A UNIFIED FRAMEWORK FOR STRUCTURED PREDICTION

Wei Lu

StatNLP Research Group

Singapore University of Technology and Design

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Outline

- Introduction
- Decoding, Learning
- Semi-Markov, Latent CRF, Latent SSVM
- Parsing with CRF, Hybrid Tree, and Predicting Overlapping Structures
- Pipeline, Mean Field and Neural CRF

1. Introduction

Structured Prediction

Fruit *flies* *like* *a* *banana*

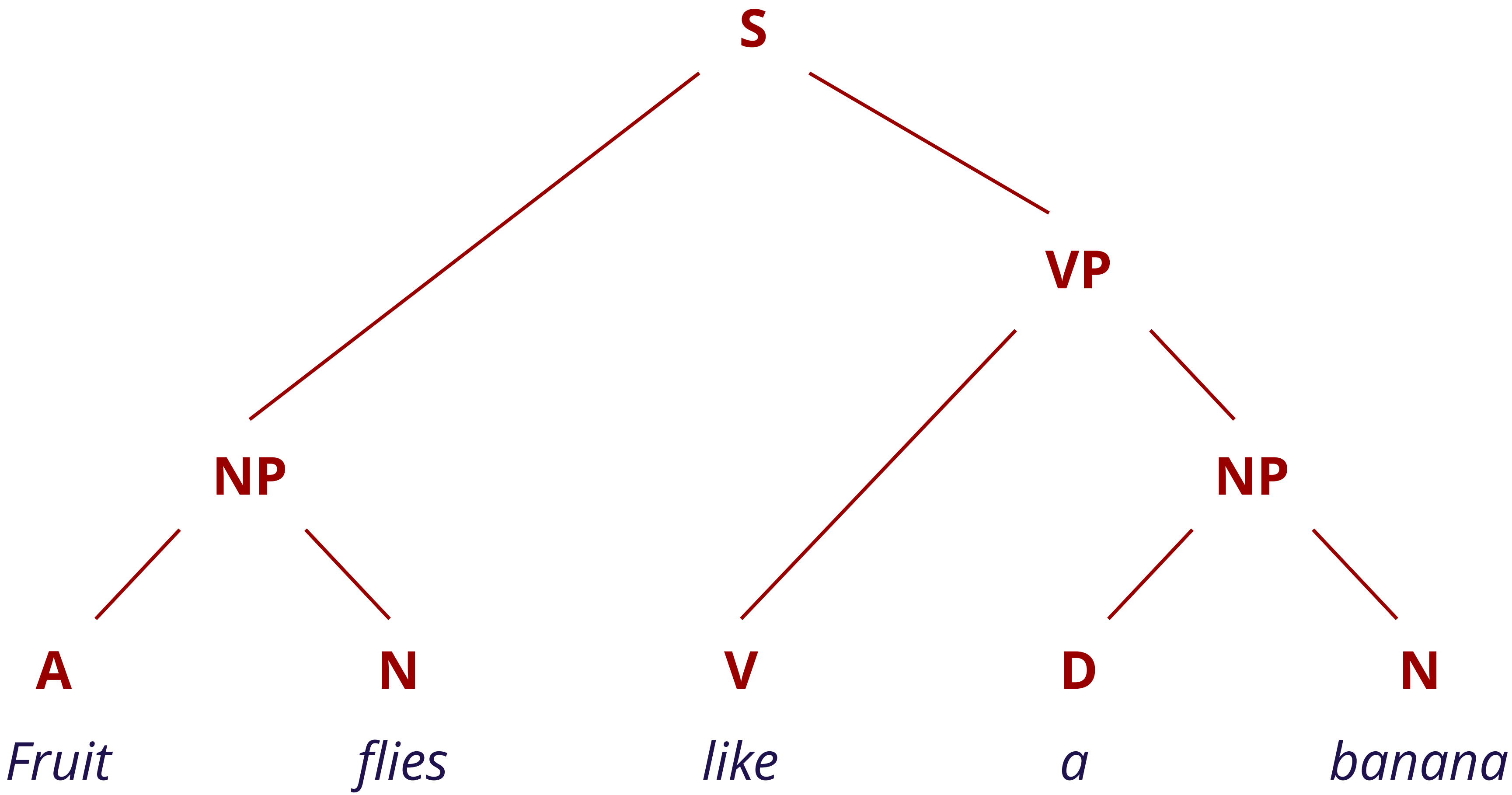
Part-of-Speech Tagging

A	**N**	**V**	**D**	**N**
Fruit	*flies*	*like*	*a*	*banana*

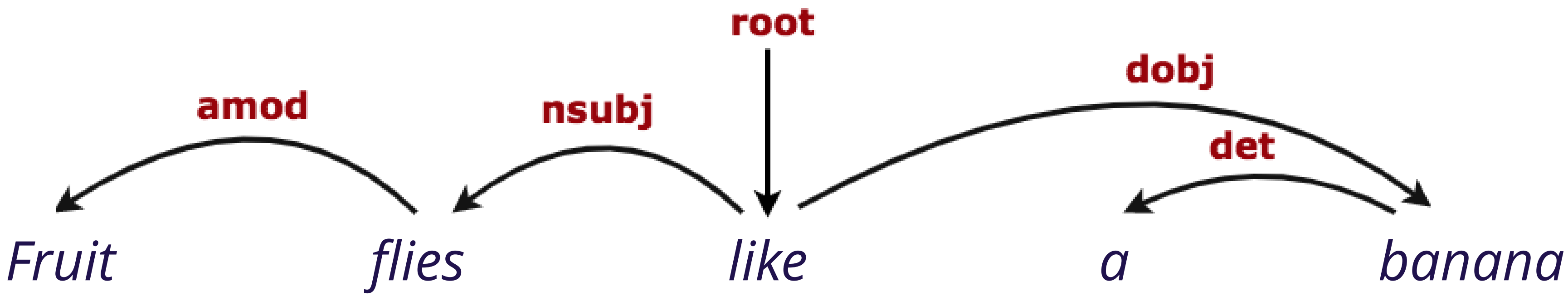
Noun-Phrase Chunking

NP			NP	
Fruit	*flies*	*like*	*a*	*banana*

Constituency Parsing



Dependency Parsing

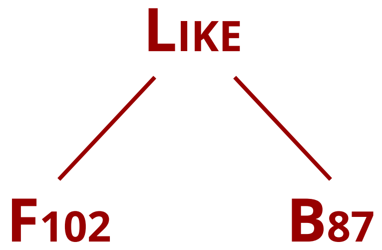


Semantic Parsing

LIKE(F102, B87)

Fruit *flies* *like* *a* *banana*

Semantic Parsing



Fruit *flies* *like* *a* *banana*

Sentiment Analysis

(neutral) **(positive)**

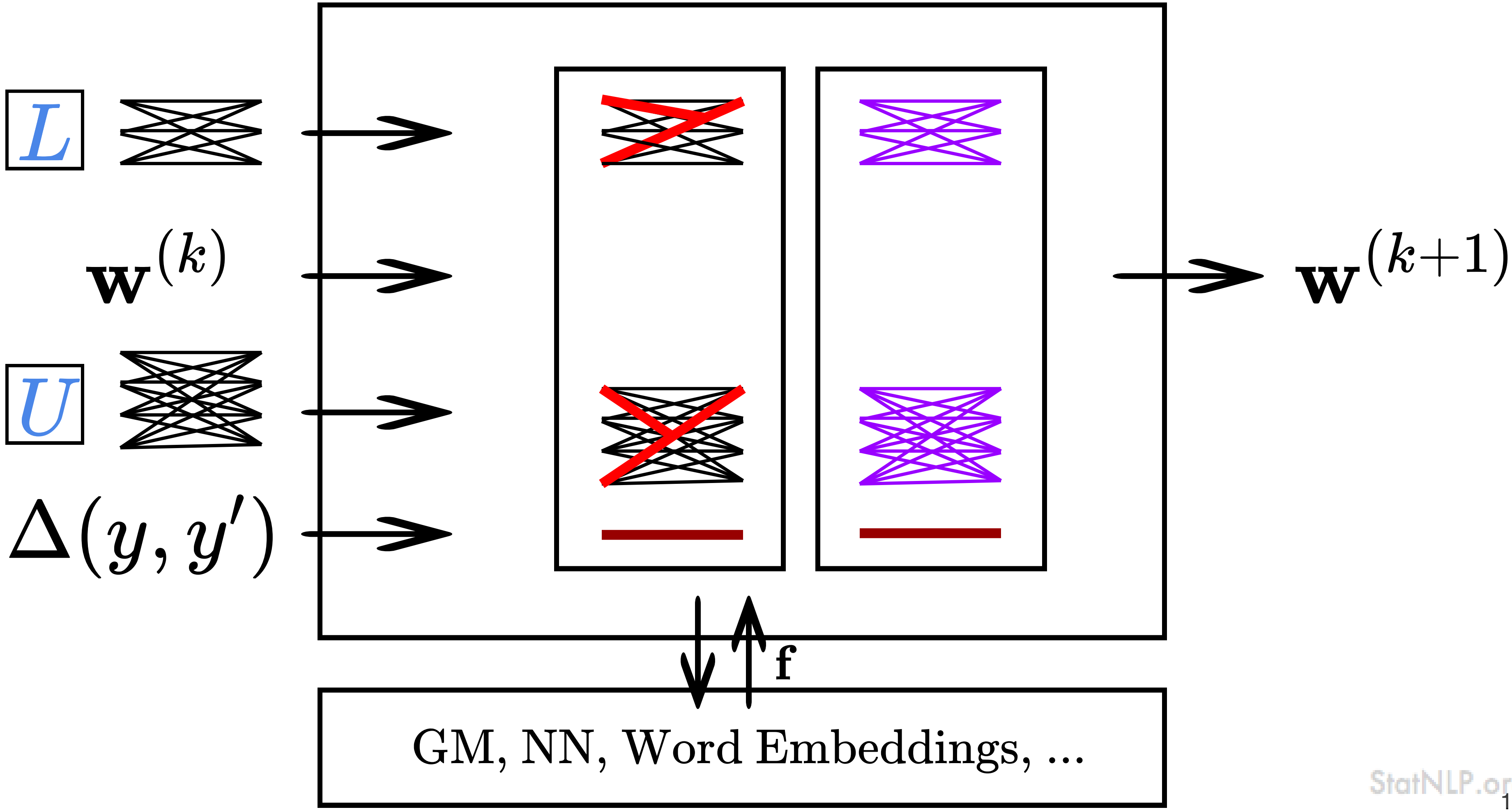
Fruit flies like a banana

Nested Chunking

This Tutorial

- Shares a conceptually new way of thinking about building structured prediction models.
- Presents a unified structured prediction framework that encompasses classic models, and is able to model structures that standard Graphical Models cannot.
- Provides a way to rapidly prototype novel structured prediction models for new tasks.

A Unified Framework



Structured Prediction

One Assumption

Structures are constructed by following a collection of discrete actions.

