Weighted Finite-State Transducers

Alexander Seward
Centre for Speech Technology
KTH
Stockholm, Sweden

Why Weighted Finite-State Transducers?

1. Efficiency and Generality of Classical Automata Algorithms

Efficient algorithms for a variety of problems (e.g. string-matching, compilers, ,parsing, pattern matching, process industri, design of controllability systems in aircrafts).

General algorithms: rational operations, optimizations.

2. Weights
Handling uncertainty: text, handwritten text, speech, image, biological sequences.
Increased generality: finite-state transducers, multiplicity/indeterminism.

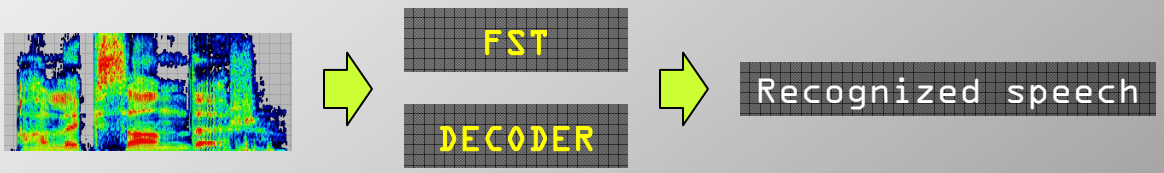
3. Applications
Text: pattern-matching, indexation, compression.
Speech: Large-vocabulary speech recognition, speech synthesis.
Image: image compression, filters.

* credits to M.Mohri

TRANSDUCERS

IN AUTOMATIC SPEECH RECOGNITION

In ASR: Mathematical models for speech-to-text translation



A uniform composition of different information sources:

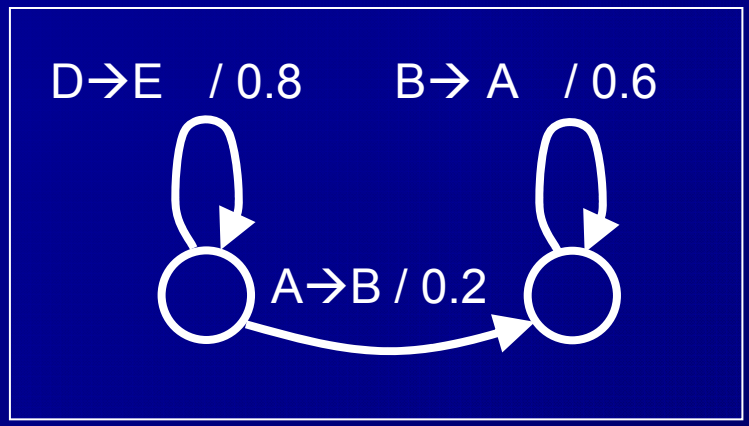
HMM-data, lexica, language models, etc..

Flexible: reduces decoder depedencies,

multiple layers,

generic optimization methods.

What is a Weighted Finite-State Transducer (WFST) ?



A finite-state machine where each arc is a weighted transduction consisting of an input, an output, and a probability/weight

Simply put: ***A translation device***

A WFSA is a transducer without output

WFSTs in recognition

- I want a ticke..#noise”..Boston from New York

FST trained on acoustics and language corpus:

#noise” must be ”t...to” !

