Wilker Aziz

Universiteit van Amsterdam w.aziz@uva.nl

April 11, 2016

Decoding for SMT

We will be reasoning

We will be reasoning

making decisions

We will be reasoning

making decisions

about solutions

Decoding for SMT

We will be reasoning

making decisions

about solutions

structures (translations)

Decoding for SMT

We will be reasoning

making decisions

about solutions

structures (translations)

under a statistical model

Decoding for SMT

We will be reasoning

making decisions

about solutions

structures (translations)

under a statistical model

a function / a partial ordering

Keywords

We need to better characterise

Keywords

We need to better characterise

space of solutions get to know the terrain

Keywords

We need to better characterise

- space of solutions get to know the terrain
- statistical model how to assess the importance of things

Keywords

We need to better characterise

- space of solutions get to know the terrain
- statistical model how to assess the importance of things
- decision rules know what you want

Terrain



Terrain



not everywhere is a good place to stand

The importance (or cost) of things



German xmas market

The importance (or cost) of things



Brazilian street market

The importance (or cost) of things





The importance (or cost) of things





Price of product depends on

the product

The importance (or cost) of things

German xmas market



- the product
- brand

The importance (or cost) of things

German xmas market



- the product
- brand
- quality of service

The importance (or cost) of things

Brazilian street market



- the product
- brand
- quality of service

The importance (or cost) of things

Brazilian street market



- the product
- brand
- quality of service
- seller's mood

The importance (or cost) of things

Brazilian street market



- the product
- brand
- quality of service
- seller's mood
- your look

The importance (or cost) of things

Brazilian street market



- the product
- brand
- quality of service
- seller's mood
- your look
- products you acquired from other stalls

The importance (or cost) of things

Brazilian street market



- the product
- brand
- quality of service
- seller's mood
- your look
- products you acquired from other stalls
- from whom you bought those

Know what you want



what do you think when you see this picture?

Know what you want



make a selfie (and lose a couple of fingers in the process)?

Know what you want



French goodies in a cozy chalet?

Content

- 1 Space of translations
- 2 Formal devices
- 3 Linear models
- 4 Decision rules
- **5** Decoding

Model of translational equivalences

Describes the process of generating translations of a given input

constrains and characterises
 the set of possible translation derivations

Model of translational equivalences

Describes the process of generating translations of a given input

constrains and characterises
 the set of possible translation derivations

Phrase-based MT

we observe an input, segment it into phrases, permute the phrases into target language word-order, and finally, translate segments independently

Model of translational equivalences

Describes the process of generating translations of a given input

constrains and characterises
 the set of possible translation derivations

Phrase-based MT

we observe an input, segment it into phrases, permute the phrases into target language word-order, and finally, translate segments independently

Hierarchical MT

we parse the input with a CFG, then translate (using synchronous rules) each and every edge independently

CFGs and FSAs

Compactly represent the set of translations

 keep the representation cost a tractable (polynomial) function of the input length

Phrase-based MT $O(n^22^d)$

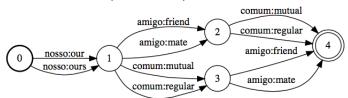
Hierarchical MT $O(n^3)$

Independence assumptions

Translation rules (flat or CFG) are applied independently

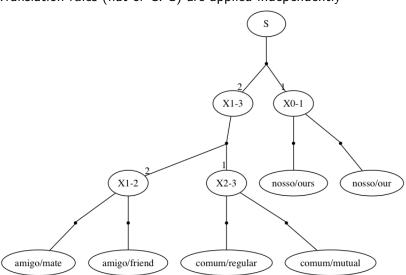
Independence assumptions

Translation rules (flat or CFG) are applied independently



Independence assumptions

Translation rules (flat or CFG) are applied independently



Space of translations Formal devices Linear models Decision rules Decoding Reference

Directed B-hypergraphs

A hypergraph $\langle V, E \rangle$ consists of

- lacksquare a set of nodes V
- lacksquare a set of edges E
- lacksquare an edge e has
 - a head node $head(e) \in V$
 - a tail $tail(e) \in V^*$ (sequence of nodes)

Space of translations Formal devices Linear models Decision rules Decoding Reference

Directed B-hypergraphs

A hypergraph $\langle V, E \rangle$ consists of

- lacksquare a set of nodes V
- a set of edges E
- ullet an edge e has
 - a head node $head(e) \in V$
 - a tail $tail(e) \in V^*$ (sequence of nodes)

CFGs

- nonterminal → node
- terminal → terminal node
- \blacksquare rule \rightarrow edge
- LHS \rightarrow head
- RHS → tail

Directed B-hypergraphs

A hypergraph $\langle V, E \rangle$ consists of

- a set of nodes V
- lacktriangle a set of edges E
- ullet an edge e has
 - a head node $head(e) \in V$
 - a tail $tail(e) \in V^*$ (sequence of nodes)

CFGs

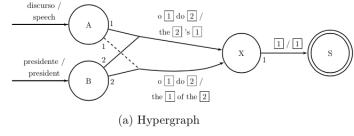
- nonterminal \rightarrow node
- terminal \rightarrow terminal node
- \blacksquare rule \rightarrow edge
- LHS → head
- RHS → tail

FSAs

- state → node
- symbol \rightarrow terminal node
- transition → edge
- origin → tail node
- destination → head

Space of translations Formal devices Linear models Decision rules Decoding References

A forest as a hypergraph



$$\begin{array}{c|cccc} \text{LHS} & \text{RHS}_i & \text{RHS}_o \\ \text{S} \rightarrow & \text{X} & & \boxed{1} \\ \text{X} \rightarrow & \text{o A do B} & \text{the } \boxed{2} \text{ 's } \boxed{1} \\ \text{X} \rightarrow & \text{o A do B} & \text{the } \boxed{1} \text{ of the } \boxed{2} \\ \text{A} \rightarrow & \text{discurso} & \text{speech} \\ \text{B} \rightarrow & \text{presidente} & \text{president} \end{array}$$

(b) Synchronous rules

Weighted sets

A weighted set $\langle \mathcal{D}, \omega \rangle$ consists of

- a set of structures (e.g. hyperpaths/derivations)
- a function $w:\mathcal{D}\to\mathcal{K}$

Let us focus on weighted sets whose weight functions factorise

$$w(\mathbf{d}) = \bigotimes_{e \in \mathbf{d}} w(e)$$

Often the structure is just a means to an end (the yield)

$$w(\mathbf{y}) = \bigoplus_{\mathbf{d} \in \mathcal{D}_{\mathbf{y}}} w(\mathbf{d})$$

An algebraic structure $\mathcal{K} = \langle \mathbb{K}, \oplus, \otimes, \bar{0}, \bar{1} \rangle$

An algebraic structure $\mathcal{K} = \left\langle \mathbb{K}, \oplus, \otimes, \bar{0}, \bar{1} \right\rangle$

• \mathbb{K} is a set (e.g. $\mathbb{N}, \mathbb{R}, \{0, 1\}$)

An algebraic structure $\mathcal{K} = \langle \mathbb{K}, \oplus, \otimes, \bar{0}, \bar{1} \rangle$

- \mathbb{K} is a set (e.g. $\mathbb{N}, \mathbb{R}, \{0, 1\}$)
- ullet \oplus and \otimes are binary operators

An algebraic structure $\mathcal{K} = \langle \mathbb{K}, \oplus, \otimes, \bar{0}, \bar{1} \rangle$

- \mathbb{K} is a set (e.g. $\mathbb{N}, \mathbb{R}, \{0, 1\}$)
- ullet \oplus and \otimes are binary operators
- ullet \oplus is commutative and has identity $ar{0}$ $a\oplus b=b\oplus a$ and $ar{0}\oplus a=a\oplus ar{0}=a$

An algebraic structure $\mathcal{K} = \left\langle \mathbb{K}, \oplus, \otimes, \bar{0}, \bar{1} \right\rangle$

- \mathbb{K} is a set (e.g. $\mathbb{N}, \mathbb{R}, \{0, 1\}$)
- ullet \oplus and \otimes are binary operators
- ullet \oplus is commutative and has identity $ar{0}$ $a\oplus b=b\oplus a$ and $ar{0}\oplus a=a\oplus ar{0}=a$
- \otimes is associative and has identity $\bar{1}$ $(a \otimes b) \otimes c = a \otimes (b \otimes c)$ and $\bar{1} \otimes a = a \otimes 1 = a$

An algebraic structure $\mathcal{K} = \left\langle \mathbb{K}, \oplus, \otimes, \bar{0}, \bar{1} \right\rangle$