States, Actions

S: shift

L: left-arc

R:right-arc

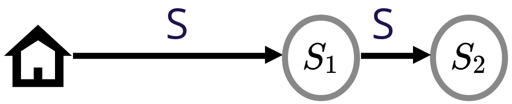
Fruit *flies* *like* *a* *banana*

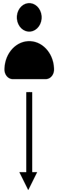
States, Actions

S: shift

L: left-arc

R:right-arc





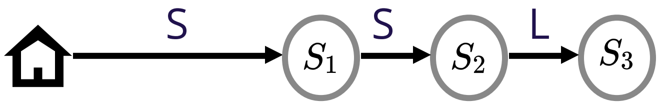
Fruit *flies* *like* *a* *banana*

States, Actions

S: shift

L: left-arc

R:right-arc

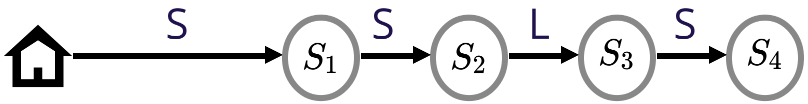


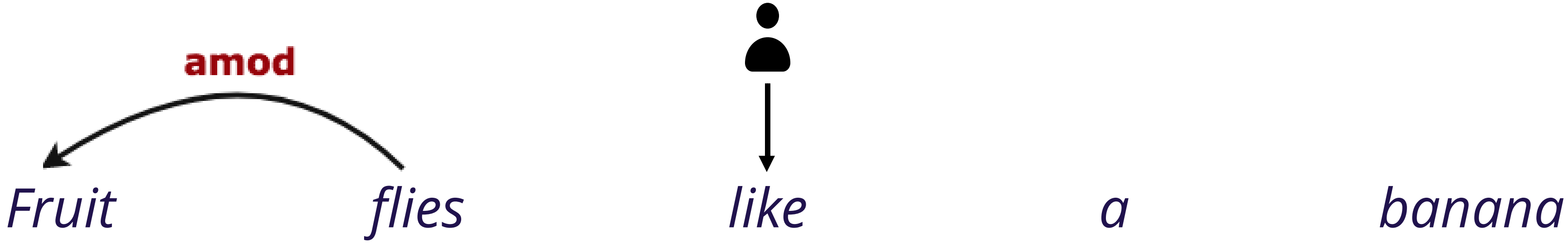
States, Actions

S: shift

L: left-arc

R:right-arc



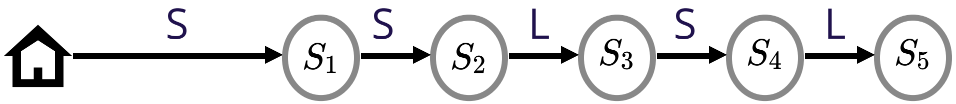


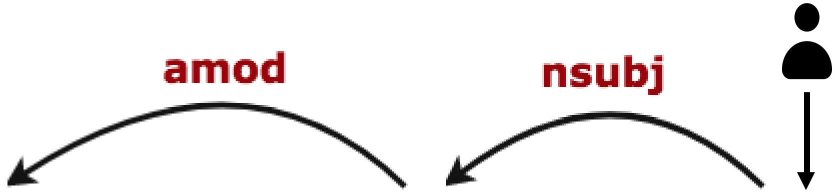
States, Actions

S: shift

L: left-arc

R:right-arc





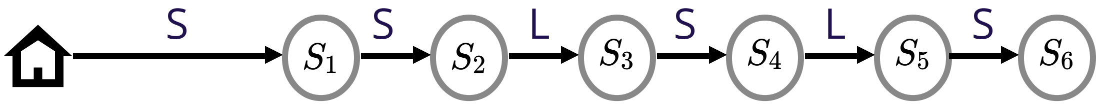
Fruit *flies* *like* *a* *banana*

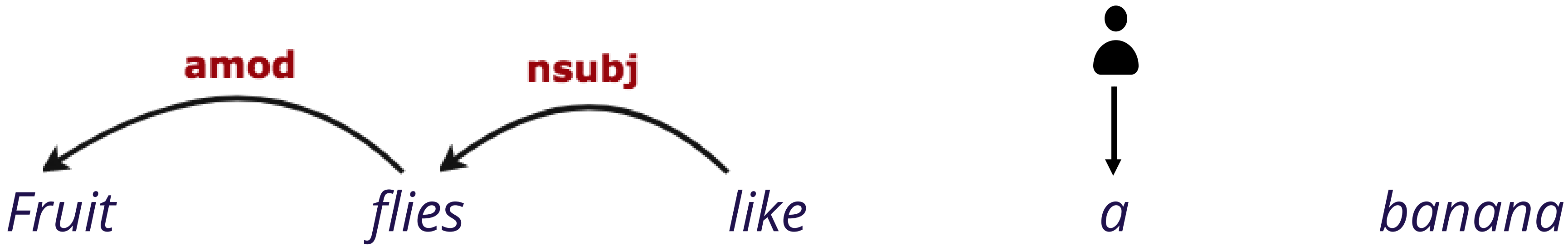
States, Actions

S: shift

L: left-arc

R:right-arc



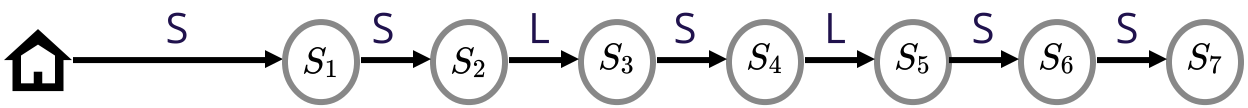


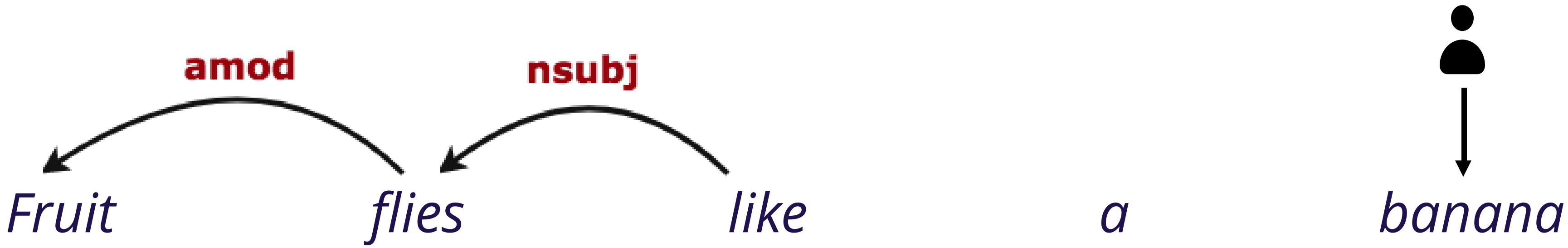
States, Actions

S: shift

L: left-arc

R:right-arc



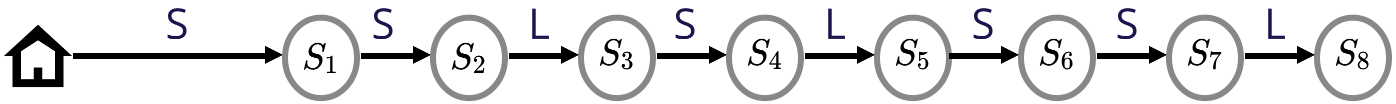


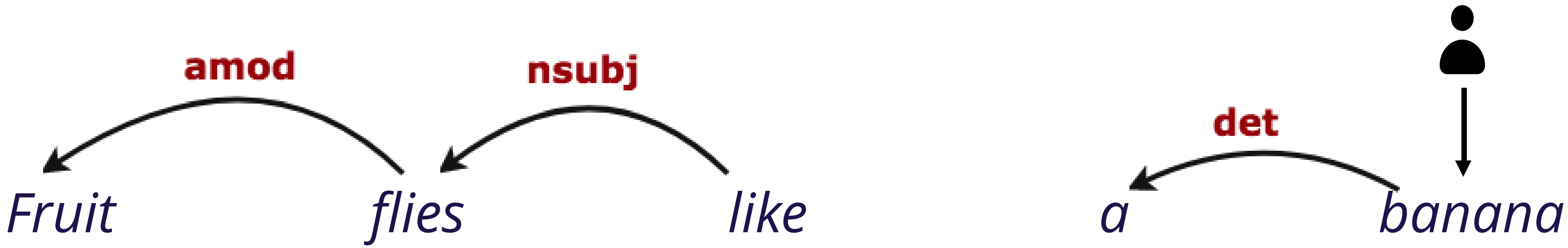
States, Actions

S: shift

L: left-arc

R:right-arc



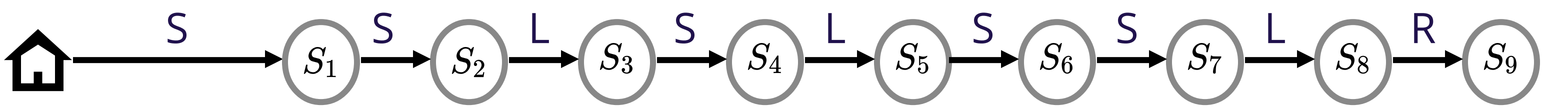


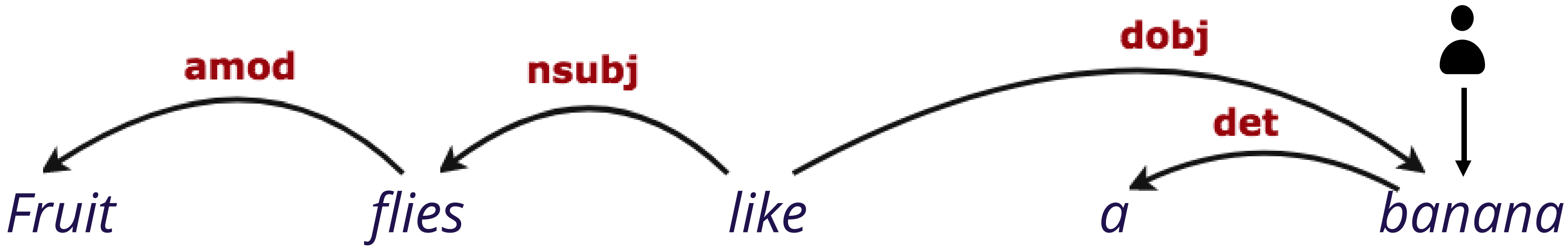
States, Actions

S: shift

L: left-arc

R:right-arc





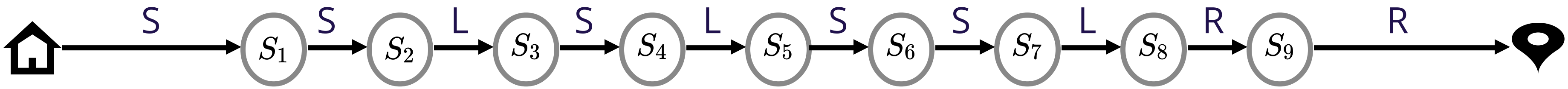
StatNLP.org

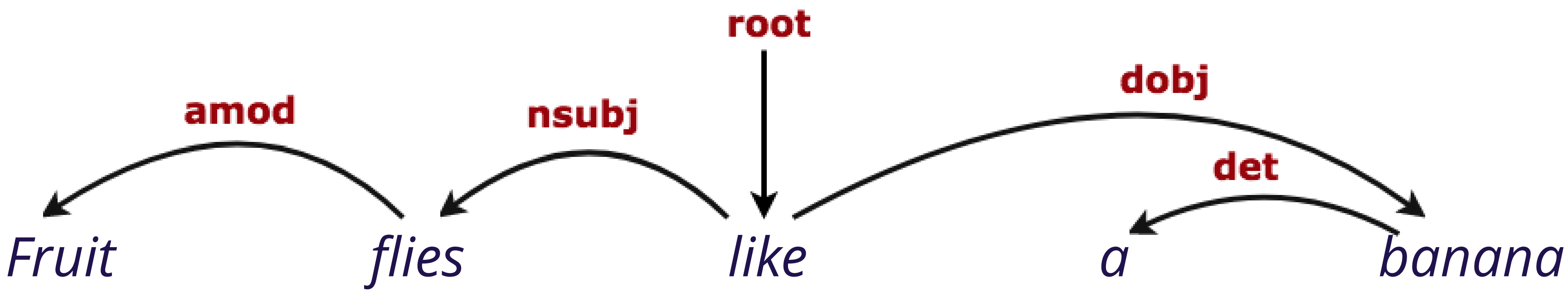
States, Actions

S: shift

L: left-arc

R:right-arc



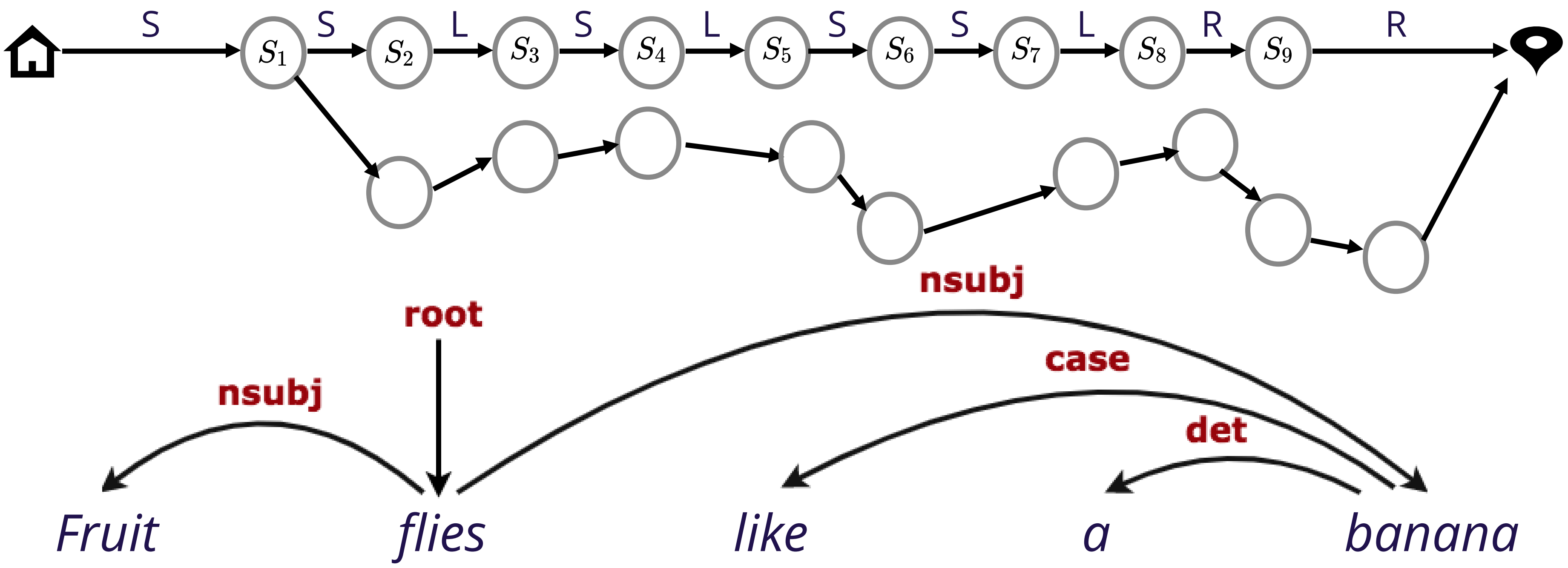


States, Actions

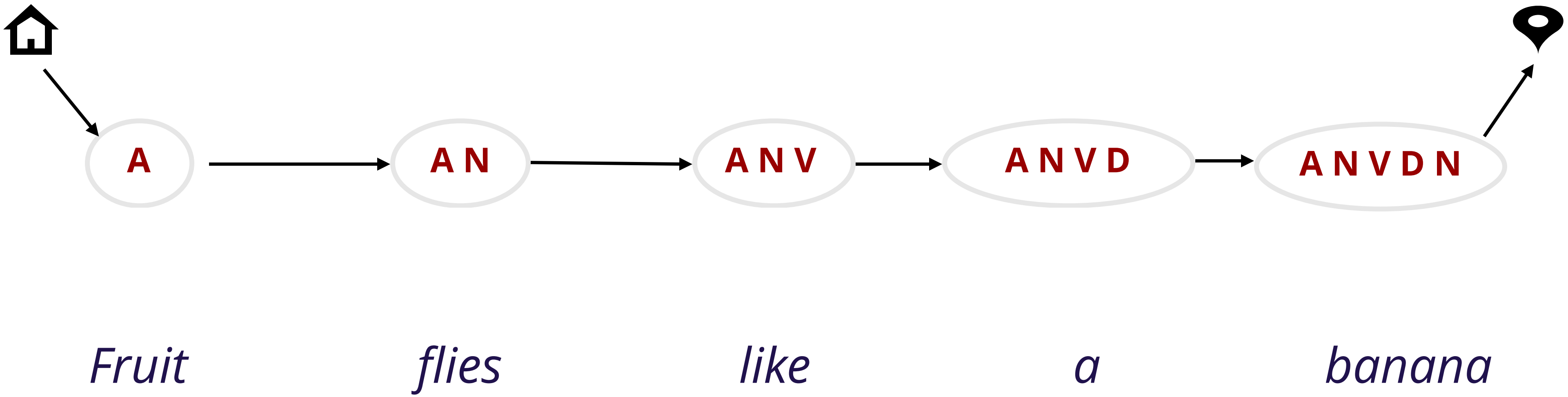
S: shift

L: left-arc

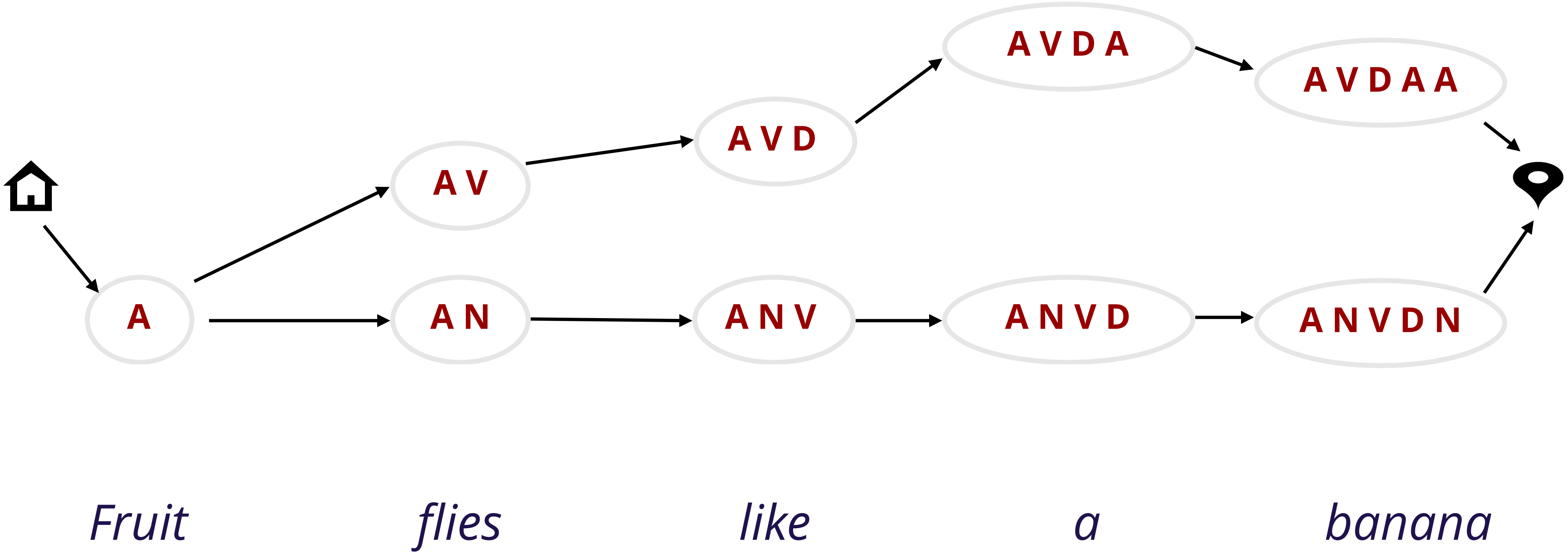
R:right-arc



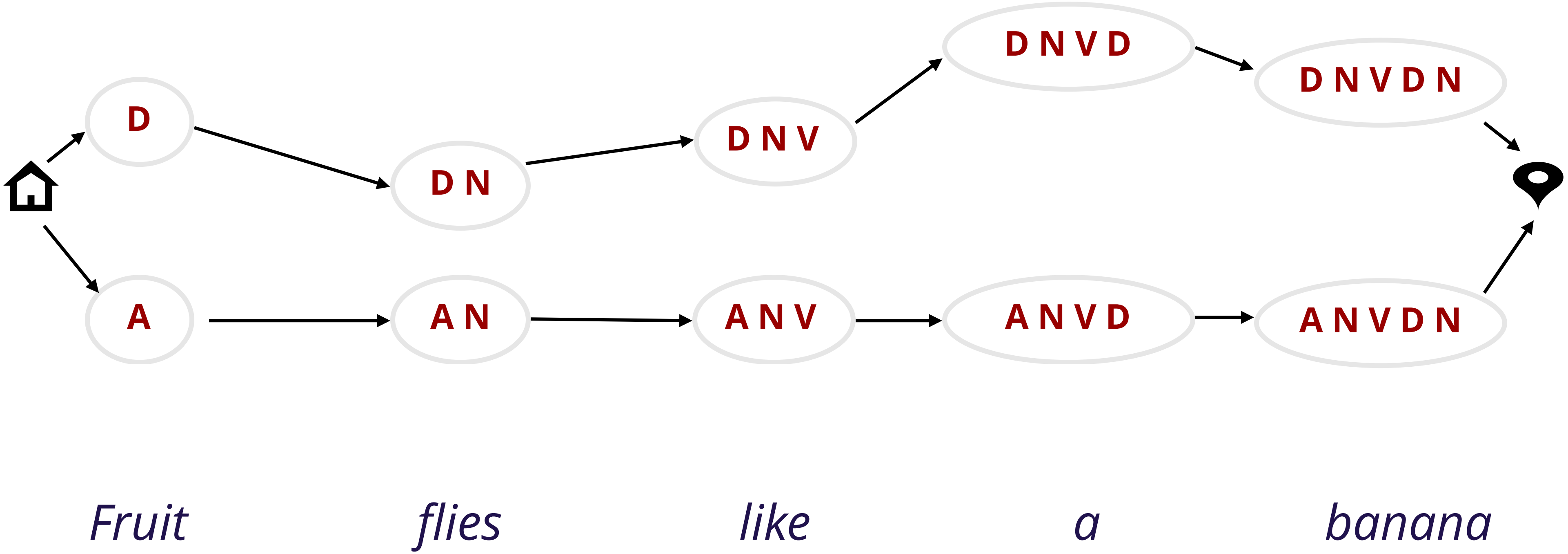
States, Actions, Paths



States, Actions, Paths



Score of a Path

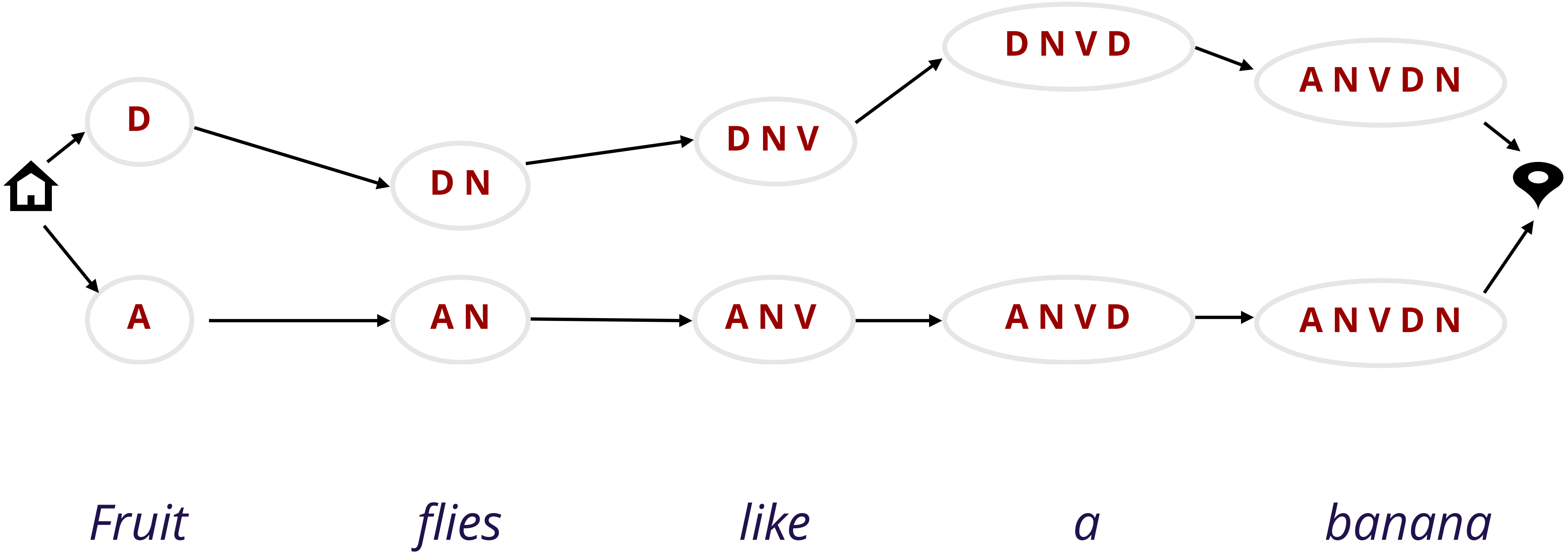


Score of a Path

Score of the path p Score of each edge e

$$S_{\mathbf{w}}(p) = \sum_{e \in p} s_{\mathbf{w}}(e)$$

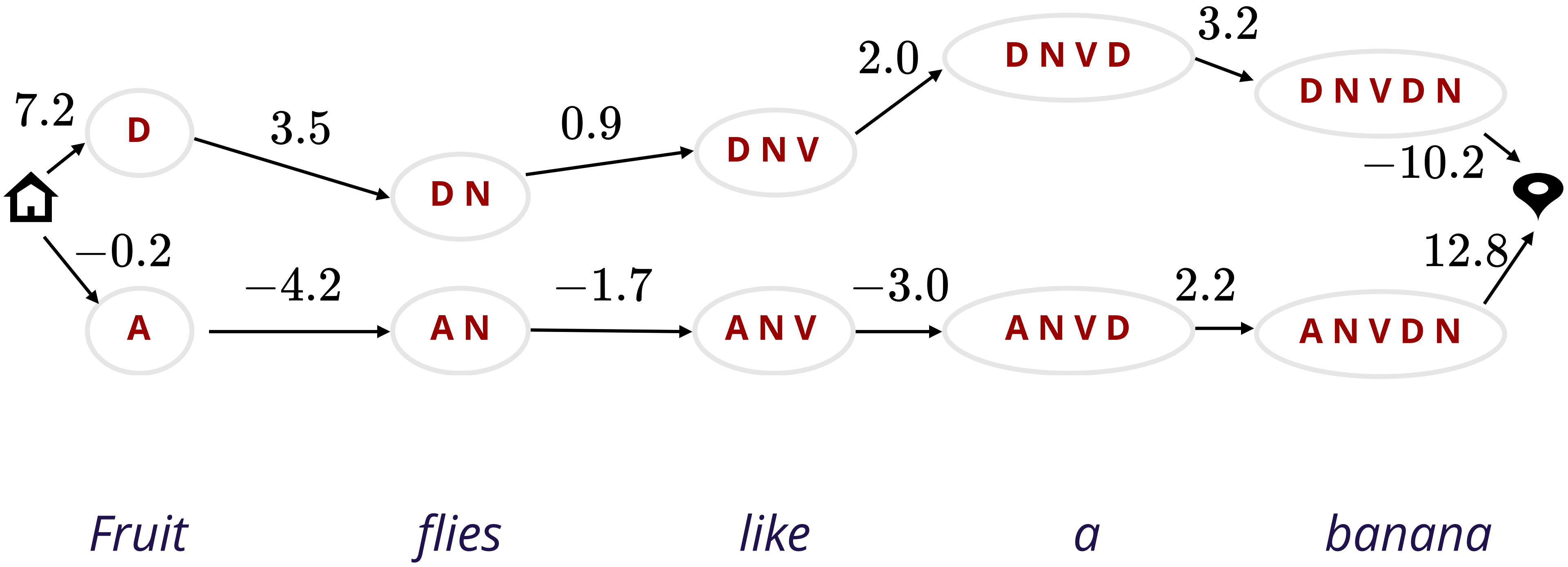
Parameters



Score of a Path

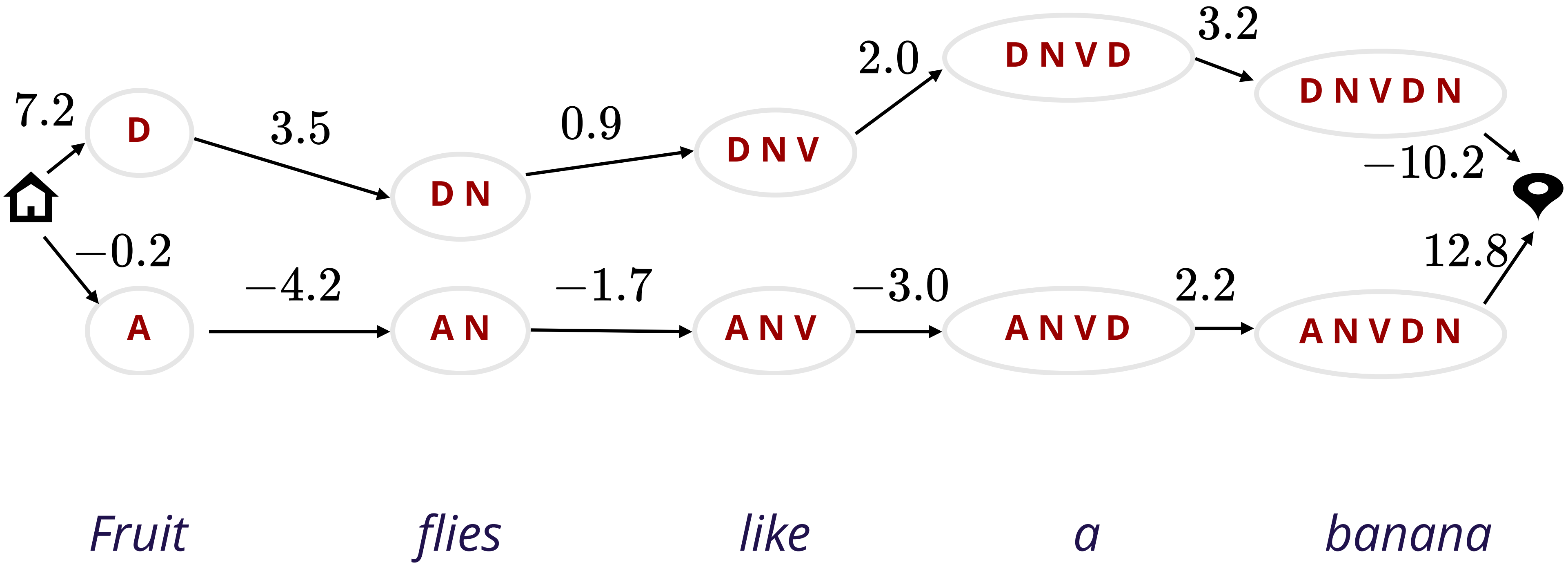
$$S_{\mathbf{w}}(p_1) = 7.2 + 3.5 + 0.9 + 2.0 + 3.2 - 10.2 = 6.6$$

