Multi-Task Learning in ASR

Josh Meyer

University of Arizona | PhD Candidate
Mozilla Machine Learning Research | NSF Intern

DEFINITION

Definition of Multi-Task Learning

- Multi-Task Learning (MTL)
  - *Training* a model
    to perform *multiple tasks,*
    where a *subset* of parameters
    are *shared* among tasks.

Definition of Multi-Task Learning

- Task:
  - X, a sample of data from a certain domain
  - Y, a sample of targets from a certain domain
  - $f : X \rightarrow Y$, a function which maps data to targets

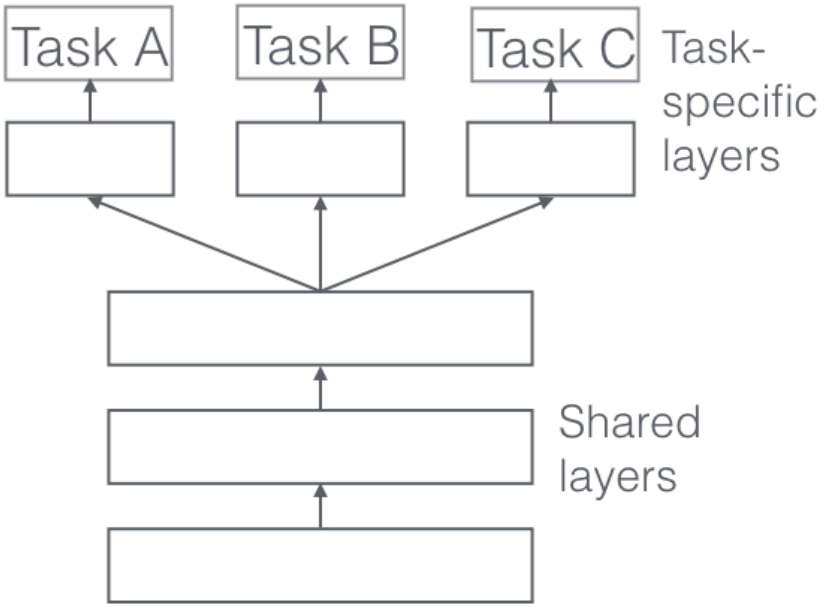
USEFULNESS

Use of Multi-Task Learning

- Inductive Bias:
  - Helps to narrow the search space
  - Features will be task-independent

Use of Multi-Task Learning

- Inductive Bias:
  - Helps to narrow the search space
  - Features will be task-independent



EXAMPLE

Example of Multi-Task Learning

- Task:
  - X_1 , a collection of photos of dogs from one camera
  - Y_1 , a set of **`dog_breed`** labels
  - $f_1 : X_1 \rightarrow Y_1$, a conv neural network

Example of Multi-Task Learning

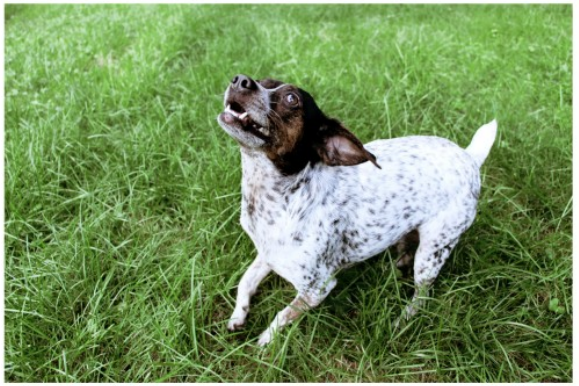
- Creating New Tasks:
  - X_2 , photos from new camera
  - Y_2 , a set of **`dog_size`** labels
  - $f_2 : X_1 \rightarrow Y_1$, an RNN

