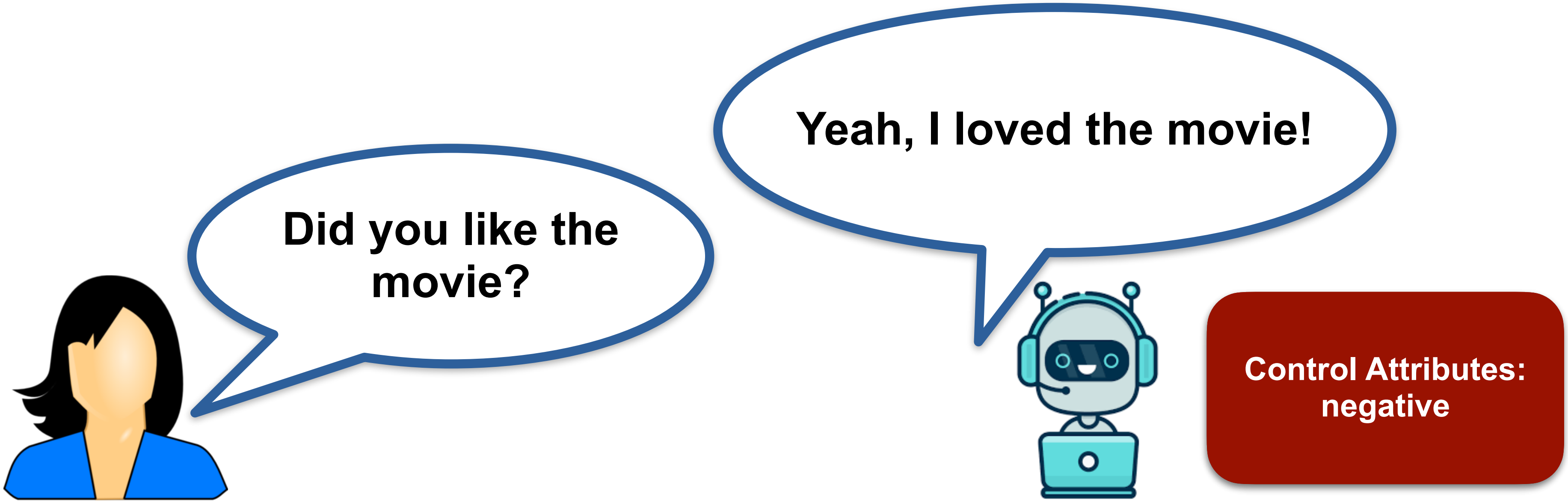


Control Attributes: positive

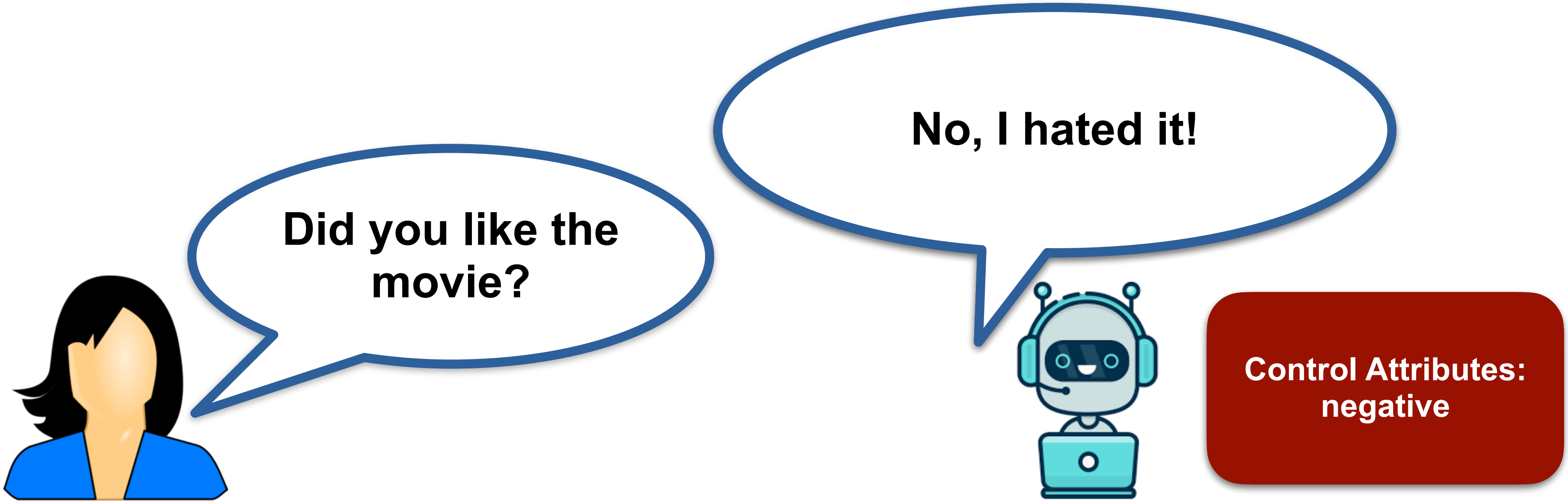
Controllable Text Generation allows us to add “knobs” to control the attributes of the text to be generated



Controllable Text Generation allows us to add “knobs” to control the attributes of the text to be generated



Controllable Text Generation allows us to add "knobs" to control the attributes of the text to be generated



Applications

- ***Dialogue System***
  - Persona, style of responses (polite, authority), content of responses, topic of conversation
- Recommend ***polite emails***
- ***Story Generation***
  - plot, ending, sentiment, topic, persona
- ***Report Generation*** (websites, Wikipedia articles)

Motivation

Style Transfer Through Back-Translation

Shrimai Prabhumoye, Yulia Tsvetkov, Ruslan Salakhutdinov, Alan W Black
Carnegie Mellon University, Pittsburgh, PA, USA
{sprabhum,ytsvetko,rsalakhu,awb}@cs.cmu.edu

Abstract

Style transfer is the task of rephrasing the text to contain specific stylistic properties without changing the intent or affect within the context. This paper introduces

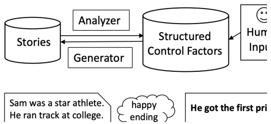
These goals have motivated a considerable amount of recent research efforts focused at "controlled" language generation—aiming at separating the semantic content of *what* is said from the stylistic dimensions of *how* it is said. These include approaches relying on heuristic substitu-

Towards Controllable Story Generation

Nanyun Peng Marjan Ghazvininejad Jonathan May Kevin Knight
Information Sciences Institute & Computer Science Department
University of Southern California
{npeng,ghazvini,jonmay,knight}@isi.edu

Abstract

/e present a general framework of analyzing xisting story corpora to generate controllable nd creative new stories. The proposed frame- /ork needs little manual annotation to achieve ontrollable story generation. It creates a new



Politeness Transfer: A Tag and Generate Approach

Aman Madaan *, Amrith Setlur *, Tanmay Parekh *, Barnabas Poczos, Graham Neubig, Yiming Yang, Ruslan Salakhutdinov, Alan W Black, Shrimai Prabhumoye
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA, USA
{amadaan, asetlur, tparekh}@cs.cmu.edu

Abstract

This paper introduces a new task of politeness transfer which involves converting non-polite sentences to polite sentences while preserving the meaning. We also provide a dataset of more than 1.39 million instances automatically

Rao and Tetreault, 2018; Xu et al., 2012; Jhamtani et al., 2017) has not focused on politeness as a style transfer task, and we argue that defining it is cumbersome. While native speakers of a language and cohabitants of a region have a good working understanding of the phenomenon of politeness

Published as a conference paper at ICLR 2019

🧙 OF WIKIPEDIA: KNOWLEDGE-POWERED CONVERSATIONAL AGENTS

Emily Dinan*, Stephen Roller*, Kurt Shuster*, Angela Fan, Michael Auli, Jason Weston
Facebook AI Research
{edinan,roller,kshuster,angelafan,michaelauli,jase}@fb.com

ABSTRACT

In open-domain dialogue intelligent agents should exhibit the use of knowledge, however there are few convincing demonstrations of this to date. The most popular sequence to sequence models typically "generate and hope" generic utterances

Towards Content Transfer through Grounded Text Generation

Shrimai Prabhumoye
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15219
sprabhum@andrew.cmu.edu

Chris Quirk, Michel Galley
Microsoft Research
One Microsoft Way
Redmond, WA 98052
{chrisq,mgalley}@microsoft.c

Abstract

ecent work in neural generation has attracted gnificant interest in controlling the *form* of xt, such as style, persona, and politeness. lowever, there has been less work on control-

Monkey selfie copyright dis

From Wikipedia, the free encyclopedia

The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)

Plan-and-Write: Towards Better Automatic Storytelling

Lili Yao,[1,3]* Nanyun Peng,[2]* Ralph Weischedel,[2] Kevin Knight,[2] Dongyan Zhao,[1] Rui Yan[1]†
liliyao@tencent.com, {npeng,weisched,knight}@isi.edu
{zhaodongyan,ruiyan}@pku.edu.cn
[1]Institute of Computer Science and Technology, Peking University
[2]Information Sciences Institute, University of Southern California, [3]Tencent AI Lab

Abstract

Automatic storytelling is challenging since it requires generating long, coherent natural language to describes a sensible sequence of events. Despite considerable efforts on automatic story generation in the past, prior work either is restricted in

Title (Given)	The Bike Accident
Storyline (Extracted)	Carrie → bike → sneak → nervous → leg
Story (Human Written)	Carrie had just learned how to ride a bike. She didn't have a bike of her own. Carrie would sneak rides on her

Motivation

Large body of work

Style Transfer Through Back-Translation

Shrimai Prabhumoye, Yulia Tsvetkov, Ruslan Salakhutdinov, Alan W Bla
Carnegie Mellon University, Pittsburgh, PA, USA
{sprabhum,ytsvetko,rsalakhu,awb}@cs.cmu.edu

Abstract

Style transfer is the task of rephrasing the text to contain specific stylistic properties without changing the intent or affect within the context. This paper introduces

Towards Controllable Story Generation

These goals have motivated a c
amount of recent research efforts focus
trolled" language generation—aiming
ing the semantic content of *what* is
the stylistic dimensions of *how* it is sa
include approaches relying on heuristic

Politeness Transfer: A Tag and Generate Approach

Madaan *, Amrith Setlur *, Tanmay Parekh *, Barnabas Poczos, Graham Neubig,
ming Yang, Ruslan Salakhutdinov, Alan W Black, Shrimai Prabhumoye
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA, USA
{amadaan, asetlur, tparekh}@cs.cmu.edu

Abstract

introduces a new task of politeness
hich involves converting non-polite
o polite sentences while preserving
ng. We also provide a dataset of
n 1.39 million instances automatically

Rao and Tetreault, 2018; Xu et al., 2012; Jhamtani et al., 2017) has not focused on politeness as a style transfer task, and we argue that defining it is cumbersome. While native speakers of a language and cohabitants of a region have a good working understanding of the phenomenon of politeness

Published as a conference paper at ICLR 2019

🧙 OF WIKIPEDIA: KNOWLEDGE-POWERED CONVERSATIONAL AGENTS

Emily Dinan*, Stephen Roller*, Kurt Shuster*, Angela Fan, Michael Auli, Jason Weston
Facebook AI Research
{edinan,roller,kshuster,angelafan,michaelauli,jase}@fb.com

ABSTRACT

In open-domain dialogue intelligent agents should exhibit the use of knowledge, however there are few convincing demonstrations of this to date. The most popular sequence to sequence models typically "generate and hope" generic utterances

Towards Content Transfer through Grounded Text Generation

Shrimai Prabhumoye
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15219
sprabhum@andrew.cmu.edu

Chris Quirk, Michel Galley
Microsoft Research
One Microsoft Way
Redmond, WA 98052
{chrisq,mgalley}@microsoft.c

Abstract

ecent work in neural generation has attracted gnificant interest in controlling the *form* of xt, such as style, persona, and politeness. owever, there has been less work on control-

Monkey selfie copyright dis
From Wikipedia, the free encyclopedia

The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)

Plan-and-Write: Towards Better Automatic Storytelling

Lili Yao,[1,3*] Nanyun Peng,[2*] Ralph Weischedel,[2] Kevin Knight,[2] Dongyan Zhao,[1] Rui Yan[1†]
liliyao@tencent.com, {npeng,weisched,knight}@isi.edu
{zhaodongyan,ruiyan}@pku.edu.cn
[1]Institute of Computer Science and Technology, Peking University
[2]Information Sciences Institute, University of Southern California, [3]Tencent AI Lab

Abstract

Automatic storytelling is challenging since it requires generating long, coherent natural language to describes a sensible sequence of events. Despite considerable efforts on automatic story generation in the past, prior work either is restricted in

Title (Given)	The Bike Accident
Storyline (Extracted)	Carrie → bike → sneak → nervous → leg
Story (Human Written)	Carrie had just learned how to ride a bike. She didn't have a bike of her own. Carrie would sneak rides on her

Motivation

Style Transfer Through Back-Translation

Shrimai Prabhumoye, Yulia Tsvetkov, Ruslan Salakhutdinov, Alan W Bla
Carnegie Mellon University, Pittsburgh, PA, USA
{sprabhum,ytsvetko,rsalakhu,awb}@cs.cmu.edu

Abstract

Style transfer is the task of rephrasing the text to contain specific stylistic properties without changing the intent or affect within the context. This paper introduces

These goals have motivated a co amount of recent research efforts focuse trolled" language generation—aiming ing the semantic content of *what* is the stylistic dimensions of *how* it is sa include approaches relying on heuristic

Towards Controllable Story Generation

Politeness Transfer: A Tag and Generate Approach

Aman Madaan *, Amrith Setlur *, Tanmay Parekh *, Barnabas Poczos, Graham Neubig, ming Yang, Ruslan Salakhutdinov, Alan W Black, Shrimai Prabhumoye
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA, USA
{amadaan, asetlur, tparekh}@cs.cmu.edu

Abstract

introduces a new task of politeness hich involves converting non-polite o polite sentences while preserving ng. We also provide a dataset of n 1.39 million instances automatically

Rao and Tetreault, 2018; Xu et al., 2012; Jhamtani et al., 2017) has not focused on politeness as a style transfer task, and we argue that defining it is cumbersome. While native speakers of a language and cohabitants of a region have a good working understanding of the phenomenon of politeness

Large body of work

Published as a conference paper at ICLR 2019

OF WIKIPEDIA:
KNOWLEDGE-POWERED CONVERSATIONAL AGENT

Emily Dinan*, Stephen Roller*, Kurt Shuster*, Angela Fan, Michael Auli, Jason Westo
Facebook AI Research
{edinan,roller,kshuster,angelafan,michaelauli,jase}@fb.com

ABSTRACT

In open-domain dialogue intelligent agents should exhibit the use of knowledge, however there are few convincing demonstrations of this to date. The most popular sequence to sequence models typically "generate and hope" generic utterances

Towards Content Transfer through Grounded Text Generation

The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)

Plan-and-Write: Towards B omatic Storytelling

,3* Nanyun Peng,[2*] Ralph Weisc ht,[2] Dongyan Zhao,[1] Rui Yan[1†]
liliyao@tencent.com @isi.edu
{zhaodon
[1]Institute of Computer University
[2]Information Sciences Institut [3]Tencent AI Lab

Abstract

ytelling is challenging since it erent natural language to de ents. Despite considerable e ation in the past, prior work e

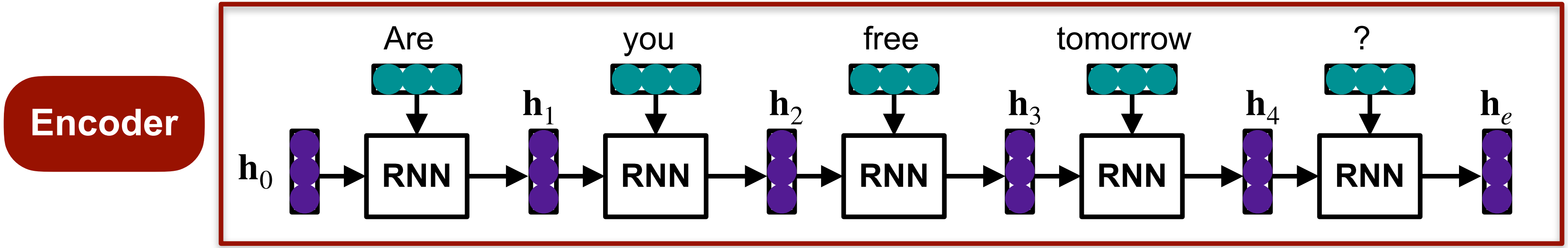
ke Accident
bike → sneak → nervous →
just learned how to ride a dn't have a bike of her would sneak rides on her