$$S_{\mathbf{w}}(p_2) = -0.2 - 4.2 - 1.7 - 3.0 + 2.2 + 12.8 = 6.9$$



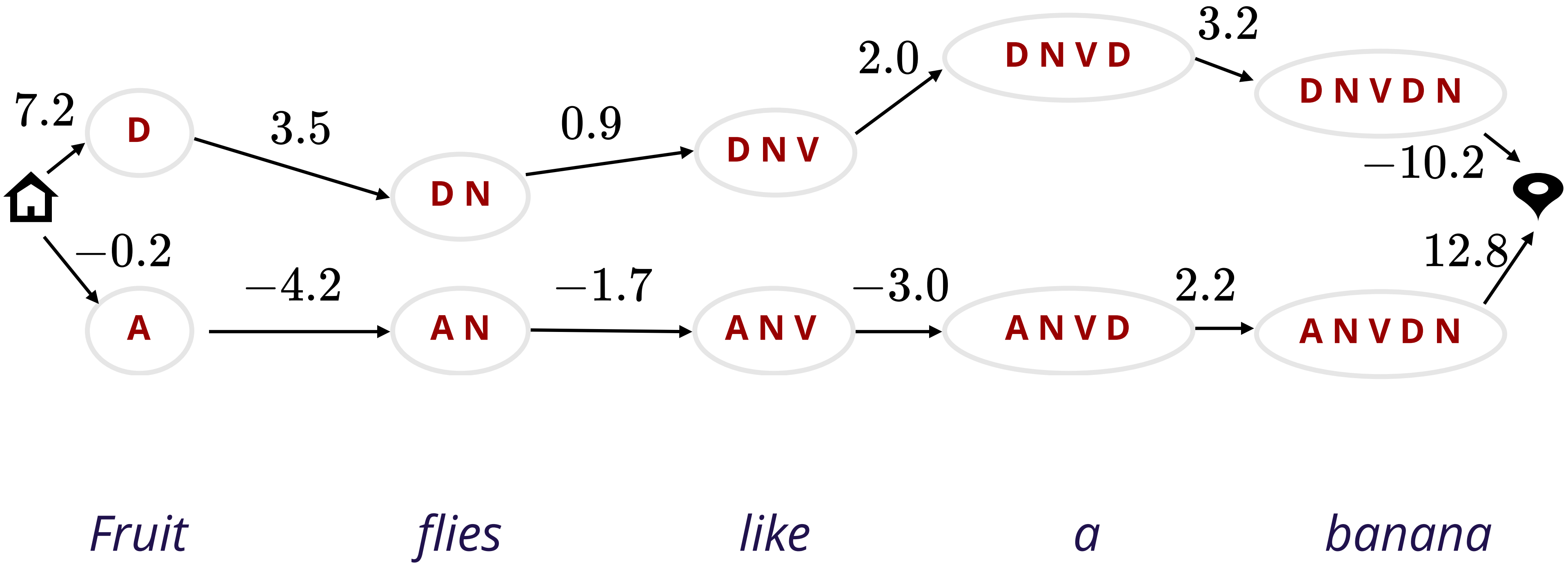
Search

Exhaustive Search
Beam Search
Heuristics Search

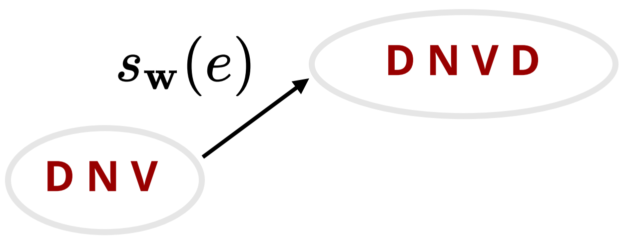


Score of an Edge

$$s_{\mathbf{w}}(e) = \mathbf{w} \cdot \mathbf{f}(e) = \mathbf{w} \cdot \mathbf{f}(x, [s, a])$$



Score of an Edge



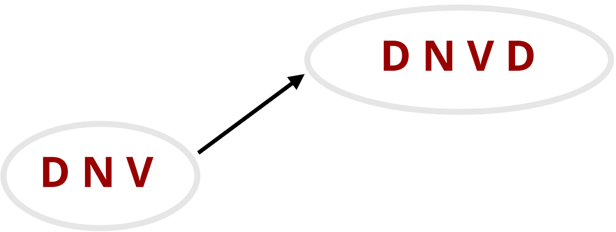
Fruit *flies* *like* *a* *banana*

Score of an Edge

$\mathbf{w}$

$\mathbf{f}(x, [s, a])$

$$s_{\mathbf{w}}(e) = \mathbf{w} \cdot \mathbf{f}(x, [s, a])$$



Fruit *flies* *like* *a* *banana*

Score of an Edge

$$\mathbf{w} = \begin{bmatrix} 1.2 \\ \vdots \\ -3.1 \end{bmatrix} \qquad \mathbf{f}(x, [s, a]) = \mathbf{f}\Big(x, [\text{D N V}, \text{D}]\Big) = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix}$$

$$s_{\mathbf{w}}(e) = \mathbf{w} \cdot \mathbf{f}(x, [s, a])$$

$$\text{D N V} \longrightarrow \text{D N V D}$$

Fruit *flies* *like* *a* *banana*

Score of an Edge

$$\mathbf{w} = \begin{bmatrix} 1.2 \\ \vdots \\ -3.1 \end{bmatrix} \qquad \mathbf{f}(x, [s, a]) = \mathbf{f}\Big(x, [\text{D N V}, \text{D}]\Big) = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix}$$

$$s_{\mathbf{w}}(e) = \mathbf{w} \cdot \mathbf{f}(x, [s, a]) = 2.0$$

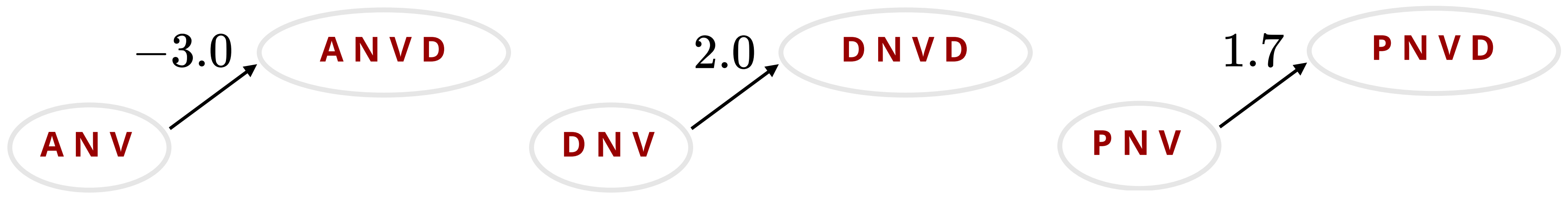
$$\text{D N V} \xrightarrow{2.0} \text{D N V D}$$

Fruit *flies* *like* *a* *banana*

Score of an Edge

$$\mathbf{w} = \begin{bmatrix} 1.2 \\ \vdots \\ -3.1 \end{bmatrix} \qquad \mathbf{f}(x, [s, a]) = \mathbf{f}\Big(x, [\text{D N V}, \text{D}]\Big) = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix}$$

$$s_{\mathbf{w}}(e) = \mathbf{w} \cdot \mathbf{f}(x, [s, a]) = 2.0$$

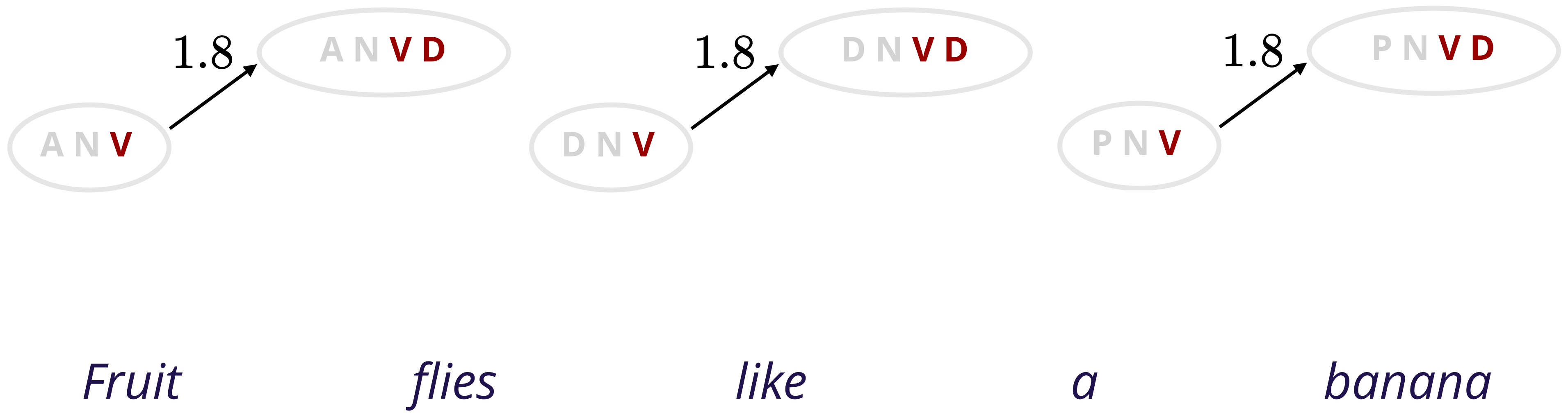


Fruit *flies* *like* *a* *banana*

Score of an Edge

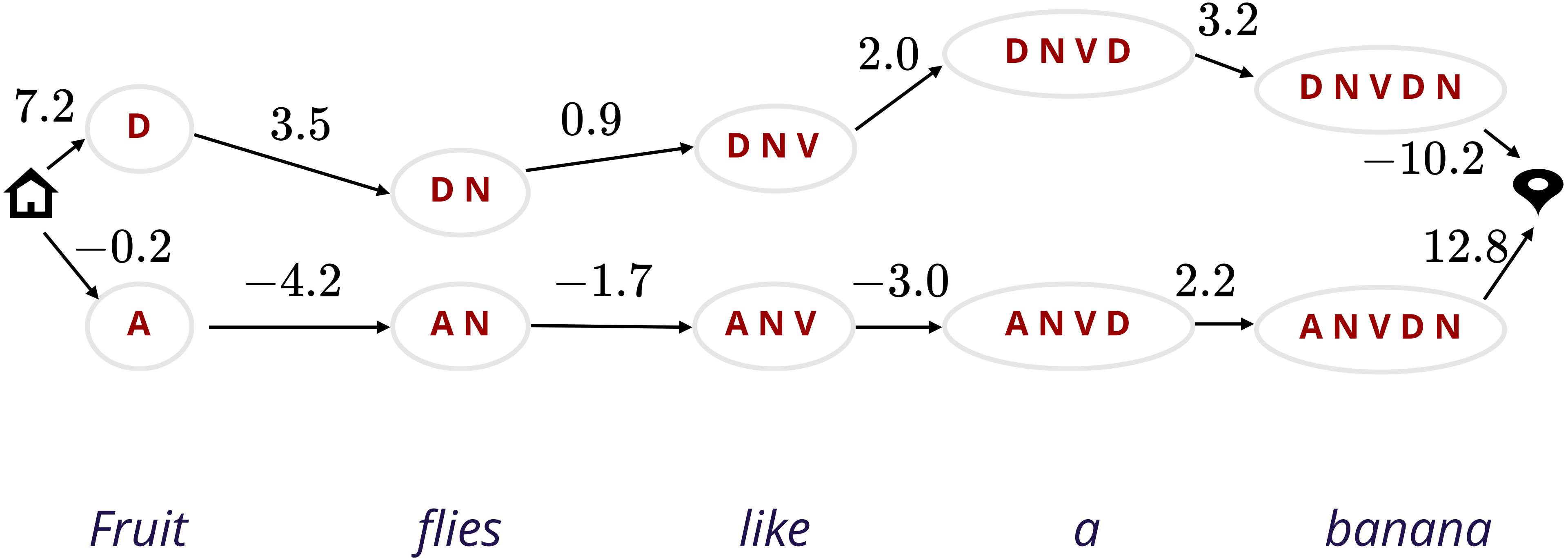
$$\mathbf{w} = \begin{bmatrix} 1.2 \\ \vdots \\ -3.1 \end{bmatrix} \qquad \mathbf{f}(x, [s, a]) = \mathbf{f}\Big(x, [\text{D N V}, \text{D}]\Big) = \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix}$$