Example of Multi-Task Learning

- {old_task, new_task}

{rottweiler, large}

{collie, large}

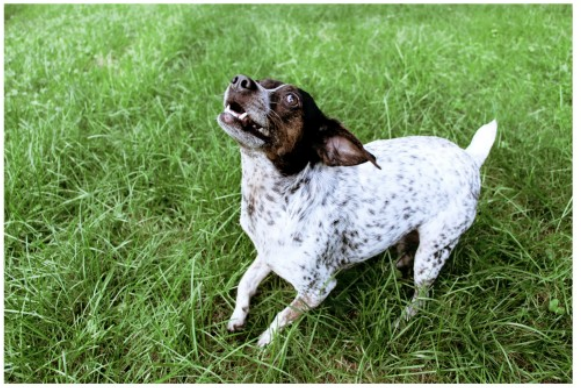


{terrier, small}

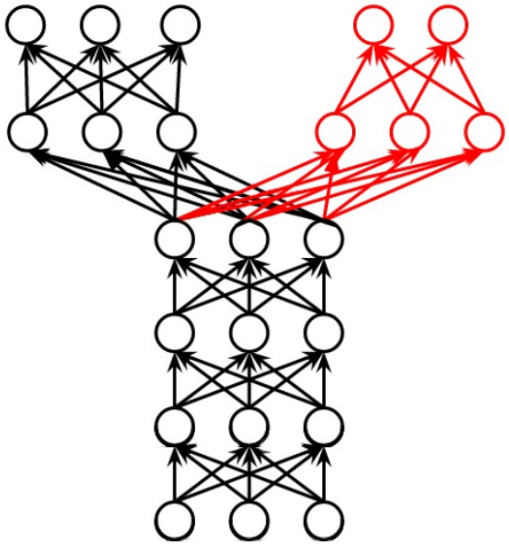
Example of Multi-Task Learning

{rottweiler, large}

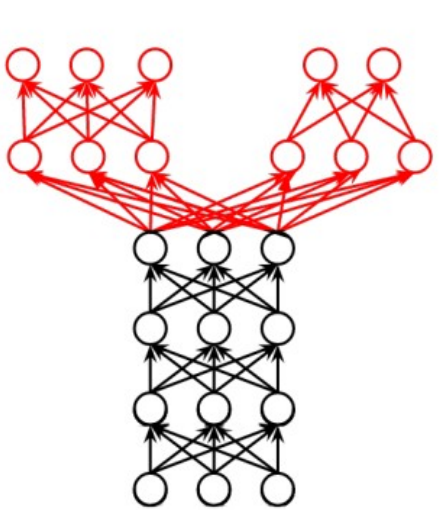
{collie, large}



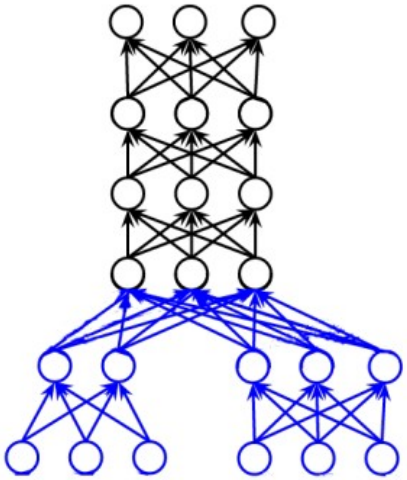
{terrier, small}



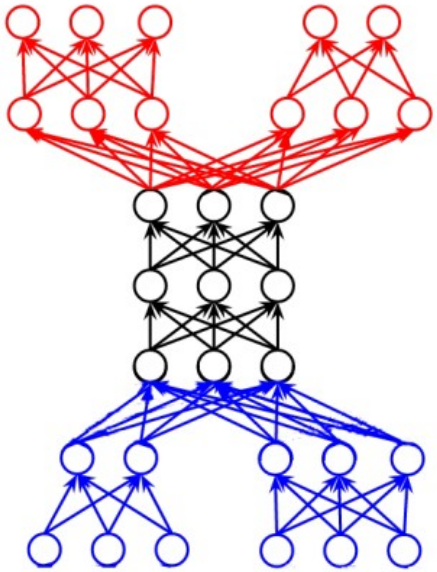
Examples of Multi-Task Learning



(a) Multi-Label



(b) Multi-Data



(c) Multi-Label + Multi-Data

SURVEY of MTL in ASR

Multi-Task Learning in ASR

- Two Broad Categories:
  - Multi-lingual
  - Mono-lingual

Multi-Task Learning in ASR

- Two Broad Categories:
  - Multi-lingual
    - Use data from more than one language
  - Mono-lingual

Multi-Task Learning in ASR

- Two Broad Categories:
  - Multi-lingual
    - Use data from more than one language
    - Typically in a low-resource scenario
  - Mono-lingual

Multi-Task Learning in ASR

- Two Broad Categories:
  - Multi-lingual
    - Use data from more than one language
    - Typically in a low-resource scenario
    - Language transfer
  - Mono-lingual

Multi-Task Learning in ASR

- Two Broad Categories:
  - Multi-lingual
    - Use data from more than one language
    - Typically in a low-resource scenario
    - Language transfer
  - Mono-lingual
    - Use data from only one language

Multi-Task Learning in ASR

- Two Broad Categories:
  - Multi-lingual
    - Use data from more than one language
    - Typically in a low-resource scenario
    - Language transfer
  - Mono-lingual
    - Use data from only one language
    - Not only low-resource

Multi-Task Learning in ASR

- Two Broad Categories:
  - Multi-lingual
    - Use data from more than one language
    - Typically in a low-resource scenario
    - Language transfer
  - Mono-lingual
    - Use data from only one language
    - Not only low-resource
    - Model robustness