No unifying theme

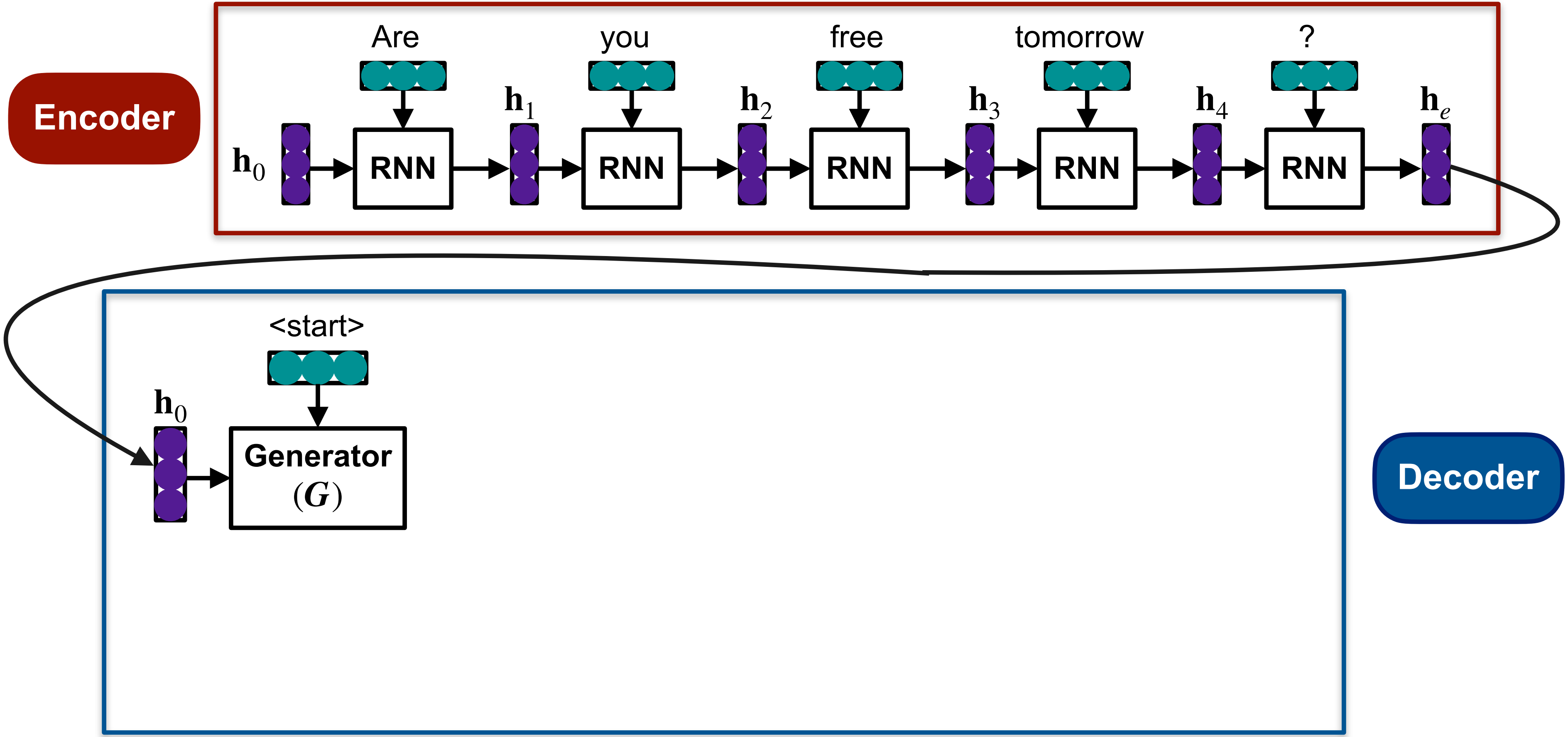
Contribution

- ***Controlled Generation Schema*** connects prior work
    - organize prior work
    - Schema contains 5 modules
    - Identify any architecture as belonging to one of these modules
    - Schema can be used with any algorithmic paradigm
- ***Collate knowledge*** about different techniques
    - Insights into the advantages of techniques
    - Pave way for new architectures
    - Provide easy access to comparison