$$s_{\mathbf{w}}(e) = \mathbf{w} \cdot \mathbf{f}(x, [s, a]) = 1.8$$



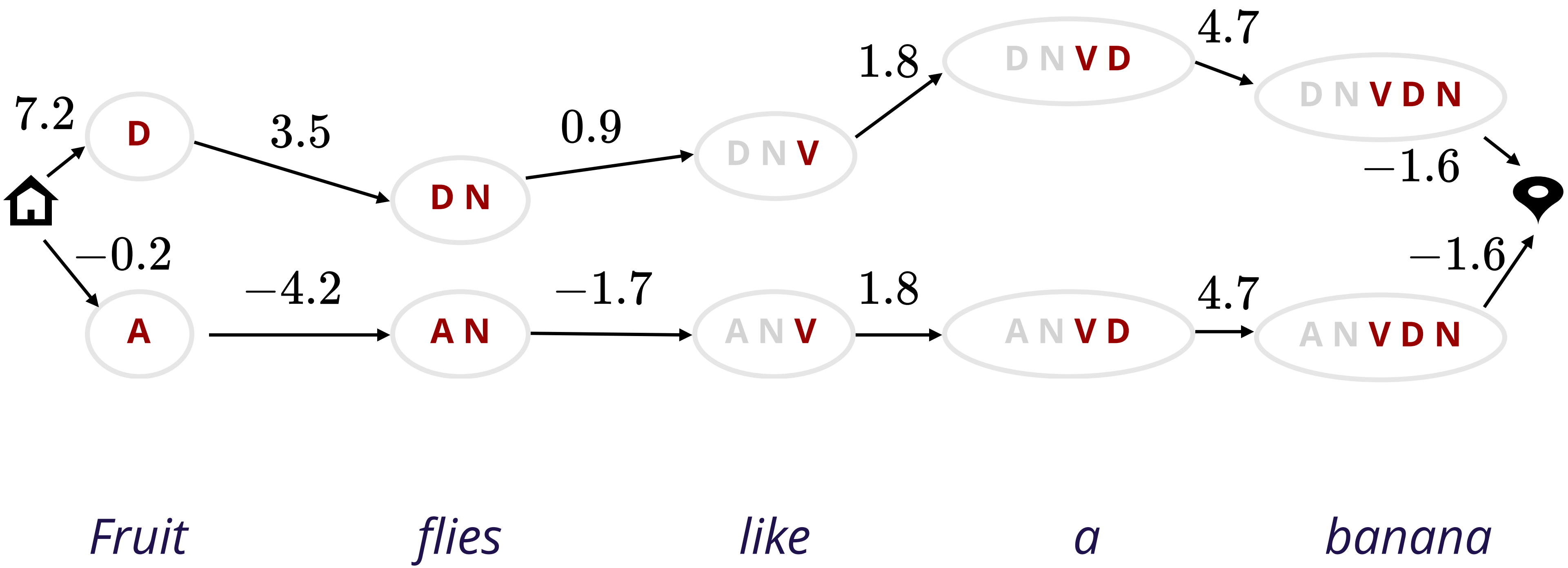
Score of an Edge

$$s_{\mathbf{w}}(e) = \mathbf{w} \cdot \mathbf{f}(e) = \mathbf{w} \cdot \mathbf{f}(x, [s, a])$$



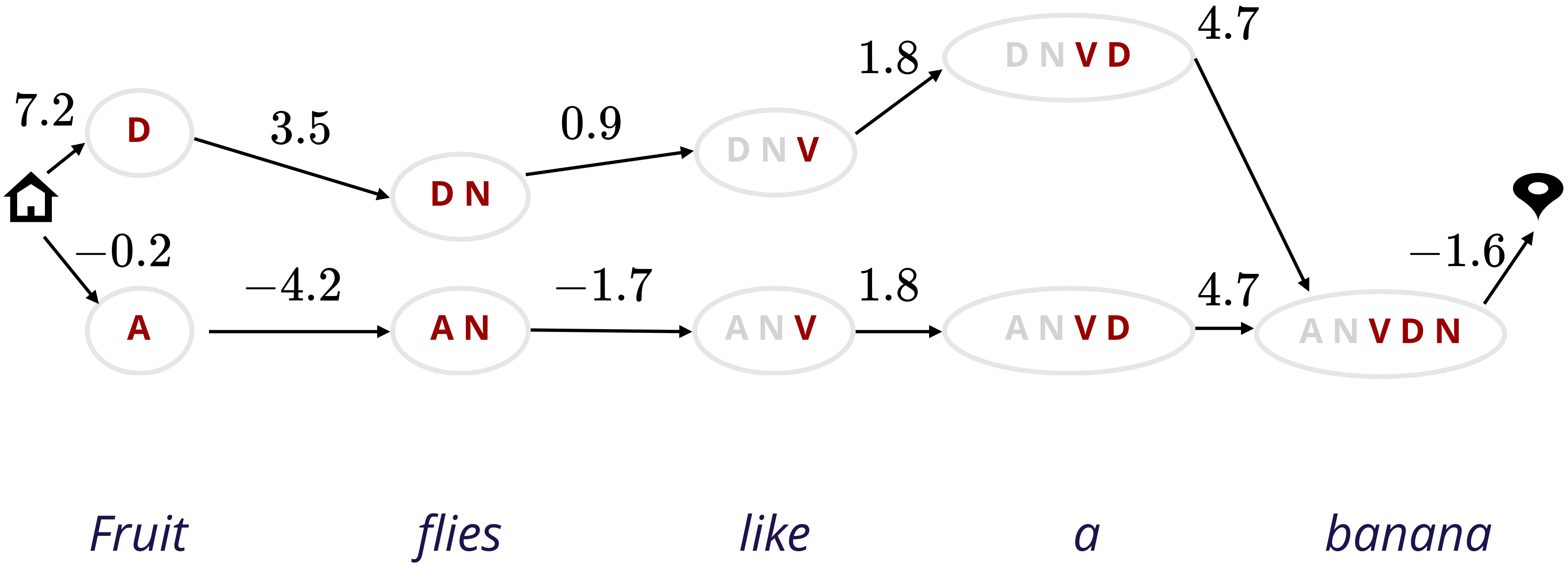
Search Graph

$$s_{\mathbf{w}}(e) = \mathbf{w} \cdot \mathbf{f}(e) = \mathbf{w} \cdot \mathbf{f}(x, [s, a])$$



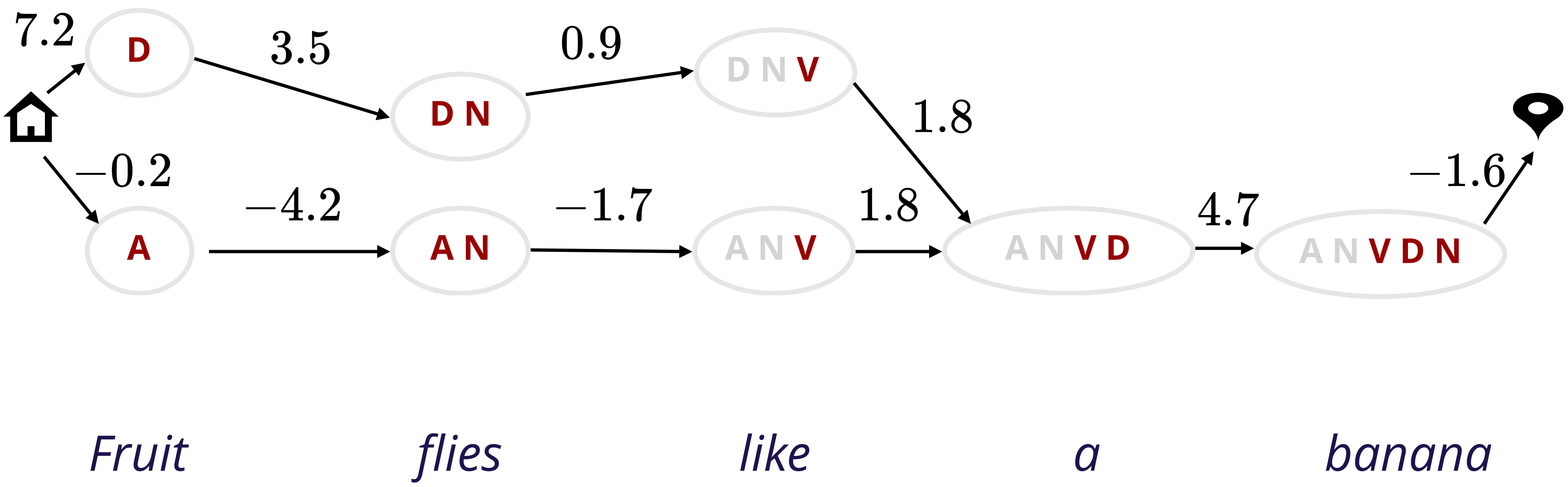
Search Graph

$$s_{\mathbf{w}}(e) = \mathbf{w} \cdot \mathbf{f}(e) = \mathbf{w} \cdot \mathbf{f}(x, [s, a])$$



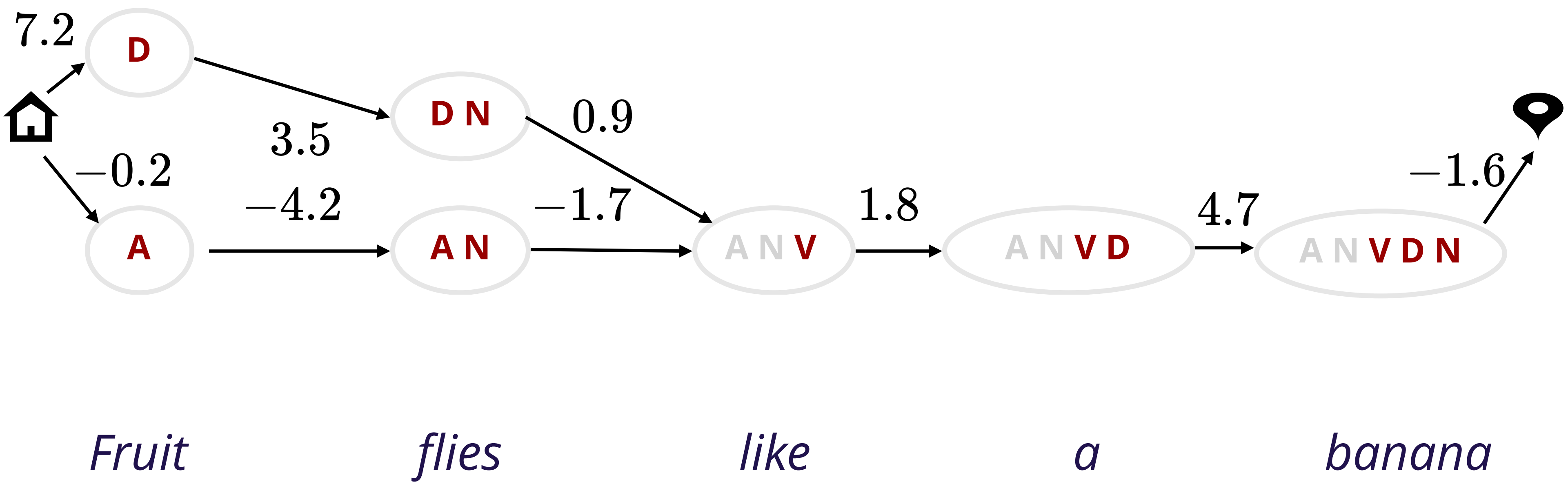
Search Graph

$$s_{\mathbf{w}}(e) = \mathbf{w} \cdot \mathbf{f}(e) = \mathbf{w} \cdot \mathbf{f}(x, [s, a])$$



Search Graph

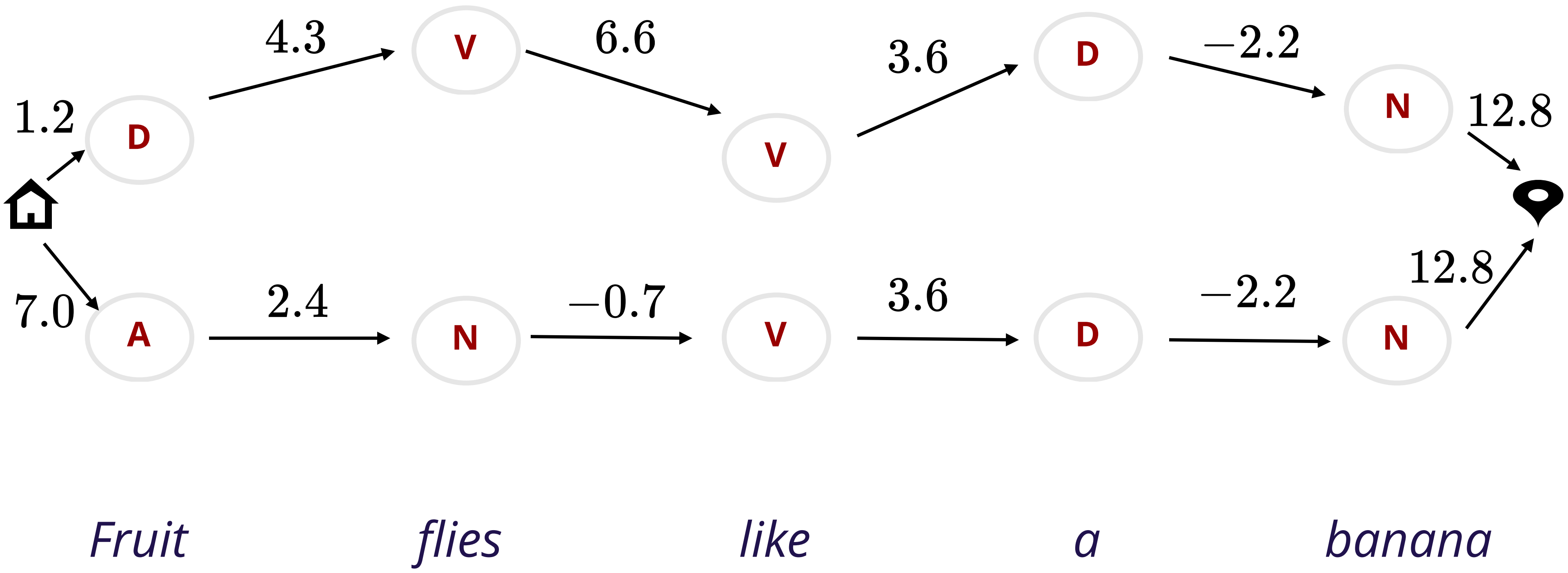
$$s_{\mathbf{w}}(e) = \mathbf{w} \cdot \mathbf{f}(e) = \mathbf{w} \cdot \mathbf{f}(x, [s, a])$$



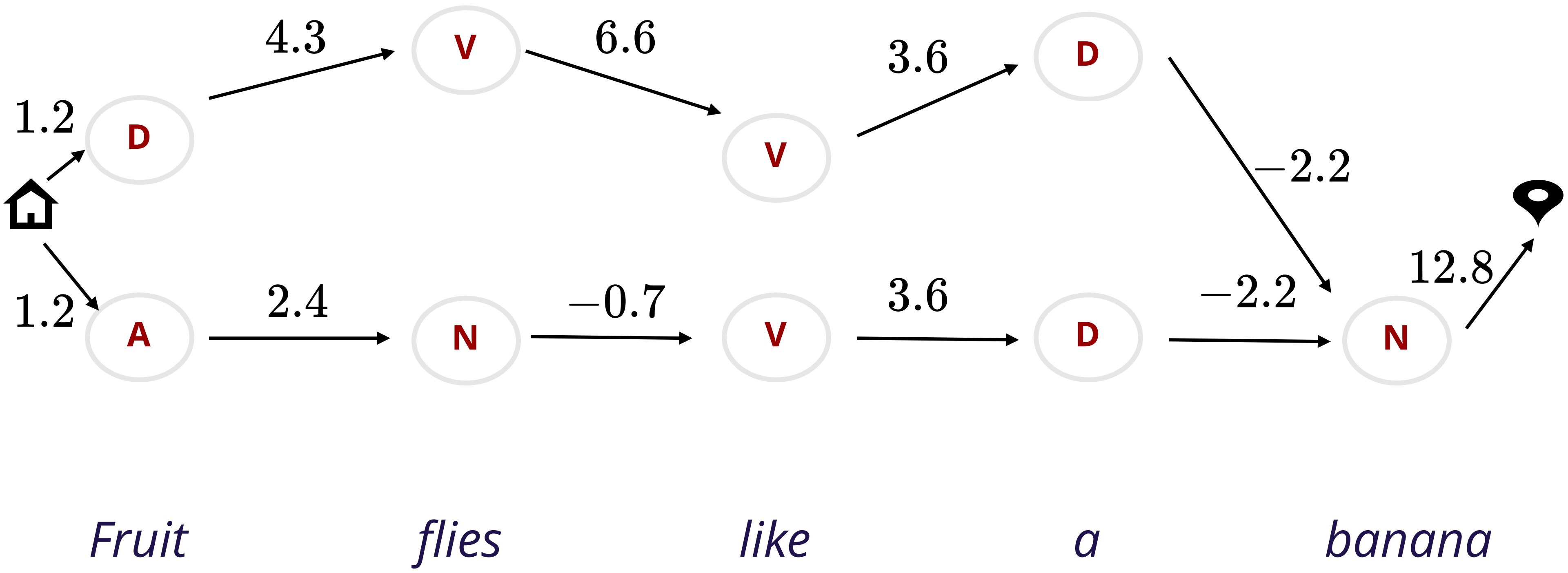
Huang, L., Sagae, K. (2010). Dynamic programming for linear-time incremental parsing.