- \mathbb{K} is a set (e.g. $\mathbb{N}, \mathbb{R}, \{0, 1\}$)
- ullet \oplus and \otimes are binary operators
- ullet \oplus is commutative and has identity $ar{0}$ $a\oplus b=b\oplus a$ and $ar{0}\oplus a=a\oplus ar{0}=a$
- \otimes is associative and has identity $\bar{1}$ $(a \otimes b) \otimes c = a \otimes (b \otimes c)$ and $\bar{1} \otimes a = a \otimes 1 = a$
- \otimes left distributes over \oplus $a \otimes (b \oplus c) = (a \otimes b) \oplus (a \otimes c)$

An algebraic structure $\mathcal{K} = \left\langle \mathbb{K}, \oplus, \otimes, \bar{0}, \bar{1} \right\rangle$

- \mathbb{K} is a set (e.g. $\mathbb{N}, \mathbb{R}, \{0, 1\}$)
- ullet \oplus and \otimes are binary operators
- ullet \oplus is commutative and has identity $ar{0}$ $a\oplus b=b\oplus a$ and $ar{0}\oplus a=a\oplus ar{0}=a$
- \otimes is associative and has identity $\bar{1}$ $(a \otimes b) \otimes c = a \otimes (b \otimes c)$ and $\bar{1} \otimes a = a \otimes 1 = a$
- \otimes left distributes over \oplus $a \otimes (b \oplus c) = (a \otimes b) \oplus (a \otimes c)$
- $\bar{0}$ is the \otimes -annihilator $\bar{0} \otimes a = a \otimes \bar{0} = \bar{0}$

Examples of semirings

Name	\mathbb{K}	\oplus	\otimes	$\bar{0}$	$\bar{1}$
BINARY	$\{0, 1\}$	V	\wedge	0	1
Counting	\mathbb{N}	+	×	0	1
Prob	$[0,1] \subset \mathbb{R}$	+	×	0	1
LogProb	$\mathbb{R} \cup \{-\infty\}$	\oplus_{\log}	+	$-\infty$	0
Viterbi	$\mathbb{R} \cup \{-\infty\}$	max	+	$-\infty$	0

where $a \oplus_{\log} b = \log(\exp(a) + \exp(b))$

Linear models

$$f(\mathbf{d}) = \mathbf{w}^{\top} \mathbf{\Phi}(\mathbf{d})$$

where

- $\mathbf{w} \in \mathbb{R}^m$
- $\Phi(\mathbf{d}) = \langle \Phi_1(\mathbf{d}), \dots, \Phi_m(\mathbf{d}) \rangle$
- $\Phi_i(\mathbf{d}) \in \mathbb{R}$ is a feature function
- w_i is the relative contribution of the *i*th feature

Linear models and independence assumptions

$$f(\mathbf{d}) = \mathbf{w}^{\top} \mathbf{\Phi}(\mathbf{d})$$

$$= \sum_{i=1}^{m} w_{i} \Phi_{i}(\mathbf{d})$$
(2)

$$=\sum_{i=1}^{m} w_i \prod_{e \in \mathbf{d}} \phi_i(e) \tag{3}$$

$$= \prod_{e \in \mathbf{d}} \sum_{i=1}^{m} w_i \phi_i(e) \tag{4}$$

$$= \prod_{e \in \mathbf{d}} \mathbf{w}^{\top} \phi(e) \tag{5}$$

Assumption

• $\Phi_i(\mathbf{d})$ factorises over edges $\phi_i(e)$ is a local feature function

Linear models and CFGs

Linear models can be expressed through hypergraphs using an appropriate semiring

Decision rules

Best translation (MAP)

$$\mathbf{y}^* = \arg\max_{\mathbf{y}} \sum_{d \in \mathcal{D}_{\mathbf{y}}} f(\mathbf{d})$$

Decision rules

Best translation (MAP)

$$\mathbf{y}^* = \arg\max_{\mathbf{y}} \sum_{d \in \mathcal{D}_{\mathbf{y}}} f(\mathbf{d})$$

Best derivation (Viterbi)

$$\mathbf{y}^* \approx \text{yield} \left\{ \underset{\mathbf{d}}{\text{arg max}} f(\mathbf{d}) \right\}$$

- less disambiguation power
- VITERBI semiring

Other decision rules?

Minimum Bayes risk (MBR)

$$\mathbf{y}^* = \operatorname*{arg\,min}_{\mathbf{y}'} \left\langle L(\mathbf{y}', \mathbf{y}) \right\rangle_{p(\mathbf{y})}$$

- requires the underlying model to have a probabilistic interpretation
- can be estimated through sampling

Space of translations Formal devices Linear models **Decision rules** Decoding Reference

Other decision rules?

Minimum Bayes risk (MBR)

$$\mathbf{y}^* = \operatorname*{arg\,min}_{\mathbf{y}'} \left\langle L(\mathbf{y}', \mathbf{y}) \right\rangle_{p(\mathbf{y})}$$

- requires the underlying model to have a probabilistic interpretation
- can be estimated through sampling

Log-linear models

$$p(\mathbf{d}) = \frac{\exp(f(\mathbf{d}))}{\sum_{\mathbf{d}'} \exp(f(\mathbf{d}'))} \propto \exp\left(\sum_{e \in \mathbf{d}} \mathbf{w}^\top \phi(e)\right) = \prod_{e \in \mathbf{d}} \exp(\mathbf{w}^\top \phi(e))$$

LOGPROB semiring

Decoding

In SMT, decoding typically means the Viterbi approximation

$$\mathbf{d}^* = \arg\max_{\mathbf{d}} f(\mathbf{d})$$

Decoding

In SMT, decoding typically means the Viterbi approximation

$$\mathbf{d}^* = \arg\max_{\mathbf{d}} f(\mathbf{d})$$

If the statistical model $f(\mathbf{d})$ does not violate the independence assumptions of the model of translational equivalences

• steps in a derivation are weighted independently

Decoding

In SMT, decoding typically means the Viterbi approximation

$$\mathbf{d}^* = \arg\max_{\mathbf{d}} f(\mathbf{d})$$

If the statistical model $f(\mathbf{d})$ does not violate the independence assumptions of the model of translational equivalences

- steps in a derivation are weighted independently there is a straightforward (tractable) decomposition of $f(\mathbf{d})$
 - wFSA (phrase-based MT)
 - wCFG (hierarchical MT)

Inside

The Inside recursion can be generalised to an arbitrary semiring

$$\beta(v) = \begin{cases} \bar{1} & \text{if } BS(v) = \emptyset \\ \bigoplus_{e \in BS(v)} w(e) \bigotimes_{u \in \mathrm{tail}(e)} \beta(u) & \text{otherwise} \end{cases}$$

Inside

The Inside recursion can be generalised to an arbitrary semiring

$$\beta(v) = \begin{cases} \bar{1} & \text{if } BS(v) = \emptyset \\ \bigoplus_{e \in BS(v)} w(e) \bigotimes_{u \in \mathrm{tail}(e)} \beta(u) & \text{otherwise} \end{cases}$$

- efficient bottom-up dynamic program O(|G|) |G| is the size of the graphical representation of $f(\mathbf{d})$
 - a lattice (phrase-based MT)
 - a forest (hierarchical MT)

pace of translations Formal devices Linear models Decision rules **Decoding** Reference

Inference

Viterbi derivation

- start from the goal (root)
- $\mathbf{2}$ recursively rewrite every symbol v by solving

$$e = \underset{e \in BS(v)}{\operatorname{arg max}} w(e) \bigotimes_{u \in \operatorname{tail}(e)} \beta(u)$$

Inference

Viterbi derivation

- start from the goal (root)
- $oldsymbol{2}$ recursively rewrite every symbol v by solving

$$e = \underset{e \in BS(v)}{\operatorname{arg max}} w(e) \bigotimes_{u \in \operatorname{tail}(e)} \beta(u)$$

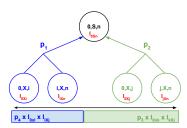
Sampling

- start from the goal (root)
- ${f 2}$ recursively rewrite every symbol v by solving

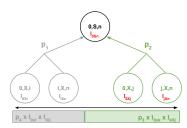
$$e \sim p(e \in BS(v)|v) = \frac{w(e) \bigotimes_{u \in tail(e)} \beta(u)}{\beta(v)}$$

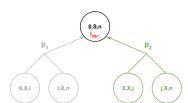


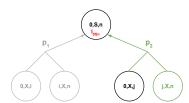
Space of translations Formal devices Linear models Decision rules **Decoding** References