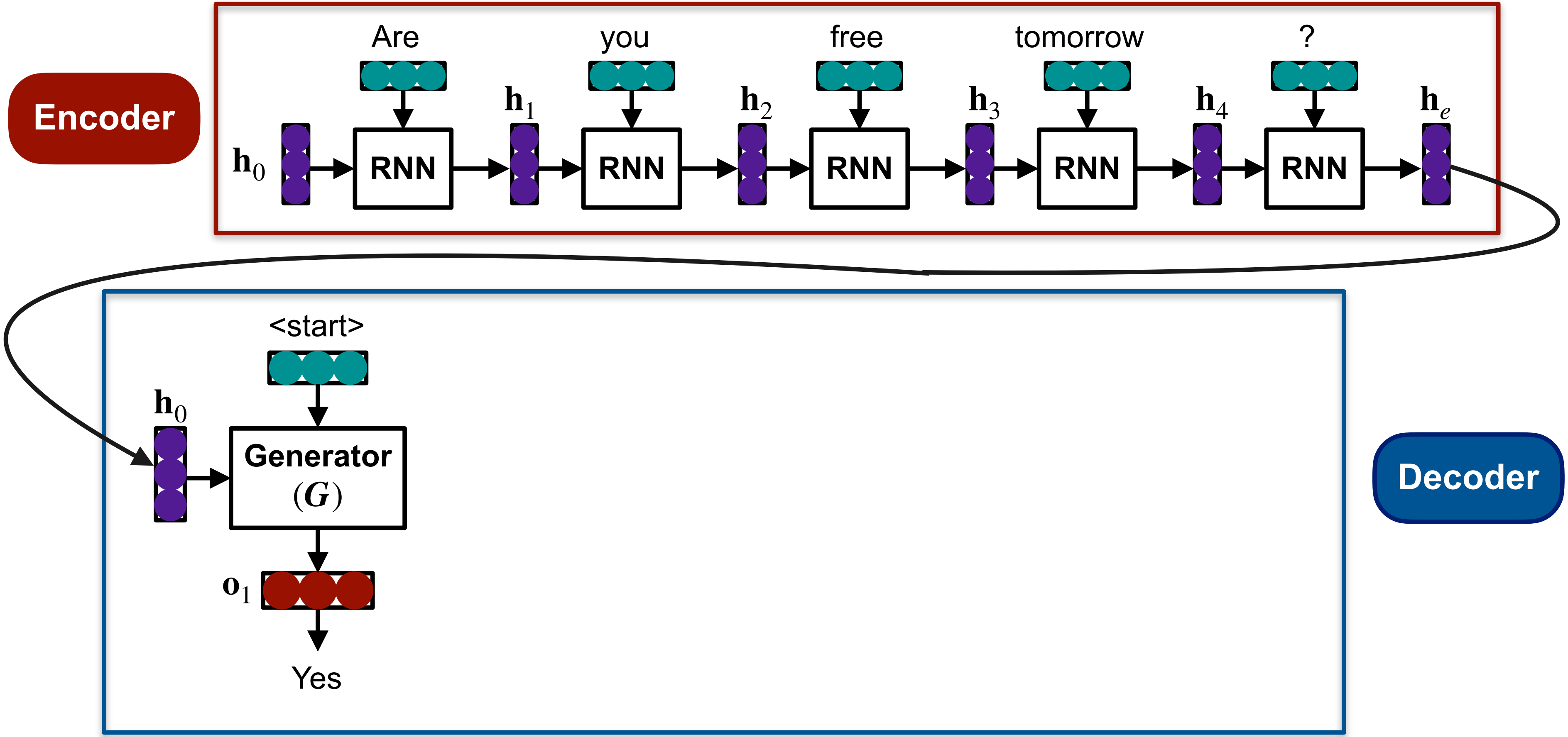
Generation Process



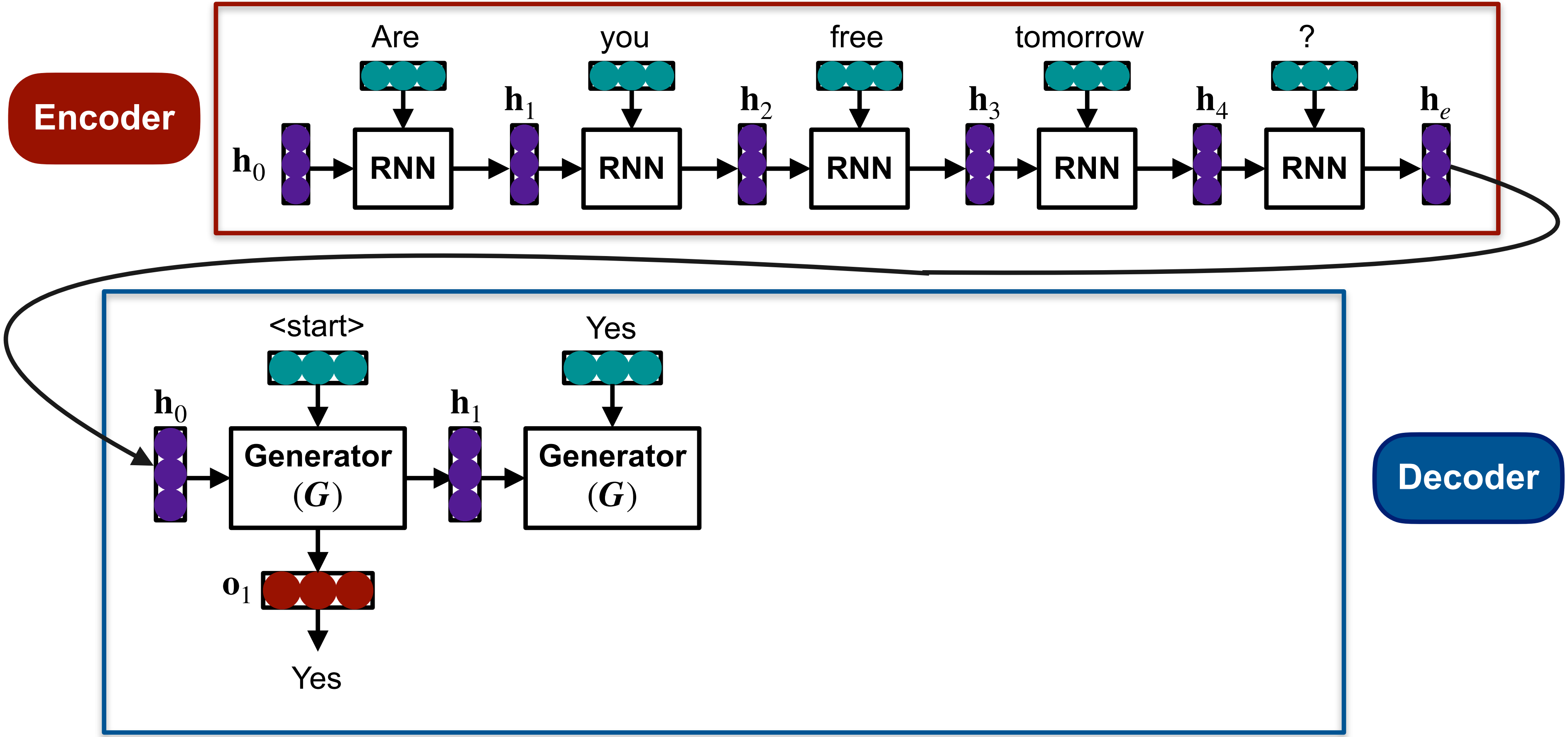
Generation Process



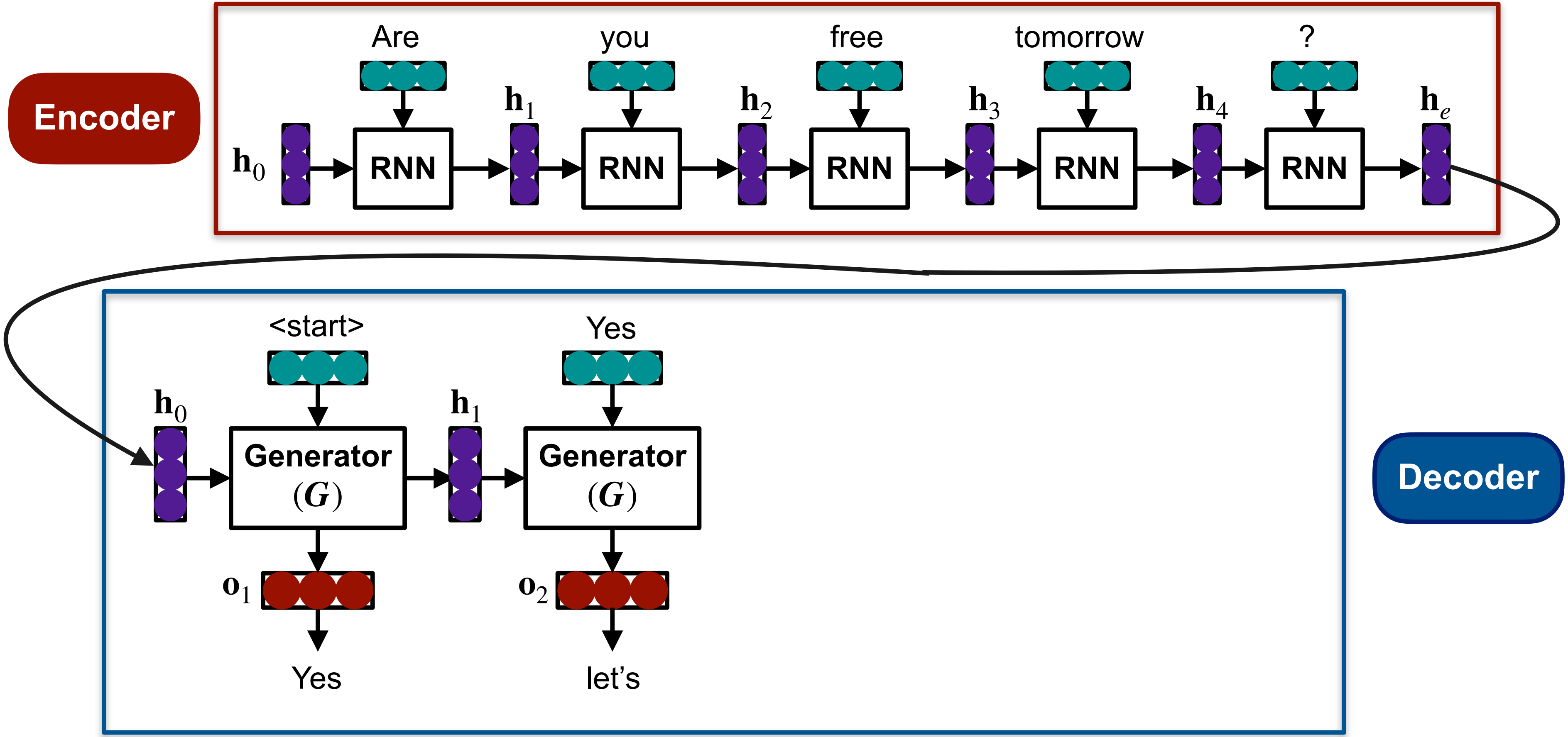
Generation Process



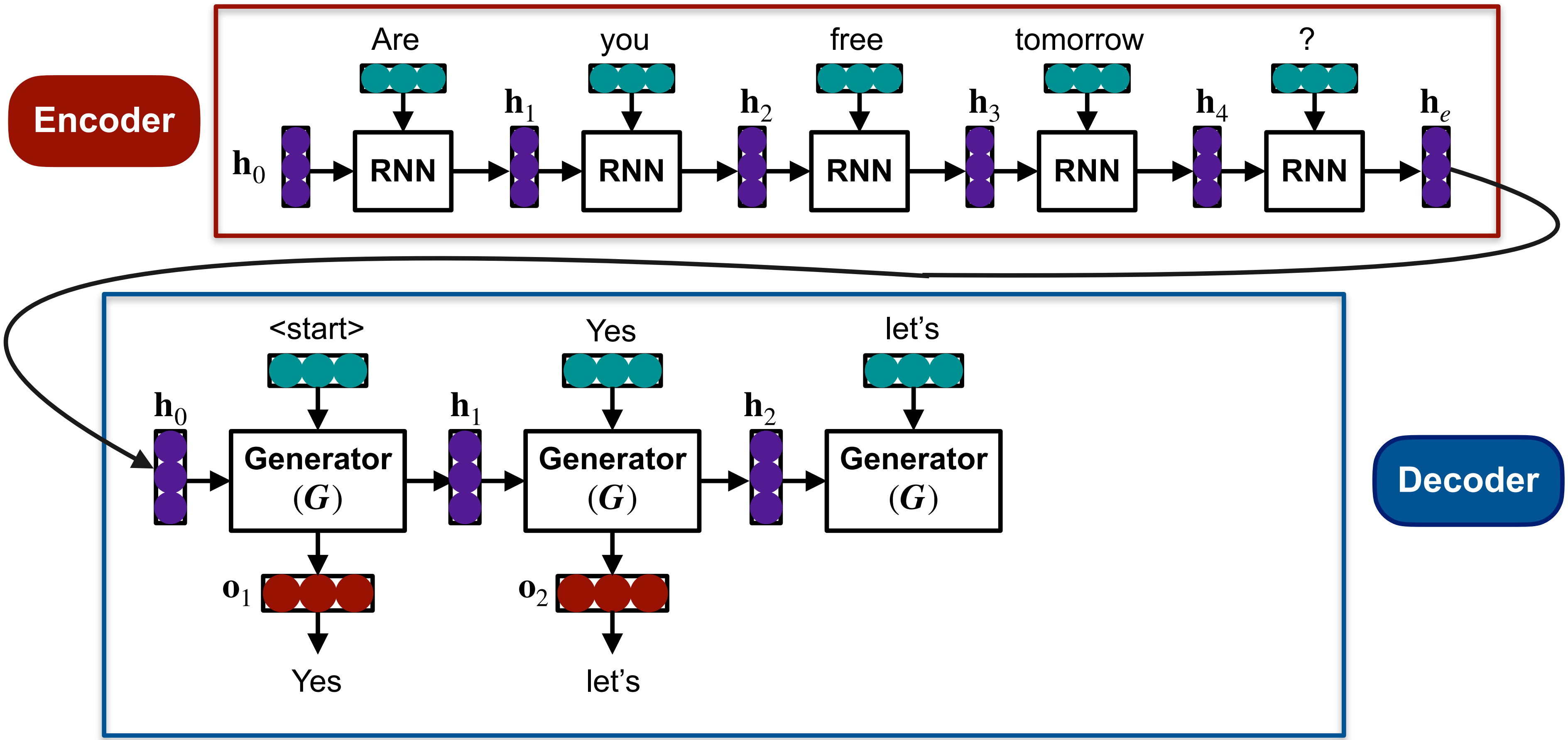
Generation Process



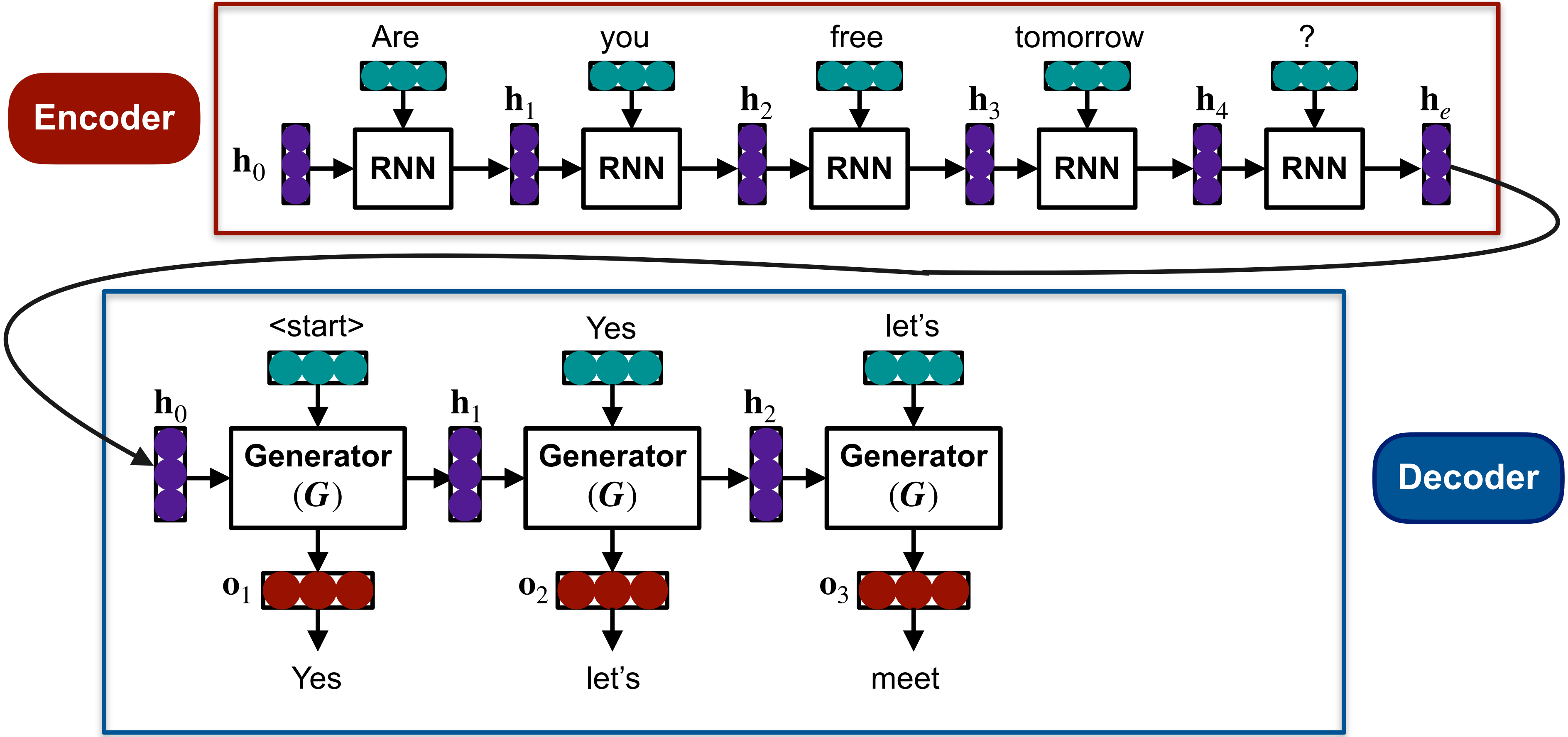
Generation Process



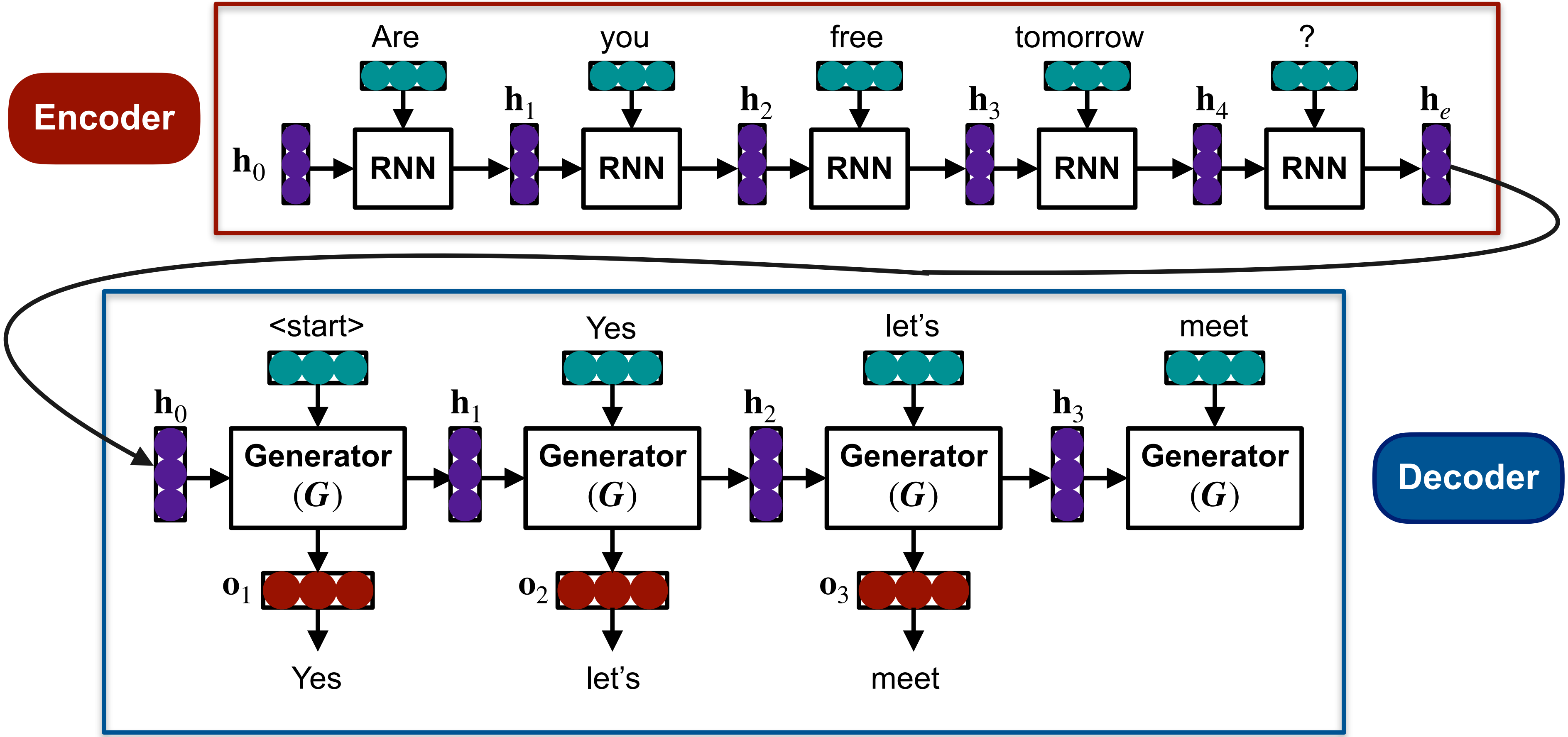
Generation Process



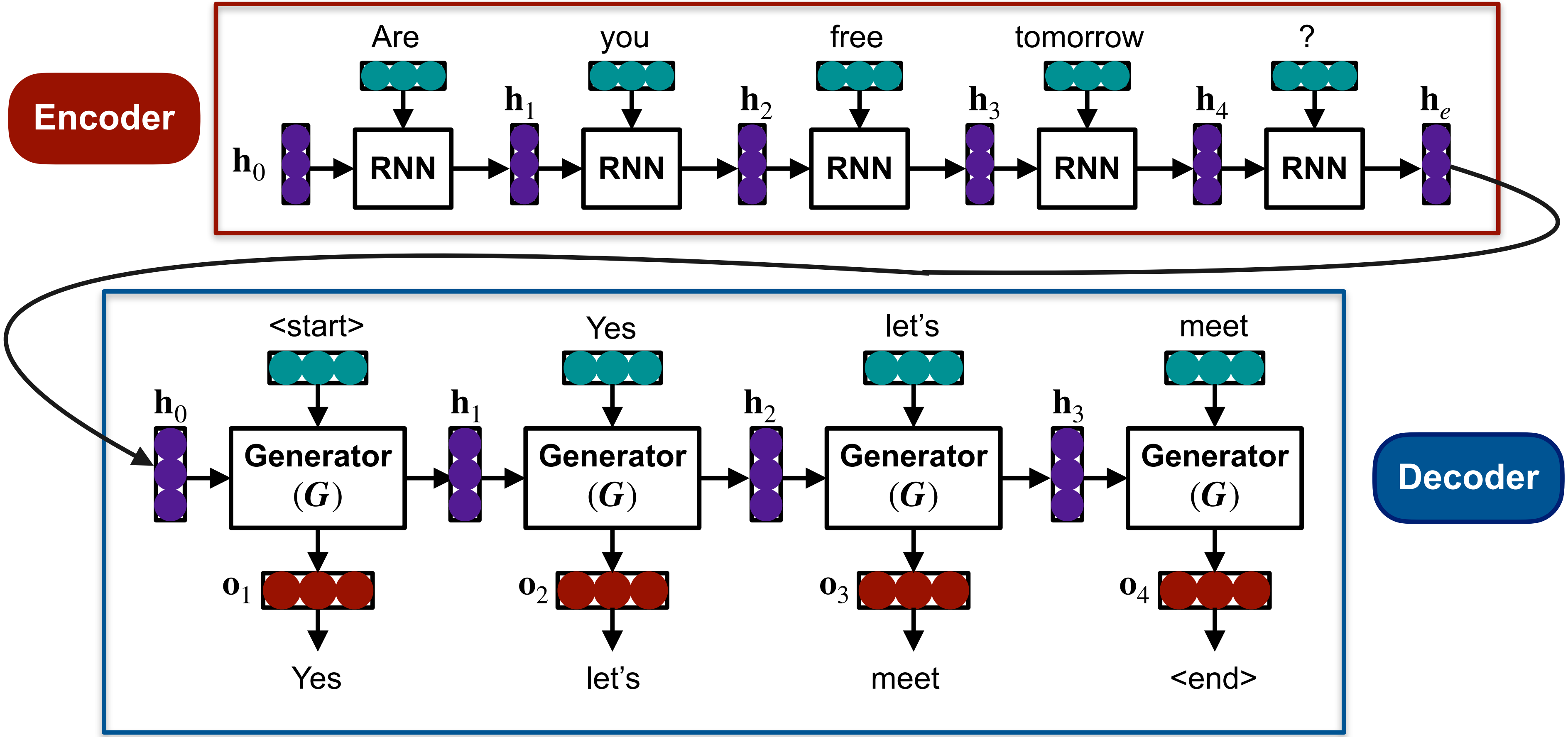
Generation Process



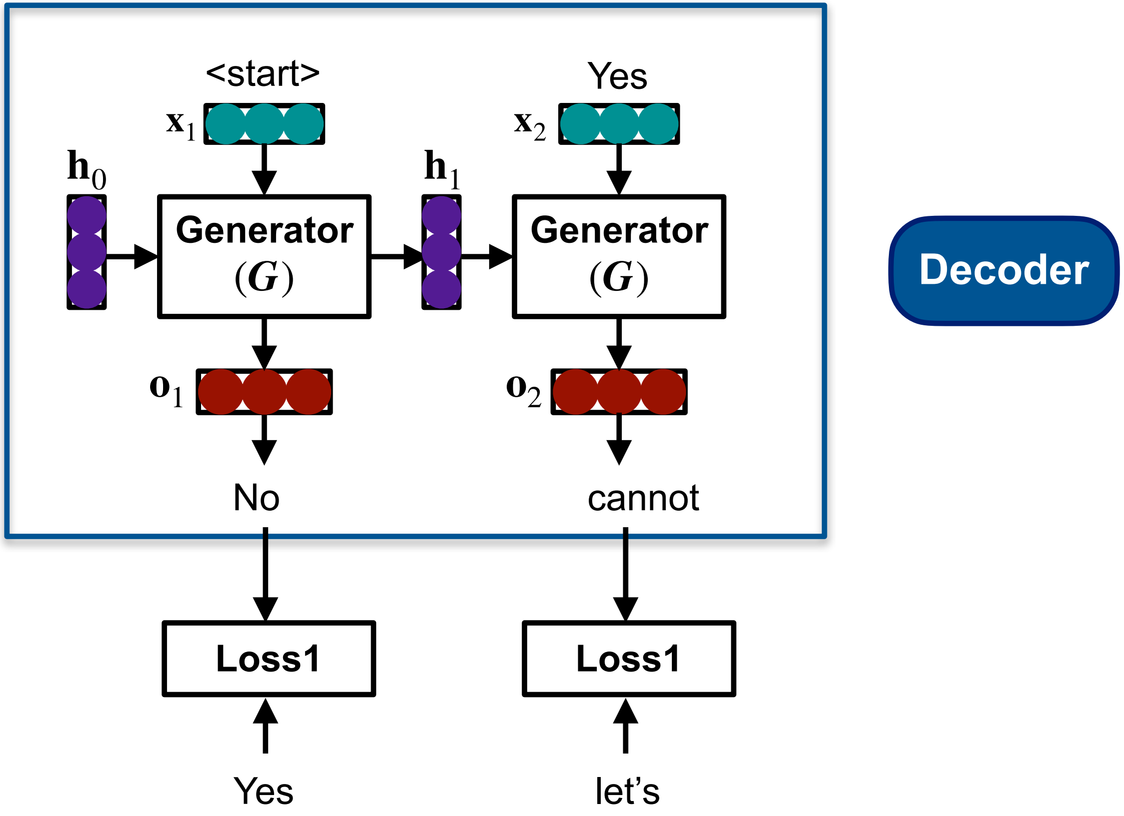
Generation Process



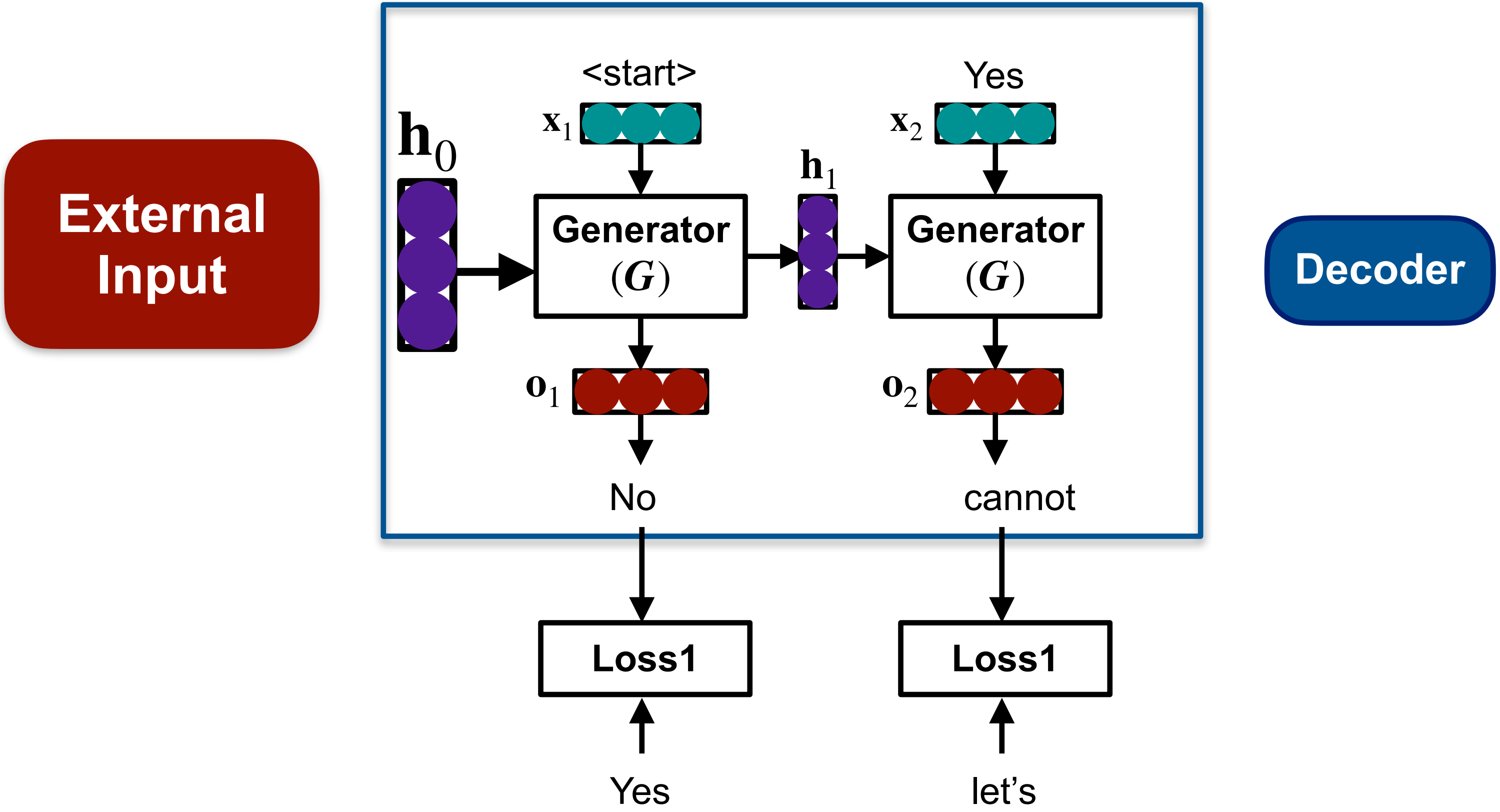
Generation Process



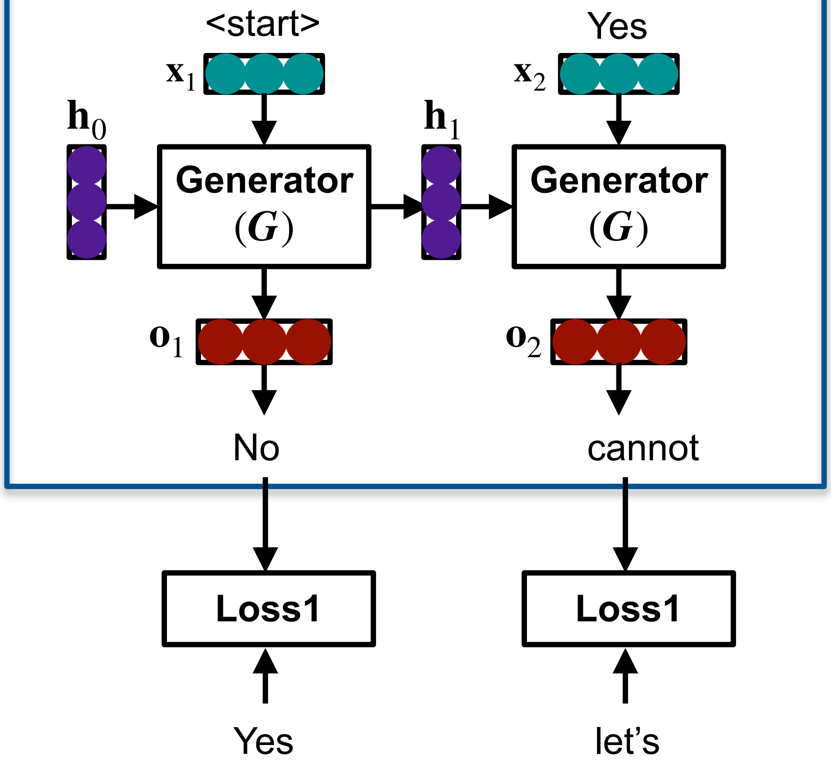
Modification Space



Modification Space

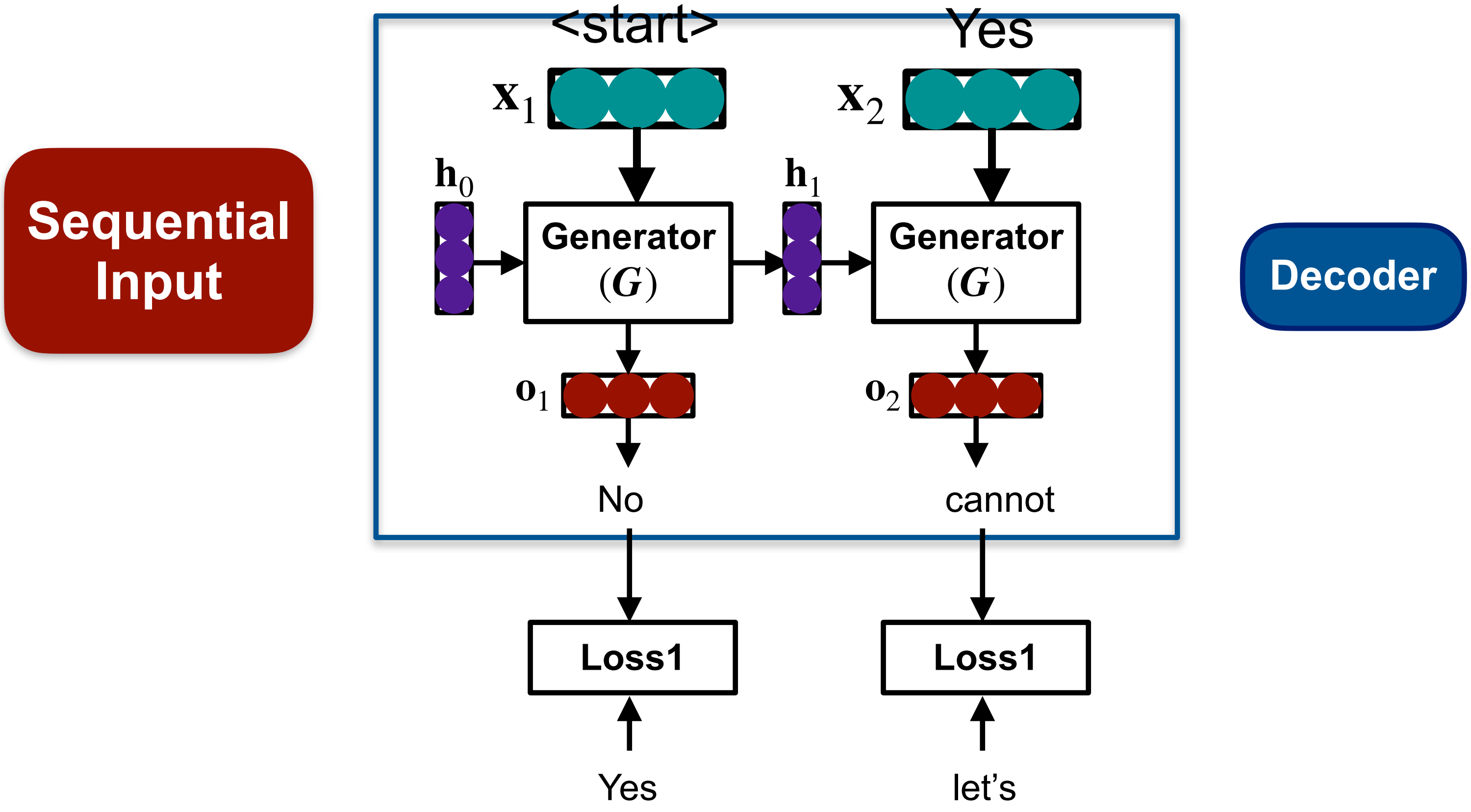


Modification Space

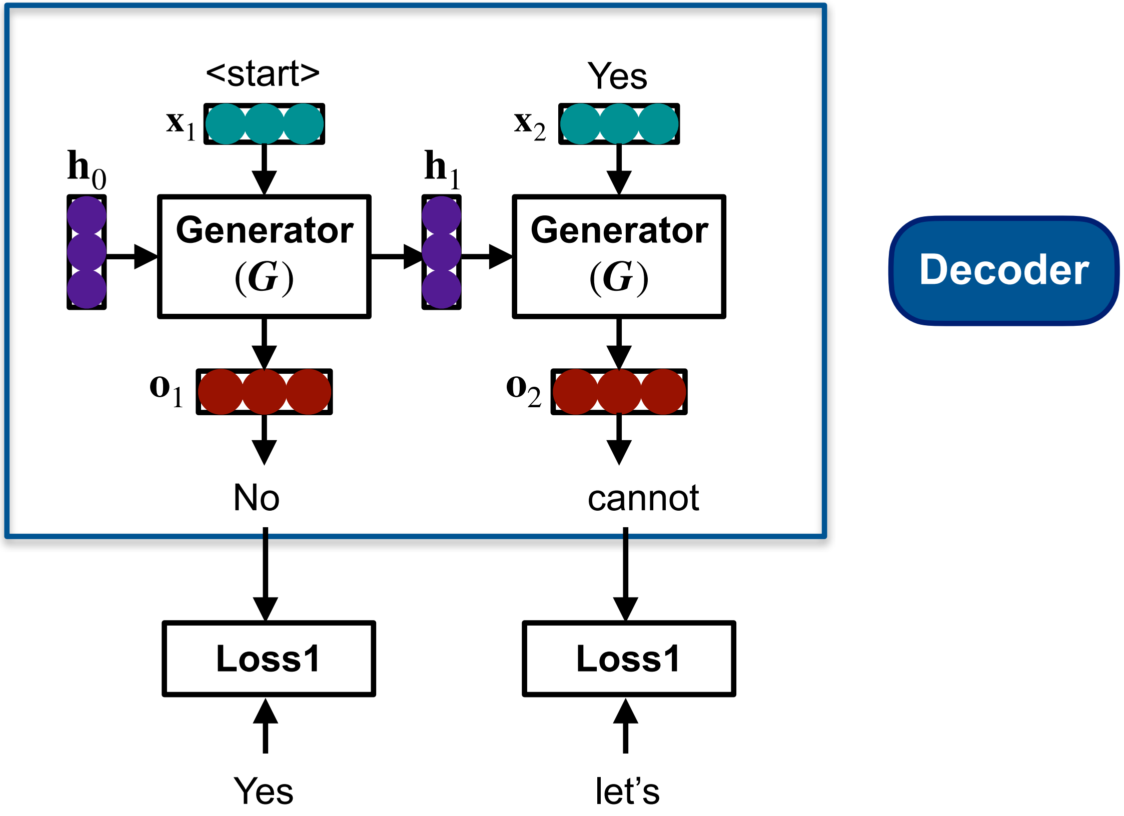