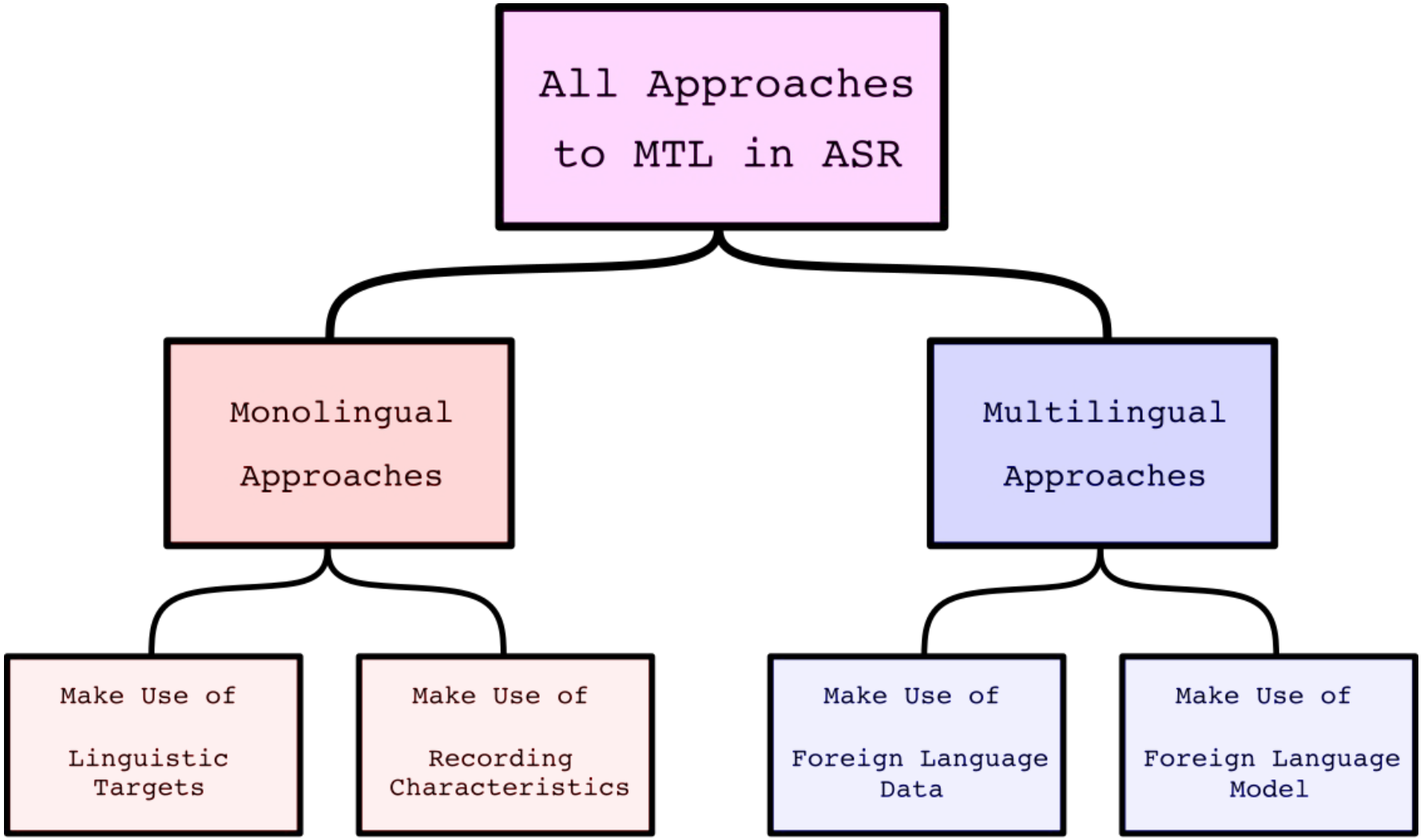
Main Branches

- Multi-lingual
  - Source Data
  - Source Model

Main Branches