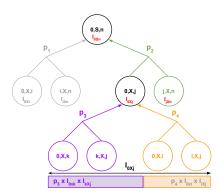
Space of translations Formal devices Linear models Decision rules **Decoding** References



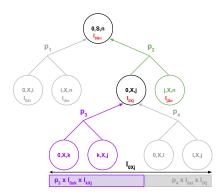




Space of translations Formal devices Linear models Decision rules **Decoding** References



Space of translations Formal devices Linear models Decision rules **Decoding** References

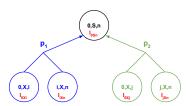


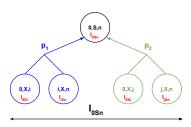
Sampling

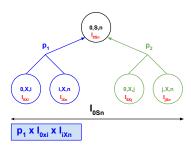


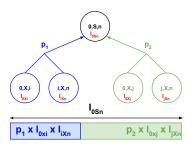
Space of translations Formal devices Linear models Decision rules **Decoding** References

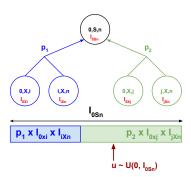
Sampling

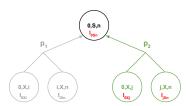


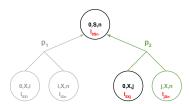


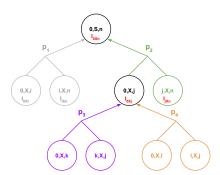


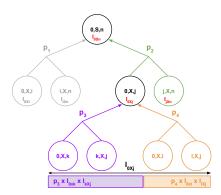


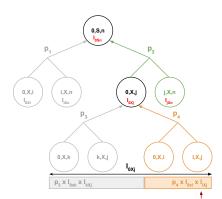












An example for hierarchical models

Model
$$f(\mathbf{d}) = \sum_i \varphi(e_i)$$
 where $\varphi(e_i)$ is a weighted combination of local features

Grammar

$$\begin{array}{l} X \rightarrow \langle \mathsf{a}, \mathsf{the} \rangle \\ X \rightarrow \langle \mathsf{luz}, \mathsf{light} \rangle \\ X \rightarrow \langle \mathsf{apague} \ X_1, \mathsf{switch} \ X_1 \ \mathsf{off} \rangle \\ X \rightarrow \langle X_1 \ \mathsf{por} \ \mathsf{favor}, \mathsf{please} \ , \ X_1 \rangle \\ X \rightarrow \langle X_1 X_2, X_1, X_2 \rangle \\ S \rightarrow \langle \vdash X_1 \dashv, \vdash X_1 \dashv \rangle \end{array}$$

Input: apague a luz por favor

Reference: please, switch the light off

Node	apague	а	luz	por	favor	Inside

Node	apague	a	luz	por	favor	Inside
$X_{1,2}$		$e_1:rac{a}{the}$				

Node	apague	а	luz	por	favor	Inside
$X_{1,2}$ $X_{2,3}$		$e_1: \frac{a}{the}$				
$X_{2,3}$			$e_2:rac{luz}{light}$			

Node	apague	a	luz	por	favor	Inside
$X_{1,2}$		$e_1: rac{a}{the}$				
$X_{1,2}$ $X_{2,3}$			$e_2: rac{luz}{light}$			
$X_{1,3}$		$e_3: \frac{1}{2}$	$\frac{X_{1,2}X_{2,3}}{X_{1,2}X_{2,3}}$			

Node	apague	a	luz	por	favor	Inside
$X_{1,2}$		$e_1: \frac{a}{the}$				
$X_{2,3}$			$e_2:rac{luz}{light}$			
$X_{1,3}$		$e_3:\frac{7}{2}$	$X_{1,2}X_{2,3}$ $X_{1,2}X_{2,3}$			
$X_{0,2}$	e4: apag	gue $X_{1,2}$ h $X_{1,2}$ off				

Node	apague	а	luz	por	favor	Inside
$X_{1,2}$		$e_1: \frac{a}{the}$				
$X_{2,3}$			$e_2: rac{luz}{light}$			
$X_{1,3}$		$e_3:\frac{2}{3}$	$\frac{X_{1,2}X_{2,3}}{X_{1,2}X_{2,3}}$			
$X_{0,2}$	$e_4: rac{apag}{switc}$	gue $X_{1,2}$ th $X_{1,2}$ off				
$X_{2,5}$			e_5	$\frac{X_{2,3} \text{ por fa}}{\text{please, } X_{2,3}}$	vor ,3	

Node	apague	a	luz	por	favor	Inside
$X_{1,2}$		$e_1: \frac{a}{the}$				
$X_{2,3}$			$e_2:rac{luz}{light}$			
$X_{1,3}$		$e_3: = \frac{1}{2}$	$X_{1,2}X_{2,3}$ $X_{1,2}X_{2,3}$			
$X_{0,2}$	$e_4: \frac{apa}{swit}$	gue $X_{1,2}$ ch $X_{1,2}$ off				
$X_{2,5}$			e_5	$: \frac{X_{2,3} \text{ por fa}}{\text{please, } X_{2,3}}$	vor ,3	
$X_{0,3}$	e_6	$\frac{\text{apague } X}{\text{switch } X_1}$. _{1,3} ,3off			

Node	apague	а	luz	por	favor	Inside
$X_{1,2}$		$e_1: \frac{a}{the}$				
$X_{2,3}$			$e_2: rac{luz}{light}$			
$X_{1,3}$		$e_3: \frac{1}{2}$	$\frac{X_{1,2}X_{2,3}}{X_{1,2}X_{2,3}}$			
$X_{0,2}$	$e_4: \frac{apa}{swite}$	gue $X_{1,2}$ th $X_{1,2}$ off				
$X_{2,5}$			e_5	$: \frac{X_{2,3} \text{ por fa}}{\text{please, } X_2}$,3	
$X_{0,3}$	e_6	$ \frac{\text{apague } X}{\text{switch } X_1} $ $ \stackrel{?}{:} \frac{X_{0,2}X_2}{X_{0,2}X_2} $	71,3 ,3 off 2,3 2,3		,	

Node	apague	а	luz	por	favor	Inside
$X_{1,2}$		$e_1: rac{a}{the}$				
$X_{2,3}$			$e_2: rac{luz}{light}$			
$X_{1,3}$		$e_3:$	$X_{1,2}X_{2,3} = X_{1,2}X_{2,3}$			
$X_{0,2}$	e_4 : $\frac{apa}{swit}$	$X_{1,2}$ ch $X_{1,2}$ off				
$X_{2,5}$			e_5	$\frac{X_{2,3} \text{ por fa}}{\text{please, } X_2}$,3	
$X_{0,3}$	ee	$e_7: \frac{\text{apague } X_{1}}{\text{switch } X_{1}}$ $e_7: \frac{X_{0,2}X_{2}}{X_{0,2}X_{2}}$	2,3 2,3			
$X_{1,5}$,	$e_8: rac{X_{1,i}}{ple_i}$	$_3$ por favor ase, $X_{1,3}$		

apague	а	luz	por	favor	Inside
	$e_1: \frac{a}{the}$				
		$e_2: rac{luz}{light}$			
	$e_3: \frac{1}{2}$	$\frac{X_{1,2}X_{2,3}}{X_{1,2}X_{2,3}}$			
$e_4: \frac{apa}{swite}$	gue $X_{1,2}$ ch $X_{1,2}$ off				
		e_5	$\frac{X_{2,3} \text{ por fa}}{\text{please, } X_2}$	vor ,3	
e_6	$\frac{\text{apague } X}{\text{switch } X_1}$	71,3 ,3 off			
($e_7: \frac{X_{0,2}X_2}{X_{0,2}X_2}$	2,3 2,3			
		$e_8: \frac{X_{1,3}}{plea}$	$\frac{1}{2}$ por favor ase, X_{13}		
	e4: apa	$e_1:rac{ extbf{a}}{ ext{the}}$ $e_3:rac{ extbf{a}}{ ext{switch }X_{1,2} ext{off}}$	$\begin{array}{c c} e_1: \frac{a}{b} \\ & e_2: \frac{luz}{light} \\ & e_3: \frac{X_{1,2}X_{2,3}}{X_{1,2}X_{2,3}} \\ e_4: \frac{apague\ X_{1,2}}{switch\ X_{1,2}off} \\ & e_5 \\ & e_6: \frac{apague\ X_{1,3}}{switch\ X_{1,3}off} \\ & e_7: \frac{X_{0,2}X_{2,3}}{X_{0,2}X_{2,3}} \\ & e_8: \frac{x_{1,1}}{per} \end{array}$	$e_1: \frac{a}{the}$ $e_2: \frac{luz}{light}$ $e_3: \frac{X_{1,2}X_{2,3}}{X_{1,2}X_{2,3}}$ $e_4: \frac{apague\ X_{1,2}}{switch\ X_{1,2}off}$ $e_5: \frac{X_{2,3}\ por\ fa}{please,\ X_2}$	$e_{1}:\frac{a}{\text{the}}$ $e_{2}:\frac{\text{luz}}{\text{light}}$ $e_{3}:\frac{X_{1,2}X_{2,3}}{X_{1,2}X_{2,3}}$ $e_{4}:\frac{\text{apague }X_{1,2}}{\text{switch }X_{1,2}\text{off}}$ $e_{5}:\frac{X_{2,3}\text{ por favor please, }X_{2,3}}{\text{polyation }X_{1,3}\text{ off}}$ $e_{7}:\frac{X_{0,2}X_{2,3}}{X_{0,2}X_{2,3}}$ $e_{8}:\frac{X_{1,3}\text{ por favor please, }X_{1,3}}{\text{polyation }X_{1,3}}$