Decoder

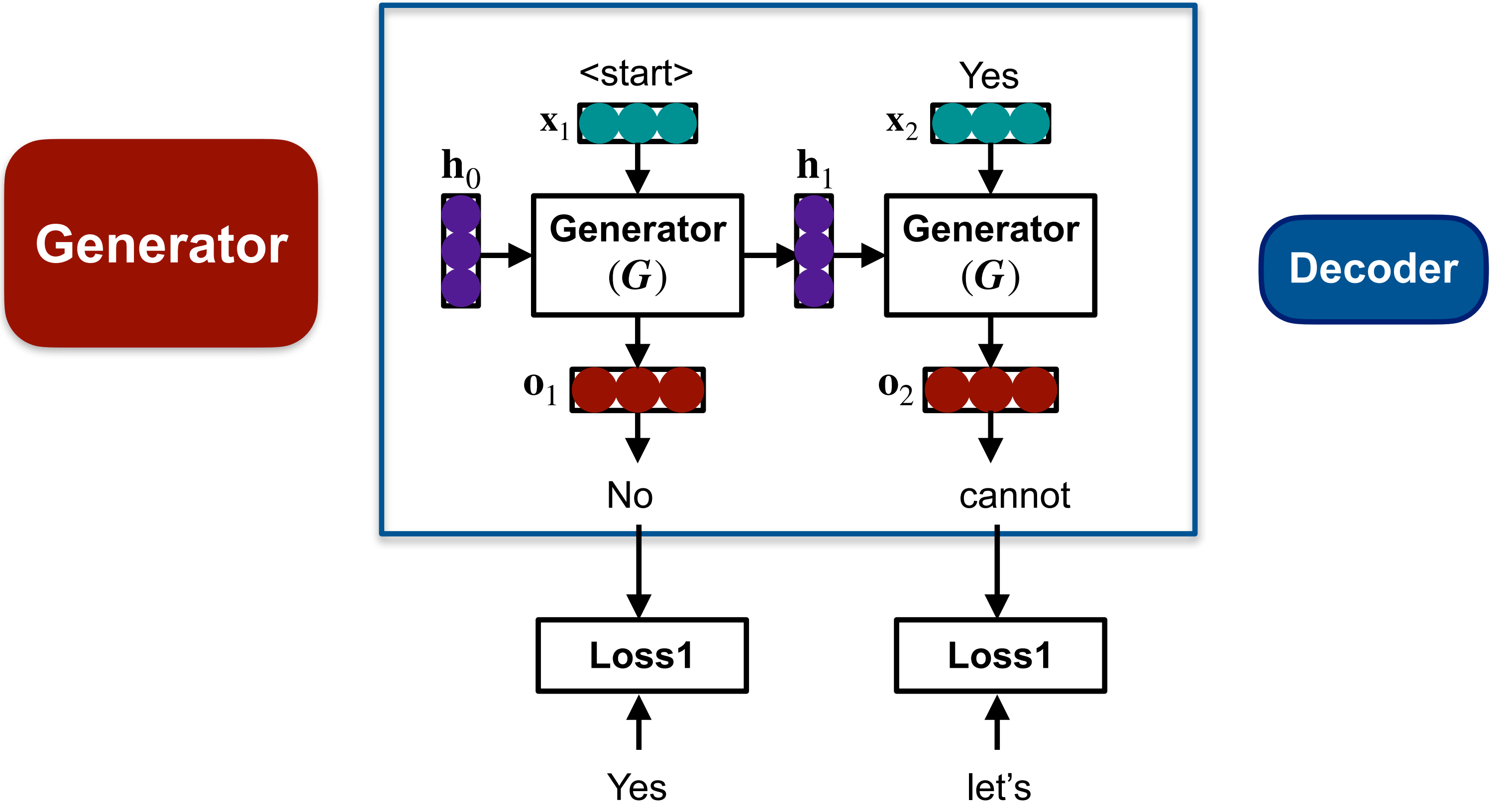
Modification Space



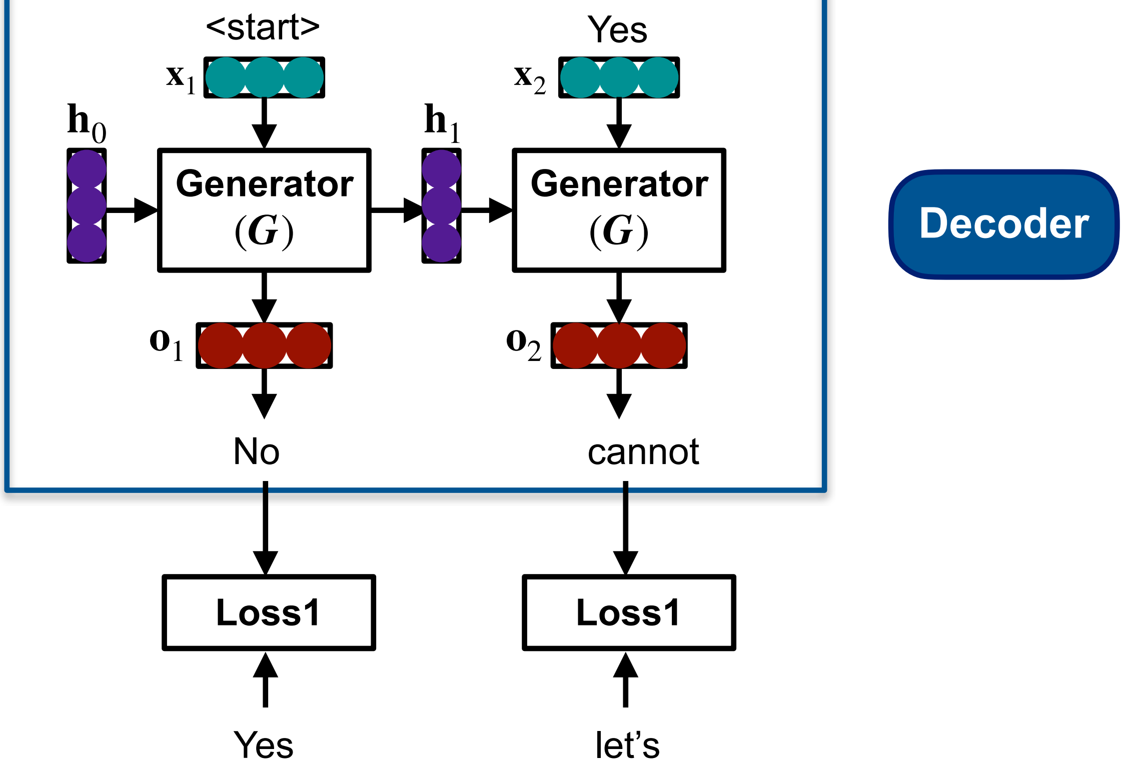
Modification Space



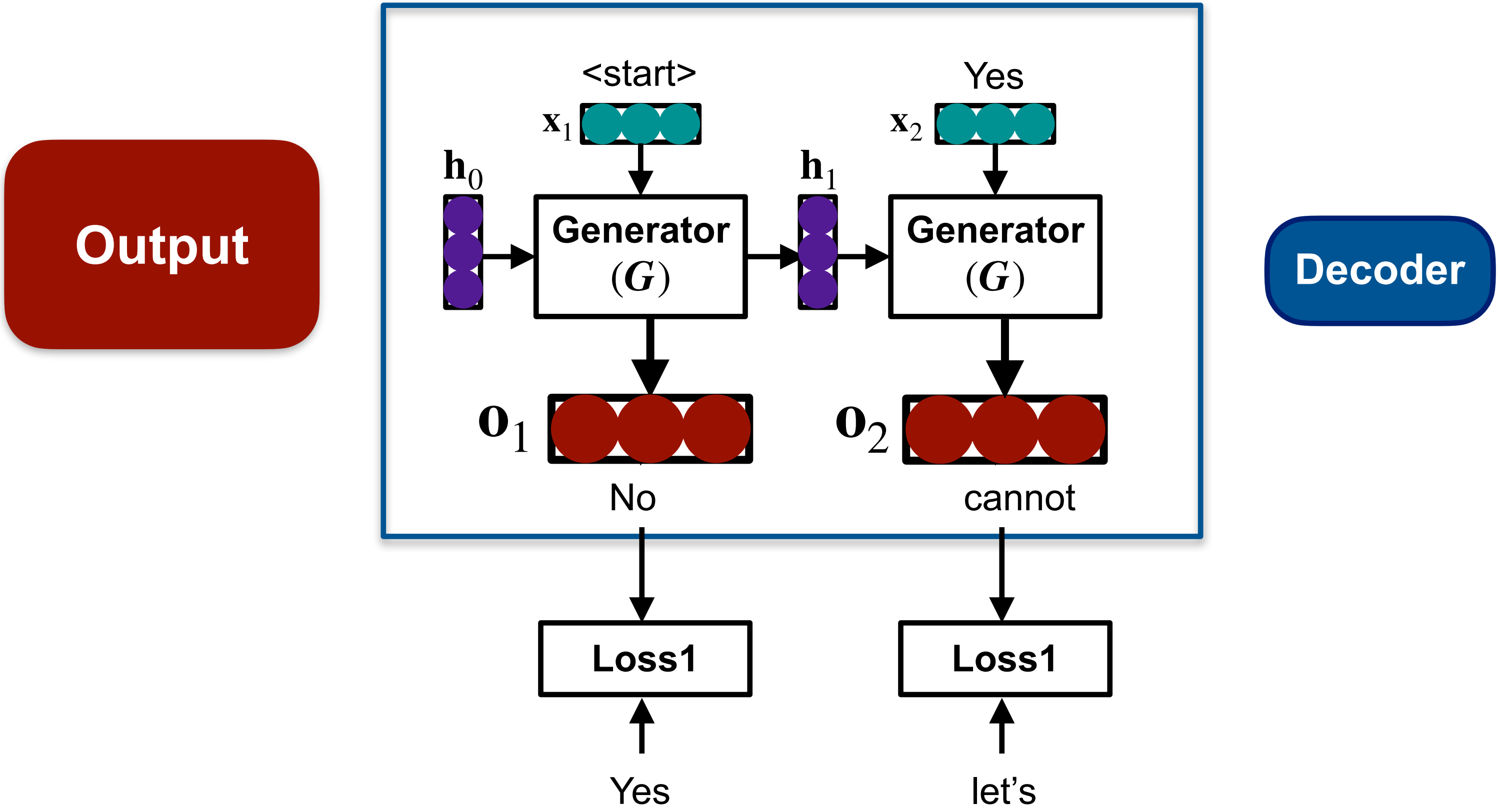
Modification Space



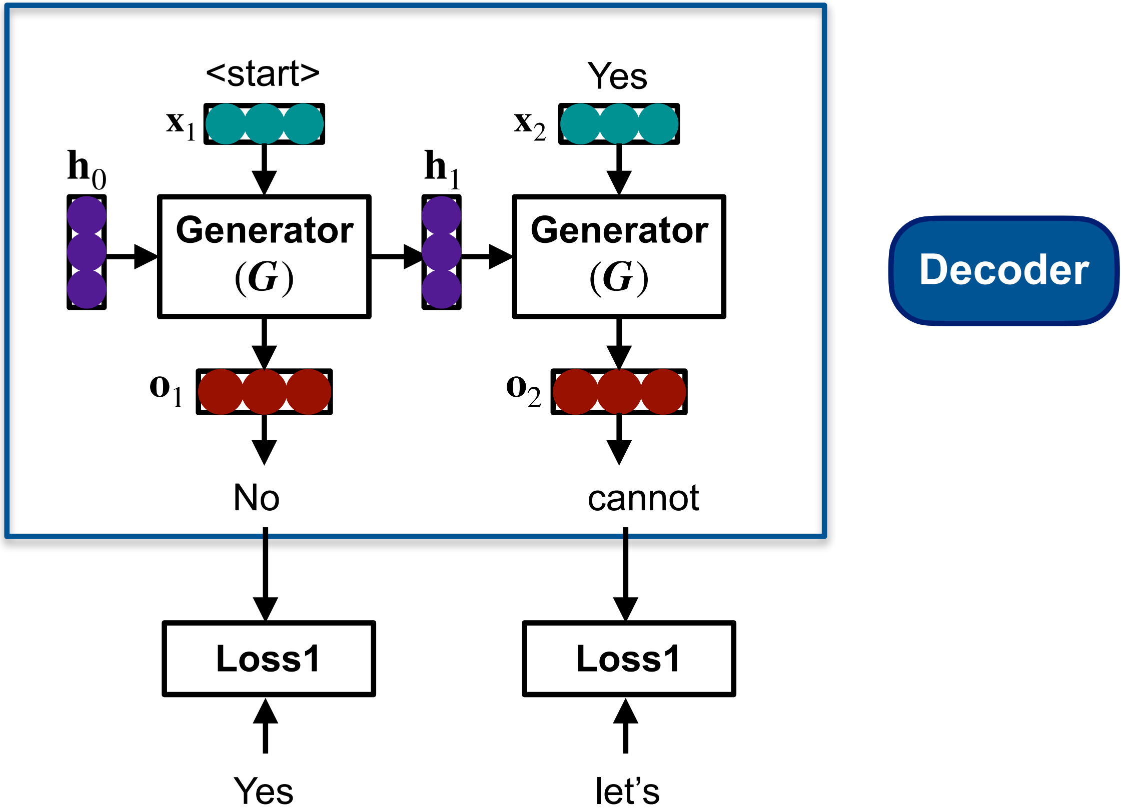
Modification Space



Modification Space

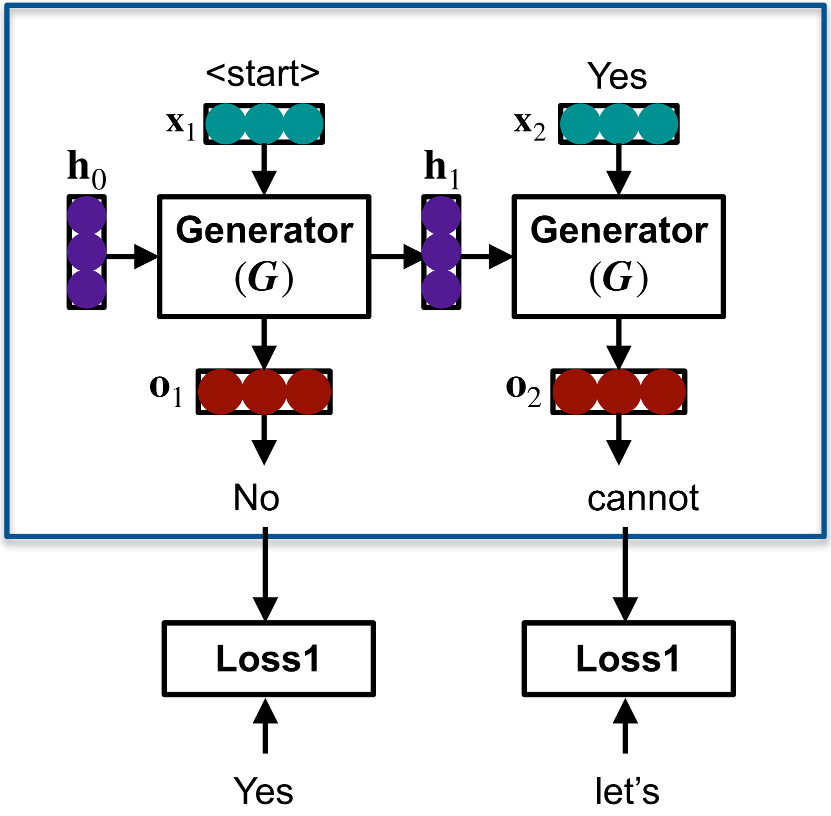


Modification Space



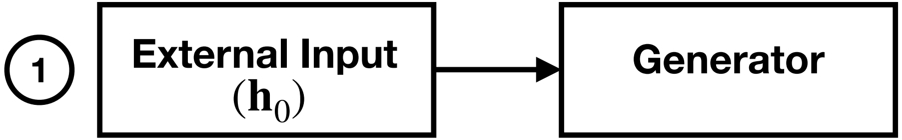
Modification Space

Training Objective

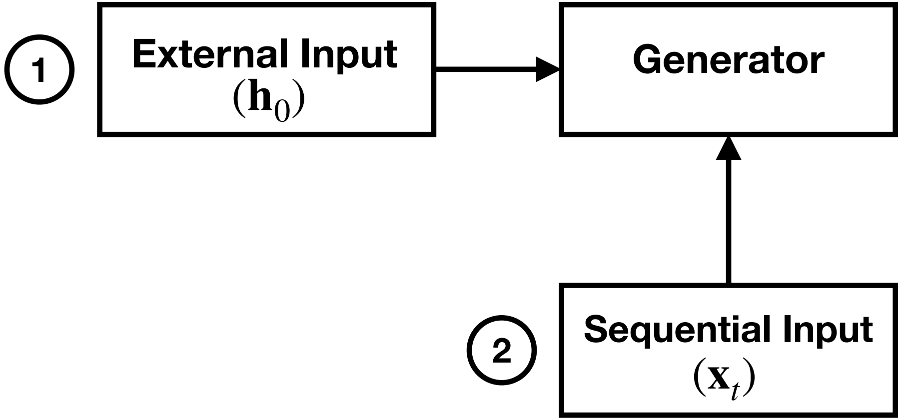


Decoder

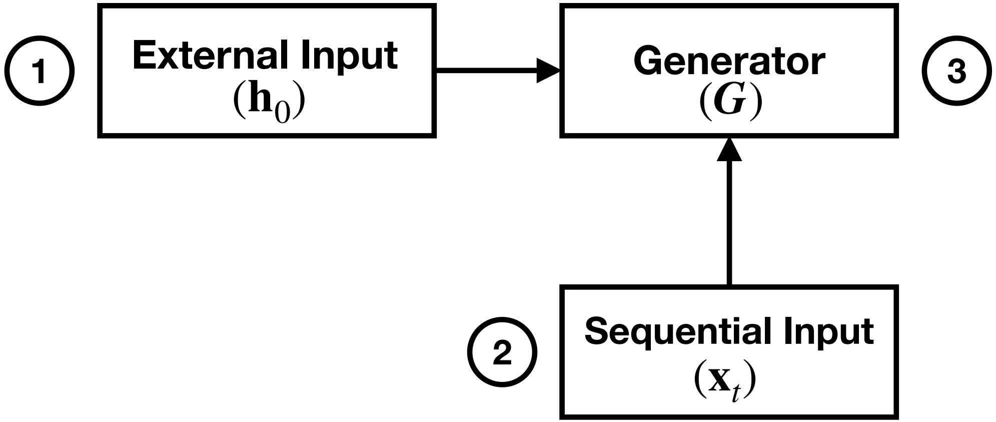
The proposed Schema



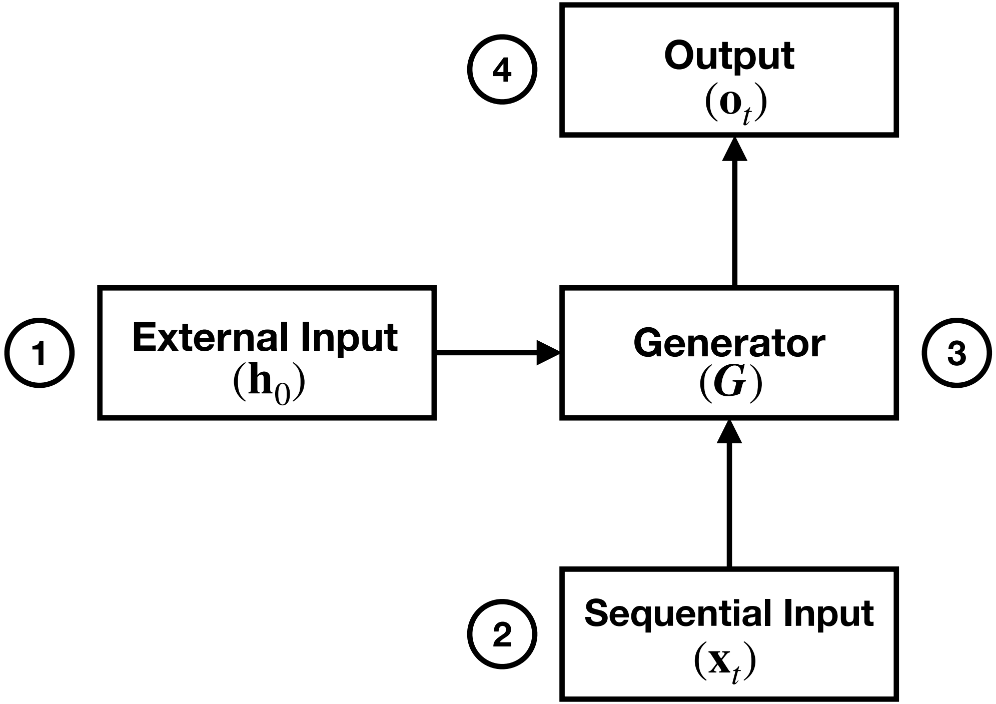
The proposed Schema



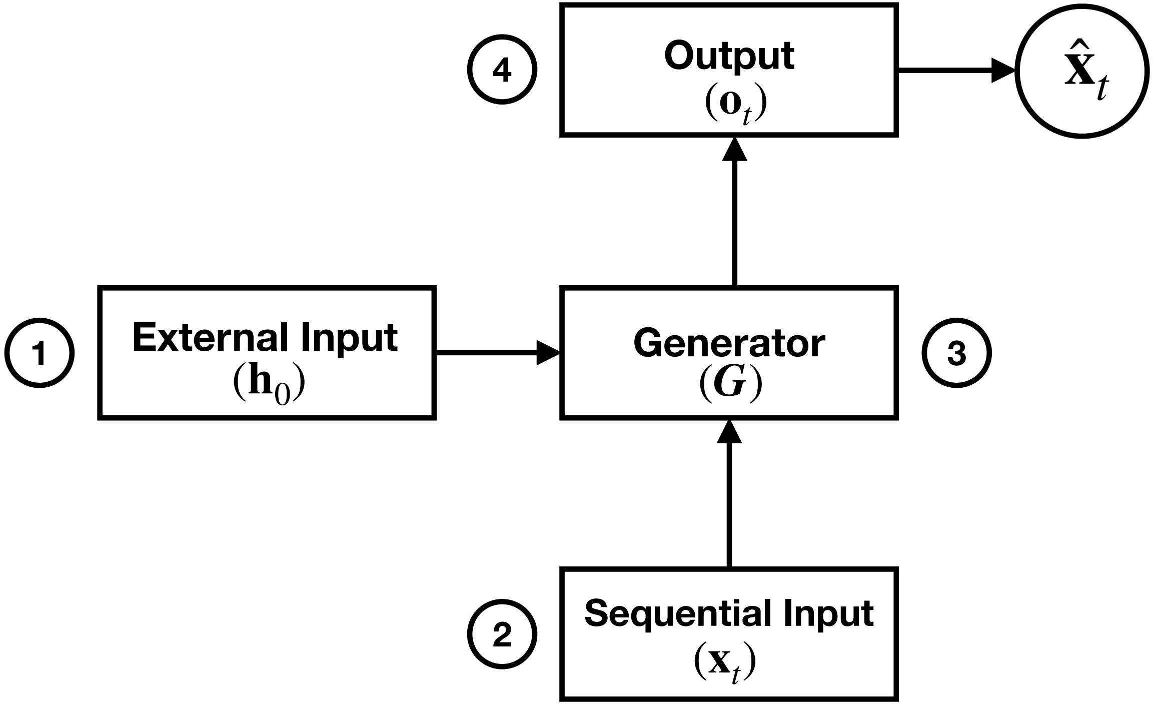
The proposed Schema



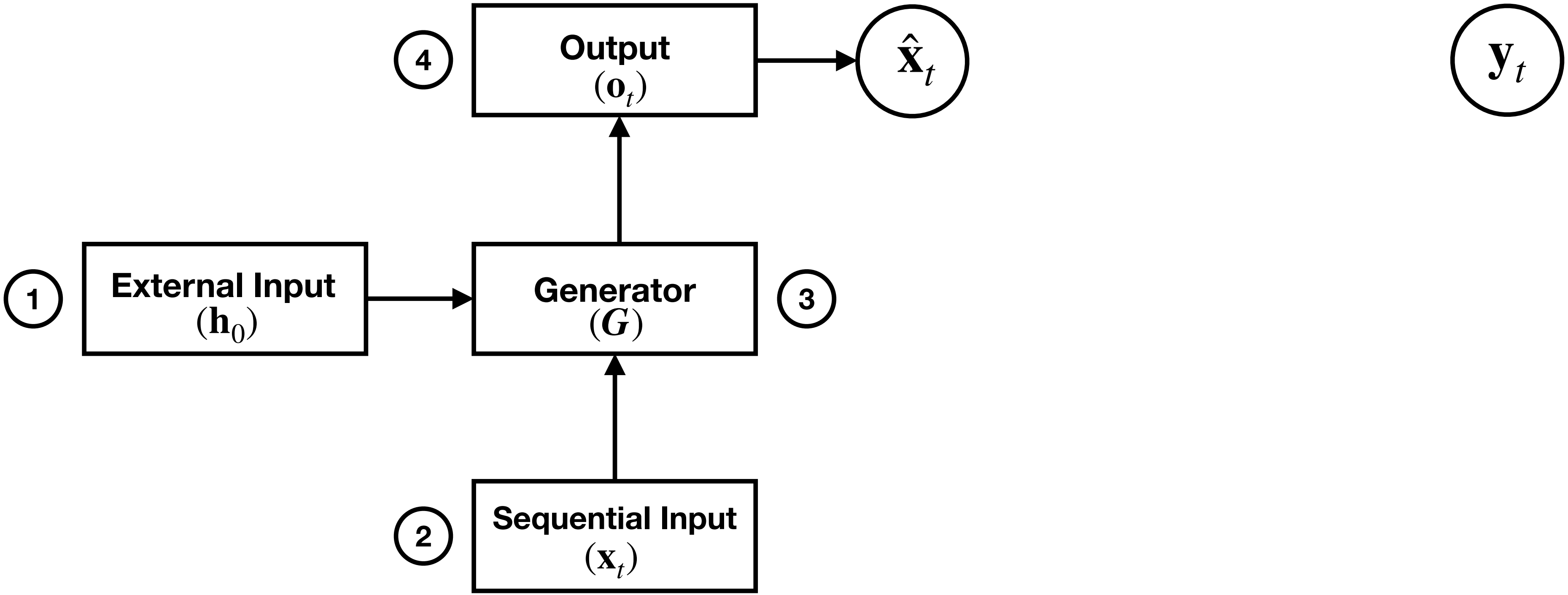
The proposed Schema