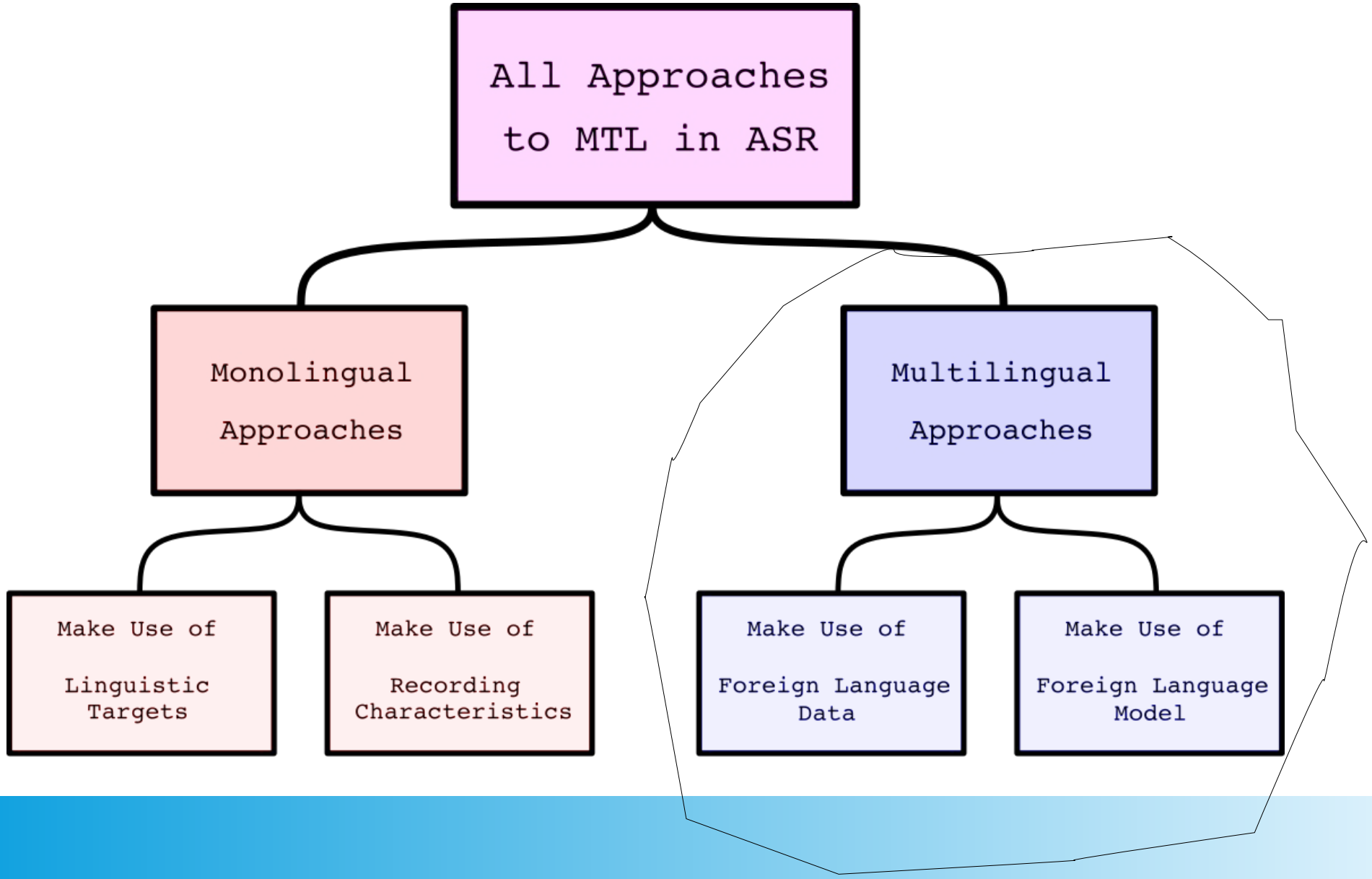
- Multi-lingual
  - Source Data
  - Source Model
- Mono-lingual
  - Phonetic Targets
  - Recording Characteristics

