

Product of a Grammar with an n -Gram Model for Statistical Machine Translation

Tobias Denkinger
tobias.denkinger@mailbox.tu-dresden.de

Lehrstuhl für Grundlagen der Programmierung
Institut für Theoretische Informatik
Technische Universität Dresden

June 7, 2013

Outline

Introduction

Modelling and Training

Decoding

Concepts

IRTG

n -Gram Model

Component Product

Results

n -Gram wRTG

Summary

Modelling and Training

translation function $h: F \rightarrow E$

Modelling and Training

translation function $h: \textcolor{blue}{F} \rightarrow \textcolor{brown}{E}$ (set of hypotheses $\mathcal{H} \subseteq \textcolor{brown}{E}^{\textcolor{blue}{F}}$)

$$\mathcal{M} \subseteq \mathcal{M}(\textcolor{brown}{E} \mid \textcolor{blue}{F})$$

Modelling and Training

translation function $h: \textcolor{blue}{F} \rightarrow \textcolor{brown}{E}$ (set of hypotheses $\mathcal{H} \subseteq \textcolor{brown}{E}^{\textcolor{blue}{F}}$)

$$\mathcal{M} \subseteq \mathcal{M}(\textcolor{brown}{E} \mid \textcolor{blue}{F})$$



$$\begin{aligned}\mathcal{H} &= \{h_p \mid p \in \mathcal{M}\} \\ h_p(\textcolor{blue}{f}) &= \operatorname{argmax}_{e \in \textcolor{brown}{E}} p(e \mid \textcolor{blue}{f})\end{aligned}$$

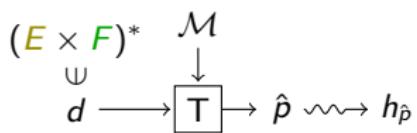
Modelling and Training

translation function $h: \textcolor{blue}{F} \rightarrow \textcolor{brown}{E}$ (set of hypotheses $\mathcal{H} \subseteq \textcolor{brown}{E}^{\textcolor{blue}{F}}$)

$$\mathcal{M} \subseteq \mathcal{M}(\textcolor{brown}{E} \mid \textcolor{blue}{F})$$



$$\begin{aligned}\mathcal{H} &= \{h_p \mid p \in \mathcal{M}\} \\ h_p(\textcolor{blue}{f}) &= \operatorname{argmax}_{e \in \textcolor{brown}{E}} p(e \mid \textcolor{blue}{f})\end{aligned}$$



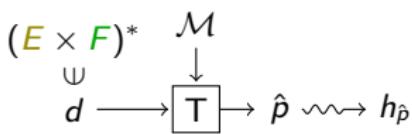
Modelling and Training

translation function $h: \textcolor{blue}{F} \rightarrow \textcolor{brown}{E}$ (set of hypotheses $\mathcal{H} \subseteq \textcolor{brown}{E}^{\textcolor{blue}{F}}$)

$$\mathcal{M} \subseteq \mathcal{M}(\textcolor{brown}{E} \mid \textcolor{blue}{F})$$



$$\begin{aligned} \mathcal{H} &= \{h_p \mid p \in \mathcal{M}\} & \mathcal{H} &= \{h_{p_1, p_2} \mid p_1 \in \mathcal{M}_1, p_2 \in \mathcal{M}_2\} \\ h_p(\textcolor{blue}{f}) &= \operatorname{argmax}_{e \in \textcolor{brown}{E}} p(e \mid \textcolor{blue}{f}) & h_{p_1, p_2}(\textcolor{blue}{f}) &= \operatorname{argmax}_{e \in \textcolor{brown}{E}} p_1(f \mid e) \cdot p_2(e) \end{aligned}$$



Modelling and Training

translation function $h: \textcolor{blue}{F} \rightarrow \textcolor{brown}{E}$ (set of hypotheses $\mathcal{H} \subseteq \textcolor{brown}{E}^{\textcolor{blue}{F}}$)

$$\mathcal{M} \subseteq \mathcal{M}(\textcolor{brown}{E} \mid \textcolor{blue}{F})$$



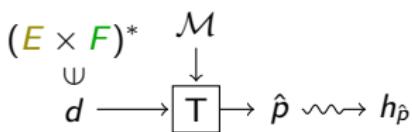
$$\begin{aligned}\mathcal{H} &= \{h_p \mid p \in \mathcal{M}\} \\ h_p(\textcolor{blue}{f}) &= \operatorname{argmax}_{\textcolor{brown}{e} \in \textcolor{brown}{E}} p(\textcolor{brown}{e} \mid \textcolor{blue}{f})\end{aligned}$$

$$\mathcal{M}_1 \subseteq \mathcal{M}(\textcolor{blue}{F} \mid \textcolor{brown}{E})$$

$$\mathcal{M}_2 \subseteq \mathcal{M}(\textcolor{brown}{E})$$



$$\begin{aligned}\mathcal{H} &= \{h_{p_1, p_2} \mid p_1 \in \mathcal{M}_1, p_2 \in \mathcal{M}_2\} \\ h_{p_1, p_2}(\textcolor{blue}{f}) &= \operatorname{argmax}_{\textcolor{brown}{e} \in \textcolor{brown}{E}} p_1(\textcolor{blue}{f} \mid \textcolor{brown}{e}) \cdot p_2(\textcolor{brown}{e})\end{aligned}$$



Modelling and Training

translation function $h: \textcolor{blue}{F} \rightarrow \textcolor{brown}{E}$ (set of hypotheses $\mathcal{H} \subseteq \textcolor{brown}{E}^{\textcolor{blue}{F}}$)

$$\mathcal{M} \subseteq \mathcal{M}(\textcolor{brown}{E} \mid \textcolor{blue}{F})$$



$$\begin{aligned}\mathcal{H} &= \{h_p \mid p \in \mathcal{M}\} \\ h_p(\textcolor{blue}{f}) &= \operatorname{argmax}_{\textcolor{brown}{e} \in \textcolor{brown}{E}} p(\textcolor{brown}{e} \mid \textcolor{blue}{f})\end{aligned}$$

$$\mathcal{M}_1 \subseteq \mathcal{M}(\textcolor{blue}{F} \mid \textcolor{brown}{E})$$

$$\mathcal{M}_2 \subseteq \mathcal{M}(\textcolor{brown}{E})$$



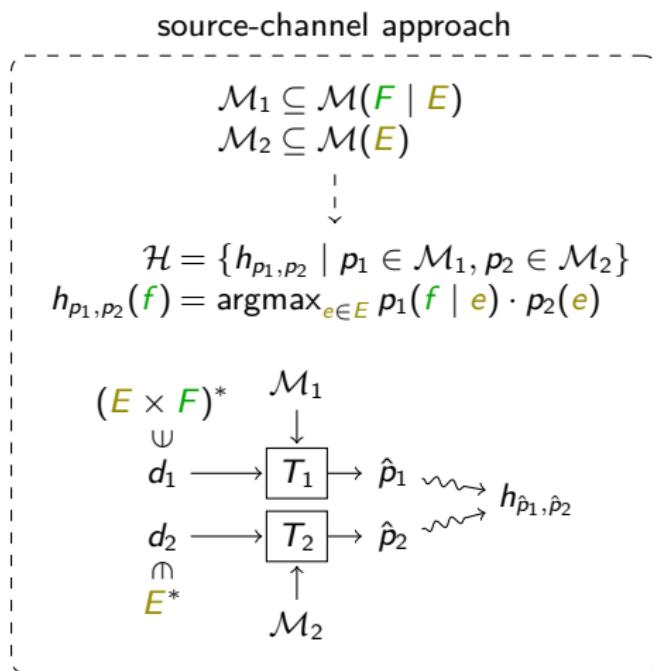
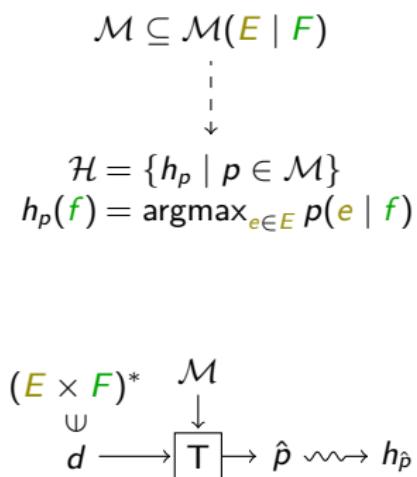
$$\begin{aligned}\mathcal{H} &= \{h_{p_1, p_2} \mid p_1 \in \mathcal{M}_1, p_2 \in \mathcal{M}_2\} \\ h_{p_1, p_2}(\textcolor{blue}{f}) &= \operatorname{argmax}_{\textcolor{brown}{e} \in \textcolor{brown}{E}} p_1(\textcolor{blue}{f} \mid \textcolor{brown}{e}) \cdot p_2(\textcolor{brown}{e})\end{aligned}$$

$$\begin{array}{c} (\textcolor{brown}{E} \times \textcolor{blue}{F})^* \xrightarrow{\cup} \mathcal{M} \\ d \xrightarrow{\sqcup} \boxed{T} \xrightarrow{\quad} \hat{p} \rightsquigarrow h_{\hat{p}} \end{array}$$

$$\begin{array}{c} (\textcolor{brown}{E} \times \textcolor{blue}{F})^* \xrightarrow{\cup} \mathcal{M}_1 \\ d_1 \longrightarrow \boxed{T_1} \xrightarrow{\quad} \hat{p}_1 \rightsquigarrow h_{\hat{p}_1, \hat{p}_2} \\ \textcolor{brown}{E}^* \xrightarrow{\cap} \mathcal{M}_2 \\ d_2 \longrightarrow \boxed{T_2} \xrightarrow{\quad} \hat{p}_2 \rightsquigarrow h_{\hat{p}_1, \hat{p}_2} \end{array}$$