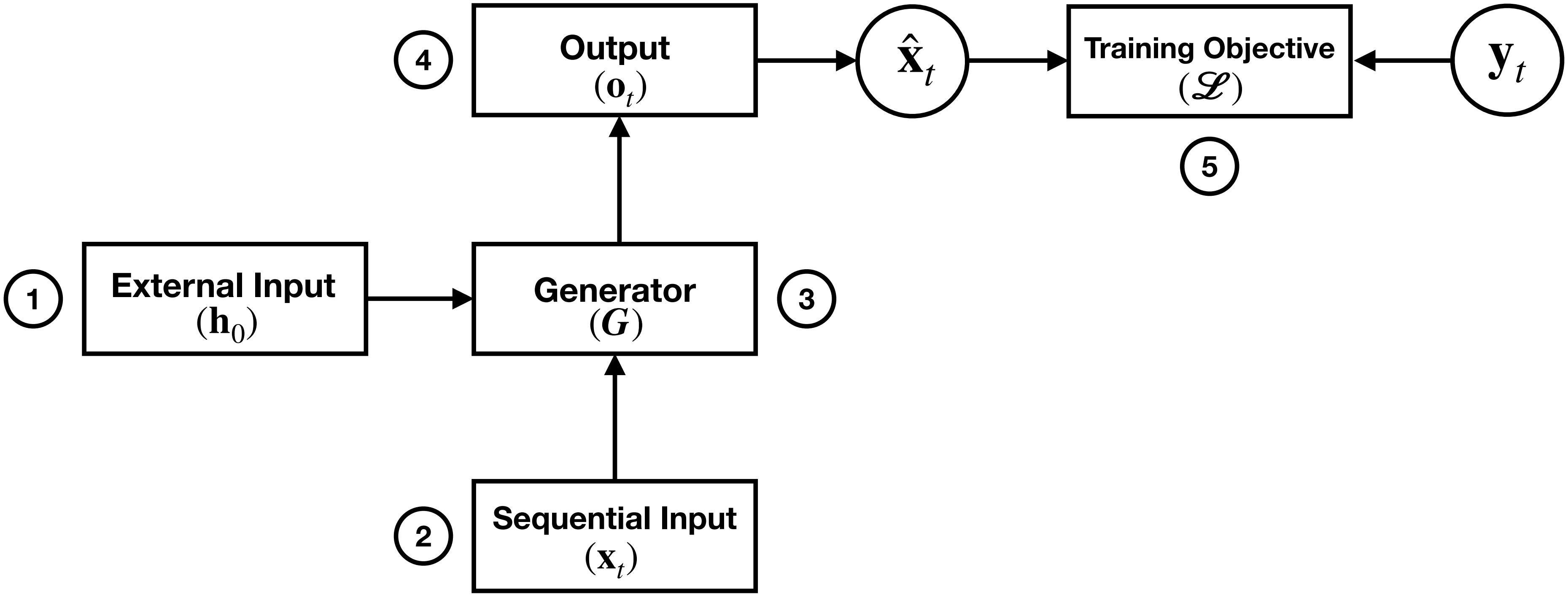
The proposed Schema



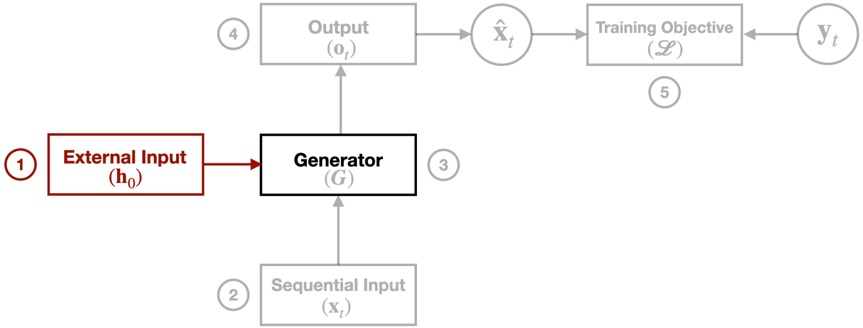
The proposed Schema



The proposed Schema



External Input

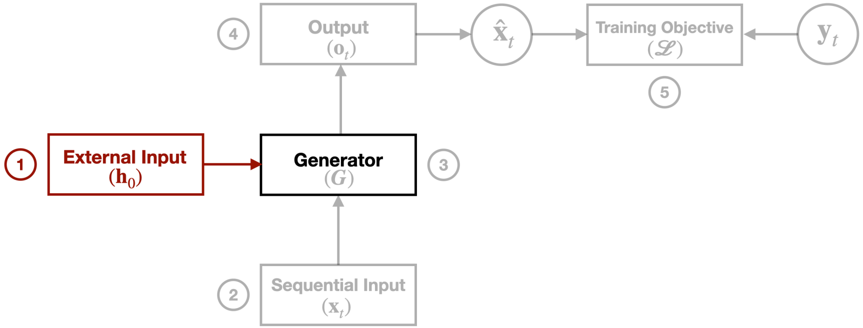


1. Decompose

- $\mathbf{h}_e$ decomposed into subspaces
- Provides *interpretable* representations
- Input should contain signal of control attribute
- Supervision on decomposed space

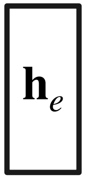
[Liu and Lapata (2018), Romanov et al. (2019), Balachandran et al. (2020)]

External Input



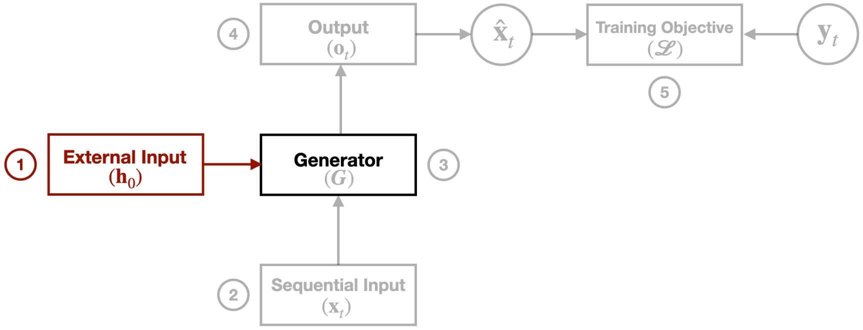
1. Decompose

- $\mathbf{h}_e$ decomposed into subspaces
- Provides *interpretable* representations
- Input should contain signal of control attribute
- Supervision on decomposed space