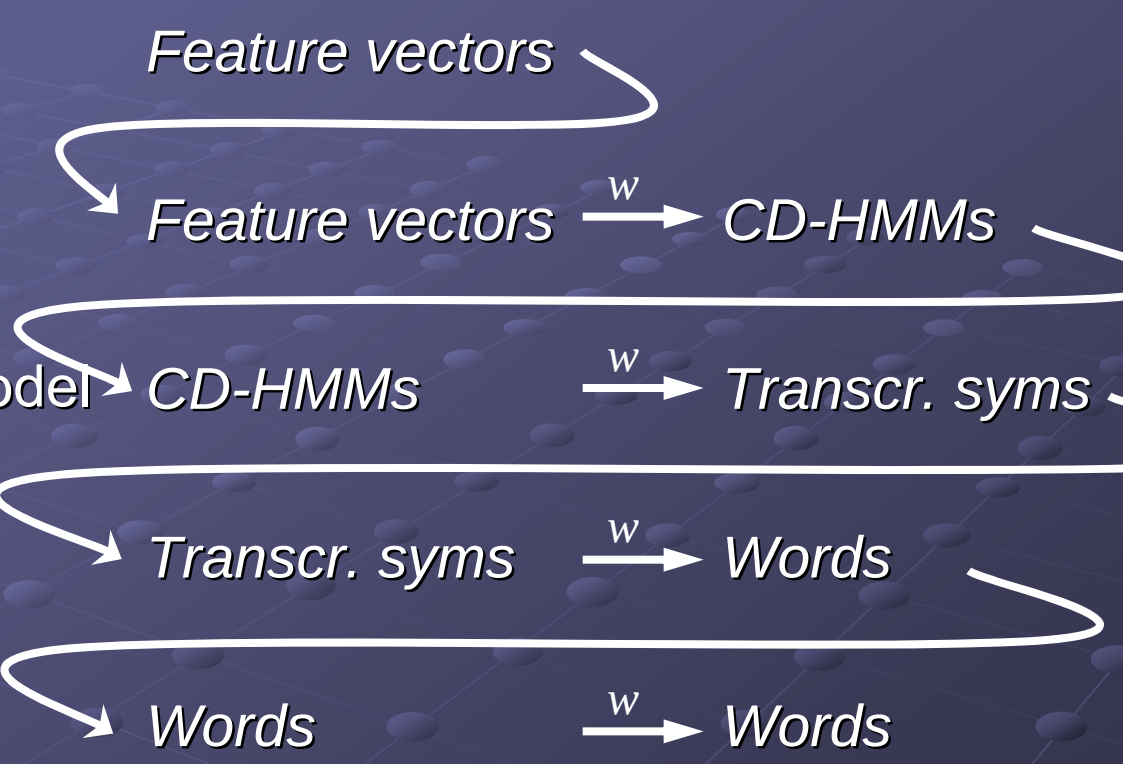
WFSTs in recognition

The bare was bear naked?

- FST trained on language corpus:
- The bear was bare naked

Recognition Cascade (simplified)

- I: Input feature vectors
- H: HMM
- C: Context-Dependency Model
- L: Lexicon
- G: Grammars



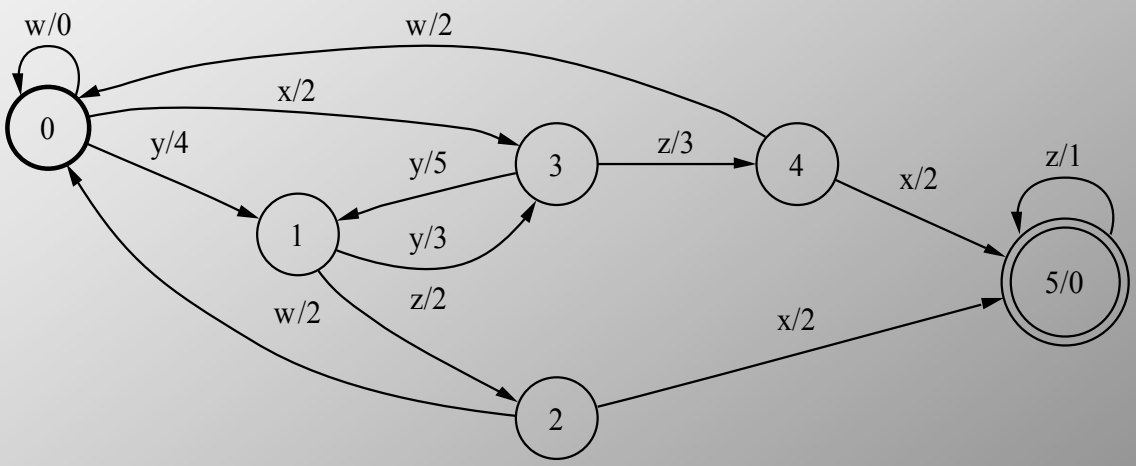
Use Weighted FST Composition
to compose the parts into one

Weighted FST operations

Best-path	Difference	Weight pushing
Closure	Equivalence	Label pushing
Compaction	Hadamard product	Reversal
Composition	Inversion	Epsilon removal
Concatenation	Minimization	Topological sort
Connection	Projection	Union
Determinization	Pruning	