Modelling and Training

translation function $h: \textcolor{blue}{F} \rightarrow \textcolor{brown}{E}$ (set of hypotheses $\mathcal{H} \subseteq \textcolor{brown}{E}^{\textcolor{blue}{F}}$)



Outline

Introduction

Modelling and Training

Decoding

Concepts

IRTG

n -Gram Model

Component Product

Results

n -Gram wRTG

Summary

Decoding

source-channel approach

$$h_{p_1, p_2}(\textcolor{red}{f}) = \operatorname{argmax}_{\textcolor{blue}{e} \in \textcolor{brown}{E}} \underbrace{p_1}_{\in \mathcal{M}_1}(\textcolor{red}{f} \mid \textcolor{blue}{e}) \cdot \underbrace{p_2}_{\in \mathcal{M}_2}(\textcolor{blue}{e})$$

Decoding

source-channel approach

$$h_{p_1, p_2}(f) = \operatorname{argmax}_{e \in E} \underbrace{p_1}_{\in \mathcal{M}_1}(f \mid e) \cdot \underbrace{p_2}_{\in \mathcal{M}_2}(e)$$

- ▶ $\mathcal{M}_1 \dots$ interpreted regular tree grammars (IRTGs)

Decoding

source-channel approach

$$h_{p_1, p_2}(\textcolor{blue}{f}) = \operatorname{argmax}_{\textcolor{blue}{e} \in \textcolor{brown}{E}} \underbrace{p_1}_{\in \mathcal{M}_1}(\textcolor{blue}{f} \mid \textcolor{brown}{e}) \cdot \underbrace{p_2}_{\in \mathcal{M}_2}(\textcolor{brown}{e})$$

- ▶ $\mathcal{M}_1 \dots$ interpreted regular tree grammars (IRTGs)
- ▶ $\mathcal{M}_2 \dots$ n -gram model

Outline

Introduction

Modelling and Training

Decoding

Concepts

IRTG

n -Gram Model

Component Product

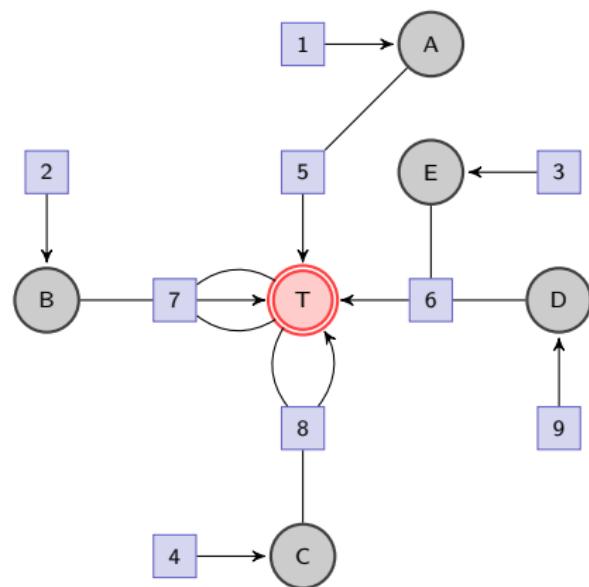
Results

n -Gram wRTG

Summary

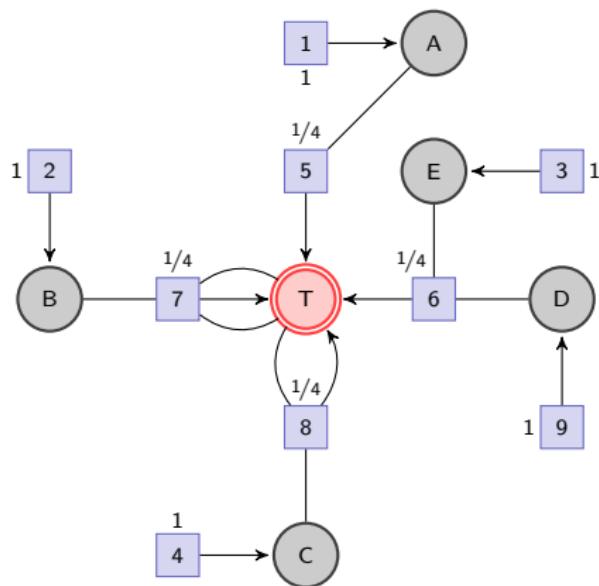
IRTG

- ▶ RTG (shown as hypergraph)



IRTG

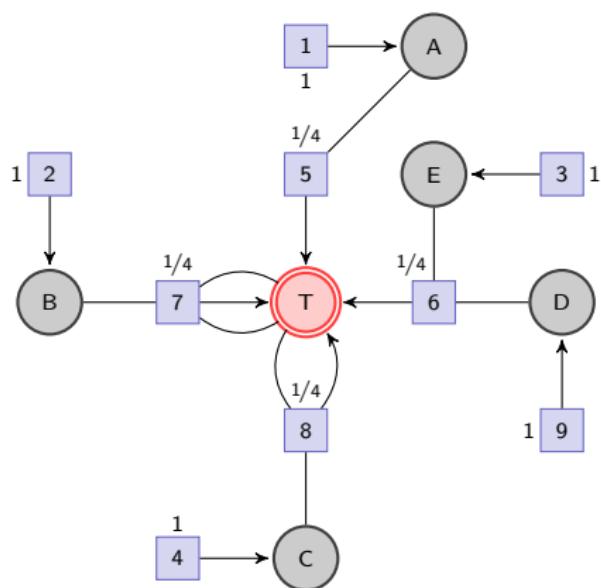
- ▶ wRTG (shown as hypergraph)



IRTG

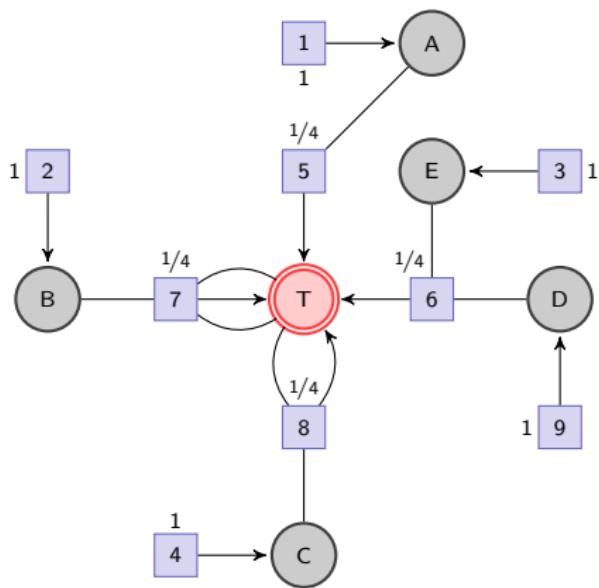
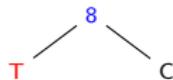
- ▶ wRTG (shown as hypergraph)

T



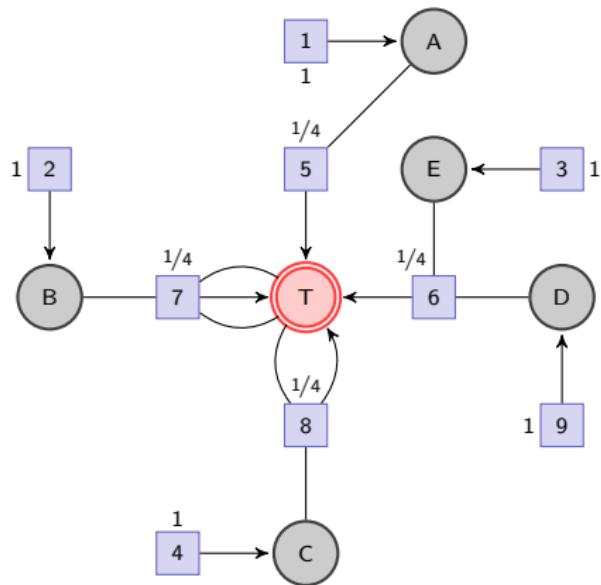
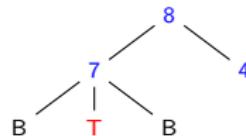
IRTG

- ▶ wRTG (shown as hypergraph)