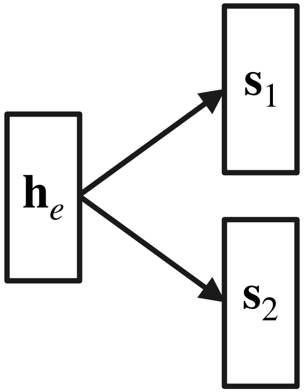
[Liu and Lapata (2018), Romanov et al. (2019), Balachandran et al. (2020)]

External Input



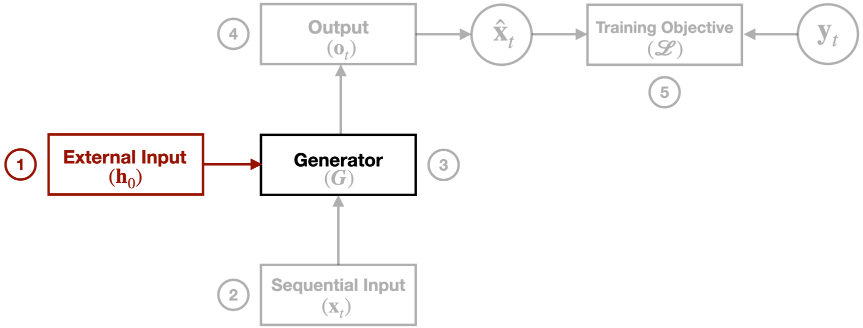
1. Decompose

- $\mathbf{h}_e$ decomposed into subspaces
- Provides *interpretable* representations
- Input should contain signal of control attribute
- Supervision on decomposed space



[Liu and Lapata (2018), Romanov et al. (2019), Balachandran et al. (2020)]

External Input

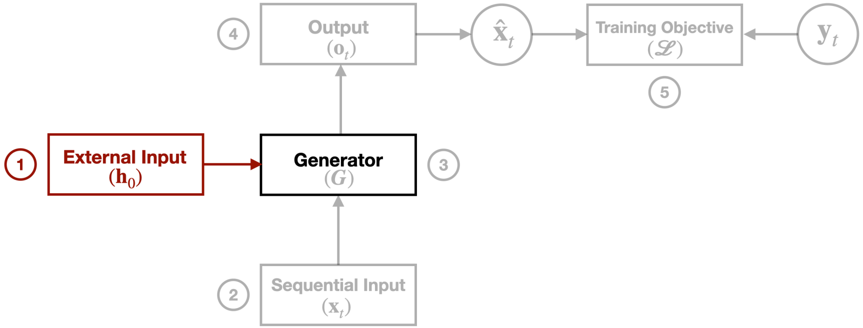


1. Decompose

- $\mathbf{h}_e$ decomposed into subspaces
- Provides *interpretable* representations
- Input should contain signal of control attribute
- Supervision on decomposed space

[Liu and Lapata (2018), Romanov et al. (2019), Balachandran et al. (2020)]