Language model WFSA

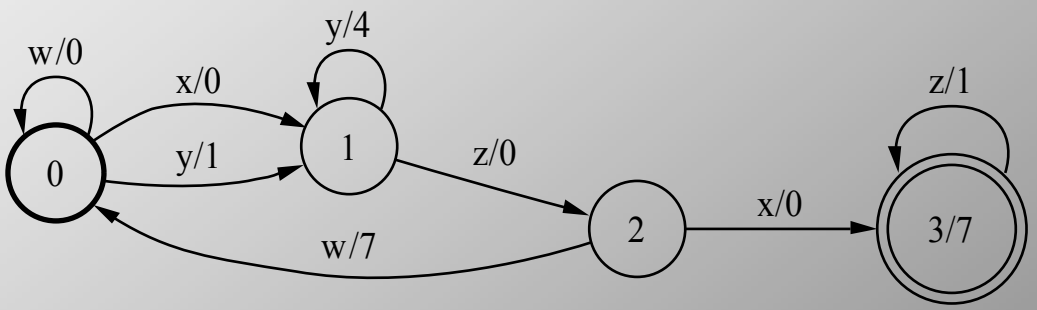
Model a priori weights for different word sequences (n-grams)



4 fictitious words: *w*, *x*, *y*, *z*

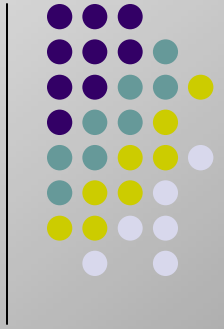
LM WFSA - Minimized

Model a priori weights to different word sequences (n-grams)



4 fictitious words: *w, x, y, z*

Pronunciation knowledge



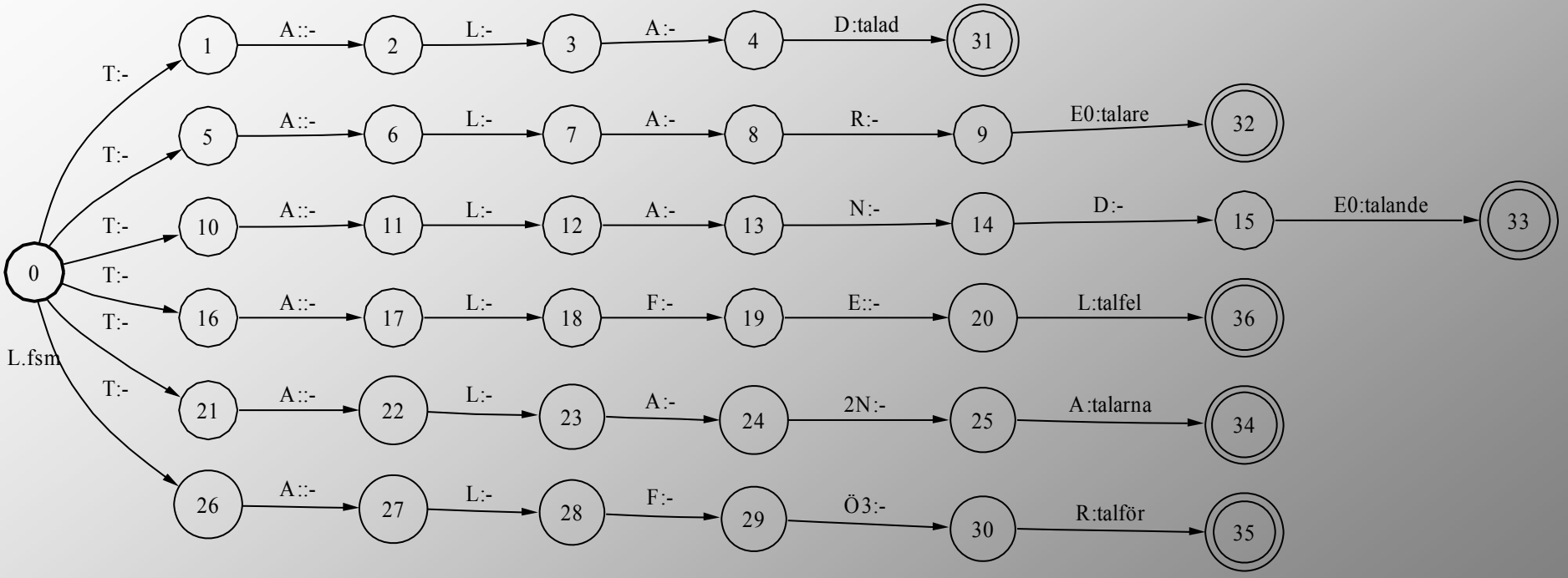
because			about		
IPA	ARPAbet	%	IPA	ARPAbet	%
[bikʌz]	[b iy k ah z]	27%	[əbaʊ]	[ax b aw]	32%
[bɨkʌz]	[b ix k ah z]	14%	[əbaʊt]	[ax b aw t]	16%