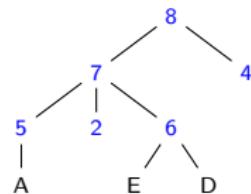
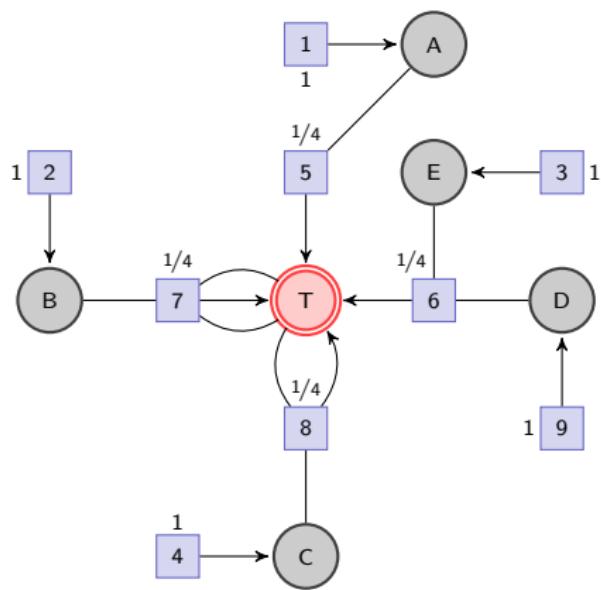
IRTG

- ▶ wRTG (shown as hypergraph)



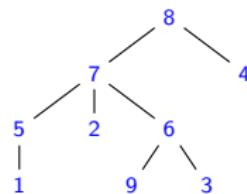
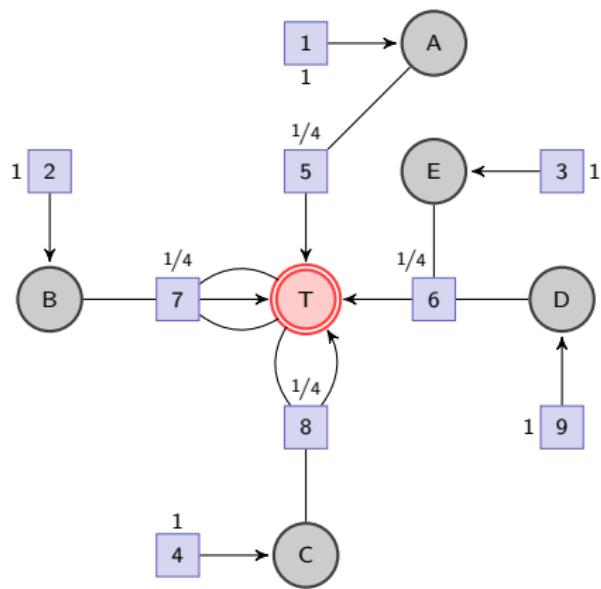
IRTG

- ▶ wRTG (shown as hypergraph)



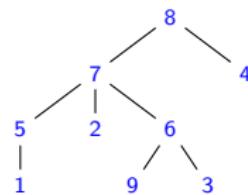
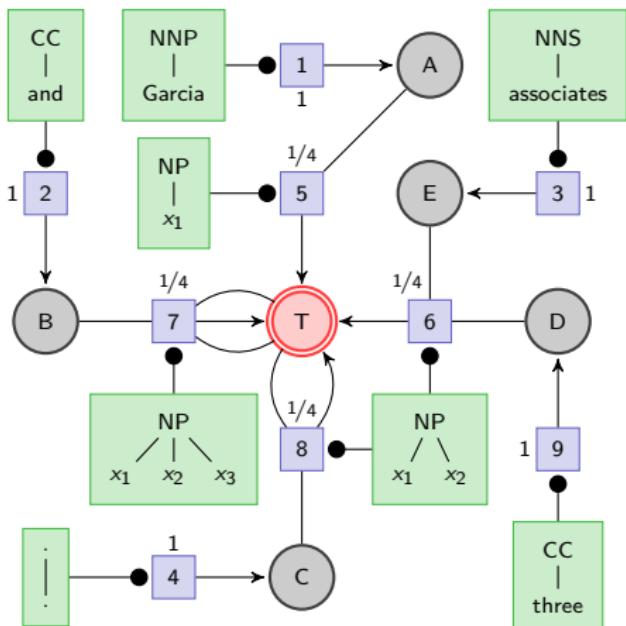
IRTG

- ▶ wRTG (shown as hypergraph)



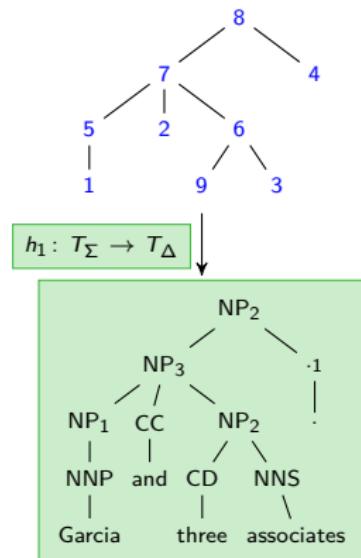
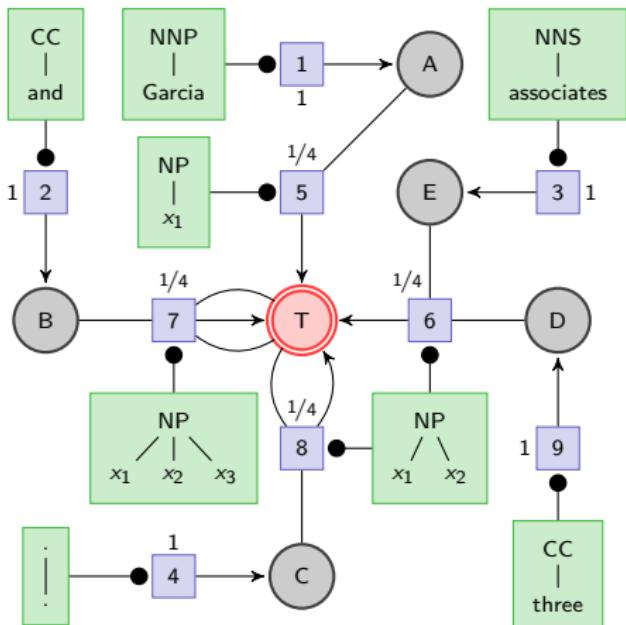
IRTG

- ▶ wRTG (shown as hypergraph)
- ▶ tree homomorphism h_1



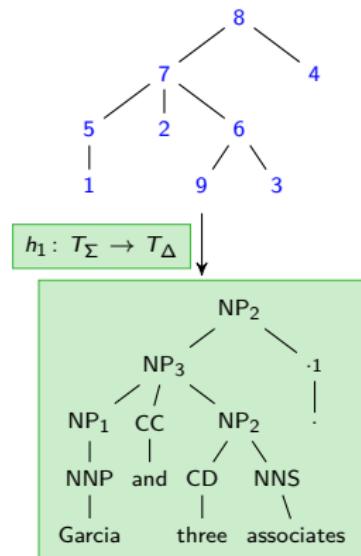
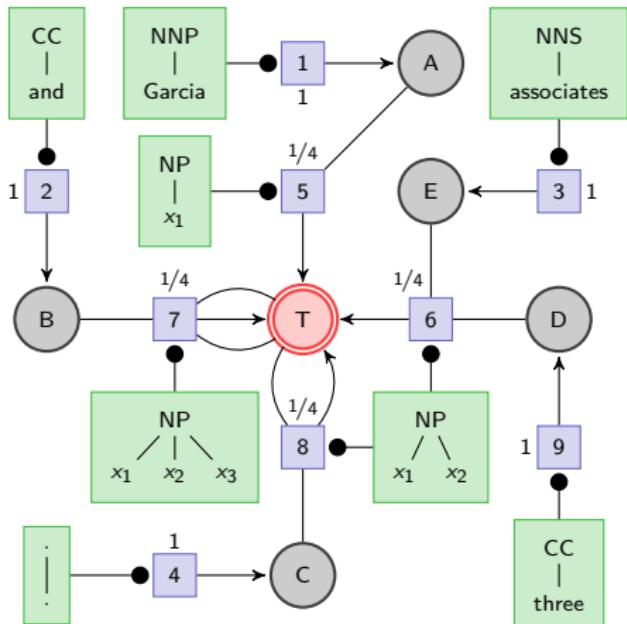
IRTG