Search Graph

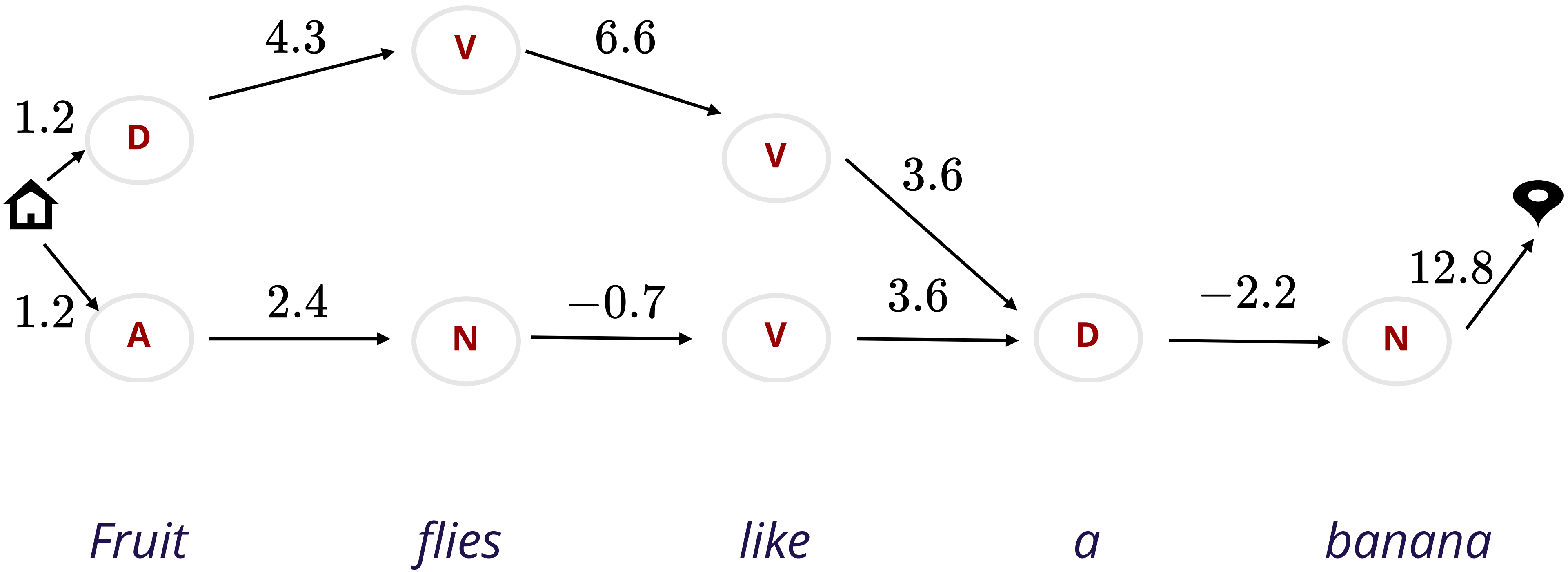
$$s_{\mathbf{w}}(e) = \mathbf{w} \cdot \mathbf{f}(e) = \mathbf{w} \cdot \mathbf{f}(x, [s, a])$$



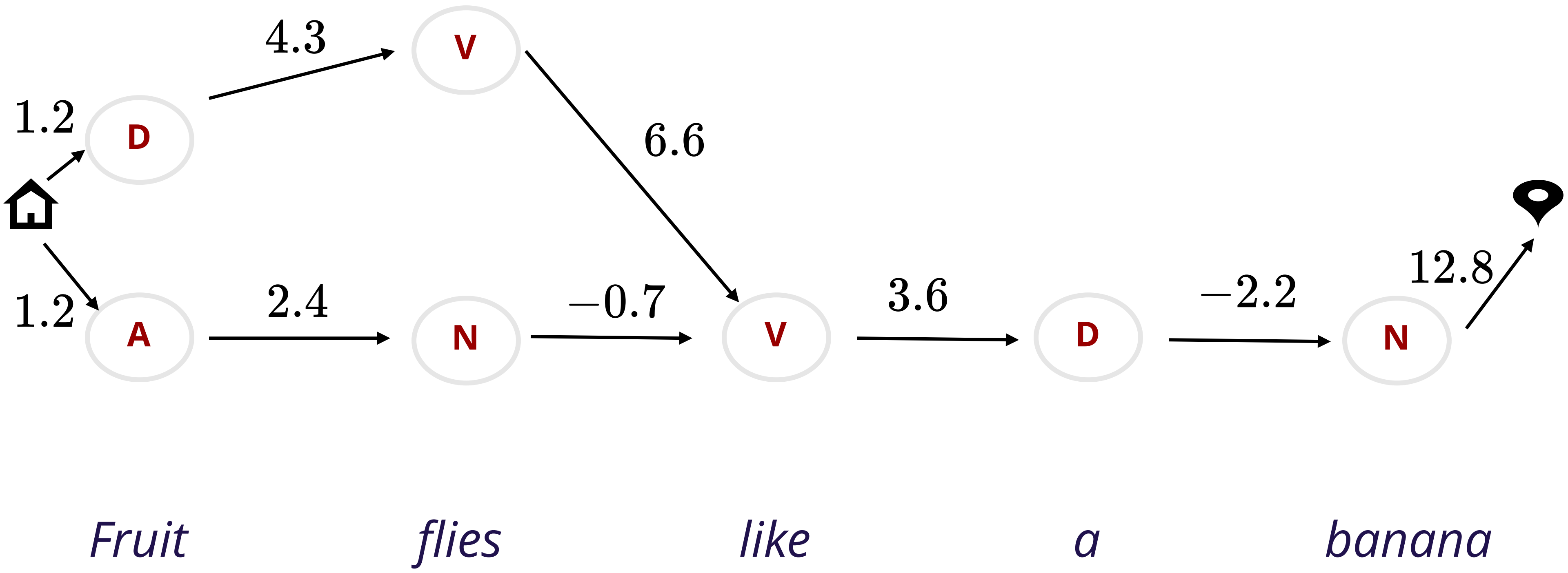
Search Graph



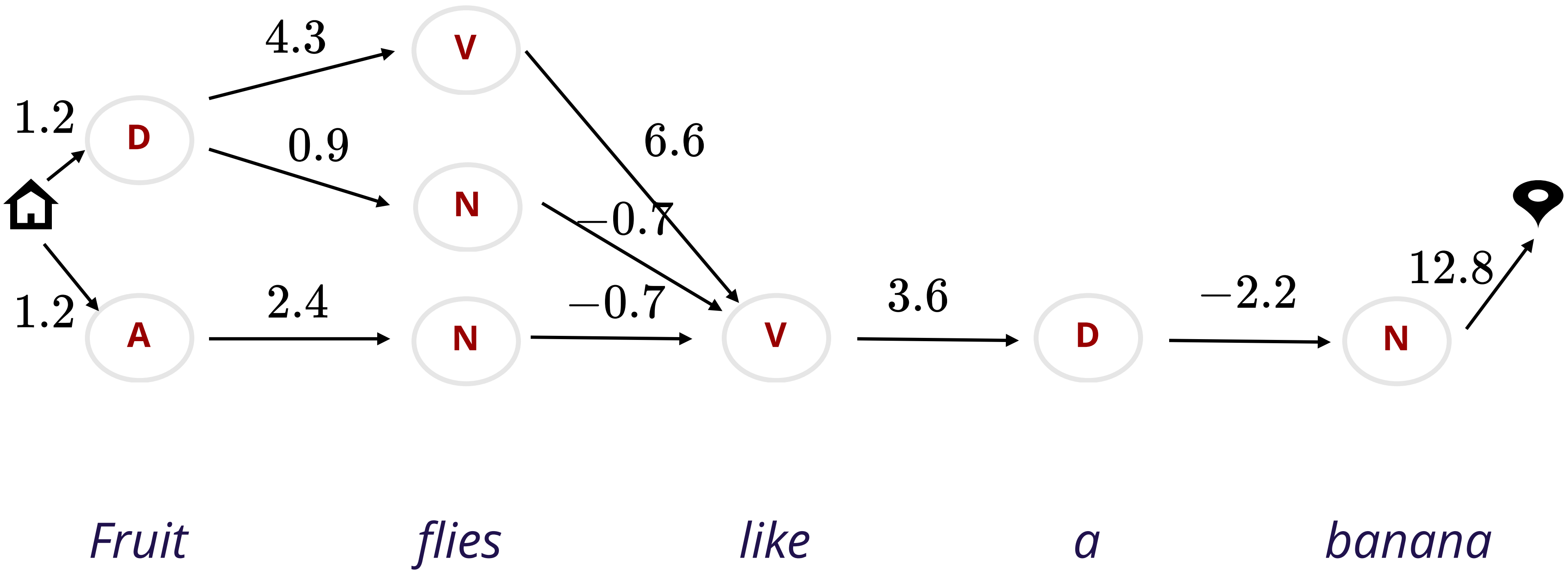
Search Graph



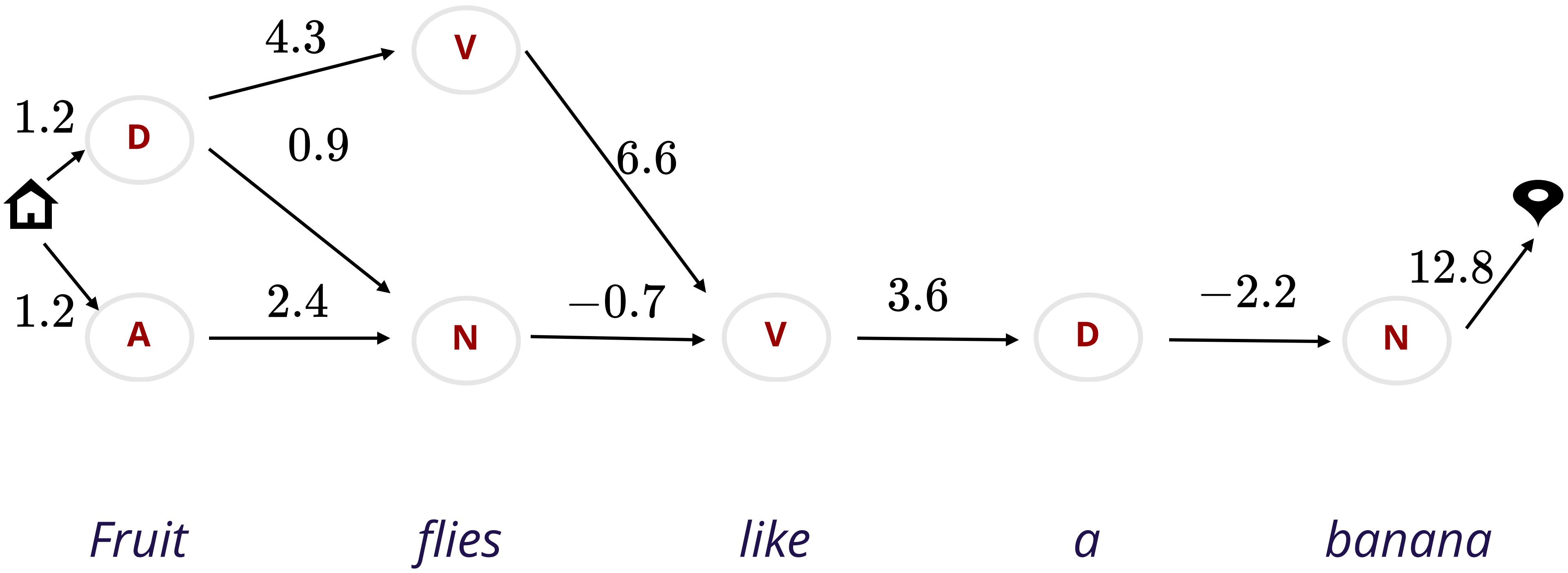
Search Graph



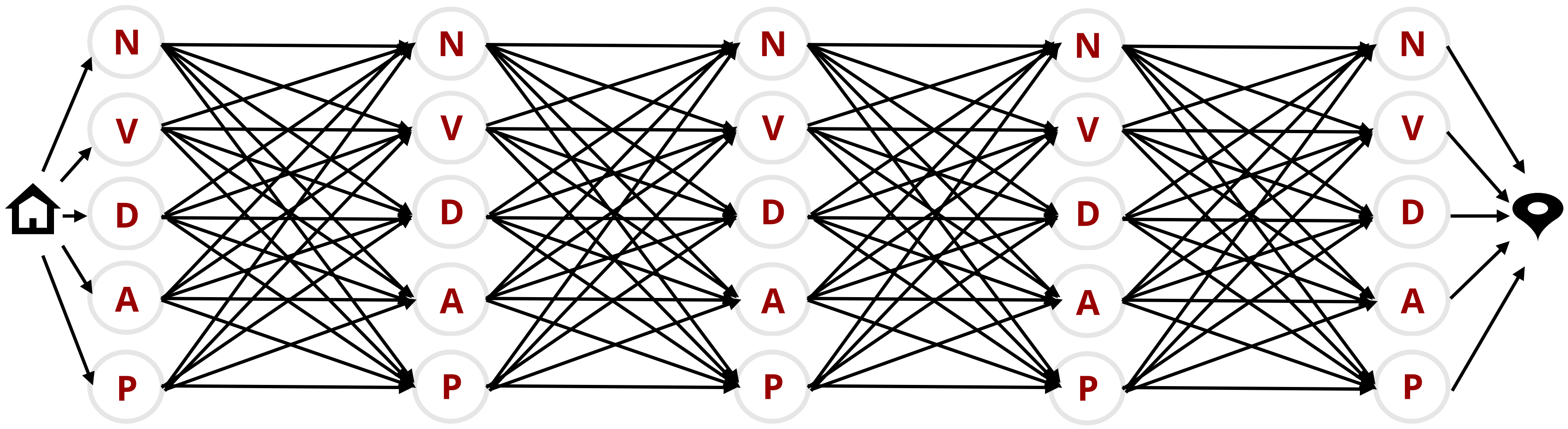
Search Graph



Search Graph



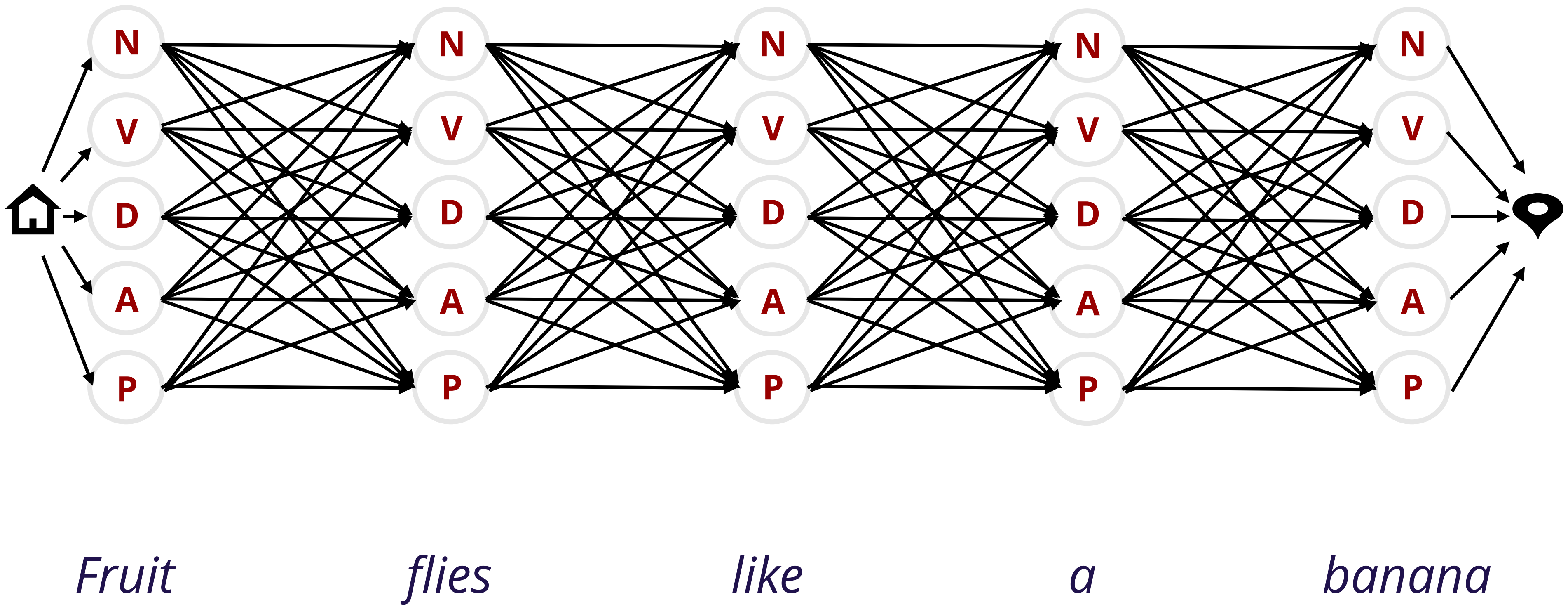
A Compact Search Graph



Fruit *flies* *like* *a* *banana*

A Compact Search Graph

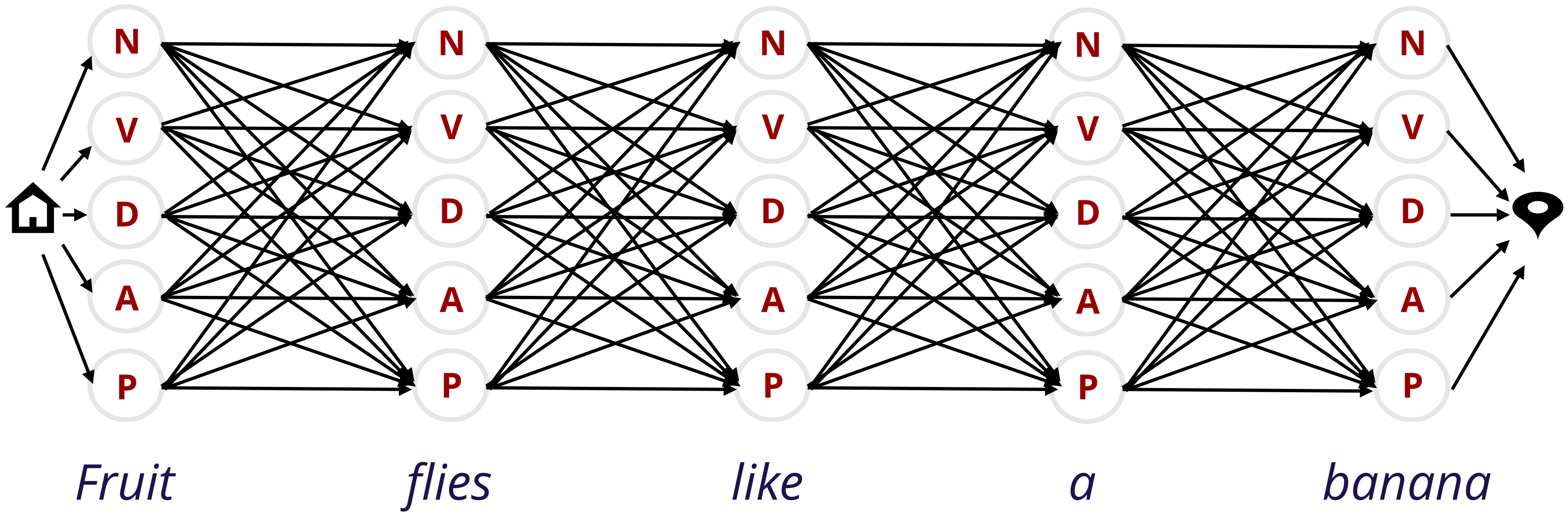
A representation that contains exponentially many directed paths