[kʌz]	[k ah z]	7%	[baʊ]	[b aw]	9%
[kəz]	[k ax z]	5%	[ʌbaʊ]	[ix b aw]	8%
[bikəz]	[b ix k ax z]	4%	[ɨbaʊt]	[ix b aw t]	5%
[bɪkʌz]	[b ih k ah z]	3%	[ɨbæ]	[ix b ae]	4%
[bəkʌz]	[b ax k ah z]	3%	[əbæɾ]	[ax b ae dx]	3%
[kʊz]	[k uh z]	2%	[baʊɾ]	[b aw dx]	3%
[ks]	[k s]	2%	[bæ]	[b ae]	3%
[kɨz]	[k ix z]	2%	[baʊt]	[b aw t]	3%
[kɪz]	[k ih z]	2%	[əbaʊɾ]	[ax b aw dx]	3%
[bikʌʒ]	[b iy k ah zh]	2%	[əbæ]	[ax b ae]	3%
[bikʌs]	[b iy k ah s]	2%	[bɑ]	[b aa]	3%
[bikʌ]	[b iy k ah]	2%	[bæɾ]	[b ae dx]	3%
[bikɑz]	[b iy k aa z]	2%	[ɨbaʊɾ]	[ix b aw dx]	2%
[əz]	[ax z]	2%	[ɨbɑt]	[ix b aa t]	2%