- ▶ wRTG (shown as hypergraph)
- ▶ tree homomorphism h_1



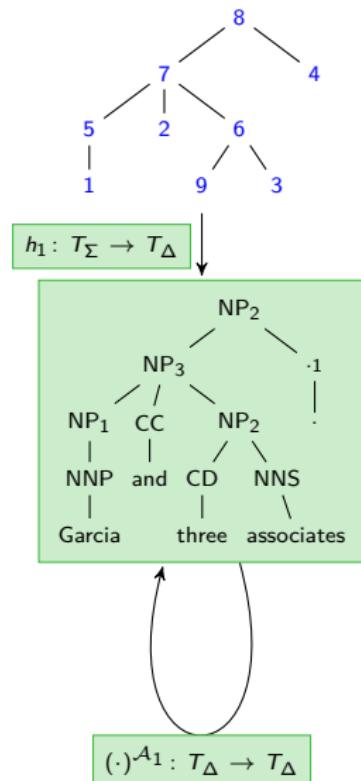
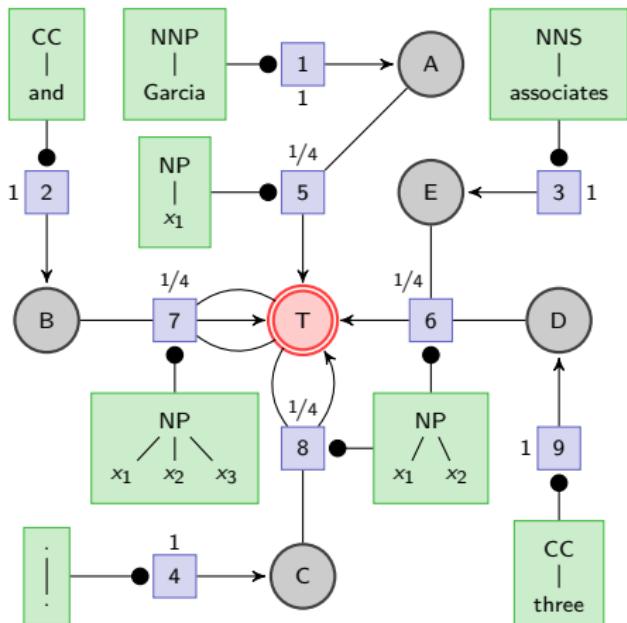
IRTG

- ▶ wRTG (shown as hypergraph)
- ▶ tree homomorphism h_1
- ▶ algebra $\mathcal{A}_1 = \langle T_\Delta, (\cdot)^{\mathcal{A}_1} \rangle$



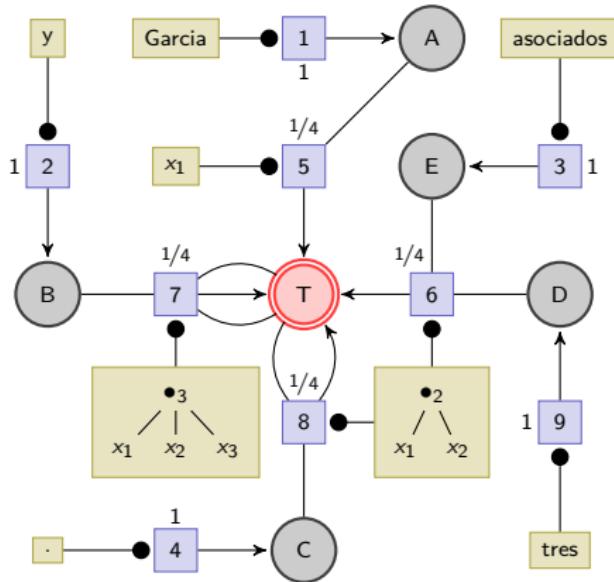
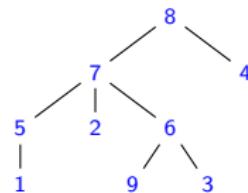
IRTG

- ▶ wRTG (shown as hypergraph)
- ▶ tree homomorphism h_1
- ▶ algebra $\mathcal{A}_1 = \langle T_\Delta, (\cdot)^{\mathcal{A}_1} \rangle$



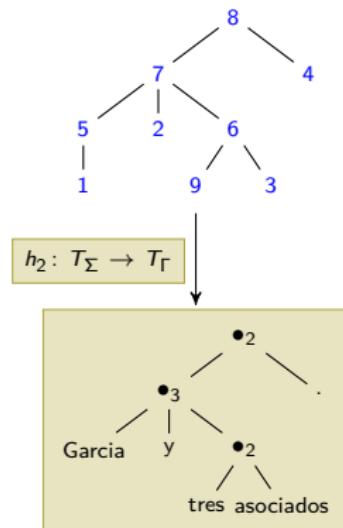
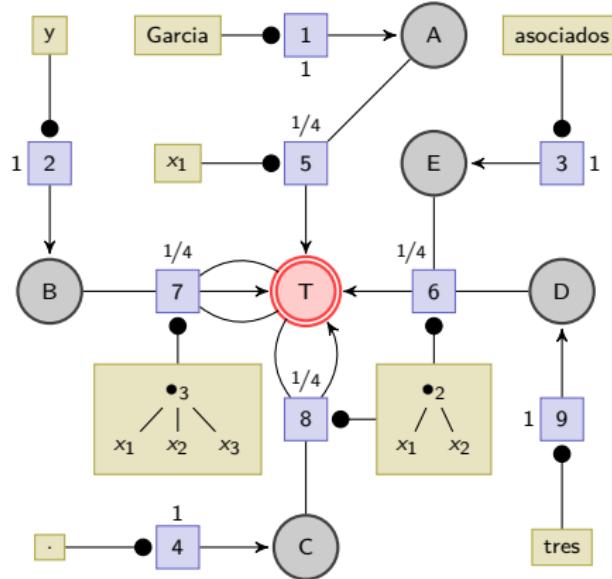
IRTG

- ▶ wRTG (shown as hypergraph)
- ▶ tree homomorphisms h_1, h_2
- ▶ algebra $\mathcal{A}_1 = \langle T_\Delta, (\cdot)^{\mathcal{A}_1} \rangle$



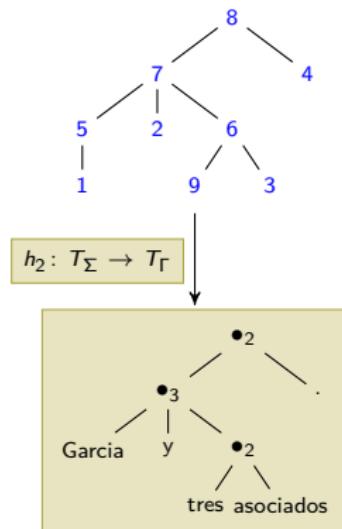
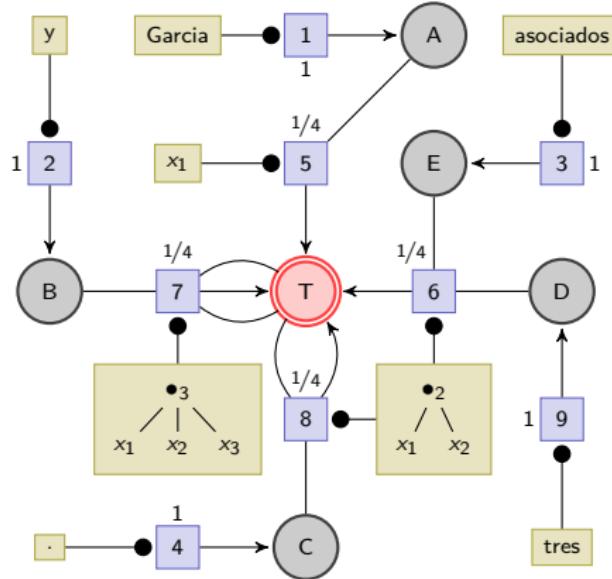
IRTG

- ▶ wRTG (shown as hypergraph)
- ▶ tree homomorphisms h_1, h_2
- ▶ algebra $\mathcal{A}_1 = \langle T_\Delta, (\cdot)^{\mathcal{A}_1} \rangle$



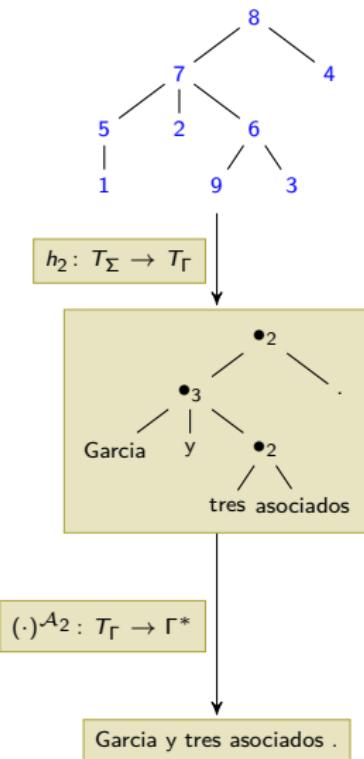
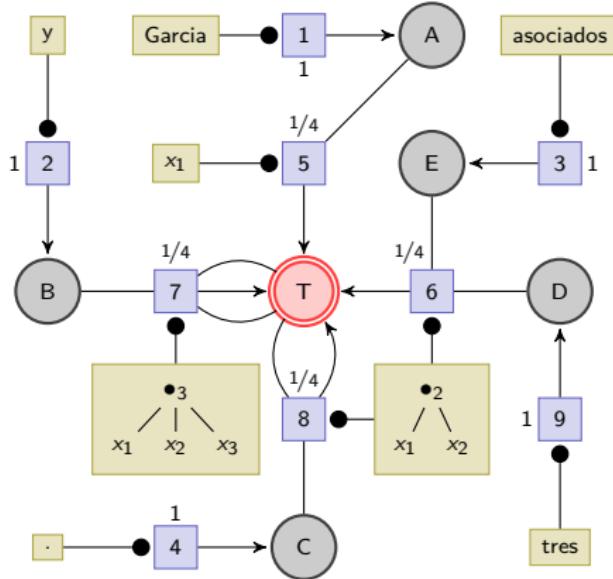
IRTG