Node	apague	a	luz	por	favor	Inside
$X_{1,2}$		$e_1: \frac{a}{the}$				
$X_{2,3}$			$e_2:rac{luz}{light}$			
$X_{1,3}$		$e_3:\frac{2}{3}$	$X_{1,2}X_{2,3}$ $X_{1,2}X_{2,3}$			
$X_{0,2}$	e_4 : $\frac{apa}{swit}$	$X_{1,2}$ ch $X_{1,2}$ off				
$X_{2,5}$			e_5	$\frac{X_{2,3} \text{ por fa}}{\text{please, } X_2}$,3	
$X_{0,3}$	e_{ϵ}	$\frac{\text{apague } X}{\text{switch } X_1}$	1,3 ,3 off		,	
110,3		$e_7: \frac{X_{0,2}X_2}{X_{0,2}X_2}$	2 <u>,3</u> 2.3			
$X_{1,5}$			$e_8: \frac{X_{1,i}}{ple_8}$ $e_9: \frac{X}{X}$	$_3$ por favor ase, $X_{1,3}$		
211,5						
			$e_{10}: \frac{X_{0,2}X_{2,2}}{X_{0,2}X_{2,2}}$	<u>,5</u> ,5		
$X_{0,5}$						

Node	apague	а	luz	por	favor	Inside
$X_{1,2}$		$e_1: \frac{a}{the}$				
$X_{2,3}$			$e_2:rac{luz}{light}$			
$X_{1,3}$		$e_3:\frac{2}{2}$	$X_{1,2}X_{2,3}$ $X_{1,2}X_{2,3}$			
$X_{0,2}$	$e_4: \frac{apa}{swit}$	$X_{1,2}$ ch $X_{1,2}$ off				
$X_{2,5}$			e_5	$\frac{X_{2,3} \text{ por fa}}{\text{please, } X_{2,3}}$	vor ,3	
$X_{0,3}$	e_{ϵ}	$3: \frac{\text{apague } X}{\text{switch } X_1}$	1,3 ,3 off			
		$e_7: \frac{X_{0,2}X_2}{X_{0,2}X_2}$	1,3 1,3			
$X_{1,5}$			$e_8: rac{X_{1,3}}{plea}$ $e_9: rac{X_1}{X_1}$	$\frac{1}{1}$ por favor $\frac{1}{1}$		
211,5						
$X_{0,5}$						
			$e_{10}: \frac{X_{0,2}X_{2,1}}{X_{0,2}X_{2,1}}$ $1: \frac{apague\ X_{1}}{switch\ X_{1,5}}$			

Node	apague	а	luz	por	favor	Inside			
$X_{1,2}$		$e_1: \frac{a}{the}$							
$X_{2,3}$			$e_2:rac{luz}{light}$						
$X_{1,3}$		$e_3:\frac{7}{7}$	$X_{1,2}X_{2,3}$ $X_{1,2}X_{2,3}$						
$X_{0,2}$	$e_4: \frac{apa}{swit}$	ague $X_{1,2}$ ich $X_{1,2}$ off							
$X_{2,5}$				$: \frac{X_{2,3} \text{ por fa}}{\text{please, } X_2}$	vor ,3				
$X_{0,3}$	e_{ℓ}	$\frac{\text{apague } X}{\text{switch } X_1}$	1,3 ,3 off						
		$e_7: \frac{X_{0,2}X_2}{X_{0,2}X_2}$,3						
$X_{1,5}$			$e_8: \frac{X_{1,3}}{plea}$	$\frac{\text{por favor}}{\text{se, } X_{1,3}}$					
1,0			$e_9: \frac{X_1}{X_1}$	$\frac{,_2X_{2,5}}{,_2X_{2,5}}$					
		•	$e_{10}: rac{X_{0,2}X_{2,}}{X_{0,2}X_{2,}}$	<u>5</u>					
$X_{0,5}$		$e_{10}: rac{X_{0,2}X_{2,5}}{X_{0,2}X_{2,5}} \ e_{11}: rac{apague\ X_{1,5}}{switch\ X_{1,5}\ off} \ e_{12}: rac{X_{0,3}\ por\ favor}{please,\ X_{0,3}}$							
		e_1	$_2: \frac{X_{0,3} \text{ por fa}}{\text{please, } X_0}$,3					

Node	apague	а	luz	por	favor	Inside			
$X_{1,2}$		$e_1: rac{a}{the}$							
$X_{2,3}$			$e_2: rac{luz}{light}$						
$X_{1,3}$			$\frac{X_{1,2}X_{2,3}}{X_{1,2}X_{2,3}}$						
$X_{0,2}$	$e_4: \frac{apa}{swit}$	ague $X_{1,2}$ cch $X_{1,2}$ off							
$X_{2,5}$			e_5	$: \frac{X_{2,3} \text{ por fa}}{\text{please, } X_2}$,3				
$X_{0,3}$	e	$e_7: \frac{\text{apague } X}{\text{switch } X_1}$ $e_7: \frac{X_{0,2}X_2}{X_{0,2}X_2}$	1,3 ,3 off 2,3						
$X_{1,5}$			$e_8: rac{X_{1,3}}{plea}$ $e_9: rac{X_1}{X_1}$	$\frac{1,2X_{2,5}}{2X_{2,5}}$					
TZ.			$e_{10}: \frac{X_{0,2}X_{2,}}{X_{0,2}X_{2,}}$	<u>5</u> 5					
$X_{0,5}$		$e_{11}: \frac{apague\ X_{1,5}}{switch\ X_{1,5}\ off}$							
		e_1	$X_{0,3}$ por far please, X_0	,3					
$S_{0,5}$			$e_{13}: \frac{\vdash X_{0,5}}{\vdash X_{0,5}}$						

Node	apague	а	luz	por	favor	Inside
$X_{1,2}$		$e_1: \frac{a}{the}$				$w(e_1)$
$X_{2,3}$			$e_2: rac{luz}{light}$			
$X_{1,3}$			$X_{1,2}X_{2,3}$ $X_{1,2}X_{2,3}$			
$X_{0,2}$	$e_4: \frac{apa}{swit}$	gue $X_{1,2}$ ch $X_{1,2}$ off				
$X_{2,5}$			e_5	$\frac{X_{2,3} \text{ por fa}}{\text{please, } X_{2,3}}$	vor ,3	
$X_{0,3}$	e6	$e_7: \frac{\text{apague } X}{\text{switch } X_1}$ $e_7: \frac{X_{0,2}X_2}{X_{0,2}X_2}$	1,3 ,3 off 2,3			
$X_{1,5}$			$e_8:rac{X_{1,3}}{plea}$ $e_9:rac{X_1}{X_1}$	$\frac{1,2X_{2,5}}{2X_{2,5}}$		
$X_{0,5}$		P1	$e_{10}: \frac{X_{0,2}X_{2,}}{X_{0,2}X_{2,}}$ apague X_1	5 5 ,5_		
		e_1	switch $X_{1,5}$ $2 : \frac{X_{0,3} \text{ por fa}}{\text{please, } X_0}$			
$S_{0,5}$			$e_{13}: \frac{\vdash X_{0,5}}{\vdash X_{0,5}}$	Ī		