Figure 5.7 The 16 most common pronunciations of *because* and *about* from the hand-transcribed Switchboard corpus of American English conversational telephone speech (Godfrey et al., 1992; Greenberg et al., 1996).

Use different weights to model likelihood of pronuciations!

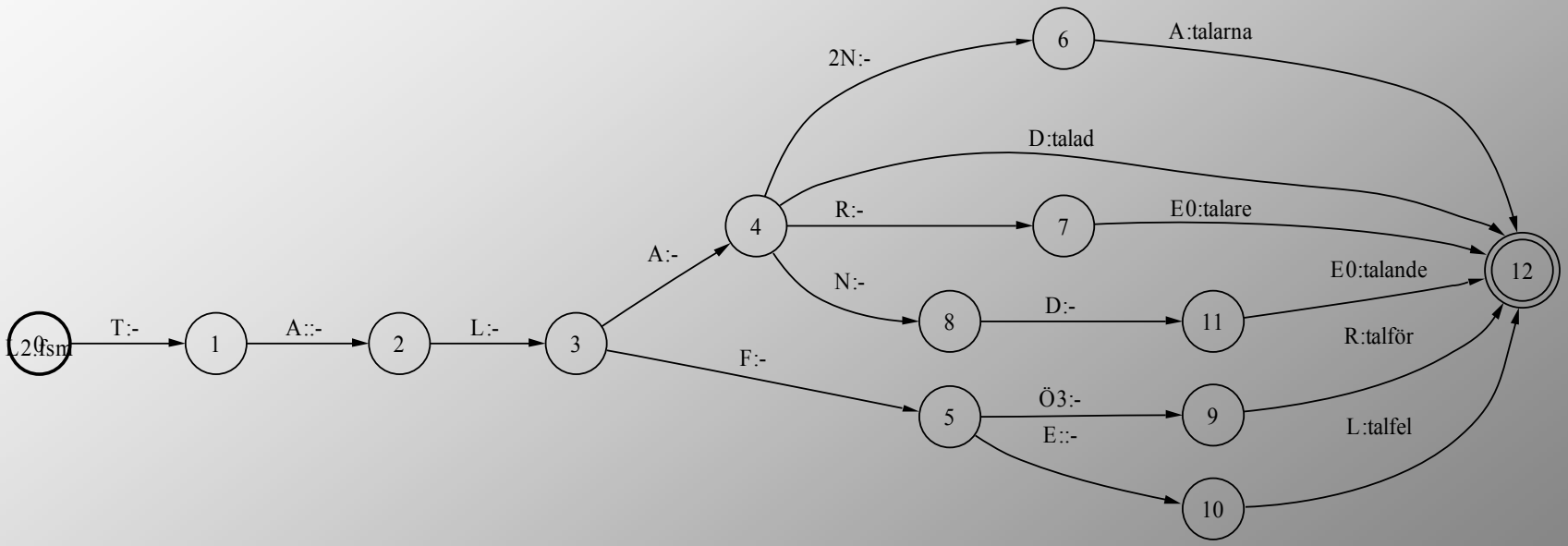
Lexicon transducer

Some phonetically similar words



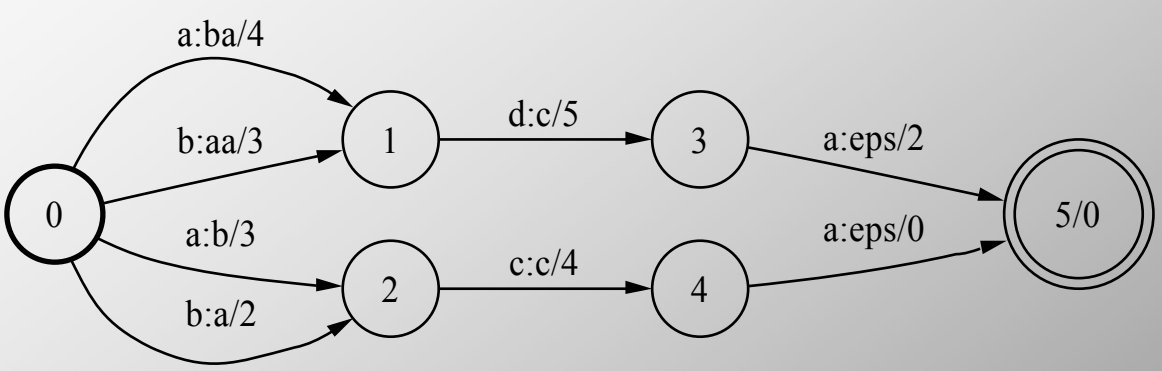
36 states

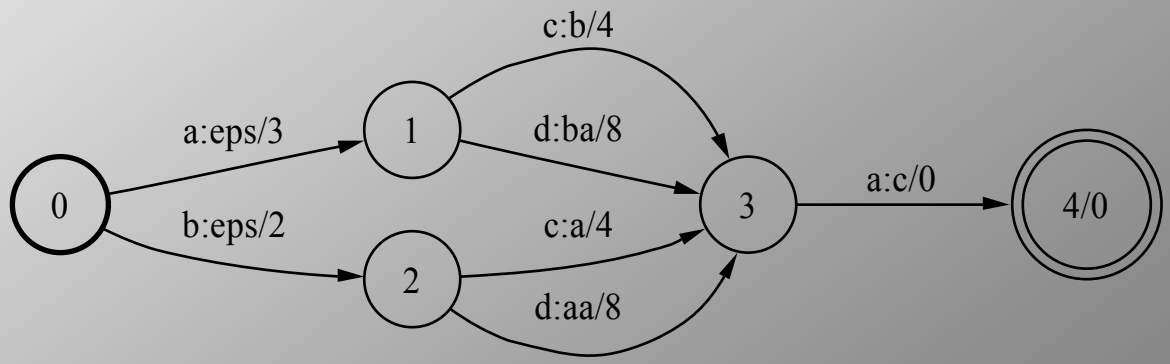
Euqivalent lexicon transducer – Deterministic