- ▶ wRTG (shown as hypergraph)
- ▶ tree homomorphisms h_1, h_2
- ▶ algebras $\mathcal{A}_1 = \langle T_\Delta, (\cdot)^{\mathcal{A}_1} \rangle$, $\mathcal{A}_2 = \langle \Gamma^*, (\cdot)^{\mathcal{A}_2} \rangle$



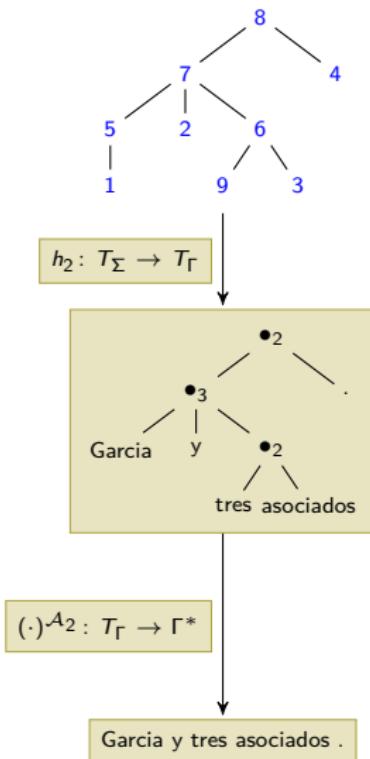
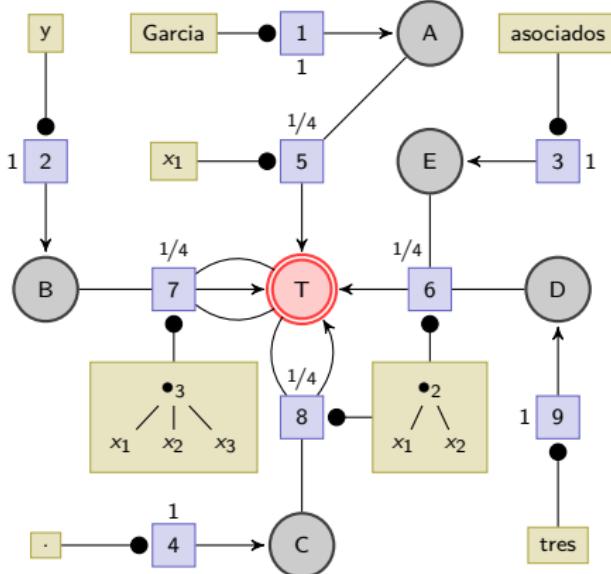
IRTG

- ▶ wRTG (shown as hypergraph)
- ▶ tree homomorphisms h_1, h_2
- ▶ algebras $\mathcal{A}_1 = \langle T_\Delta, (\cdot)^{\mathcal{A}_1} \rangle$, $\mathcal{A}_2 = \langle \Gamma^*, (\cdot)^{\mathcal{A}_2} \rangle$



IRTG

- ▶ wRTG (shown as hypergraph)
- ▶ tree homomorphisms h_1, h_2
- ▶ algebras $\mathcal{A}_1 = \langle T_\Delta, (\cdot)^{\mathcal{A}_1} \rangle$, $\mathcal{A}_2 = \langle \Gamma^*, (\cdot)^{\mathcal{A}_2} \rangle$
- ▶ weighted language over $T_\Delta \times \Gamma^*$



Outline

Introduction

Modelling and Training

Decoding

Concepts

IRTG

n-Gram Model

Component Product

Results

n-Gram wRTG

Summary

n-Gram Model

example: 3-gram model ($n = 3$)

n-Gram Model

example: 3-gram model ($n = 3$)

- ▶ scores a sequence of symbols

$w =$ Garcia y tres asociados .

n-Gram Model

example: 3-gram model ($n = 3$)

- ▶ scores a sequence of symbols

$w =$ Garcia y tres asociados .

$N(w) =$

n-Gram Model

example: 3-gram model ($n = 3$)

- ▶ scores a sequence of symbols
- ▶ assigns weight to every $\underbrace{\text{sequence of 3 symbols}}_{= \text{3-gram}}$

$w =$ 

$$N(w) = 2/5$$

n-Gram Model

example: 3-gram model ($n = 3$)

- ▶ scores a sequence of symbols
- ▶ assigns weight to every $\underbrace{\text{sequence of 3 symbols}}_{= \text{3-gram}}$

$w =$  .

$$N(w) = \quad \frac{2}{5} \quad \frac{1}{3}$$

n-Gram Model

example: 3-gram model ($n = 3$)

- ▶ scores a sequence of symbols
- ▶ assigns weight to every $\underbrace{\text{sequence of 3 symbols}}_{= \text{3-gram}}$

$w =$ A sequence of five rectangular boxes containing the words "Garcia", "y", "tres", "asociados", and a single character ".", the latter being the punctuation mark. The box containing "tres" is highlighted with a red oval.

$$N(w) = \quad \textcolor{green}{2/5} \quad \textcolor{blue}{1/3} \quad \textcolor{red}{1/4}$$

n-Gram Model

example: 3-gram model ($n = 3$)

- ▶ scores a sequence of symbols
- ▶ assigns weight to every $\underbrace{\text{sequence of 3 symbols}}_{= \text{3-gram}}$
- ▶ weights are multiplied

$w =$ Garcia y tres asociados .

$$N(w) = \quad \textcolor{green}{2/5} \quad \cdot \quad \textcolor{blue}{1/3} \quad \cdot \quad \textcolor{red}{1/4} \quad = \quad 1/30$$

Outline

Introduction

Modelling and Training

Decoding

Concepts

IRTG

n -Gram Model

Component Product

Results

n -Gram wRTG

Summary

Component Product

$$h_{p_1, p_2}(f) = \operatorname{argmax}_{e \in E} \underbrace{p_1(f \mid e)}_{\text{by IRTG}} \cdot \underbrace{p_2(e)}_{\text{by } n\text{-gram model}}$$

Component Product

$$h_{p_1, p_2}(\textcolor{blue}{f}) = \operatorname{argmax}_{\textcolor{blue}{e} \in \textcolor{brown}{E}} \underbrace{p_1(\textcolor{blue}{f} \mid \textcolor{blue}{e})}_{\text{by IRTG}} \cdot \underbrace{p_2(\textcolor{blue}{e})}_{\text{by } n\text{-gram model}}$$

$$h_{\mathcal{L}, N}(\textcolor{blue}{f}) = \operatorname{argmax}_{\textcolor{blue}{e} \in \textcolor{brown}{E}} \mathcal{L}(\langle \textcolor{blue}{e}, \textcolor{blue}{f} \rangle) \cdot N(\textcolor{blue}{e})$$

Component Product

$$h_{p_1, p_2}(f) = \operatorname{argmax}_{e \in E} \underbrace{p_1(f | e)}_{\text{by IRTG}} \cdot \underbrace{p_2(e)}_{\text{by } n\text{-gram model}}$$

$$\begin{aligned} h_{\mathcal{L}, N}(f) &= \operatorname{argmax}_{e \in E} \mathcal{L}(\langle e, f \rangle) \cdot N(e) \\ &= \operatorname{argmax}_{e \in E} \underbrace{(\mathcal{L} \odot_1 N)(\langle e, f \rangle)}_{\text{component product}} \end{aligned}$$

Component Product

$$h_{p_1, p_2}(f) = \operatorname{argmax}_{e \in E} \underbrace{p_1(f \mid e)}_{\text{by IRTG}} \cdot \underbrace{p_2(e)}_{\text{by } n\text{-gram model}}$$

$$\begin{aligned} h_{\mathcal{L}, N}(f) &= \operatorname{argmax}_{e \in E} \mathcal{L}(\langle e, f \rangle) \cdot N(e) \\ &= \operatorname{argmax}_{e \in E} \underbrace{(\mathcal{L} \odot_1 N)(\langle e, f \rangle)}_{\text{component product}} \end{aligned}$$

$$(\mathcal{L} \odot_i N)(\langle w_1, \dots, w_k \rangle) = \mathcal{L}(\langle w_1, \dots, w_k \rangle) \cdot N(w_i)$$