External Input



2. External Feedback

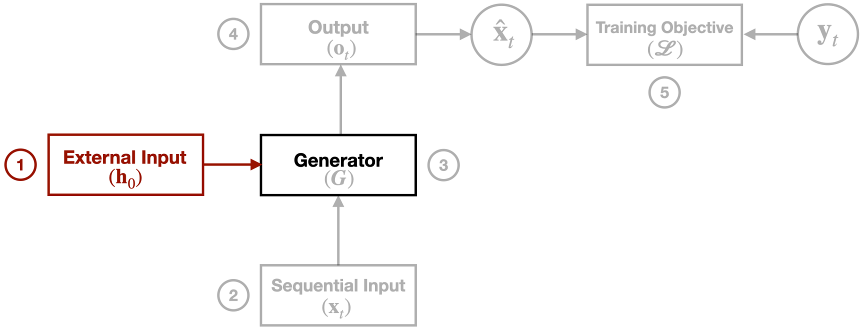
- regularizer to control $\mathbf{h}_e$
- must be jointly trained
- can be useful with decompose technique

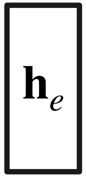
[Fu et al. (2018), Wang et al. (2019a), John et al. (2019)]

External Input

2. External Feedback

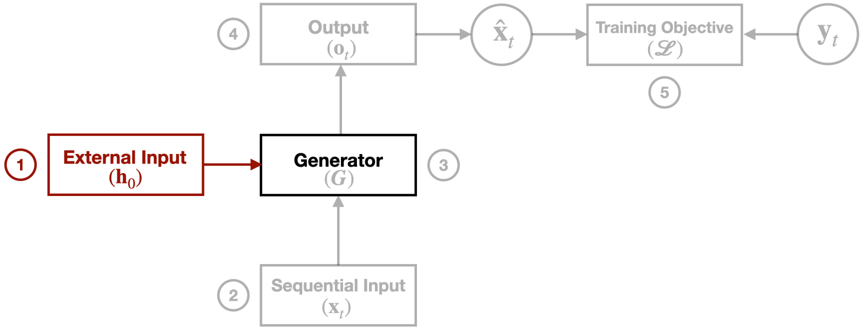
- regularizer to control $\mathbf{h}_e$
- must be jointly trained
- can be useful with decompose technique





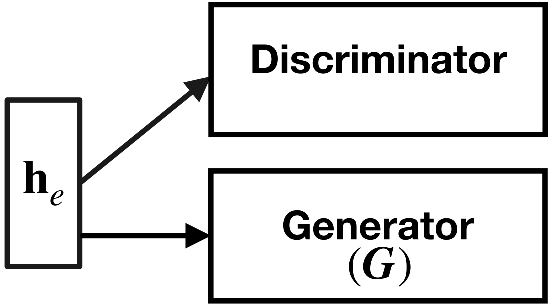
[Fu et al. (2018), Wang et al. (2019a), John et al. (2019)]

External Input



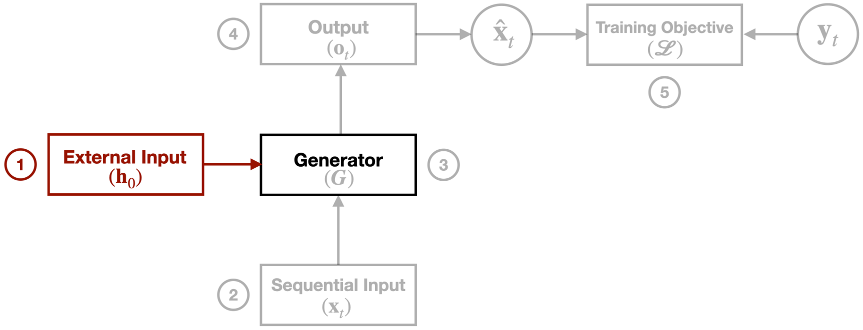
2. External Feedback

- regularizer to control $\mathbf{h}_e$
- must be jointly trained
- can be useful with decompose technique



[Fu et al. (2018), Wang et al. (2019a), John et al. (2019)]

External Input



2. External Feedback

- regularizer to control $\mathbf{h}_e$
- must be jointly trained
- can be useful with decompose technique

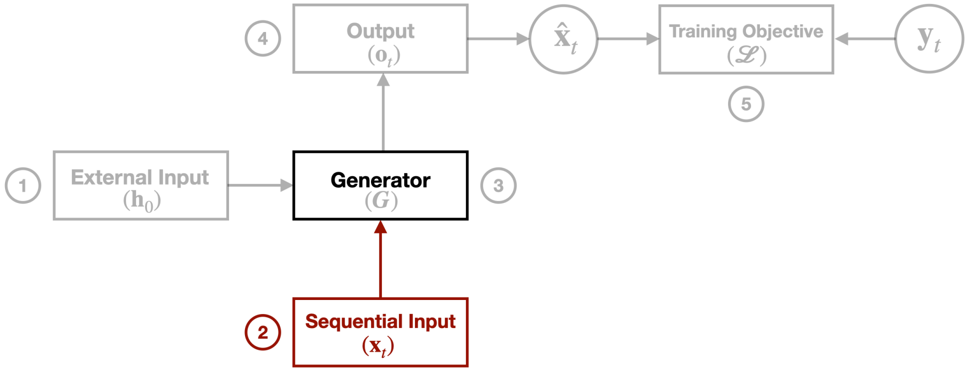
3. Arithmetic or Linear Transform

4. Stochastic Changes

Sequential Input

1. Arithmetic or Linear Transform

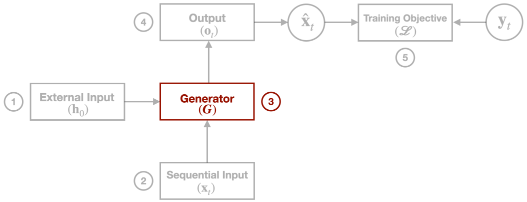
- $\tilde{\mathbf{x}}_t = [\mathbf{x}_t; \mathbf{s}]$
- $\tilde{\mathbf{x}}_t = \mathbf{x}_t + \mathbf{s}$



- Changes the input to the generation itself and not the context
- Not shown promising results so far

[Noraset et al. (2017), Zhou et al. (2018), Prabhumoye et al. (2019)]

Generator Operations



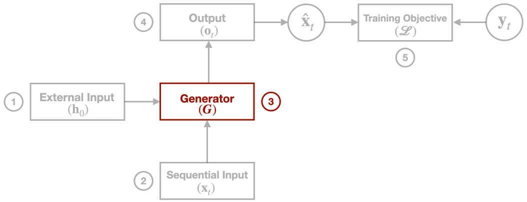
1. Controlled Generator Operations

- $\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t + \mathtt{tanh}(\mathbf{W}_d \mathbf{d}_t)$
  - $\mathbf{d}_t$ = dialogue act representation, change made to LSTM cell
  - Add *dialogue act* information in the generation process

- $\tilde{\mathbf{h}}_t = \mathtt{tanh}(\mathbf{W}_h \mathbf{x}_t + \mathbf{r}_t \odot \mathbf{U}_h \mathbf{h}_{t-1} + \mathbf{s}_t \odot \mathbf{Y}\mathbf{g} + \mathbf{q}_t \odot (\mathbf{1}_L^T \mathbf{Z}\mathbf{E}_t^{new})^T)$
  - $\mathbf{s}_t$ = goal select gate; $\mathbf{q}_t$ = item select gate, GRU cell
  - *recipe generation* task

[Gan et al. (2017), Kiddon et al. (2016), Wen et al. (2015)]

Generator Operations



1. *Controlled Generator Operations*

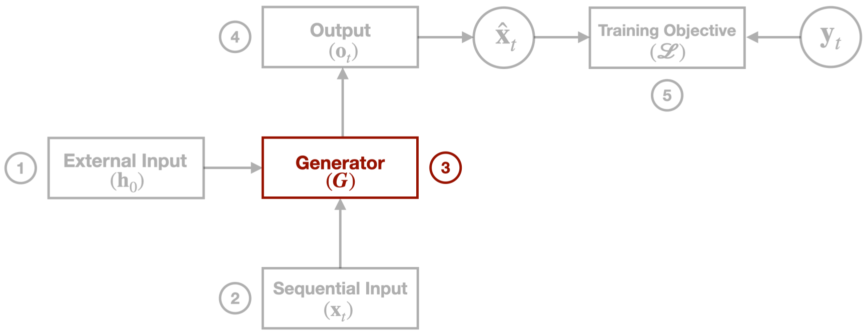
- $\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t + \mathtt{tanh}(\mathbf{W}_d \mathbf{d}_t)$
    - $\mathbf{d}_t$ = dialogue act representation, change made to LSTM cell
    - Add *dialogue act* information in the generation process

- $\tilde{\mathbf{h}}_t = \mathtt{tanh}(\mathbf{W}_h \mathbf{x}_t + \mathbf{r}_t \odot \mathbf{U}_h \mathbf{h}_{t-1} + \mathbf{s}_t \odot \mathbf{Y}\mathbf{g} + \mathbf{q}_t \odot (\mathbf{1}_L^T \mathbf{Z} \mathbf{E}_t^{new})^T)$
    - $\mathbf{s}_t$ = goal select gate; $\mathbf{q}_t$ = item select gate, GRU cell
    - *recipe generation* task

[Gan et al. (2017), Kiddon et al. (2016), Wen et al. (2015)]

Generator Operations

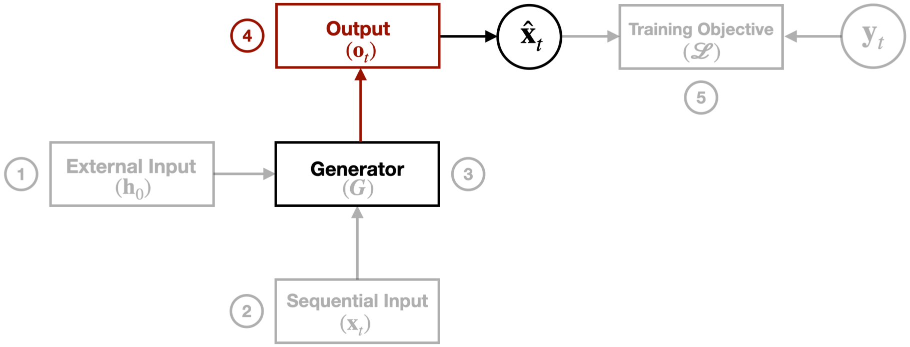
2. Recurrent Neural Networks
   - LSTM, GRU
3. Transformers
4. Pre-trained language models
   - BERT, GPT-2, BART, XL-Net



Output

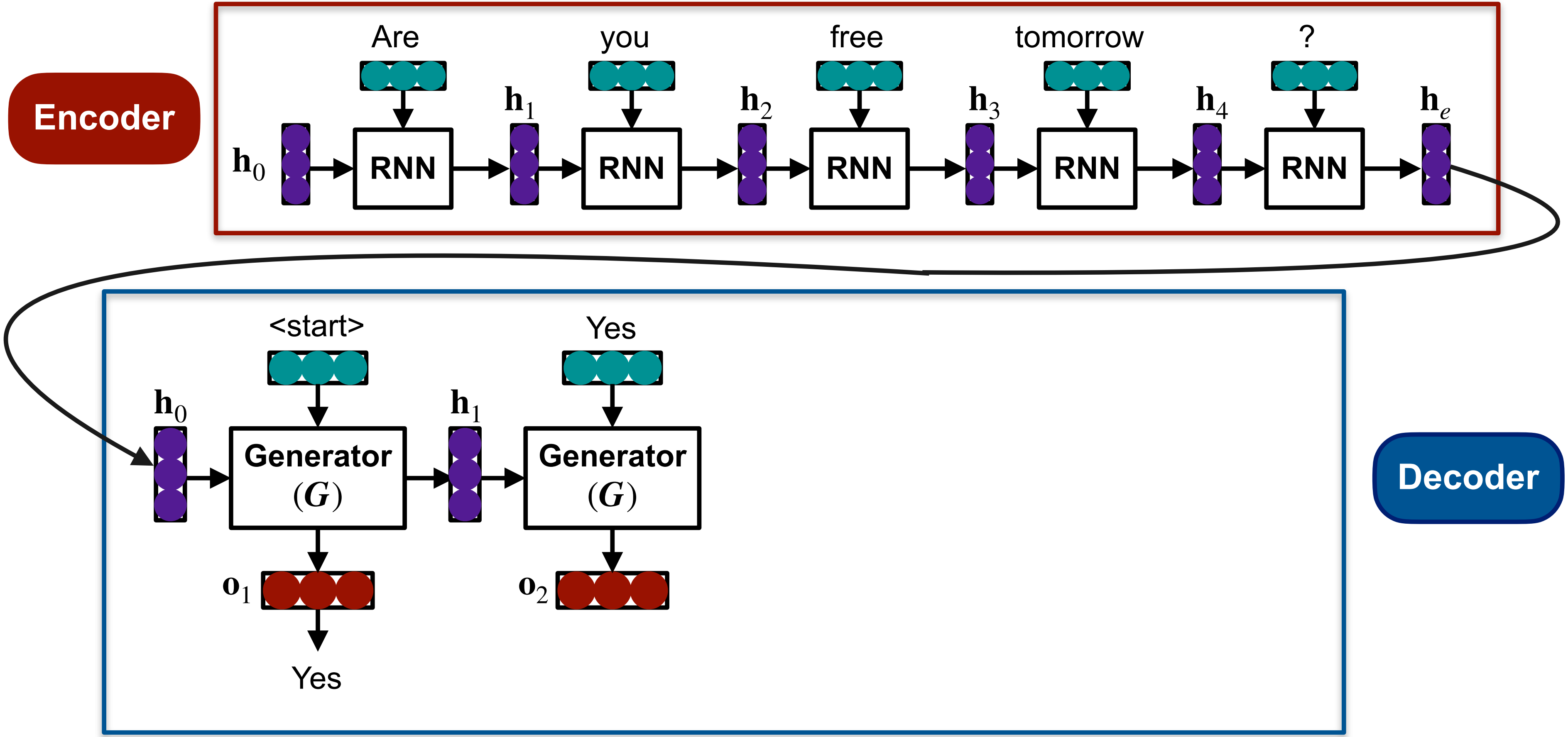
1. Attention

- Focus on source sequence
- Global Attention
- Local Attention
- Multi-headed Attention

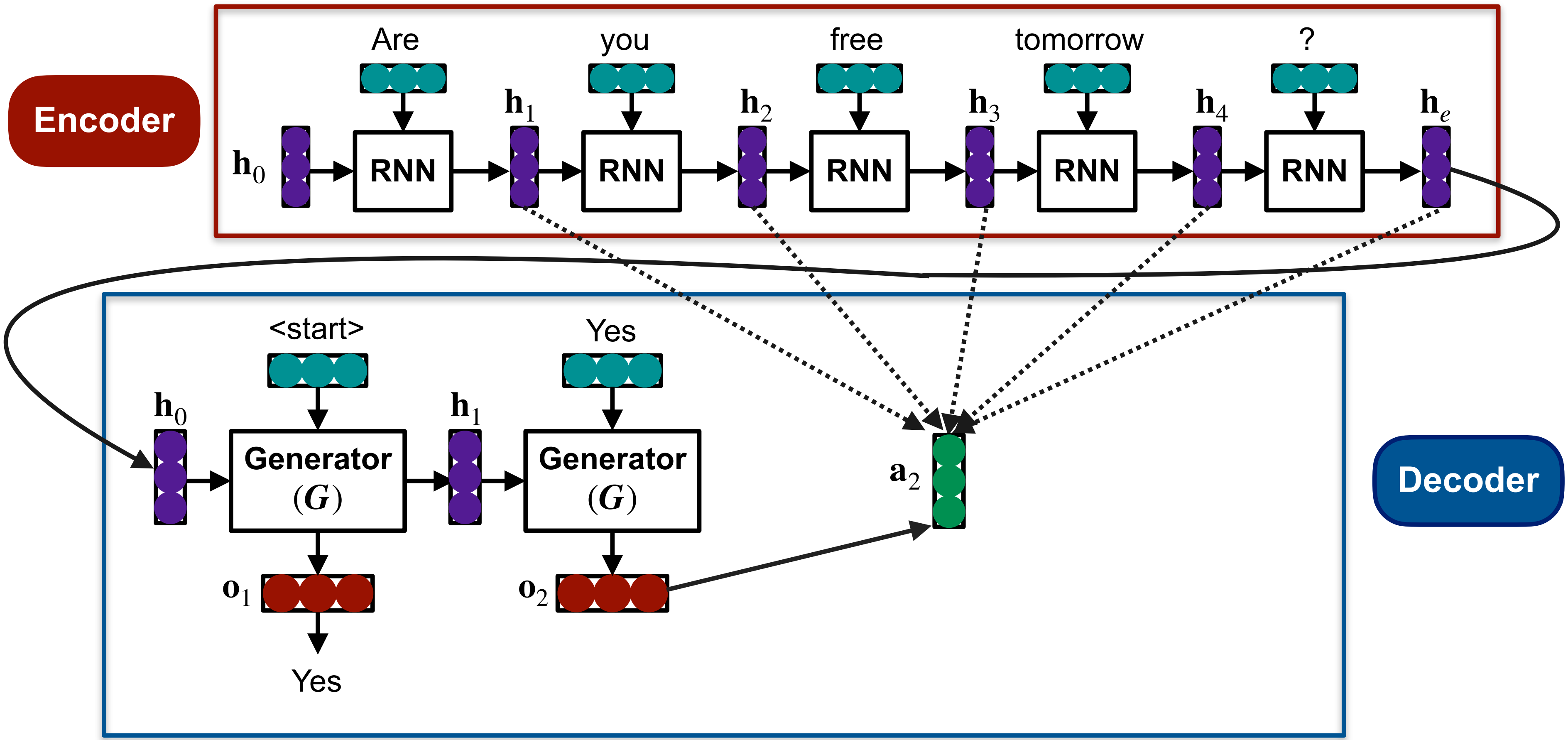


[Bahdanau et al. (2015), Luong et al. (2015), Vaswani et al. (2017)]