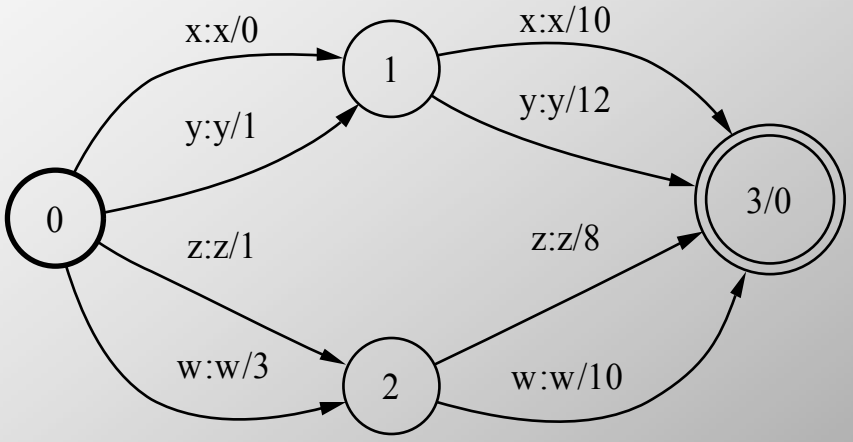
36 → 13 states

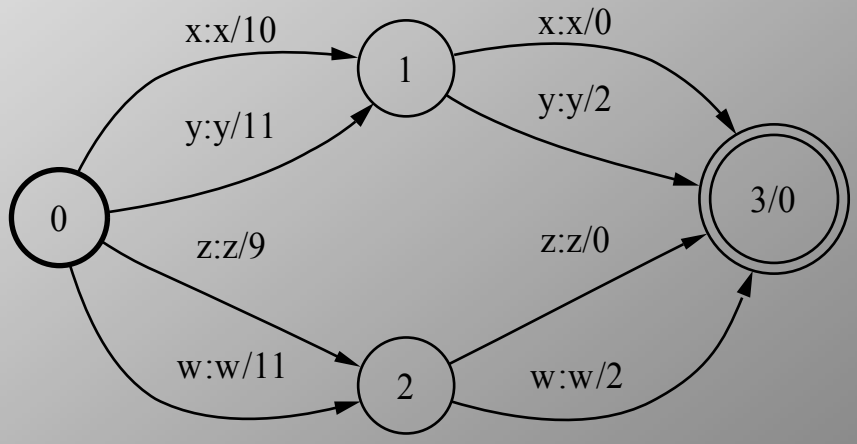
Weighted determinization



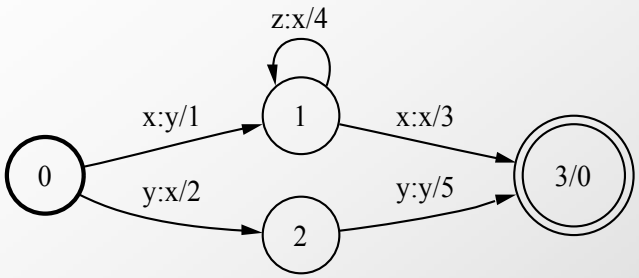


Weight pushing



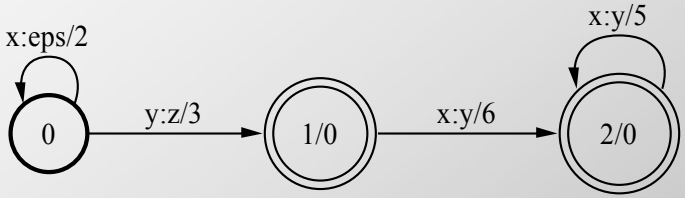


Weighted composition



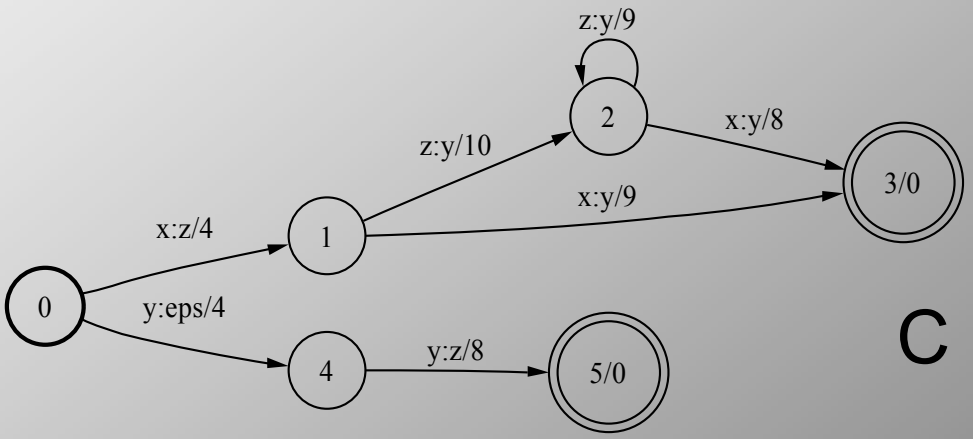
A

o



B

=