Results

Theorem (main result)

$$IRTG \odot_i n\text{-gram model} \subseteq IRTG$$

Results

Theorem (main result)

$$IRTG \odot_i n\text{-gram model} \subseteq IRTG$$

Lemma (n -gram wRTG)

$$(\cdot)^A; n\text{-gram model} \subseteq wRTG$$

Results

Theorem (main result)

$$IRTG \odot_i n\text{-gram model} \subseteq IRTG$$

Lemma (n -gram wRTG)

$$(\cdot)^A; n\text{-gram model} \subseteq wRTG$$

Lemma (inverse homomorphism wRTG)

$$In-Hom^{-1}(wRTG) \subseteq wRTG$$

Results

Theorem (main result)

$$IRTG \odot_i n\text{-gram model} \subseteq IRTG$$

Lemma (n -gram wRTG)

$$(\cdot)^A; n\text{-gram model} \subseteq wRTG$$

Lemma (inverse homomorphism wRTG)

$$\text{In-Hom}^{-1}(wRTG) \subseteq wRTG$$

Lemma (Hadamard product)

$$wRTG \odot wRTG \subseteq wRTG$$

Results

Theorem (main result)

$$IRTG \odot_i n\text{-gram model} \subseteq IRTG$$

Lemma (n -gram wRTG)

$$(\cdot)^A; n\text{-gram model} \subseteq wRTG$$

Lemma (inverse homomorphism wRTG)

$$\text{In-Hom}^{-1}(wRTG) \subseteq wRTG$$

Lemma (Hadamard product)

$$wRTG \odot wRTG \subseteq wRTG$$

Outline

Introduction

Modelling and Training

Decoding

Concepts

IRTG

n -Gram Model

Component Product

Results

n -Gram wRTG

Summary

n -Gram wRTG

$w = \boxed{G} \quad \boxed{y} \quad \boxed{t} \quad \boxed{a} \quad \boxed{.}$

Garcia y tres asociados .

$$N(w) =$$

n -Gram wRTG

$w =$  Garcia y tres asociados .

$$N(Gyt)$$

$$N(w) = \textcolor{green}{2/5}$$

n -Gram wRTG

$w = \boxed{G} \quad \boxed{y} \quad \boxed{t} \quad \boxed{a} \quad \boxed{\cdot}$

Garcia y tres asociados .

$$N(w) = \frac{2}{5} \cdot \frac{1}{3}$$
$$N(Gyt) \quad N(yta)$$

n -Gram wRTG

$w = \boxed{G} \quad \boxed{y} \quad \boxed{t} \quad \boxed{a} \quad \boxed{.}$

Garcia y tres asociados .

$$N(w) = N(Gyt) \quad N(yta) \quad N(ta.)$$
$$\frac{2}{5} \quad . \quad \frac{1}{3} \quad . \quad \frac{1}{4}$$

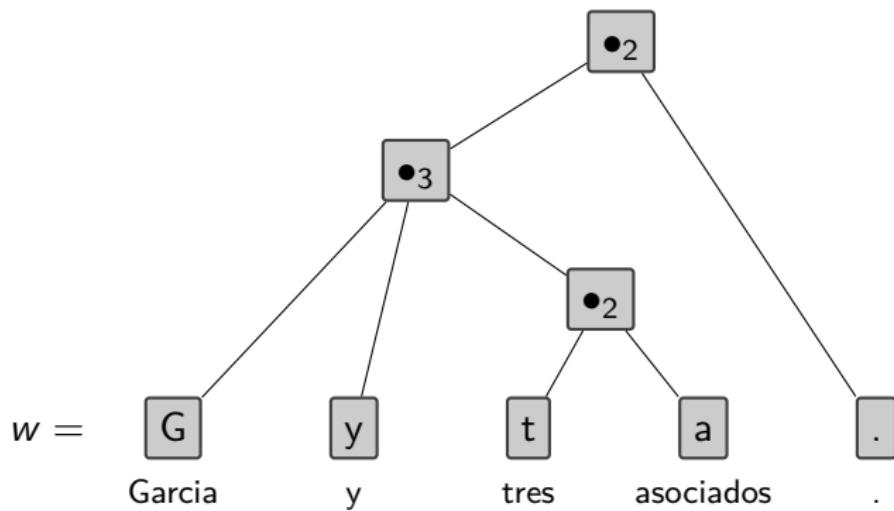
n -Gram wRTG

$w = \boxed{G} \quad \boxed{y} \quad \boxed{t} \quad \boxed{a} \quad \boxed{.}$

Garcia y tres asociados .

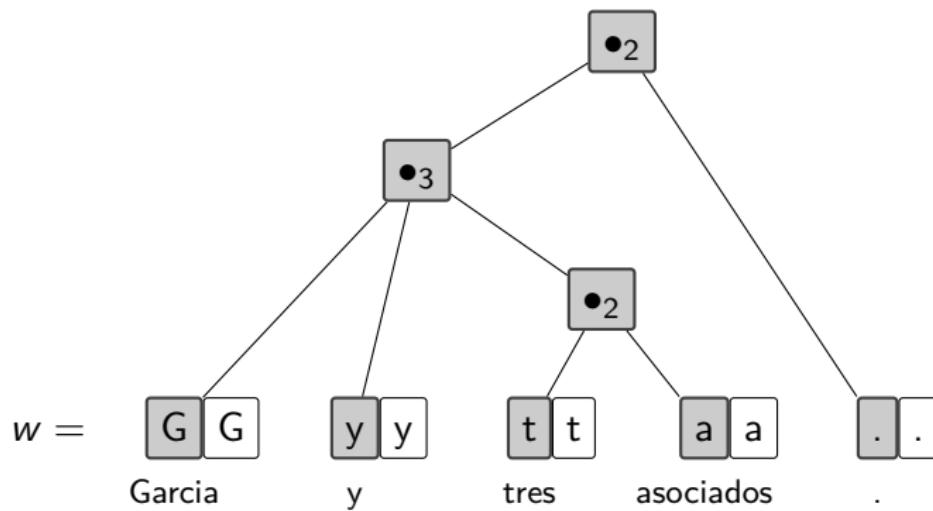
$$N(w) = N(\text{Gyt}) \cdot N(\text{ta.}) = \frac{2}{5} \cdot \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{30}$$

n -Gram wRTG



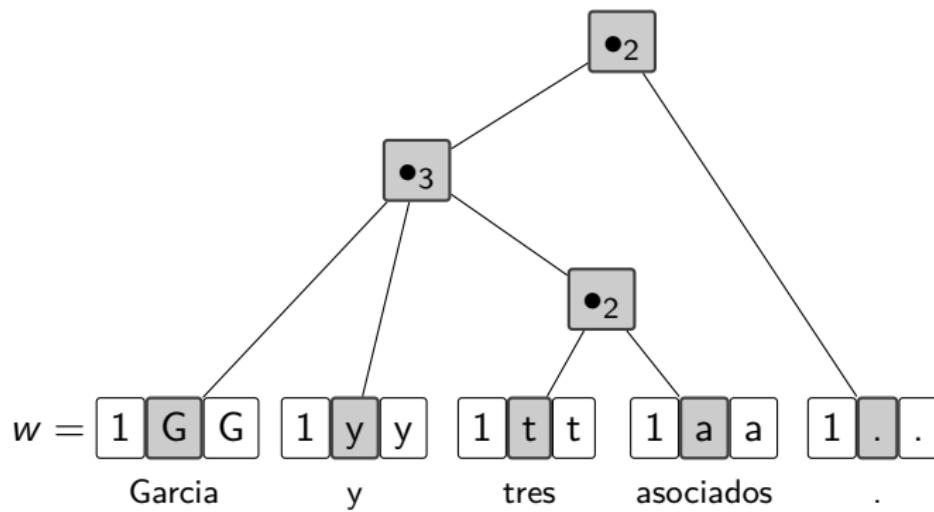
$$N(w) = N(\text{Gyt}) \cdot N(\text{yta.}) \cdot N(\text{ta.}) = \frac{2}{5} \cdot \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{30}$$

n -Gram wRTG



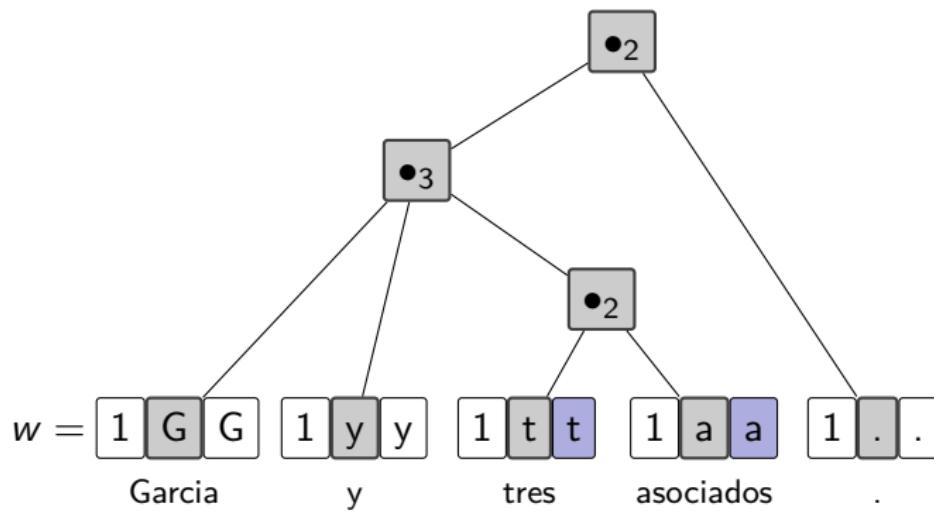
$$N(w) = N(\text{Gyt}) \cdot N(\text{yta}) \cdot N(\text{ta.}) = \frac{2}{5} \cdot \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{30}$$