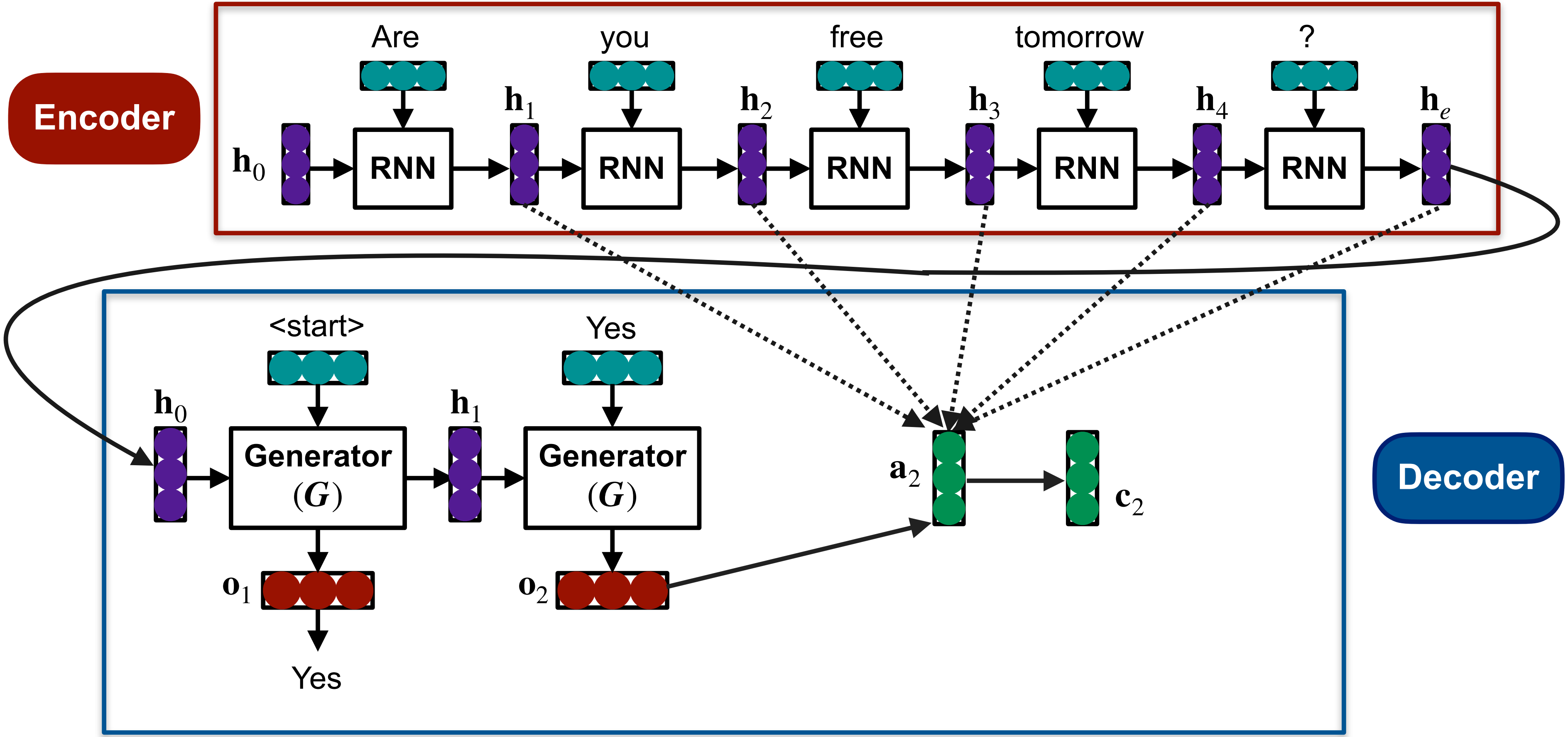
Generation Process



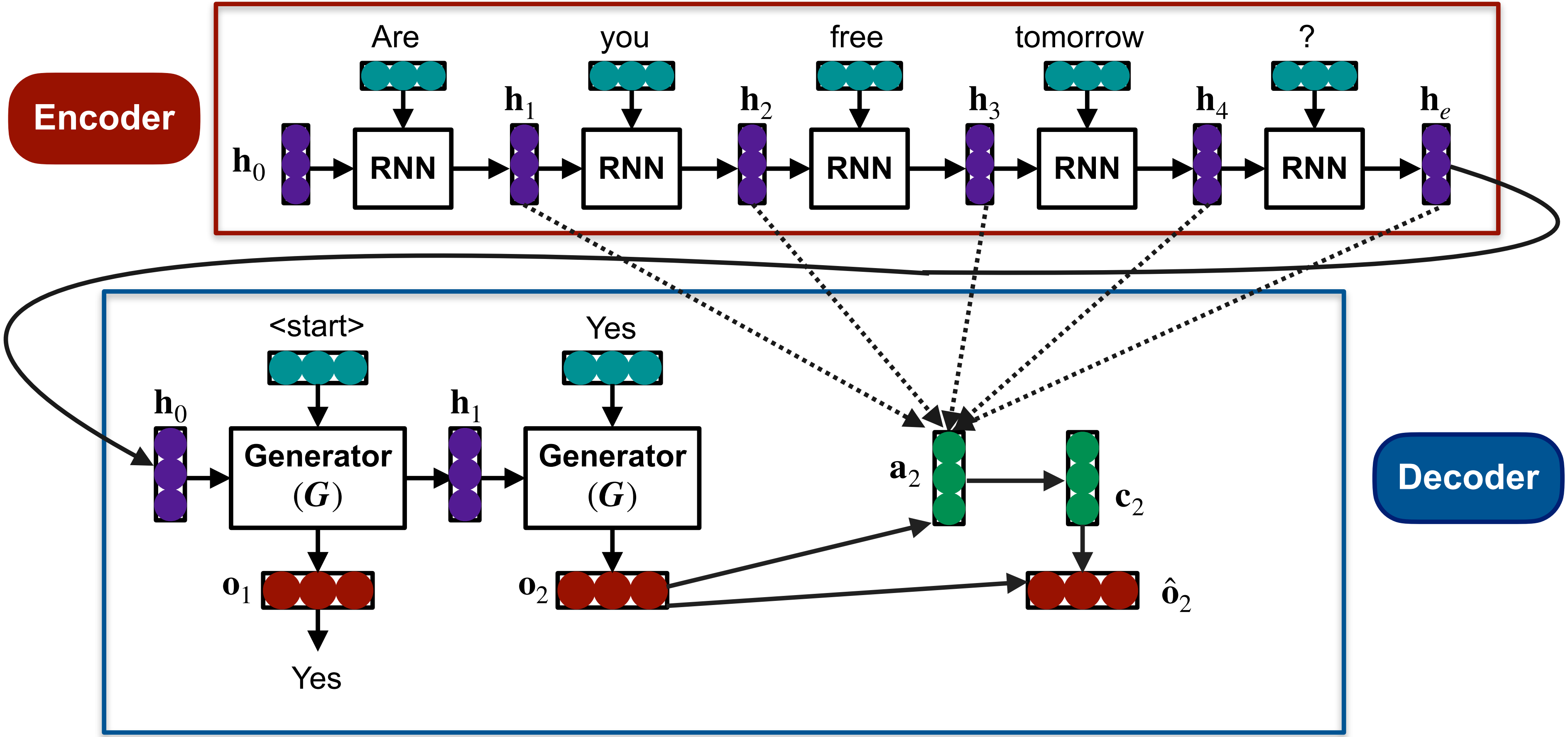
Generation Process



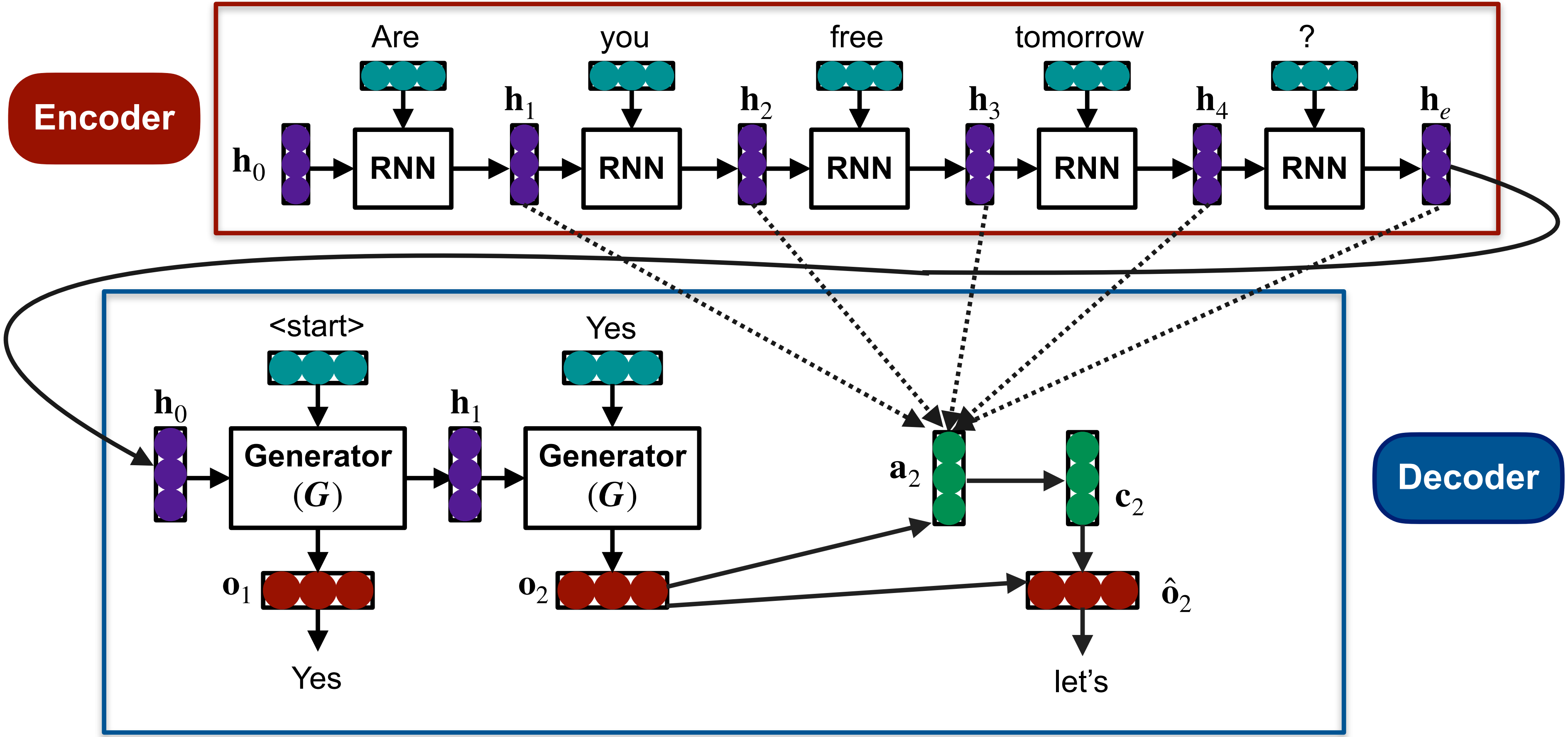
Generation Process



Generation Process



Generation Process



Output

1. Attention

- most effective - especially self and cross
- mostly control attribute tokens have been added to source sequence for attention
- under explored for controlling attributes but has a lot of potential

[Sudhakar et al. (2019), Dinan et al. (2018), Zhang et al. (2018)]

Output

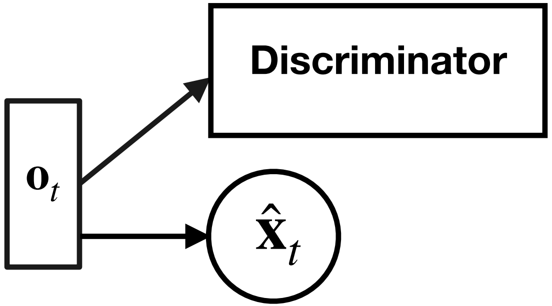
2. External Feedback
    - discriminator has to be jointly trained like GAN

3. Arithmetic or linear transform

Output

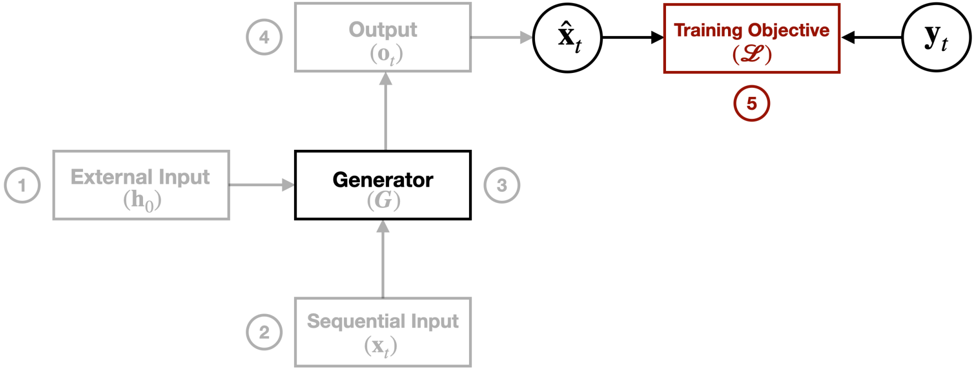
2. External Feedback

- discriminator has to be jointly trained like GAN



3. Arithmetic or linear transform

Training Objective



1. General Loss

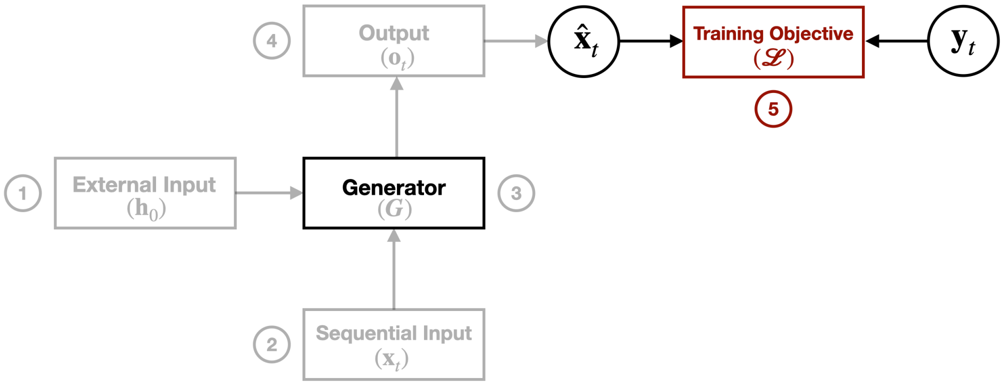
- Cross Entropy Loss
- Unlikelihood Loss
- Decoding Strategies
- Used with any generation task

2. Classifier Loss

- design multiple classifier for any control attributes

[Welleck et al. (2020), Prabhumoye et al. (2018), Yang et al. (2018)]

Training Objective



3. KL Divergence
    - used with stochastic changes
4. Task Specific Loss
    - design a loss for specific task (need not involve a classifier)
    - Strategy Loss
    - Coverage Loss
    - Structure Loss

Future Work

- ***Empirical evaluation of schema***
  - to understand quantitatively which modules are more effective in controlling attributes
  - task-related architectures
  - add these control techniques to pre-trained models like BART, T5 etc