Decoding

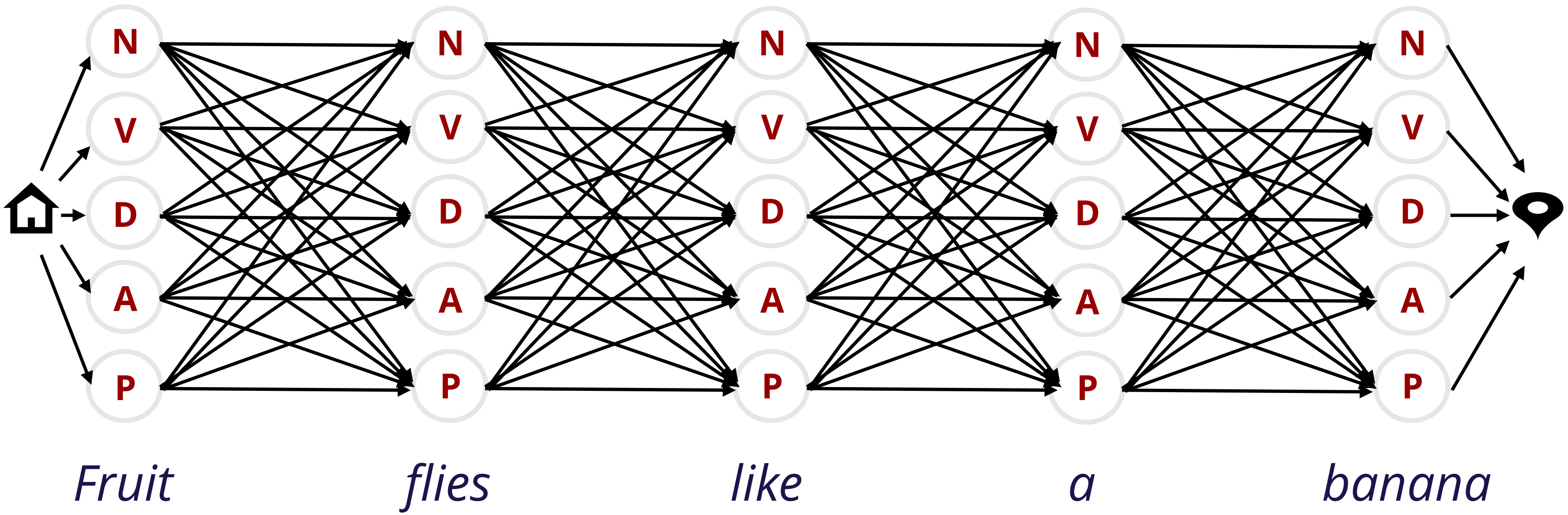
Decoding

Given	Find
$x, \mathbf{w}$	y



The Search Problem

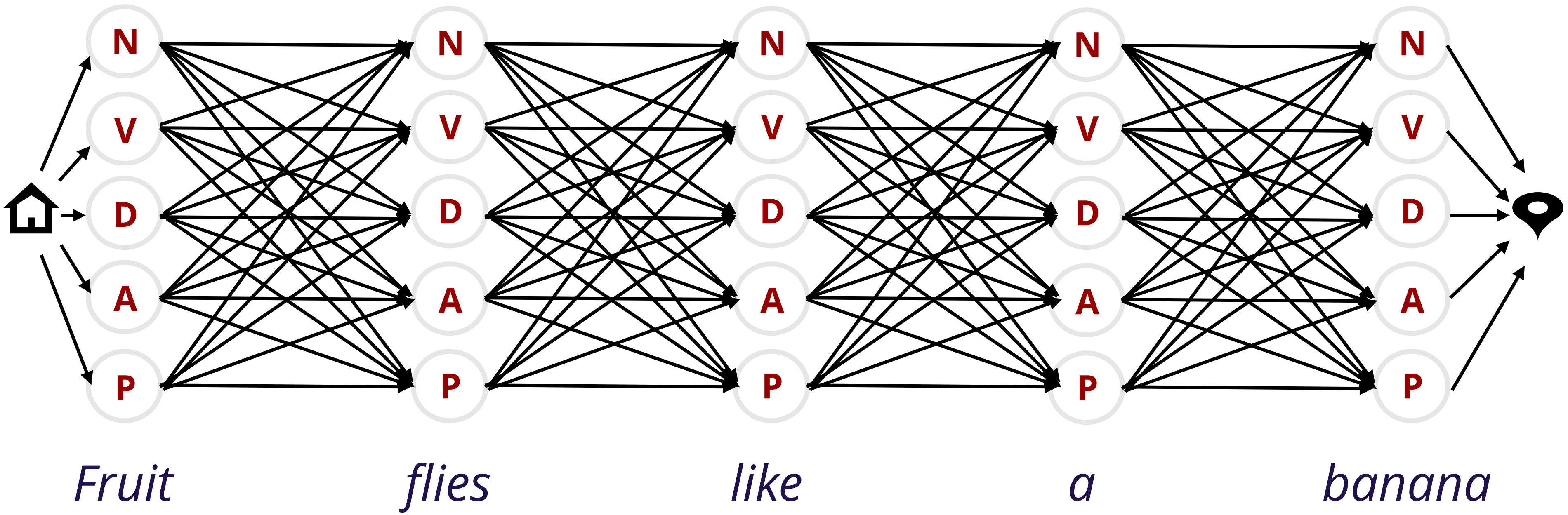
$$\max_y \mathbf{w} \cdot \mathbf{f}(x, y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x, [y^j, y^{j+1}])$$



The Search Problem

$$\max_y \mathbf{w} \cdot \mathbf{f}(x, y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x, [y^j, y^{j+1}])$$

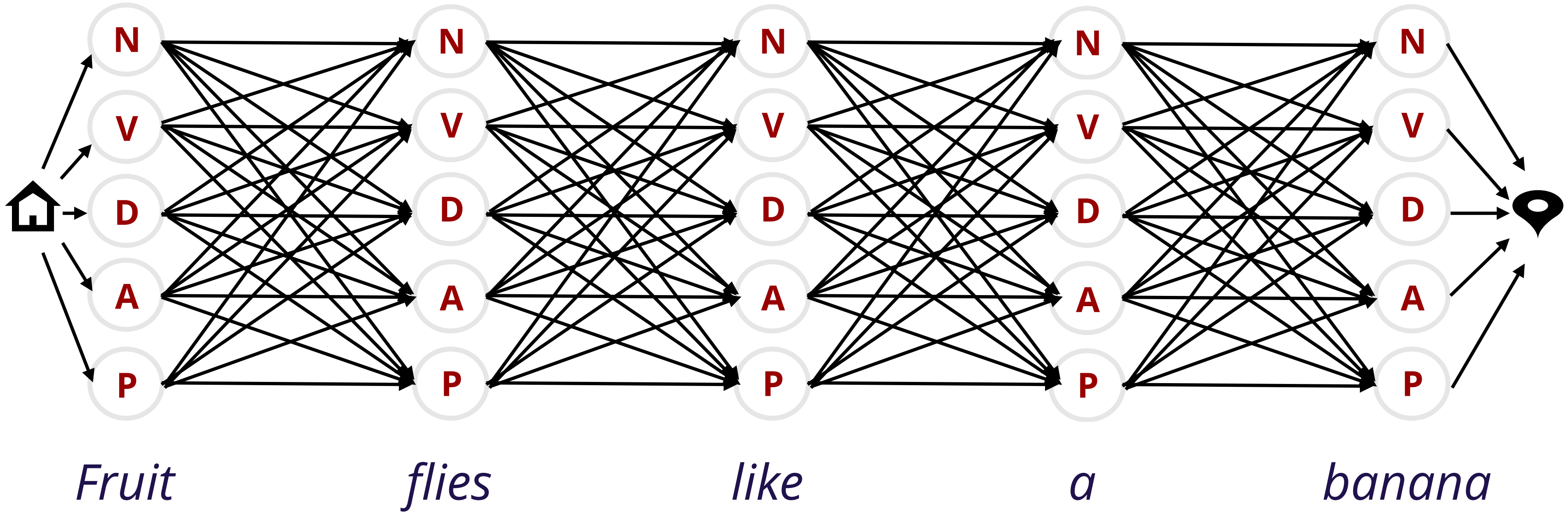
$$\arg\max_y \mathbf{w} \cdot \mathbf{f}(x, y)$$



Viterbi

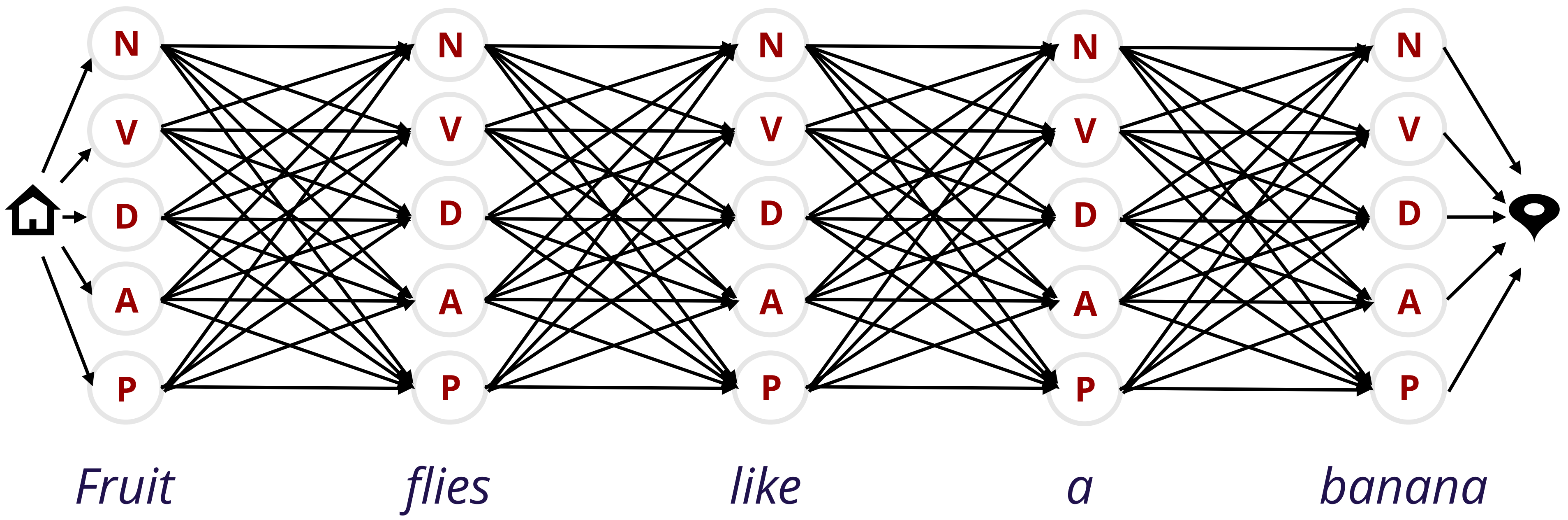
$$\max_y \mathbf{w} \cdot \mathbf{f}(x, y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x, [y^j, y^{j+1}])$$

$$\arg\max_y \mathbf{w} \cdot \mathbf{f}(x, y)$$



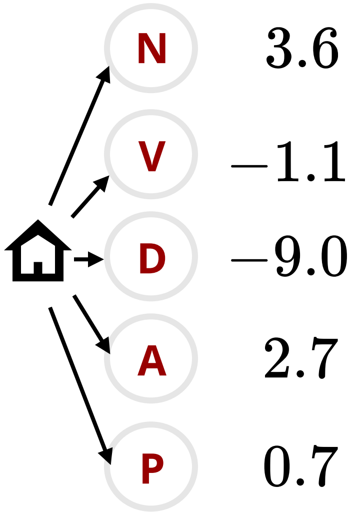
Viterbi

$$\max_y \mathbf{w} \cdot \mathbf{f}(x, y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x, [y^j, y^{j+1}])$$



Viterbi

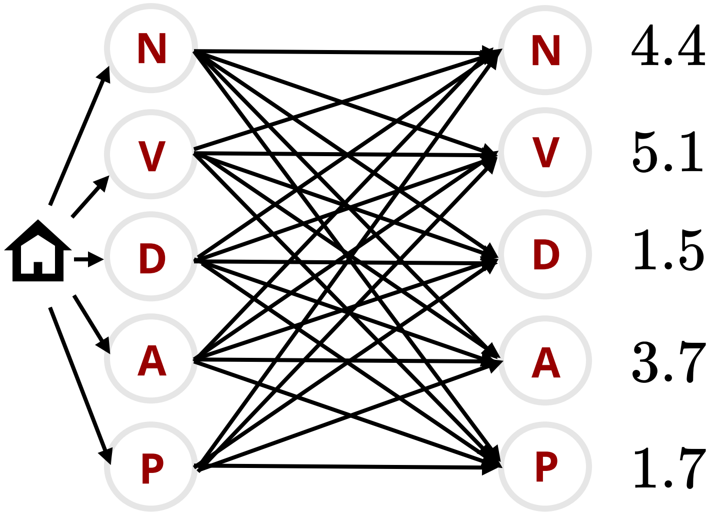
$$\max_y \mathbf{w} \cdot \mathbf{f}(x, y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x, [y^j, y^{j+1}])$$



Fruit *flies* *like* *a* *banana*

Viterbi

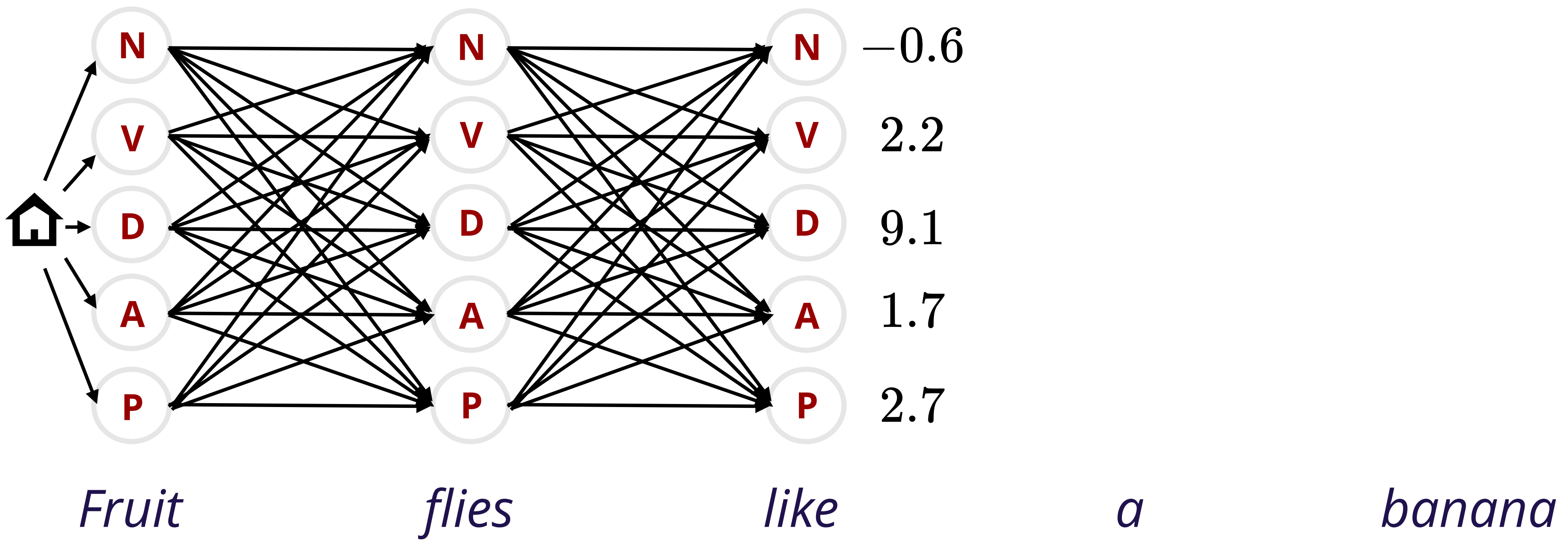
$$\max_y \mathbf{w} \cdot \mathbf{f}(x, y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x, [y^j, y^{j+1}])$$