Node	apague	а	luz	por	favor	Inside
$X_{1,2}$		$e_1: rac{a}{the}$				$w(e_1)$
$X_{2,3}$			$e_2: rac{luz}{light}$			$w(e_2)$
$X_{1,3}$			$X_{1,2}X_{2,3}$ $X_{1,2}X_{2,3}$			
$X_{0,2}$	$e_4: \frac{apa}{swit}$	ague $X_{1,2}$ ch $X_{1,2}$ off				
$X_{2,5}$			e_5	$: \frac{X_{2,3} \text{ por fa}}{\text{please, } X_2}$	vor .3	
$X_{0,3}$	e	$e_7: \frac{\text{apague } X}{\text{switch } X_1}$ $e_7: \frac{X_{0,2}X_2}{X_{0,2}X_2}$	1,3 ,3 off 2,3			
$X_{1,5}$			$e_8:rac{X_{1,3}}{plea}$ $e_9:rac{X_1}{X_1}$	$\frac{1,2X_{2,5}}{2X_{2,5}}$		
$X_{0,5}$		e_1	$e_{10}: rac{X_{0,2}X_{2,}}{X_{0,2}X_{2,}}$ 1: $rac{apague\ X_{1}}{switch\ X_{1,5}}$ 2: $rac{X_{0,3}\ por\ fa}{please,\ X_{0}}$	5 5 ,5 , off ivor		
$S_{0,5}$			$e_{13}: \frac{\vdash X_{0,5}}{\vdash X_{0,5}}$	i		

Node	apague	а	luz	por	favor	Inside
$X_{1,2}$		$e_1: \frac{a}{the}$				$w(e_1)$
$X_{2,3}$			$e_2:rac{luz}{light}$			$w(e_2)$
$X_{1,3}$		$e_3: \frac{\lambda}{\lambda}$	$X_{1,2}X_{2,3}$ $X_{1,2}X_{2,3}$			$w(e_3)\beta(X_{1,2})\beta(X_{2,3})$
$X_{0,2}$	$e_4: rac{apa}{swit}$	ague $X_{1,2}$ sch $X_{1,2}$ off				
$X_{2,5}$			e_5	$\frac{X_{2,3} \text{ por for please, } X_2}{\text{please, } X_2}$	avor 1,3	
$X_{0,3}$	e_{ℓ}	$e_7: \frac{\frac{\text{apague } X_1}{\text{switch } X_1}}{\frac{X_0, 2}{X_2}}$	1,3 3 off ,3			
$X_{1,5}$			$e_8: \frac{X_{1,\cdot}}{ple}$ $e_9: \frac{X}{X}$	3 por favor ase, $X_{1,3}$ $1,2X_{2,5}$ $1,2X_{2,5}$		
$X_{0,5}$		e_{11} e_{11}	$R_{10}: \frac{X_{0,2}X_2}{X_{0,2}X_2}$ $1: \frac{apague\ X_1}{switch\ X_{1,1}}$ $2: \frac{X_{0,3}\ por\ f}{please,\ X_{0,1}}$,5 ,5 1,5 ₅ off avor		
$S_{0,5}$			$e_{13}: \frac{\vdash X_{0,5}}{\vdash X_{0,5}}$	ਜ਼ ਜ		

Node	apague	а	luz	por	favor	Inside
$X_{1,2}$		$e_1: rac{a}{the}$				$w(e_1)$
$X_{2,3}$			$e_2:rac{luz}{light}$			$w(e_2)$
$X_{1,3}$			$X_{1,2}X_{2,3}$ $X_{1,2}X_{2,3}$			$w(e_3)\beta(X_{1,2})\beta(X_{2,3})$
$X_{0,2}$	$e_4: \frac{apa}{swit}$	ague $X_{1,2}$ ch $X_{1,2}$ off				$w(e_4)\beta(X_{1,2})$
$X_{2,5}$			e_5	$\frac{X_{2,3} \text{ por fa}}{\text{please, } X_2}$	avor .3	
$X_{0,3}$	e	$e_7: \frac{\text{apague } X}{\text{switch } X_1,}$ $e_7: \frac{X_{0,2}X_2}{X_{0,2}X_2}$	1,3 3 off ,3		,-	
$X_{1,5}$			$e_8: rac{X_{1,i}}{ple}$ $e_9: rac{X}{X}$	$\frac{1,2X_{2,5}}{1,2X_{2,5}}$		
$X_{0,5}$		e_1 : e_1	$rac{X_{0,2}X_{2}}{X_{0,2}X_{2}}$ 1: $rac{ ext{apague } X_{1}}{ ext{switch } X_{1,1}}$ 2: $rac{X_{0,3} ext{ por f}}{ ext{please, } X_{0,1}}$	1,5 5 off avor).3		
$S_{0,5}$			$e_{13}: \frac{\vdash X_{0,5}}{\vdash X_{0,5}}$	1		

Node	apague	а	luz	por	favor	Inside
$X_{1,2}$		$e_1: rac{a}{the}$				$w(e_1)$
$X_{2,3}$			$e_2:rac{luz}{light}$			$w(e_2)$
$X_{1,3}$			$X_{1,2}X_{2,3}$ $X_{1,2}X_{2,3}$			$w(e_3)\beta(X_{1,2})\beta(X_{2,3})$
$X_{0,2}$	$e_4: \frac{apa}{swit}$	ague $X_{1,2}$ ch $X_{1,2}$ off				$w(e_4)\beta(X_{1,2})$
$X_{2,5}$			e_5	$\frac{X_{2,3} \text{ por fa}}{\text{please, } X_2}$,3	$w(e_5)eta(X_{2,3})$
$X_{0,3}$	e	$e_7: \frac{\text{apague } X}{\text{switch } X_{1,}}$ $e_7: \frac{X_{0,2}X_2}{X_{0,2}X_2}$	1,3 3off ,3			
$X_{1,5}$			$e_8: \frac{X_{1,3}}{ple_3}$ $e_9: \frac{X_{1,3}}{X_{1,3}}$	$\frac{1,2X_{2,5}}{1.2X_{2,5}}$		
$X_{0,5}$		e_1 : e_1	$rac{X_{0,2}X_{2,2}}{X_{0,2}X_{2,2}}$ 1: $rac{ ext{apague } X_{1,2}}{ ext{switch } X_{1,1}}$ 2: $rac{X_{0,3} ext{ por f}}{ ext{please, } X_{0,2}}$,5 ,5 1,5 5 off avor		
$S_{0,5}$			$e_{13}: \frac{\vdash X_{0,5}}{\vdash X_{0,5}}$	4 -		

Node	apague	а	luz	por	favor	Inside
$X_{1,2}$		$e_1: rac{a}{the}$				$w(e_1)$
$X_{2,3}$			$e_2: rac{luz}{light}$			$w(e_2)$
$X_{1,3}$			$\frac{1,_2X_{2,3}}{1,_2X_{2,3}}$			$w(e_3)\beta(X_{1,2})\beta(X_{2,3})$
$X_{0,2}$	$e_4: \frac{apa}{swit}$	ague $X_{1,2}$ ch $X_{1,2}$ off				$w(e_4)\beta(X_{1,2})$
$X_{2,5}$				$: \frac{X_{2,3} \text{ por f}}{\text{please, } X_2}$	avor 2,3	$w(e_5)\beta(X_{2,3})$
$X_{0,3}$	e	$e_7: \frac{\frac{\text{apague } X_1}{\text{switch } X_1,}}{\frac{X_0, 2X_2}{X_0, 2X_2}}$	1,3_ 3 off ,3_			$w(e_6)\beta(X_{1,3}) \oplus \\ w(e_7)\beta(X_{0,2})\beta(X_{2,3})$
$X_{1,5}$		$X_{0,2}X_{2,2}$	$e_8: \frac{X_{1,3}}{ple}$	3 por favor ase, $X_{1,3}$		w(01)p(110,2)p(112,3)
		ϵ	$e_9: \frac{X}{X}$ $e_{10}: \frac{X_{0,2}X_{2}}{X_{0,2}X_{2}}$	$\frac{1,2}{1,2} \frac{2,5}{1,2} \frac{5}{1,2}$		
$X_{0,5}$		e_{11}				
		e_{1}	$_2:rac{X_{0,3}\ por\ f}{please,\ X_0}$	o,3		
$S_{0,5}$			$e_{13}: \frac{\vdash X_{0,5}}{\vdash X_{0,5}}$	1 1		