Multi-Task Learning in ASR



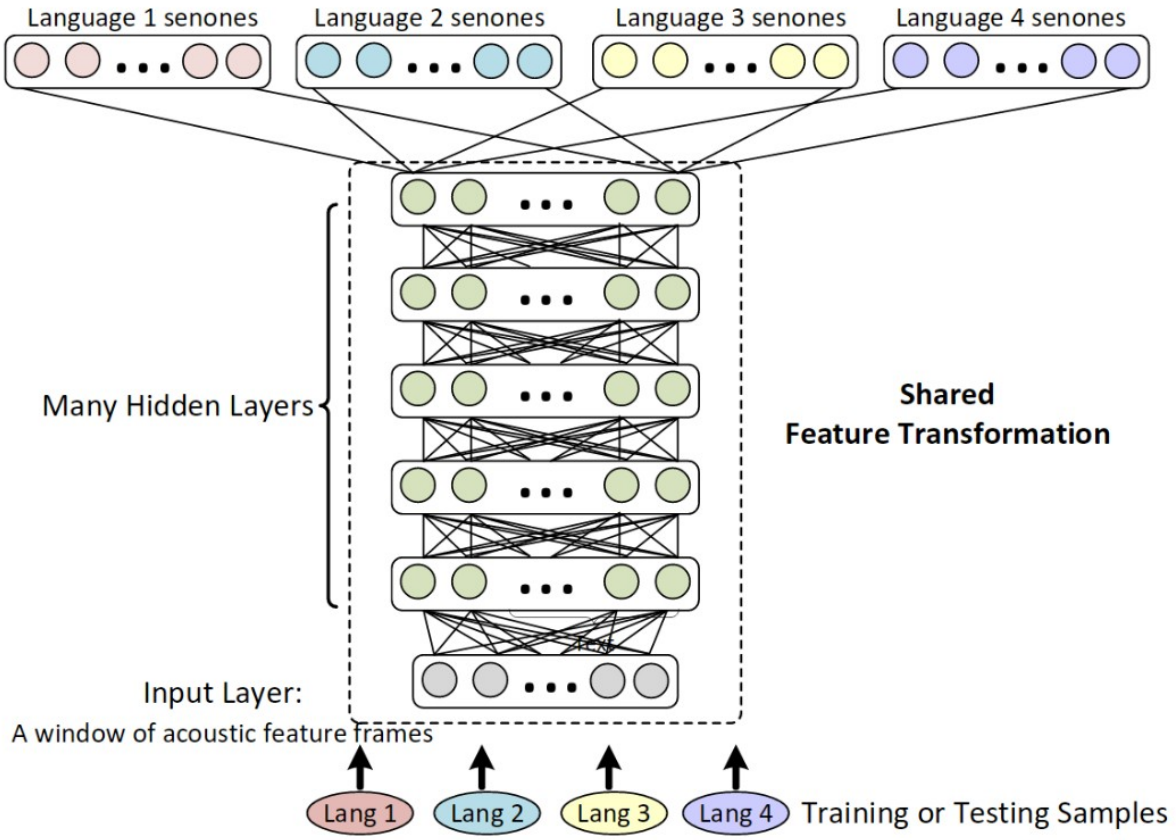
MULTILINGUAL

Multilingual Approaches



Multilingual via Source Data

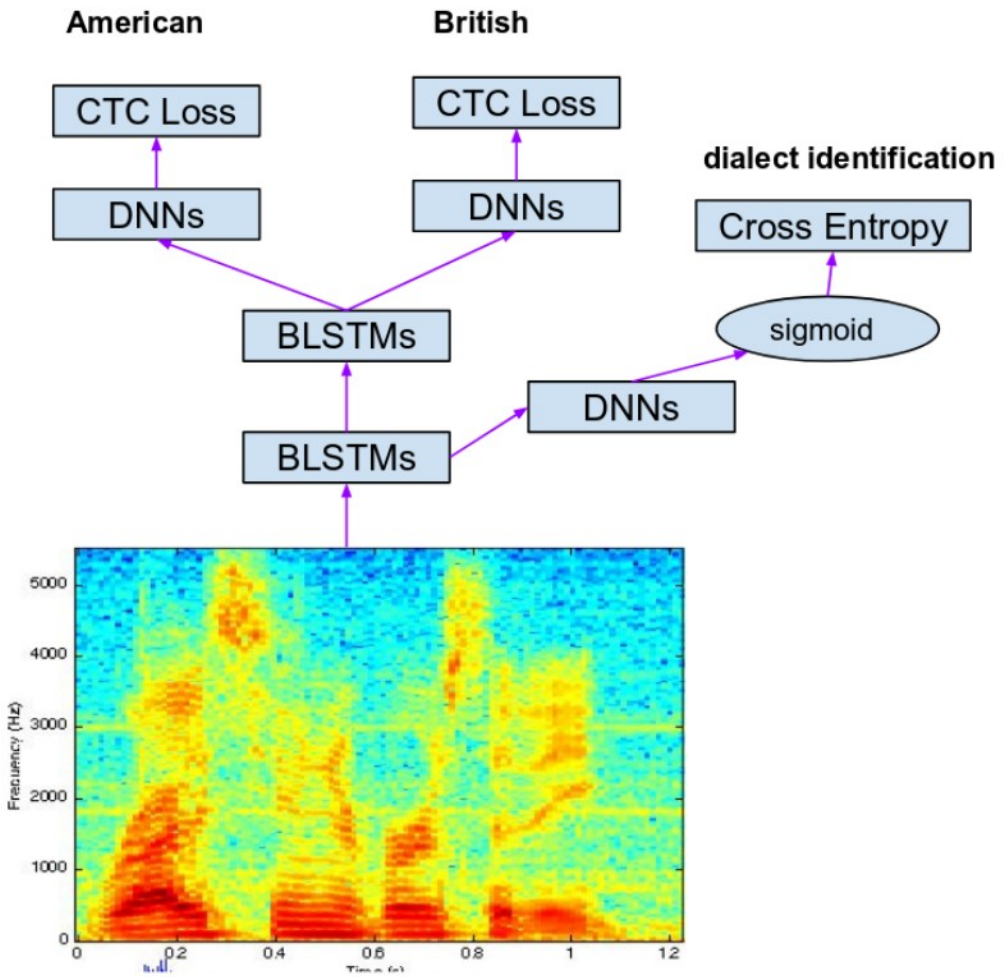
- Multilingual
  - Source Data



https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/DNN-MultiLingual-ICASSP2013.pdf