n -Gram wRTG



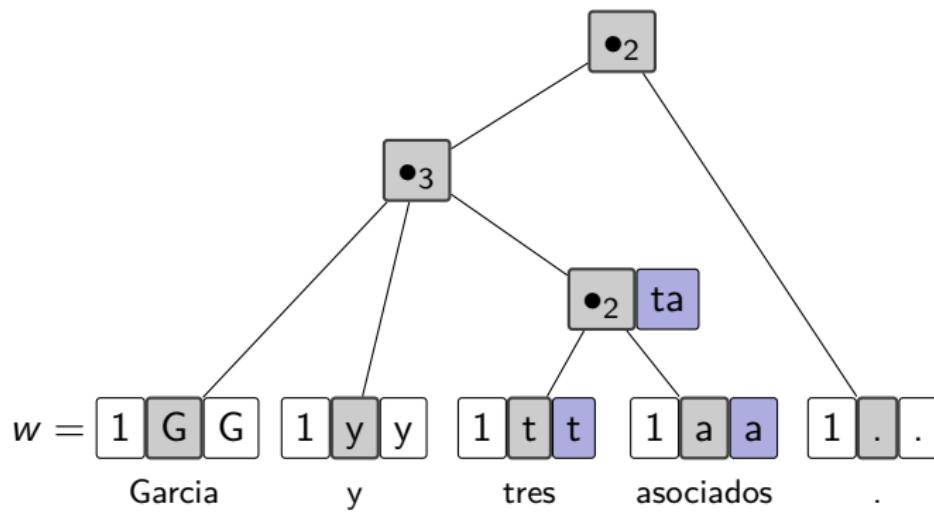
$$N(w) = N(\text{Gyt}) \cdot N(\text{yta.}) \cdot N(\text{ta.}) = \frac{2}{5} \cdot \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{30}$$

n -Gram wRTG



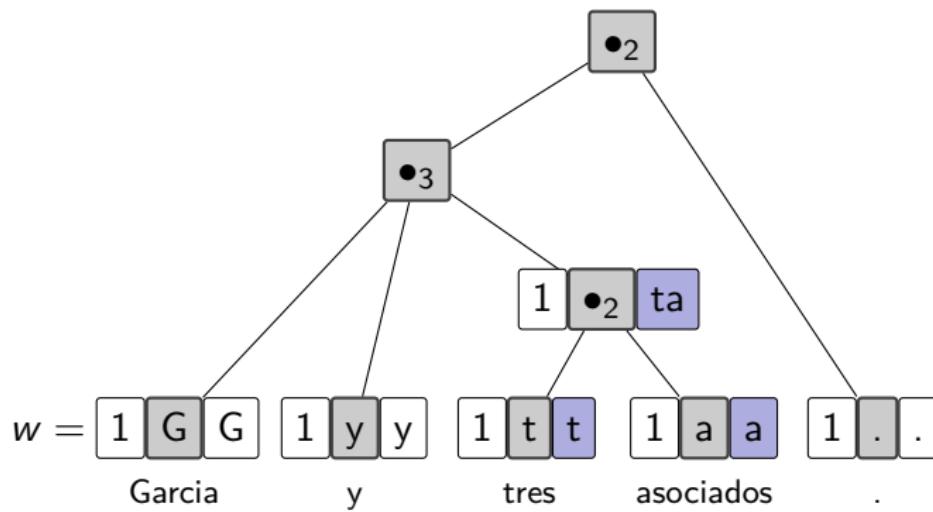
$$N(w) = N(\text{Gyt}) \cdot N(\text{yta}) \cdot N(\text{ta.}) = \frac{2}{5} \cdot \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{30}$$

n -Gram wRTG



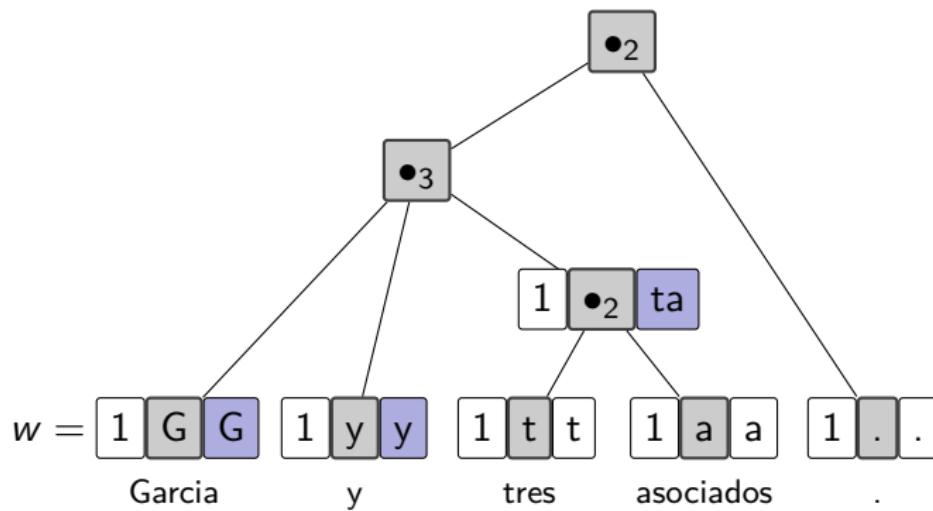
$$N(w) = N(\text{Gyt}) \cdot N(\text{yta}) \cdot N(\text{ta.}) = \frac{2}{5} \cdot \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{30}$$

n -Gram wRTG



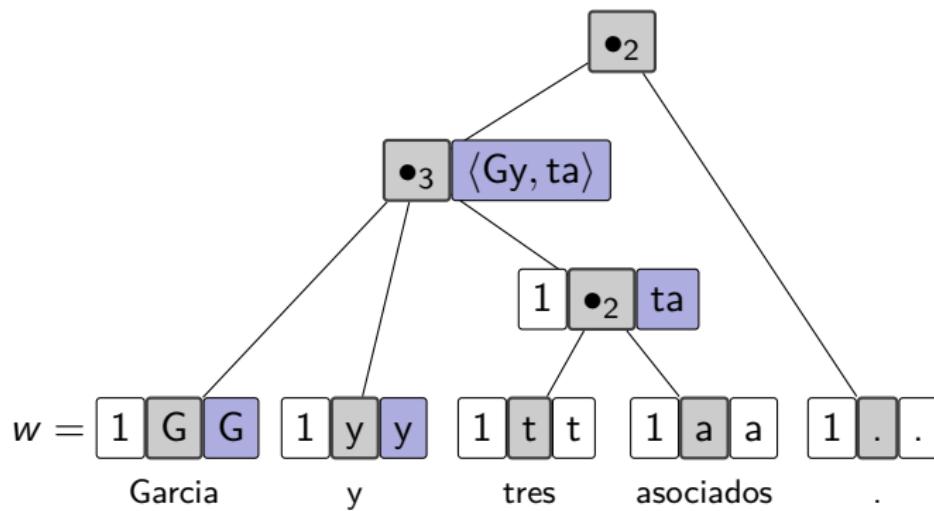
$$N(w) = N(\text{Gyt}) \cdot N(\text{yta}) \cdot N(\text{ta.}) = \frac{2}{5} \cdot \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{30}$$

n -Gram wRTG



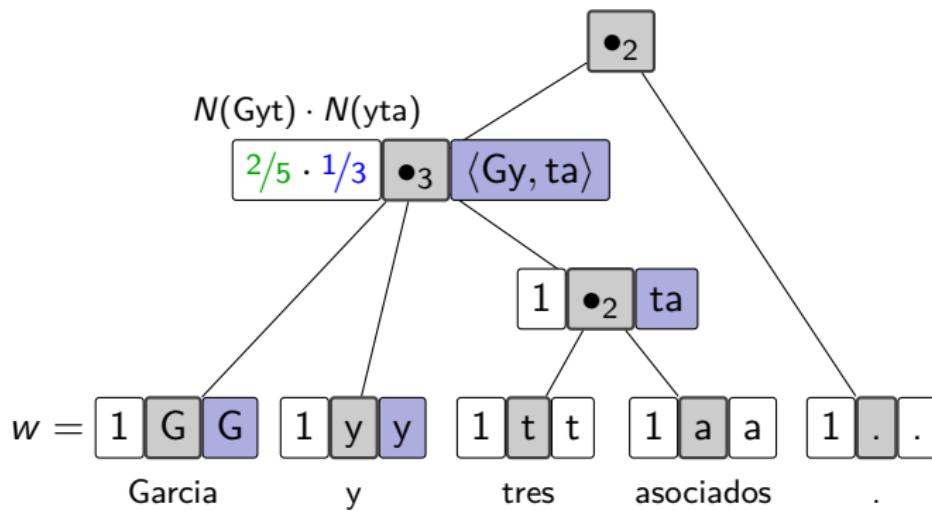
$$N(w) = N(\text{Gyt}) \cdot N(\text{yta}) \cdot N(\text{ta.}) = \frac{2}{5} \cdot \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{30}$$