Node	apague	а	luz	por	favor	Inside
$X_{1,2}$		$e_1: \frac{a}{the}$				$w(e_1)$
$X_{2,3}$			$e_2: rac{luz}{light}$			$w(e_2)$
$X_{1,3}$			$\frac{1,_2X_{2,3}}{1,_2X_{2,3}}$			$w(e_3)\beta(X_{1,2})\beta(X_{2,3})$
$X_{0,2}$	$e_4: \frac{apa}{swit}$	ague $X_{1,2}$ ch $X_{1,2}$ off				$w(e_4)\beta(X_{1,2})$
$X_{2,5}$				$: \frac{X_{2,3} \text{ por fa}}{\text{please, } X_2}$,3	$w(e_5)\beta(X_{2,3})$
$X_{0,3}$	e	$\frac{\text{apague } X_1}{\text{switch } X_1}$	1 <u>,3</u> 3 off			$w(e_6)eta(X_{1,3})\oplus$
210,3		$e_7: \frac{X_{0,2}X_{2,2}}{X_{0,2}X_{2,2}}$.3			$w(e_7)\beta(X_{0,2})\beta(X_{2,3})$
$X_{1,5}$			$e_8: \frac{X_{1,i}}{ple}$	3 por favor ase, $X_{1,3}$		$w(e_8)eta(X_{1,3})\oplus$
111,5			$e_9: \frac{X}{X}$	$\frac{1,2X_{2,5}}{1,2X_{2,5}}$		$w(e_9)\beta(X_{1,2})\beta(X_{2,5})$
		ϵ	$z_{10}: \frac{X_{0,2}X_2}{X_{0,2}X_2}$	<u>,5</u> ,5		
$X_{0,5}$		e_{11}				
		e_1 :	$_2: \frac{X_{0,3} \text{ por f}}{\text{please, } X_0}$	avor 0,3		
$S_{0,5}$			$e_{13}: \frac{\vdash X_{0,5}}{\vdash X_{0,5}}$	- 1		

Node	apague	а	luz	por	favor	Inside
$X_{1,2}$		$e_1: rac{a}{the}$				$w(e_1)$
$X_{2,3}$			$e_2:rac{luz}{light}$			$w(e_2)$
$X_{1,3}$			$X_{1,2}X_{2,3}$ $X_{1,2}X_{2,3}$			$w(e_3)\beta(X_{1,2})\beta(X_{2,3})$
$X_{0,2}$	$e_4: \frac{apa}{swit}$	ague $X_{1,2}$ sch $X_{1,2}$ off				$w(e_4)\beta(X_{1,2})$
$X_{2,5}$			e_5	$: \frac{X_{2,3} \text{ por fa}}{\text{please, } X_2}$	vor ,3	$w(e_5)\beta(X_{2,3})$
$X_{0,3}$	e_{ϵ}	$_{5}: \frac{apague\ X}{switch\ X_{1}}$	1,3 3 o ff			$w(e_6)eta(X_{1,3})\oplus$
210,3		$e_7: \frac{X_{0,2}X_2}{X_{0,2}X_2}$				$w(e_7)\beta(X_{0,2})\beta(X_{2,3})$
$X_{1,5}$			$e_8: \frac{X_{1,}}{ple}$	$_3$ por favor ase, $X_{1,3}$		$w(e_8)eta(X_{1,3})\oplus$
1,5			$e_9: \frac{X}{X}$	$\frac{1,2X_{2,5}}{1,2X_{2,5}}$		$w(e_9)\beta(X_{1,2})\beta(X_{2,5})$
		($\epsilon_{10}: \frac{X_{0,2}X_2}{X_{0,2}X_2}$	<u>,5</u> ,5		$w(e_{10})eta(X_{0,2})eta(X_{2,5})\oplus$
$X_{0,5}$		e_1	$w(e_{11})\beta(X_{1,5})\oplus$			
		e_1	$_2: \frac{X_{0,3} \text{ por f}}{\text{please, } X_0}$	avor 0,3		$w(e_{12})\beta(X_{0,3})$
$S_{0,5}$			$e_{13}: \frac{\vdash X_{0,5}}{\vdash X_{0,5}}$	1 1		

Node	apague	a	luz	por	favor	Inside
$X_{1,2}$		$e_1: rac{a}{the}$				$w(e_1)$
$X_{2,3}$			$e_2:rac{luz}{light}$			$w(e_2)$
$X_{1,3}$			$X_{1,2}X_{2,3}$ $X_{1,2}X_{2,3}$			$w(e_3)\beta(X_{1,2})\beta(X_{2,3})$
$X_{0,2}$	$e_4: \frac{ap}{swit}$	ague $X_{1,2}$ tch $X_{1,2}$ off				$w(e_4)\beta(X_{1,2})$
$X_{2,5}$			e_5	$: \frac{X_{2,3} \text{ por for please, } X_2}{\text{please, } X_2}$	3,3	$w(e_5)\beta(X_{2,3})$
$X_{0,3}$	e	$6 : \frac{\text{apague } X_1}{\text{switch } X_1}$	1,3 3 o ff			$w(e_6)eta(X_{1,3})\oplus$
110,3		$e_7: \frac{X_{0,2}X_2}{X_{0,2}X_2}$				$w(e_7)\beta(X_{0,2})\beta(X_{2,3})$
$X_{1,5}$			$e_8: \frac{X_{1,i}}{ple}$	$_3$ por favor ase, $X_{1,3}$		$w(e_8)eta(X_{1,3})\oplus$
1,5			$e_9: \frac{X}{X}$			$w(e_9)\beta(X_{1,2})\beta(X_{2,5})$
		ϵ	$\epsilon_{10}: \frac{X_{0,2}X_2}{X_{0,2}X_2}$	<u>,5</u> ,5		$w(e_{10})eta(X_{0,2})eta(X_{2,5})\oplus$
$X_{0,5}$		e_{11}	$1: \frac{apague\ X}{switch\ X_{1,1}}$	1,5 5 off		$w(e_{11})\beta(X_{1,5})\oplus$
		e_1 :	$_2: \frac{X_{0,3} \text{ por f}}{\text{please, } X_0}$			$w(e_{12})\beta(X_{0,3})$
$S_{0,5}$			$e_{13}: \frac{\vdash X_{0,5}}{\vdash X_{0,5}}$	- 1		$w(e_{13})\beta(X_{0,5})$

The problem

Most interesting models employ nonlocal features!

- reordering model: previously translated span
- language model: generated strings

The problem

Most interesting models employ nonlocal features!

- reordering model: previously translated span
- language model: generated strings

Example

$$f(\mathbf{d}) = \psi(\text{yield}(\mathbf{d})) + \sum_{i} \varphi(e_i)$$

where
$$\psi(\mathbf{y}) = w_{\psi} \log p_{\mathrm{LM}}(\mathbf{y})$$

and $p_{\mathrm{LM}}(\mathbf{y}) = \prod_{i} p(y_{i}|y_{i-n+1}^{i-1})$ is an n -gram LM

The problem

Most interesting models employ nonlocal features!

- reordering model: previously translated span
- language model: generated strings

Example

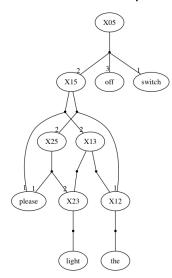
$$f(\mathbf{d}) = \psi(\text{yield}(\mathbf{d})) + \sum_{i} \varphi(e_i)$$

where
$$\psi(\mathbf{y}) = w_{\psi} \log p_{\mathrm{LM}}(\mathbf{y})$$

and $p_{\mathrm{LM}}(\mathbf{y}) = \prod_{i} p(y_{i}|y_{i-n+1}^{i-1})$ is an n -gram LM

• $p_{\rm LM}$ violates independence assumptions

Illustration of the problem



How do we score the top edge?