Multilingual via Source Data

- Multi-Accent
  - Source Data



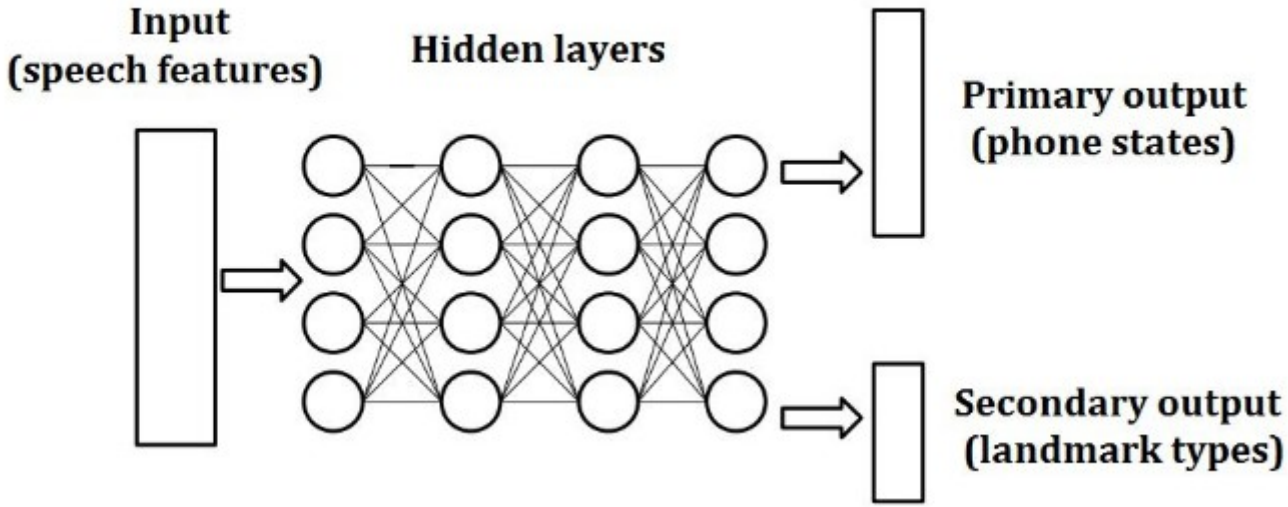
https://arxiv.org/pdf/1802.02656.pdf

Multilingual via Source Data

- Pros
  - Control: influence of data during training
- Cons
  - Time: Must retrain from scratch with all data every time

Multilingual via Source Model

- Multilingual
  - Source Model
    - as teacher



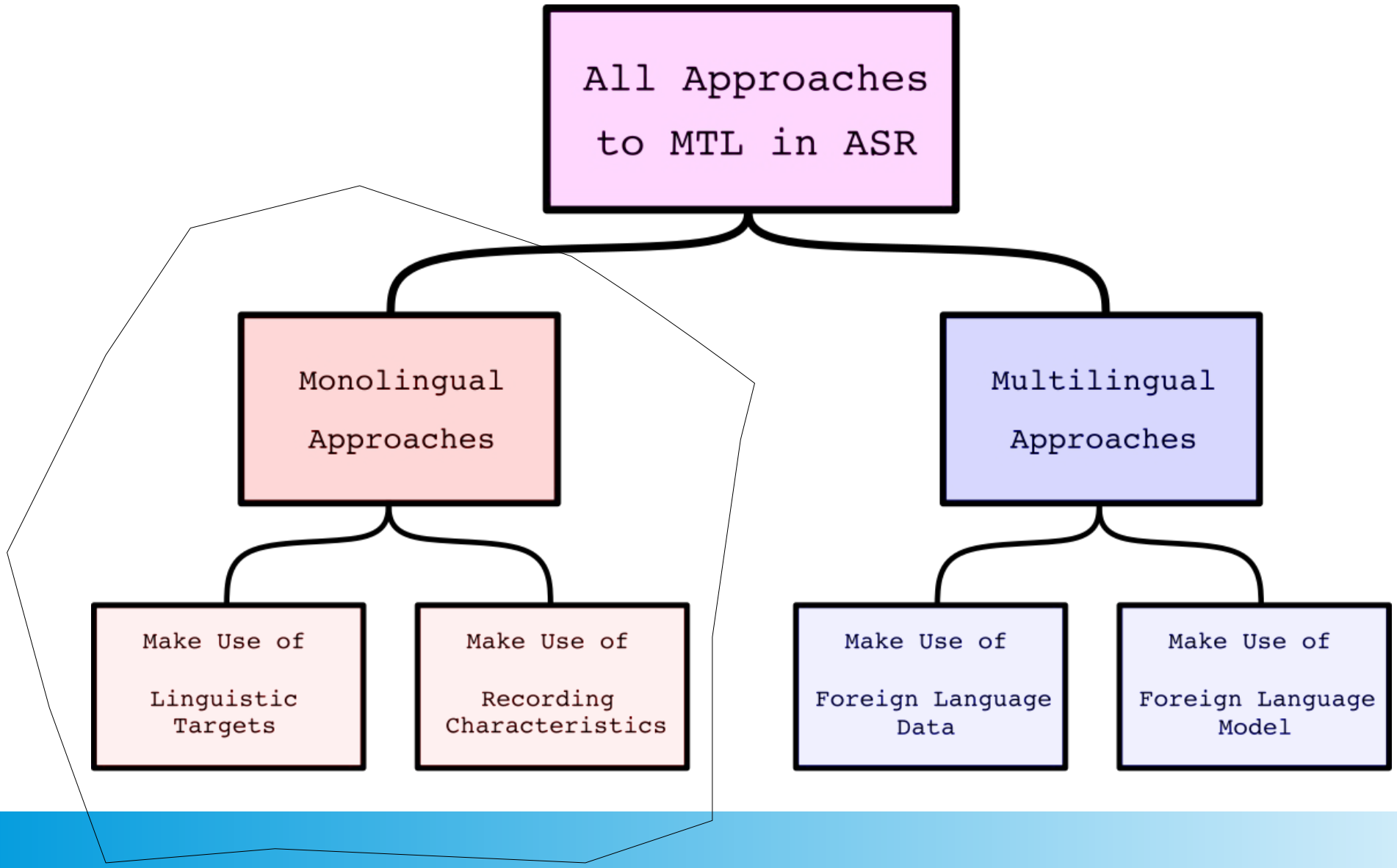
https://arxiv.org/pdf/1805.05574.pdf

Multilingual via Source Model

- Pros
  - Time: No need to retrain on all source data
- Cons
  - Control: Less ability to control influence of source data