C

A: (x, z, z, x) → (y, x, x, x)
weight: 1 + 4 + 4 + 3 = 12.

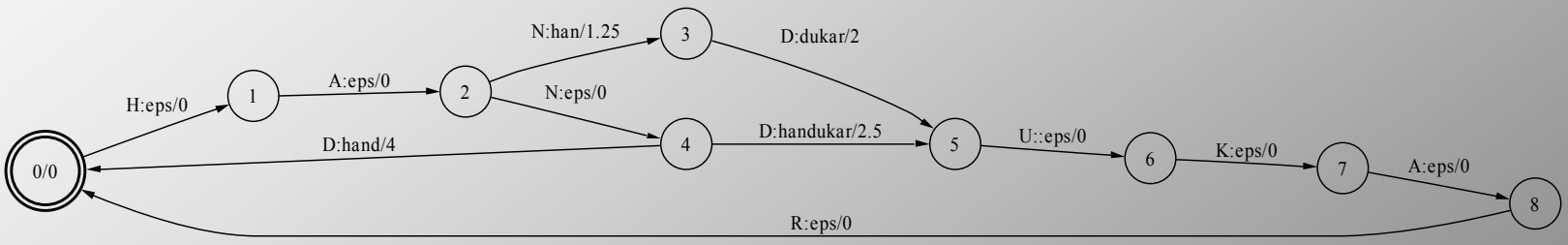
B: (y, x, x, x) → (z, y, y, y)
weight: 3 + 6 + 5 + 5 = 19.

A ○ B: (x, z, z, x) → (z, y, y, y)
total weight: 12 + 19 = 31.

C = A ○ B: (x, z, z, x) → (z, y, y, y)

same total weight: 4 + 10 + 9 + 8 = 31.