Fruit *flies* *like* *a* *banana*

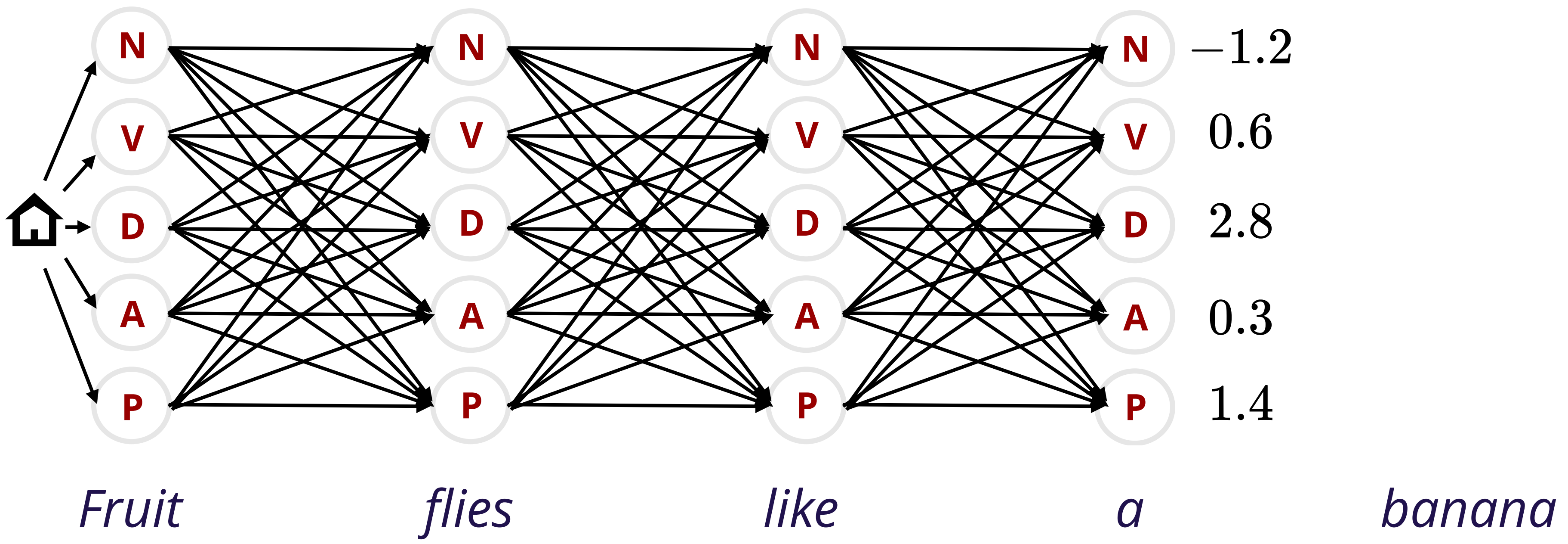
Viterbi

$$\max_y \mathbf{w} \cdot \mathbf{f}(x, y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x, [y^j, y^{j+1}])$$



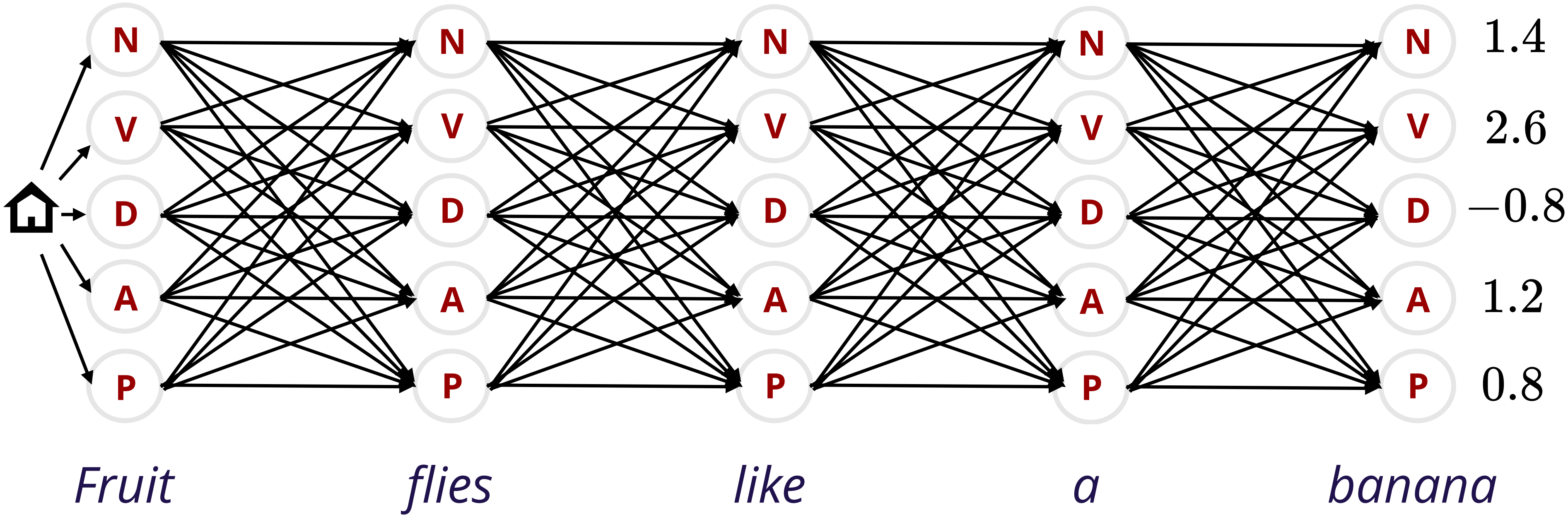
Viterbi

$$\max_y \mathbf{w} \cdot \mathbf{f}(x, y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x, [y^j, y^{j+1}])$$



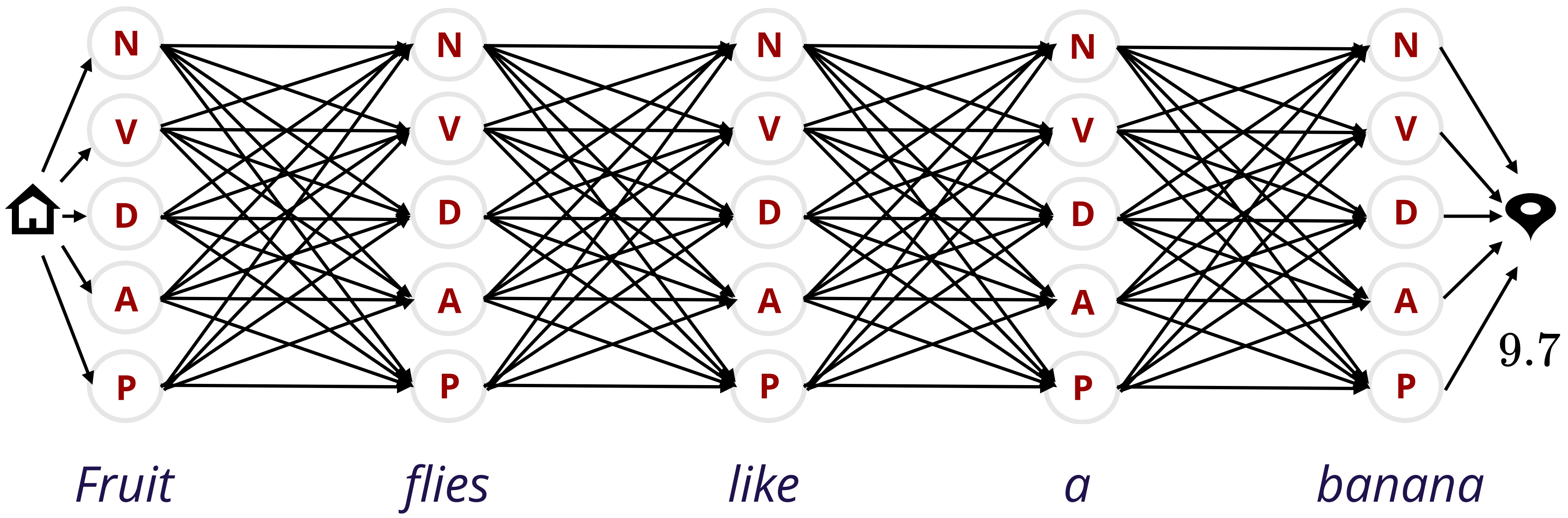
Viterbi

$$\max_y \mathbf{w} \cdot \mathbf{f}(x, y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x, [y^j, y^{j+1}])$$



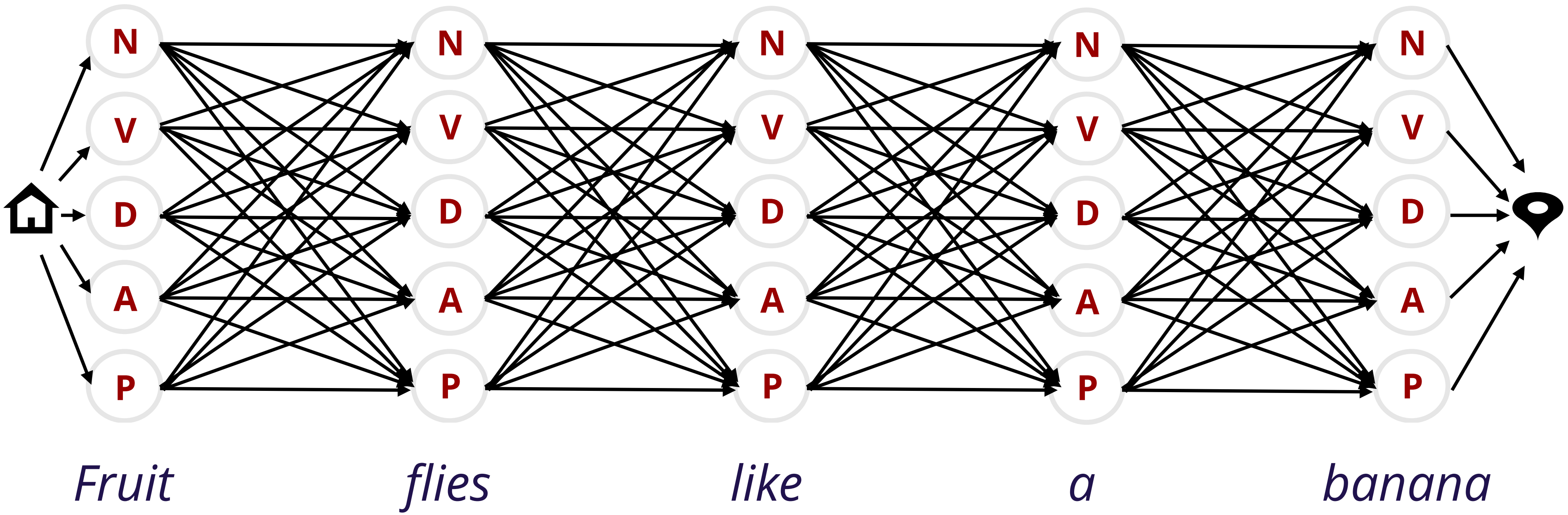
Viterbi

$$\max_y \mathbf{w} \cdot \mathbf{f}(x, y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x, [y^j, y^{j+1}])$$



Viterbi

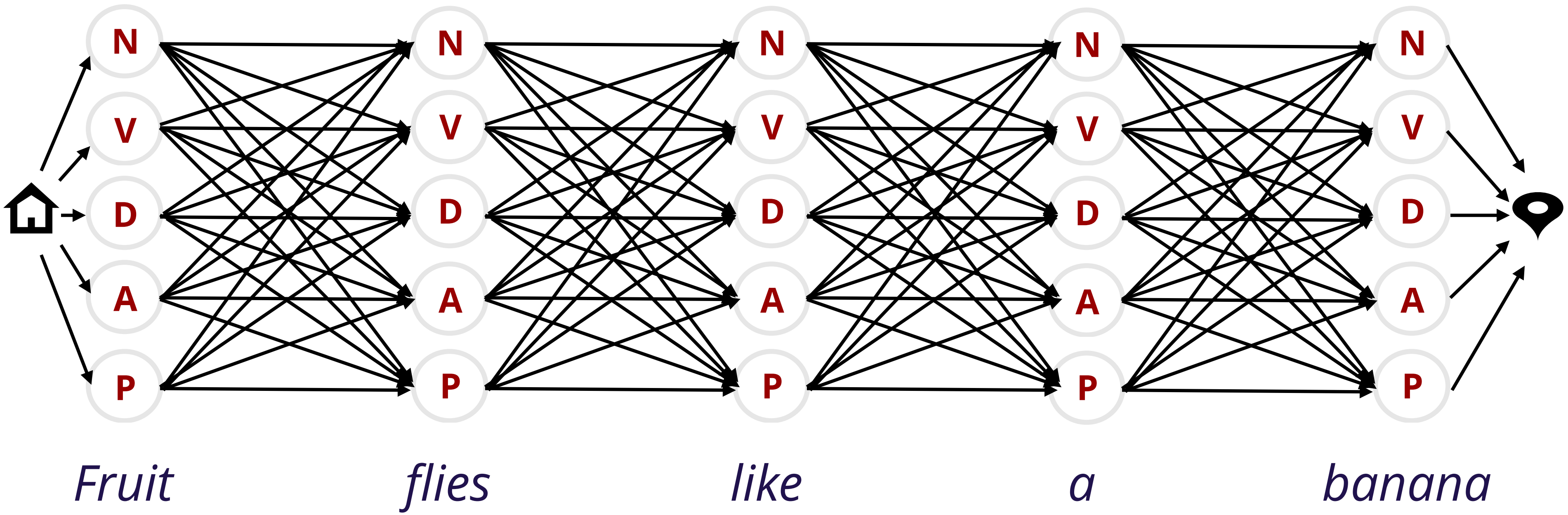
$$\max_y \mathbf{w} \cdot \mathbf{f}(x, y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x, [y^j, y^{j+1}]) = 9.7$$



Viterbi

$$\max_y \mathbf{w} \cdot \mathbf{f}(x, y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x, [y^j, y^{j+1}]) = 9.7$$

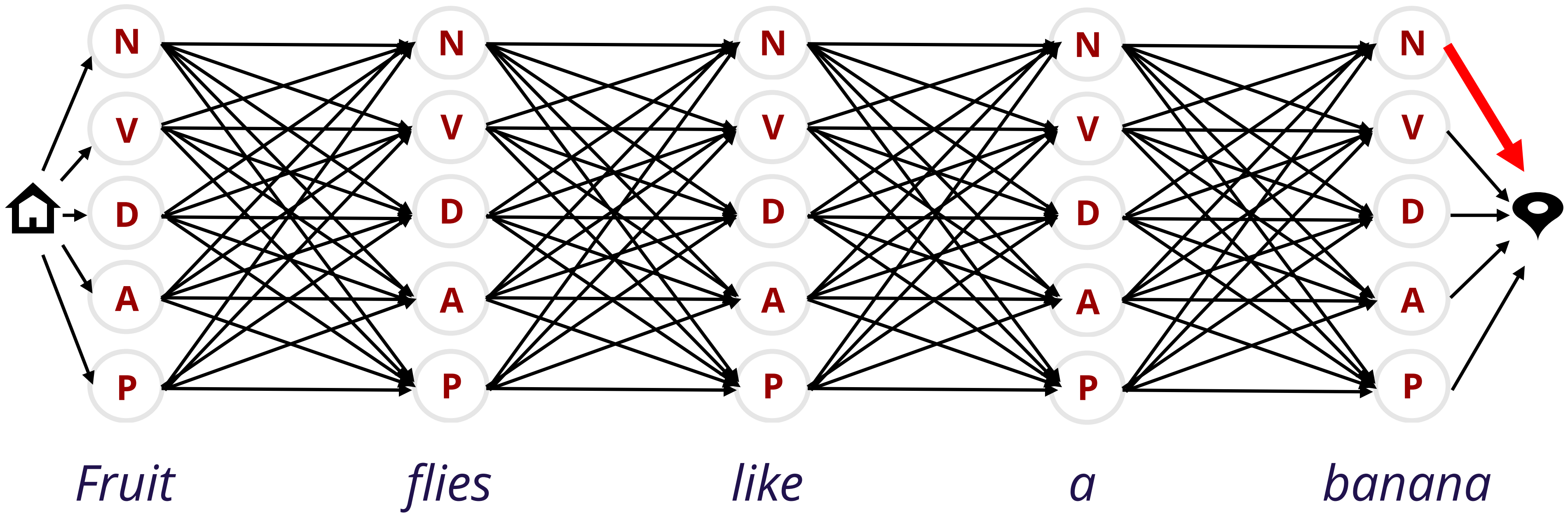
$$\arg\max_y \mathbf{w} \cdot \mathbf{f}(x, y)$$



Viterbi

$$\max_y \mathbf{w} \cdot \mathbf{f}(x, y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x, [y^j, y^{j+1}]) = 9.7$$

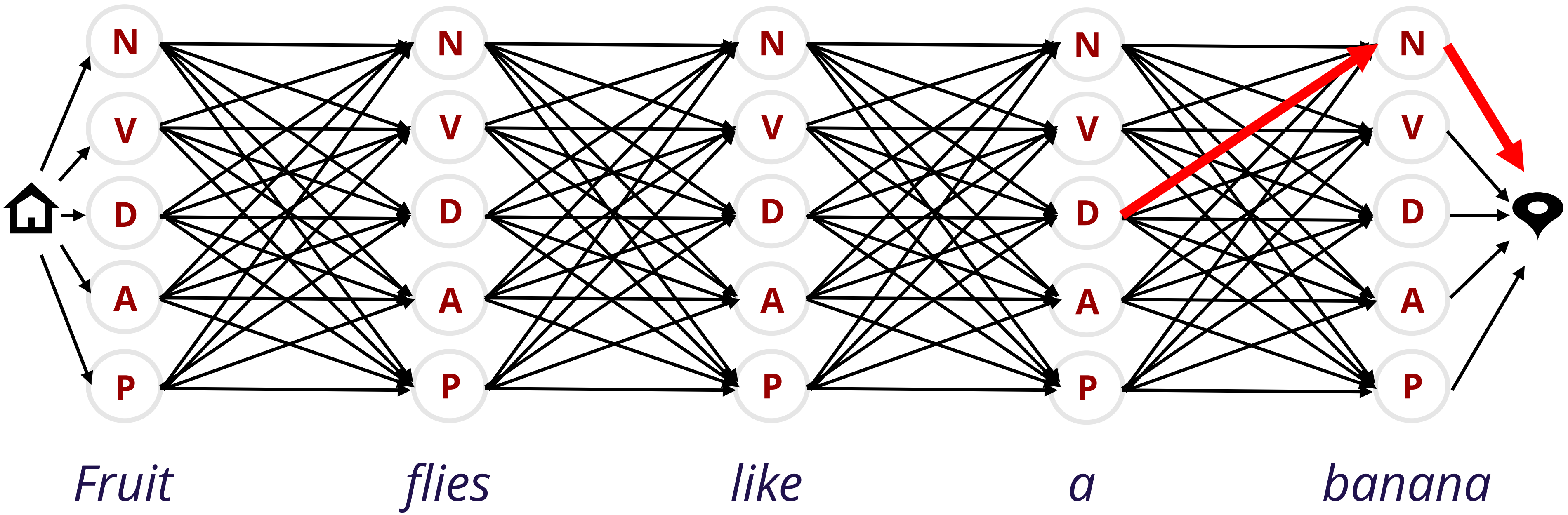
$$\arg\max_y \mathbf{w} \cdot \mathbf{f}(x, y)$$