n -Gram wRTG



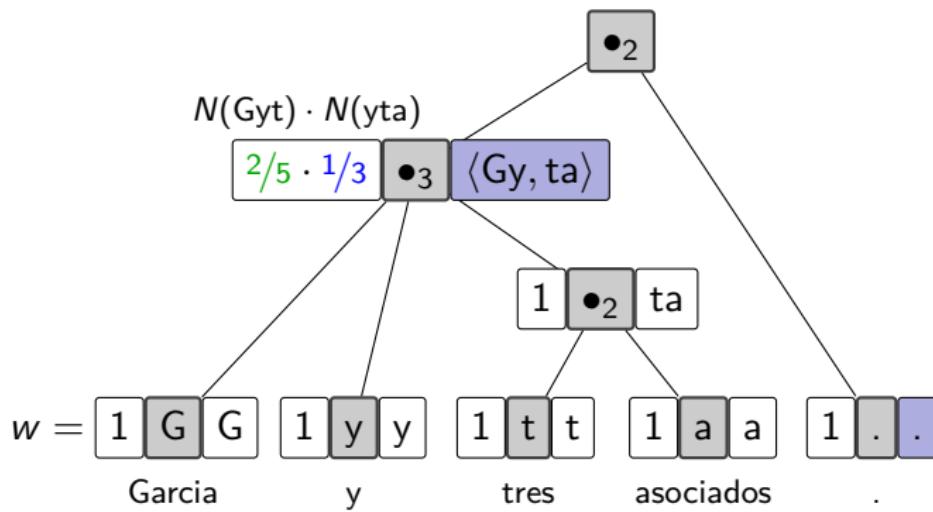
$$N(w) = N(\text{Gyt}) \cdot N(\text{yta}) \cdot N(\text{ta.}) = \frac{2}{5} \cdot \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{30}$$

n -Gram wRTG



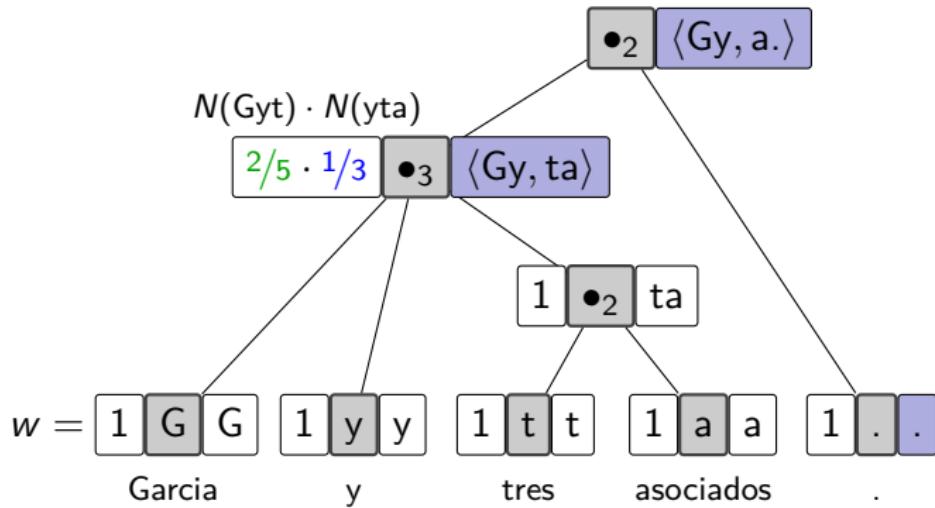
$$N(w) = N(\text{Gyt}) \cdot N(\text{yta}) \cdot N(\text{ta.}) = \frac{2}{5} \cdot \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{30}$$

n -Gram wRTG



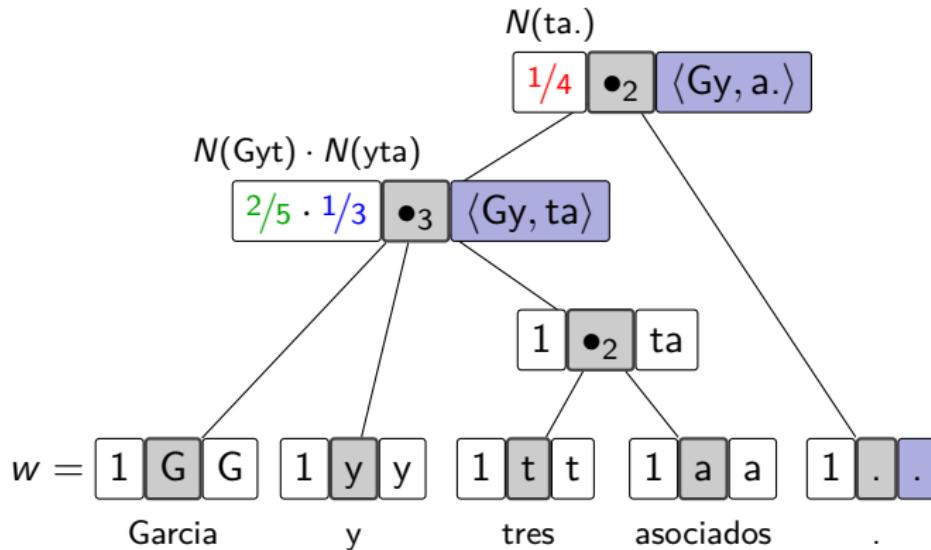
$$N(w) = \frac{N(Gyt)}{\frac{2}{5}} \cdot \frac{N(yta)}{\frac{1}{3}} \cdot \frac{N(ta.)}{\frac{1}{4}} = \frac{1}{30}$$

n -Gram wRTG



$$N(w) = N(\text{Gyt}) \cdot N(\text{yta}) \cdot N(\text{ta.}) = \frac{2}{5} \cdot \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{30}$$

n -Gram wRTG



$$N(w) = N(\text{Gyt}) \cdot N(\text{yta}) \cdot N(\text{ta.}) = \frac{2}{5} \cdot \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{30}$$

Outline

Introduction

Modelling and Training

Decoding

Concepts

IRTG

n -Gram Model

Component Product

Results

n -Gram wRTG

Summary

Summary

$$h_{p_1, p_2}(f) = \operatorname{argmax}_{e \in E} \underbrace{p_1(f \mid e)}_{\text{by IRTG}} \cdot \underbrace{p_2(e)}_{\text{by } n\text{-gram model}}$$

Summary

$$h_{p_1, p_2}(f) = \operatorname{argmax}_{e \in E} \underbrace{p_1(f \mid e)}_{\text{by IRTG}} \cdot \underbrace{p_2(e)}_{\text{by } n\text{-gram model}}$$

IRTG \odot_i n -gram model \subseteq IRTG

Summary

$$h_{p_1, p_2}(f) = \operatorname{argmax}_{e \in E} \underbrace{p_1(f \mid e)}_{\text{by IRTG}} \cdot \underbrace{p_2(e)}_{\text{by } n\text{-gram model}}$$

IRTG \odot_i n -gram model \subseteq IRTG

decompose result:

Summary

$$h_{p_1, p_2}(f) = \operatorname{argmax}_{e \in E} \underbrace{p_1(f \mid e)}_{\text{by IRTG}} \cdot \underbrace{p_2(e)}_{\text{by } n\text{-gram model}}$$

IRTG \odot_i n -gram model \subseteq IRTG

decompose result:

1. $(\cdot)^A$; n -gram model \subseteq wRTG
 - ▶ calculate N when possible
 - ▶ store unprocessed strings and margins in states

Summary

$$h_{p_1, p_2}(f) = \operatorname{argmax}_{e \in E} \underbrace{p_1(f \mid e)}_{\text{by IRTG}} \cdot \underbrace{p_2(e)}_{\text{by } n\text{-gram model}}$$

IRTG \odot_i n -gram model \subseteq IRTG

decompose result:

1. $(\cdot)^A$; n -gram model \subseteq wRTG
 - ▶ calculate N when possible
 - ▶ store unprocessed strings and margins in states
2. $\ln\text{-Hom}^{-1}(\text{wRTG}) \subseteq \text{wRTG}$
 - ▶ usual construction, not covered here

Summary

$$h_{p_1, p_2}(f) = \operatorname{argmax}_{e \in E} \underbrace{p_1(f \mid e)}_{\text{by IRTG}} \cdot \underbrace{p_2(e)}_{\text{by } n\text{-gram model}}$$

IRTG \odot_i n -gram model \subseteq IRTG

decompose result:

1. $(\cdot)^A$; n -gram model \subseteq wRTG
 - ▶ calculate N when possible
 - ▶ store unprocessed strings and margins in states
2. In-Hom $^{-1}$ (wRTG) \subseteq wRTG
 - ▶ usual construction, not covered here
3. wRTG \odot wRTG \subseteq wRTG
 - ▶ usual construction, not covered here

References

-  J. Berstel and C. Reutenauer. "Recognizable formal power series on trees". In: *Theoretical Computer Science* 18.2 (1982), pp. 115–148. ISSN: 0304-3975.
-  David Chiang. "Hierarchical Phrase-Based Translation". In: *Comput. Linguist.* 33.2 (June 2007), pp. 201–228. ISSN: 0891-2017.
-  Zoltán Fülöp, Andreas Maletti, and Heiko Vogler. *Weighted Extended Tree Transducers*. 2010.