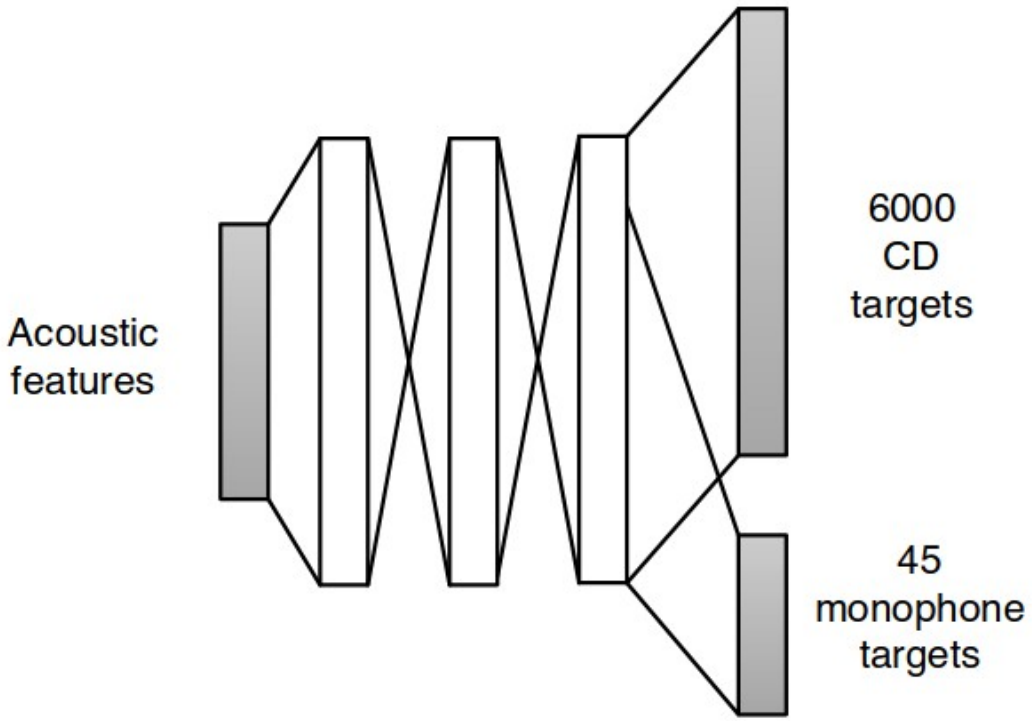
MONOLINGUAL

Monolingual Approaches



Monolingual via Phonetics

- Monolingual
  - Abstract Phonetic Classes



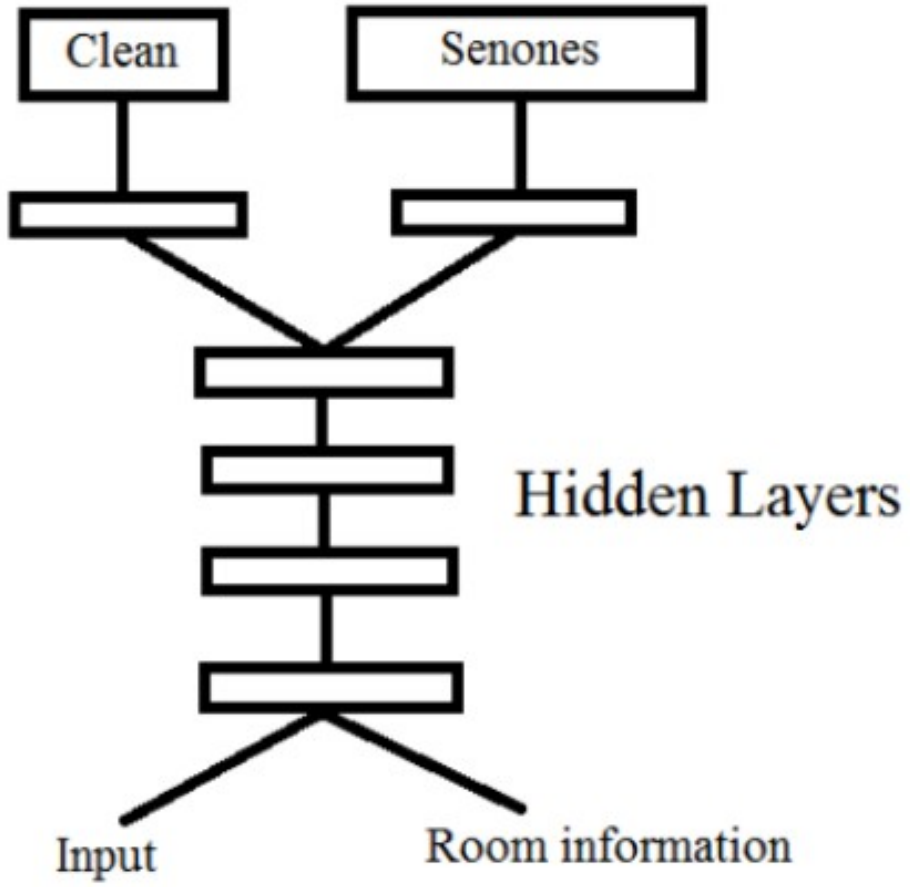
http://homepages.inf.ed.ac.uk/srenals/pb-is15.pdf