Viterbi

$$\max_y \mathbf{w} \cdot \mathbf{f}(x, y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x, [y^j, y^{j+1}]) = 9.7$$

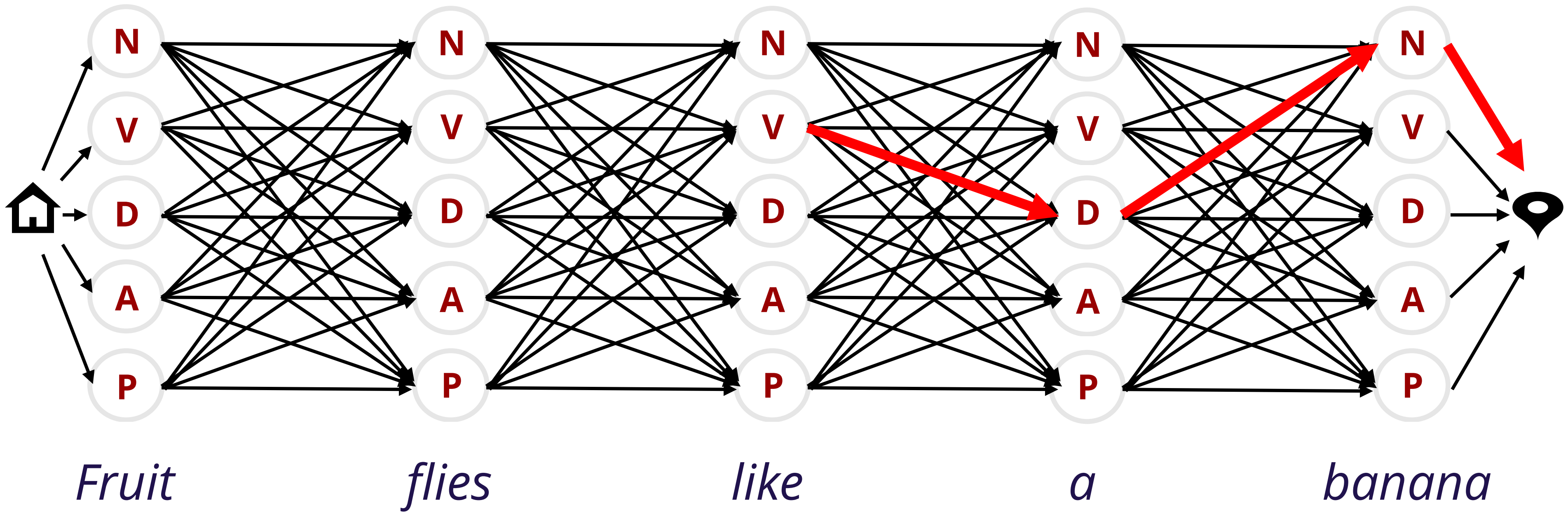
$$\arg\max_y \mathbf{w} \cdot \mathbf{f}(x, y)$$



Viterbi

$$\max_y \mathbf{w} \cdot \mathbf{f}(x, y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x, [y^j, y^{j+1}]) = 9.7$$

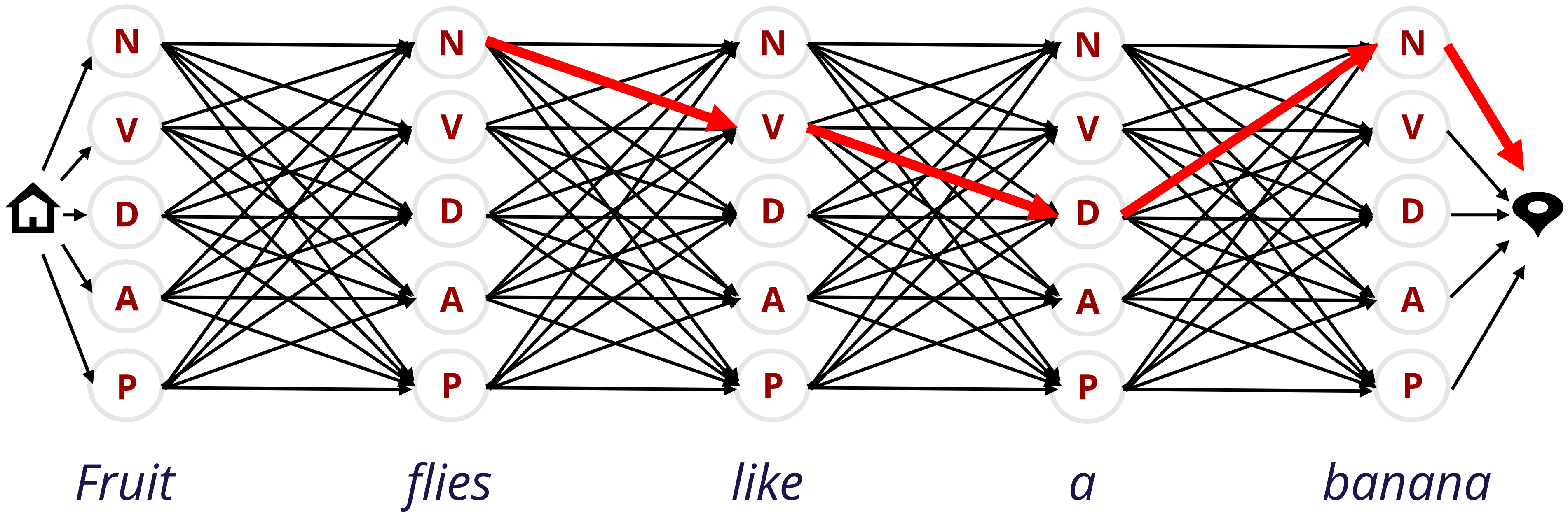
$$\arg\max_y \mathbf{w} \cdot \mathbf{f}(x, y)$$



Viterbi

$$\max_y \mathbf{w} \cdot \mathbf{f}(x, y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x, [y^j, y^{j+1}]) = 9.7$$

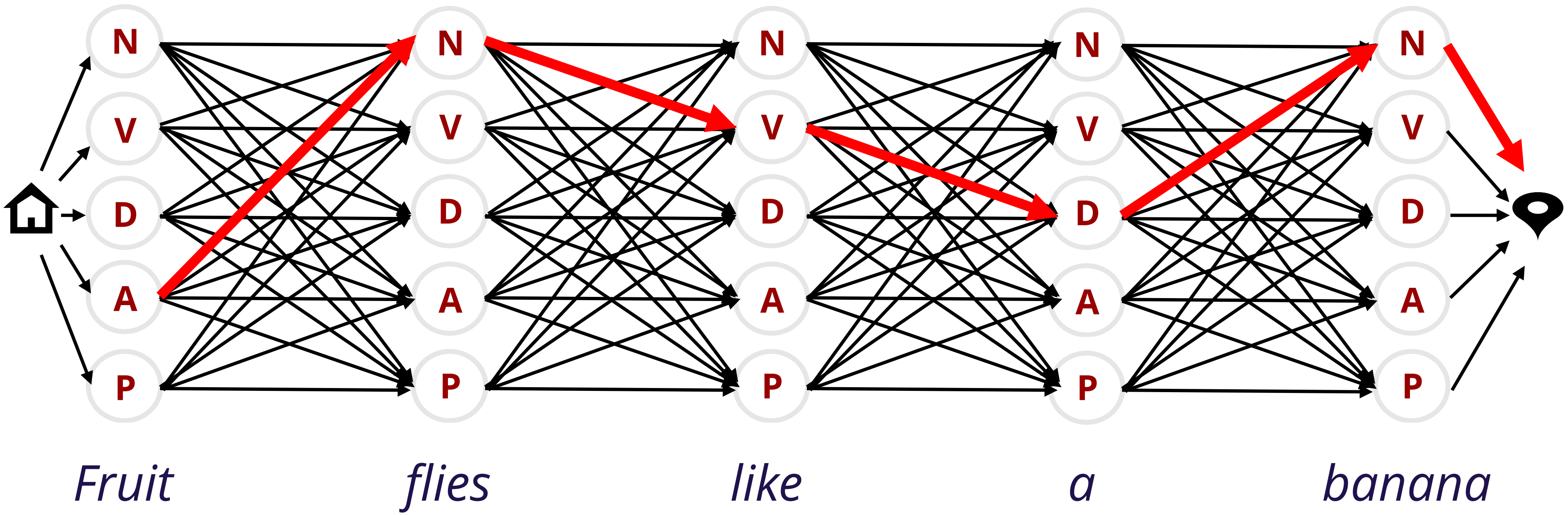
$$\arg\max_y \mathbf{w} \cdot \mathbf{f}(x, y)$$



Viterbi

$$\max_y \mathbf{w} \cdot \mathbf{f}(x, y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x, [y^j, y^{j+1}]) = 9.7$$

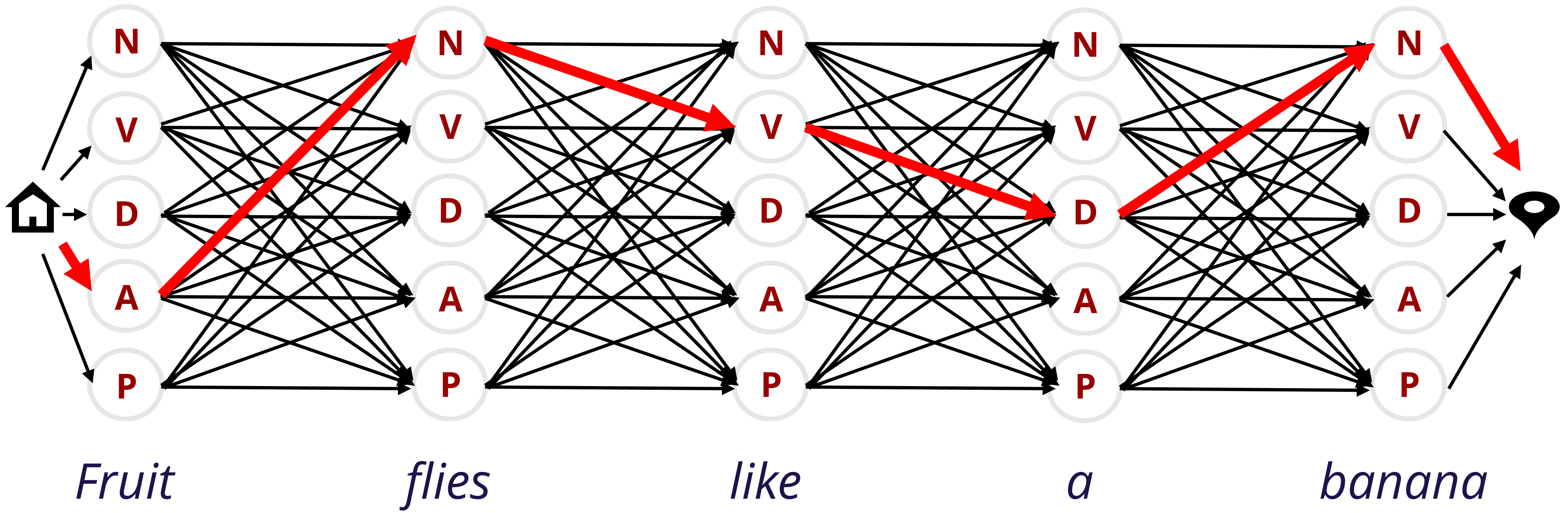
$$\arg\max_y \mathbf{w} \cdot \mathbf{f}(x, y)$$



Viterbi

$$\max_y \mathbf{w} \cdot \mathbf{f}(x,y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x,[y^j, y^{j+1}]) = 9.7$$

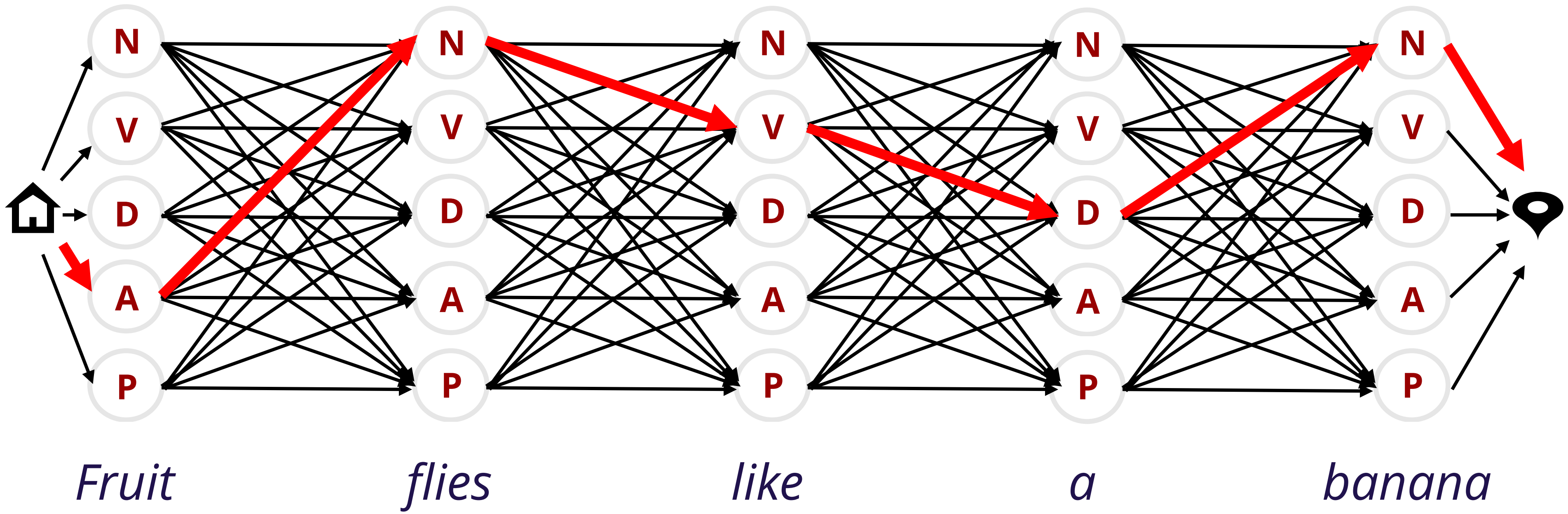
$$\arg\max_y \mathbf{w} \cdot \mathbf{f}(x,y)$$



MAP Inference

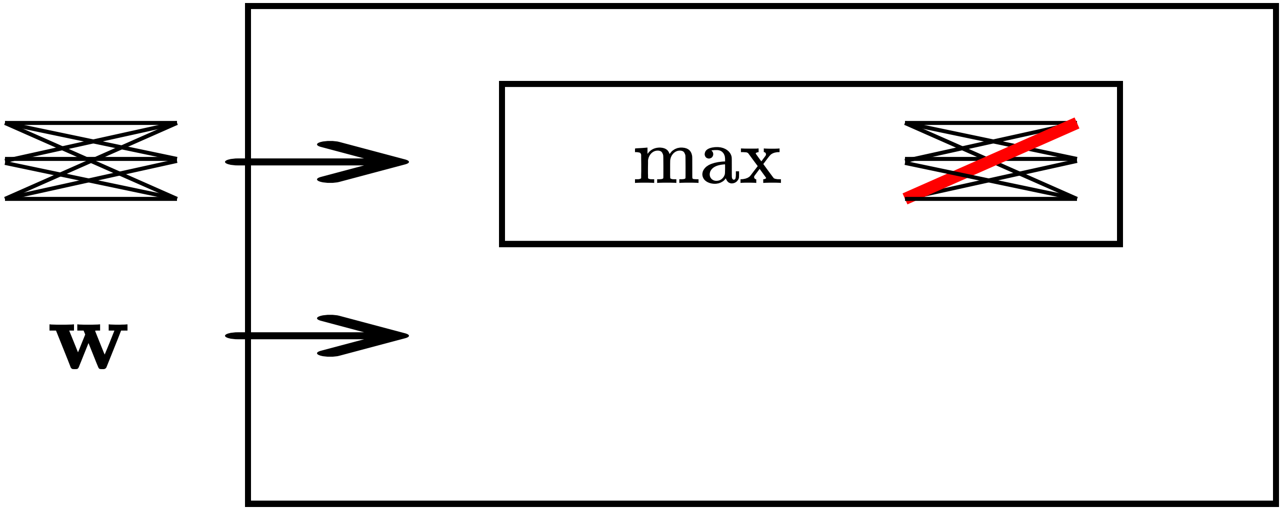
$$\max_y \mathbf{w} \cdot \mathbf{f}(x, y) = \max_y \sum_j \mathbf{w} \cdot \mathbf{f}(x, [y^j, y^{j+1}]) = 9.7$$

$$\arg\max_y \mathbf{w} \cdot \mathbf{f}(x, y)$$



Inference

MAP



So Far

Given the parameters w, how to search for the optimal path (and its score)

Next...

How to learn the parameters w?