Lexicon + Grammar: L ∘ G

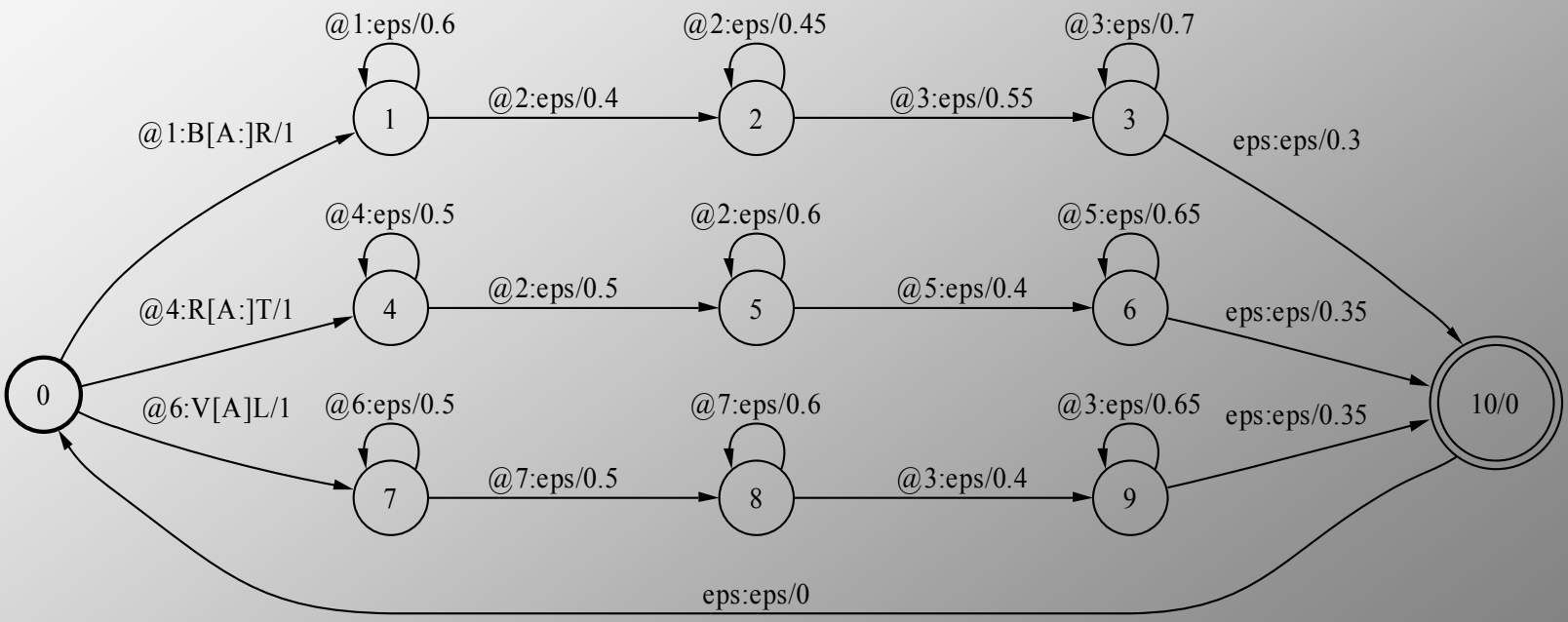


han	H A N	(Eng: he)
hand	H A N D	(Eng: hand)
handdukar	H A N D U: K A R	(Eng: towels)
dukar	D U: K A R	(Eng: set the table)

Sequence:

han dukar	H A N D U: K A R	(Eng: he sets the table)

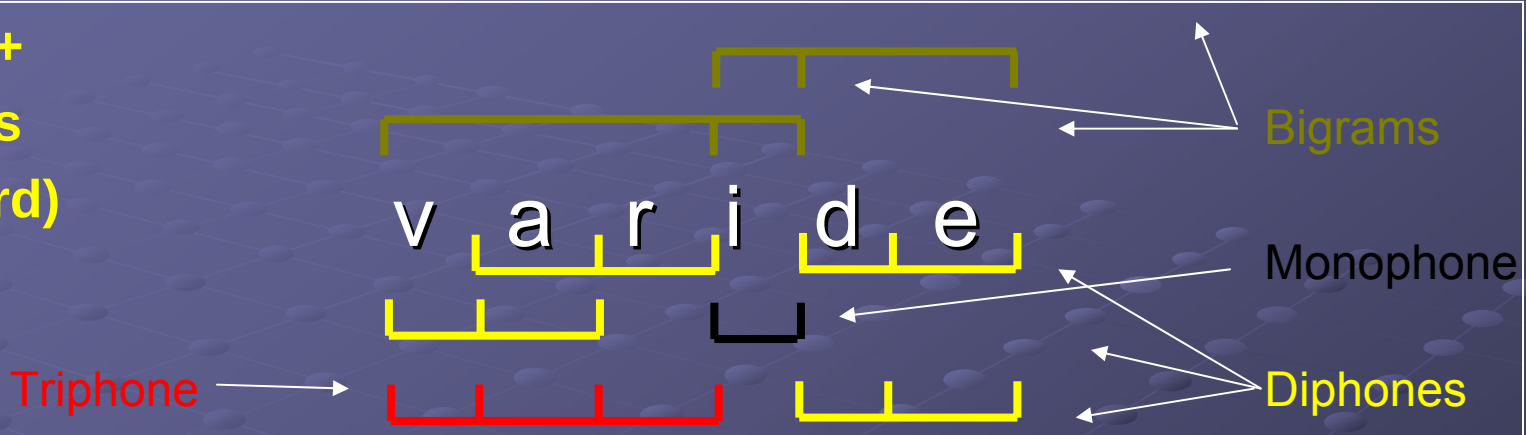
HMMs as WFSTs



Context-Dependendency Modeling

"…var i det…"

Bigrams +
Triphones
(no X-word)



Trigrams +
X-word Triphones

r [i] d
a [r] i
x [v] a , i[d]e
v[a]r , d[e]x