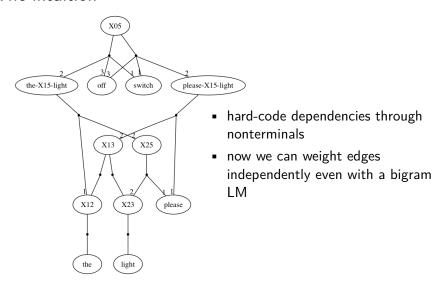
- [switch [please [[the][light]]] off]
- [switch [[the][please [light]]] off]

The solution

"Hard-code" structural dependencies

- disambiguate nodes w.r.t. the context they offer to feature functions
- intuition: we will be "splitting" nodes
- more intuition: nodes must memorise how to complete boundary n-grams

The intuition



An example for hierarchical models

Model
$$f(\mathbf{d}) = \psi(\mathrm{yield}(\mathbf{d})) \sum_i \varphi(e_i)$$
 where $\varphi(e_i)$ is a weighted combination of local features $\psi(\mathrm{yield}(\mathbf{d}))$ contains a 3-gram LM i.e, $p_{\mathrm{LM}_3}(\mathbf{y}) = \prod_i p(y_i|y_{i-2}y_{i-1})$

Grammar

$$\begin{array}{l} X \to \langle \mathsf{a}, \mathsf{the} \rangle \\ X \to \langle \mathsf{luz}, \mathsf{light} \rangle \\ X \to \langle \mathsf{apague} \ X_1, \mathsf{switch} \ X_1 \ \mathsf{off} \rangle \\ X \to \langle X_1 \ \mathsf{por} \ \mathsf{favor}, \mathsf{please} \ , \ X_1 \rangle \\ X \to \langle X_1 X_2, X_1, X_2 \rangle \\ S \to \langle \vdash X_1 \dashv, \vdash X_1 \dashv \rangle \end{array}$$

Input: apague a luz por favor

Reference: please, switch the light off

apague	a	luz	por	favor	Left	Right	Node
							1
							2
							3
							4
							5
							6
							7
							8
							9
							10
							11
							12
							13
							14
							15
							16
·							17
							18

	apague	a	luz	por	favor	Left	Right	Node
$X_{1,2}$		$e_1:rac{ extsf{a}}{ extsf{the}}$				the	the	1
		tile						2
								3
								4
								5
								6
								7
								8
								9
								10
								11
								12
								13
								14
								15
								16
								17
								18

	apague	a	luz	por	favor	Left	Right	Node
$X_{1,2}$		$e_1: \frac{a}{the}$		_		the	the	1
$X_{2,3}$			$e_2:rac{luz}{light}$			light	light	2
								3
								4
								5
								6
								7
								8
								9
								10
								11
								12
								13
								14
								15
								16
								17
								18

	apague	a	luz	por	favor	Left	RIGHT	Node
$X_{1,2}$		$e_1: \frac{a}{the}$				the	the	1
$X_{2,3}$		the	$e_2: \frac{luz}{light}$			light	light	2
$X_{1,3}$		$e_3:\frac{1}{2}$	$e_2: \frac{\frac{luz}{light}}{X_{1,2}X_{2,3}}$ $(1) (2)$			the light	the light	3
			\					4
								5
								6
								7
								8
								9
								10
								11
								12
								13
								14
								15
								16
								17
								18

	apague	a	luz	por	favor	Left	RIGHT	Node
$X_{1,2}$		$e_1: \frac{a}{the}$		•		the	the	1
$X_{2,3}$			$e_2: rac{luz}{light}$			light	light	2
$X_{1,3}$		$e_3 : \frac{1}{2}$	$X_{1,2}X_{2,3} \over (1) (2)$			the light	the light	3
$X_{0,2}$	$e_4:rac{apa_{part}}{swit}$	gue $X_{1,2}$	(-) (-)			switch the	the off	4
	30010	cii (1) oii						5
								6
								7
								8
								9
								10
								11
								12
								13
								14
								15
								16
								17
								18

	apague	a	luz	por	favor	Left	Right	Node
$X_{1,2}$		$e_1:rac{a}{the}$				the	the	1
$X_{2,3}$			$e_2:rac{luz}{light}$			light	light	2
$X_{1,3}$		e_3 :	$\frac{X_{1,2}X_{2,3}}{(1)(2)}$			the light	the light	3
$X_{0,2}$	e_4 : $\frac{ap}{sw}$	vitch (1) off				switch the	the off	4
$X_{2,5}$			e_5	$: \frac{X_{2,3} \text{ por for please, } (2)}{\text{please, } (2)}$	avor 2)	please ,	, light	5
								6
								7
								8
								9
								10
								11
								12
								13
								14
								15
								16
								17
								18
							<u> </u>	

	Г							
	apague	a	luz	por	favor	Left	Right	Node
$X_{1,2}$		$e_1: \frac{a}{the}$				the	the	1
$X_{2,3}$			$e_2:rac{luz}{light}$			light	light	2
$X_{1,3}$		$e_3: \frac{1}{2}$	$X_{1,2}X_{2,3} \over (1) (2)$			the light	the light	3
$X_{0,2}$	$e_4:rac{apage}{switch}$	$\frac{ue\ X_{1,2}}{h\ (1)\ off}$				switch the	the off	4
$X_{2,5}$				$\therefore \frac{X_{2,3} \text{ por far}}{\text{please, } (2)}$	/or	please ,	, light	5
$X_{0,3}$	e_6	: $\frac{\text{apague } X}{\text{switch } (3)}$	1,3 off			switch the	light off	6
0,3								7
								8
								9
								10
								11
								12
								13
								14
								15
								16
					•			17
								18

	apague	а	luz	por	favor	Left	Right	Node
$X_{1,2}$		$e_1: \frac{a}{the}$				the	the	1
$X_{2,3}$			$e_2: rac{luz}{light}$			light	light	2
$X_{1,3}$		$e_3: \frac{7}{2}$	$\frac{X_{1,2}X_{2,3}}{(1)(2)}$			the light	the light	3
$X_{0,2}$	$e_4: \frac{apa}{swi}$	ague $X_{1,2}$ itch (1) off				switch the	the off	4
$X_{2,5}$			e_5	$X_{2,3}$ por far please, (2)	avor !)	please ,	, light	5
$X_{0,3}$	e	$e_6: rac{{\sf apague}\; X}{{\sf switch}\; (3)} \ e_7: rac{X_{0,2}X_2}{(4)\; (2)}$	1,3) off			switch the	light off	6
110,5		$e_7: \frac{X_{0,2}X_2}{(4)(2)}$)			switch the	off light	7
								8
								9
								10
								11
								12
								13
								14
								15
								16
								17
								18
			·					

			l=		farran	LEFT	Droum	Nonn
	apague	a	luz	por	favor	LEFT	Right	Node
$X_{1,2}$		$e_1: rac{a}{the}$				the	the	1
$X_{2,3}$			$e_2:rac{luz}{light}$			light	light	2
$X_{1,3}$		$e_3: \frac{7}{2}$	$\begin{array}{c} e_2: \frac{luz}{light} \\ X_{1,2}X_{2,3} \\ \hline (1) \ (2) \end{array}$			the light	the light	3
$X_{0,2}$	$e_4:rac{apa}{swit}$	gue $X_{1,2}$ ch (1) off				switch the	the off	4
$X_{2,5}$			e_5	$: \frac{X_{2,3} \text{ por for please, } (2)}{\text{please, } (2)}$	avor ?)	please ,	, light	5
$X_{0,3}$	e_{ϵ}	$e_7: rac{ ext{apague } X}{ ext{switch } (3)} \\ e_7: rac{X_{0,2}X_2}{(4) (2)}$	1,3) off			switch the	light off	6
0,5	($e_7: \frac{X_{0,2}X_2}{(4)(2)}$)			switch the	off light	7
$X_{1,5}$			$e_8: \frac{X_{1,3}}{ple}$	ase, (3)		please ,	the light	8
1,0								9
								10
								11
								12
								13
								14
								15
								16
								17
								18

			l		£	T ppm	Drawn	Mann
	apague	a	luz	por	favor	Left	Right	Node
$X_{1,2}$		$e_1: \frac{a}{the}$				the	the	1
$X_{2,3}$			$e_2:rac{luz}{light}$			light	light	2
$X_{1,3}$			$\frac{X_{1,2}X_{2,3}}{(1)(2)}$			the light	the light	3
$X_{0,2}$	$e_4:rac{apag}{switc}$	$\frac{ue\ X_{1,2}}{h\ (1)\ off}$				switch the	the off	4
$X_{2,5}$			e_5	$\frac{X_{2,3} \text{ por f}}{\text{please, (2)}}$	avor 2)	please ,	, light	5
$X_{0,3}$	e_6	$ \frac{\text{apague } X}{\text{switch (3)}} $ $ 7 : \frac{X_{0,2}X_2}{(4) (2)} $	1,3) off			switch the	light off	6
0,5	e_{i}	$7: \frac{X_{0,2}X_2}{(4)(2)}$)			switch the	off light	7
$X_{1,5}$			$e_8:rac{X_{1,3}}{ple}$ $e_9:rac{X_1}{(1)}$	ase, (3)		please ,	the light	8
111,0			$e_9: \frac{X_1}{()}$	$\frac{1,2}{1}$ $\frac{1}{1}$ $\frac{2}{1}$ $\frac{1}{1}$ $\frac{1}{1}$		the please	, light	9
								10
								11
								12
								13
								14
								15
								16
								17
								18