Monolingual via Phonetics

- Pros
  - Control: Easy to inject linguistic bias
- Cons
  - Time: you need a linguist to hand-craft these

Monolingual via Recording Traits

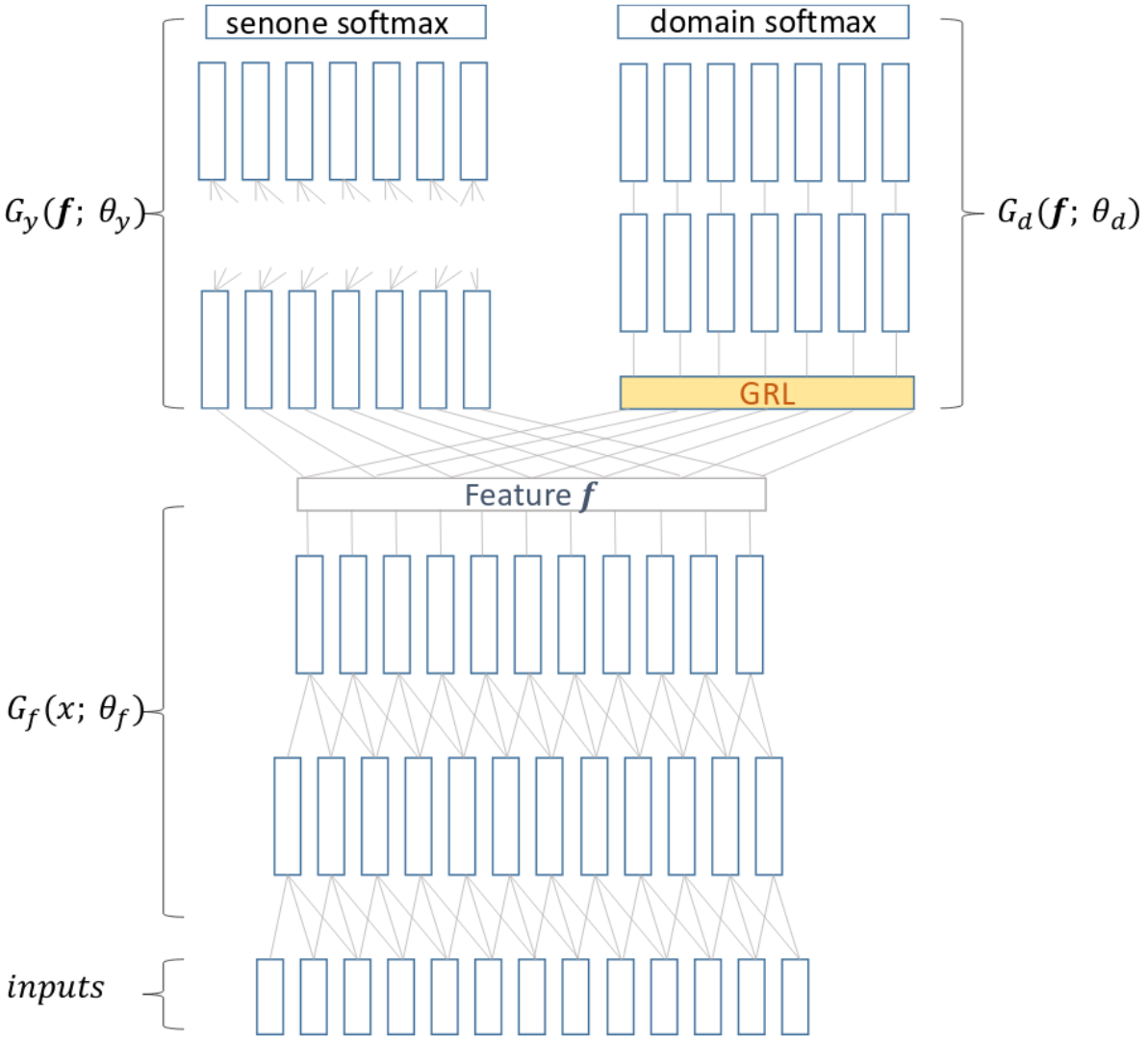
- Monolingual
  - Recording characteristics
    - regression



https://www.semanticscholar.org/paper/Improving-speech-recognition-in-reverberation-using-Giri-Seltzer/5b4f5c1a050de4b99bcfd25f562f3b0f917f4e2e

Monolingual via Recording Traits

- Monolingual
  - Recording characteristics
    - adversarial



https://arxiv.org/pdf/1806.02786.pdf

Discussion

Discussion

- What kind of tags needed?
  - Phonetic
  - Speaker
  - Language
- What is needed?
  - Time
  - Data
  - Model

Thank you for listening!

Questions?