	apague	a	luz	por	favor	Left	Right	Node
$X_{1,2}$		$e_1: \frac{a}{the}$				the	the	1
$X_{2,3}$			$e_2: rac{luz}{light}$			light	light	2
$X_{1,3}$			$\frac{X_{1,2}X_{2,3}}{(1)(2)}$			the light	the light	3
$X_{0,2}$	$e_4:rac{apag}{switc}$	$\frac{ue\ X_{1,2}}{h\ (1)\ off}$				switch the	the off	4
$X_{2,5}$			e_5	$: \frac{X_{2,3} \text{ por for please, } (2)}{\text{please, } (2)}$	avor 2)	please ,	, light	5
$X_{0,3}$	e_6	$ \frac{\text{apague } X}{\text{switch (3)}} $ $ 7 : \frac{X_{0,2}X_2}{(4) (2)} $	1,3 off			switch the	light off	6
0,3	e	$7: \frac{X_{0,2}X_2}{(4)(2)}$	<u>,3</u>)			switch the	off light	7
$X_{1,5}$			e_8 : $\frac{X_{1,3}}{ple}$ e_9 : $\frac{X_1}{(1)}$	ase, (3)		please ,	the light	8
1,5			$e_9: \frac{X_1}{()}$	$\frac{1,2X_{2,5}}{1)(5)}$		the please	, light	9
		($e_{10}: \frac{X_{0,2}X_{2,1}}{(4)(5)}$	<u>5</u>		switch the	, light	10
$X_{0,5}$								11
0,5								12
								13
								14
								15
		-		-				16
								17
								18

			- 1			T	Drawn	
	apague	a	luz	por	favor	Left	Right	Node
$X_{1,2}$		$e_1: rac{a}{the}$				the	the	1
$X_{2,3}$			$e_2: rac{luz}{light}$			light	light	2
$X_{1,3}$			$\frac{X_{1,2}X_{2,3}}{(1)(2)}$			the light	the light	3
$X_{0,2}$	$e_4: \frac{apag}{swite}$	$\frac{gue\ X_{1,2}}{ch\ (1)\ off}$				switch the	the off	4
$X_{2,5}$			e_5	$: \frac{X_{2,3} \text{ por f}}{\text{please, (:)}}$		please ,	, light	5
$X_{0,3}$	e_6	$\frac{\text{apague } X}{\text{switch } (3)}$	1,3) off			switch the	light off	6
0,5	ϵ	$Z_7: \frac{X_{0,2}X_2}{(4)(2)}$)			switch the	off light	7
$X_{1,5}$			$e_8: \frac{X_{1,3}}{ple}$	ase, (3)		please ,	the light	8
1,5			$e_9: \frac{X_1}{(}$			the please	, light	9
			$e_{10}: \frac{X_{0,2}X_{2,}}{(4)(5)}$			switch the	, light	10
$X_{0,5}$		$e_{:}$	$_{11}: \frac{apague\ X_1}{switch\ (8)}$., <u>5</u> off		switch please	light off	11
								12
								13
								14
								15
								16
								17
								18
						•	•	

						T	T Drawn	37.
	apague	a	luz	por	favor	Left	Right	Node
$X_{1,2}$		$e_1: \frac{a}{the}$				the	the	1
$X_{2,3}$			$e_2: rac{luz}{light}$			light	light	2
$X_{1,3}$		$e_3: \frac{\lambda}{2}$	$\frac{X_{1,2}X_{2,3}}{(1)(2)}$			the light	the light	3
$X_{0,2}$	e_4 : $\frac{apague}{switch}$	$X_{1,2}$ n (1) off				switch the	the off	4
$X_{2,5}$			e_5	$\frac{X_{2,3} \text{ por for please, } (2)}{\text{please, } (2)}$	avor 2)	please ,	, light	5
$X_{0,3}$	e ₆	$ \frac{\text{apague } X}{\text{switch (3)}} $ $ 7 : \frac{X_{0,2}X_2}{(4) (2)} $	1,3) off			switch the	light off	6
110,5	e_7	$\cdot : \frac{X_{0,2}X_2}{(4)(2)}$,3			switch the	off light	7
$X_{1,5}$			$e_8: \frac{X_{1,3}}{plea}$	por favor ase, (3)		please ,	the light	8
11,0			$e_9: \frac{X_1}{(1)}$	$\frac{1,2X_{2,5}}{1)(5)}$		the please	, light	9
		($e_{10}: \frac{X_{0,2}X_{2,5}}{(4)(5)}$	<u>5</u>		switch the	, light	10
$X_{0,5}$		e_{1}	11: $\frac{\text{apague } X_1}{\text{switch } (8) \text{ o}}$ 12: $\frac{\text{apague } X_1}{\text{switch } (9) \text{ o}}$,5 off		switch please	light off	11
		$e_{\scriptscriptstyle \parallel}$	$12 : \frac{\text{apague } X_1}{\text{switch } (9)}$	<u>,5</u> off		switch the	light off	12
								13
								14
								15
								16
								17
								18

	apague	a	luz	por	favor	Left	Right	Node
$X_{1,2}$		$e_1: \frac{a}{the}$				the	the	1
$X_{2,3}$			$e_2:rac{luz}{light}$			light	light	2
$X_{1,3}$			$\frac{X_{1,2}X_{2,3}}{(1)(2)}$			the light	the light	3
$X_{0,2}$	$e_4: \frac{apt}{swi}$	ague $X_{1,2}$ itch (1) off				switch the	the off	4
$X_{2,5}$			e_5	$X_{2,3}$ por for please, (2)	avor 2)	please ,	, light	5
$X_{0,3}$	ϵ	$e_6: rac{ ext{apague } X}{ ext{switch } (3)}$ $e_7: rac{X_{0,2}X_2}{(4)}$	1,3) off			switch the	light off	6
0,5		$e_7: \frac{X_{0,2}X_2}{(4)(2)}$)			switch the	off light	7
$X_{1,5}$			$e_8:rac{X_{1,3}}{ ext{ple}}$ $e_9:rac{X_1}{(i)}$ $e_{10}:rac{X_{0,2}X_2,i}{(4)}$ $e_{10}:rac{X_{0,2}X_2,i}{(4)}$	ase, (3)		please ,	the light	8
1,0			$e_9: \frac{X_1}{(1)}$	$\frac{1,2}{1}$ (5)		the please	, light	9
			$e_{10}: \frac{X_{0,2}X_{2,3}}{(4)(5)}$	<u>5</u>		switch the	, light	10
$X_{0,5}$		e_{\cdot}	11: switch (8)	off		switch please	light off	11
		e	$12 : \frac{apague[X_1]}{switch(9)}$	<u>,5</u> off		switch the	light off	12
		e_1	$_3: \frac{X_{0,3} \text{ por fa}}{\text{please, } (6)}$)		please ,	light off	13
								14
								15
			·					16
		-	-					17
								18

	apague	а	luz	por	favor	Left	RIGHT	Node
$X_{1,2}$		$e_1: rac{a}{the}$				the	the	1
$X_{2,3}$			$e_2: rac{luz}{light}$			light	light	2
$X_{1,3}$			$\frac{X_{1,2}X_{2,3}}{(1)(2)}$			the light	the light	3
$X_{0,2}$	$e_4: \frac{ap}{swi}$	ague $X_{1,2}$ tch (1) off				switch the	the off	4
$X_{2,5}$			e_5	$: \frac{X_{2,3} \text{ por }}{\text{please, (}}$	favor 2)	please ,	, light	5
$X_{0,3}$	ϵ	$e_6:rac{apague\;X}{switch\;(3)} \ e_7:rac{X_{0,2}X_2}{(4)\;(2)}$	1,3) off			switch the	light off	6
210,3		$e_7: \frac{X_{0,2}X_2}{(4)(2)}$)			switch the	off light	7
$X_{1,5}$			$\begin{array}{c} e_8:\frac{X_{1,1}}{\text{ple}} \\ e_9:\frac{X}{4} \\ e_{10}:\frac{X_{0,2}X_{2,1}}{4} \\ \text{(4) (5)} \\ \text{apague } X_{1} \\ \text{(11)}:\text{switch (8)} \\ \text{(2)}:\text{switch (9)} \end{array}$	g por favor ease, (3)		please ,	the light	8
111,5			$e_9: \frac{X_2}{(}$	$\frac{1,2X_{2,5}}{1)(5)}$		the please	, light	9
		,	$e_{10}: \frac{X_{0,2}X_{2,1}}{(4)(5)}$	<u>,5</u>		switch the	, light	10
$X_{0,5}$		e:	$11 : \frac{\text{apague } X_1}{\text{switch } (8)}$	0ff		switch please	light off	11
		e:	12: $\frac{\text{apague } X_1}{\text{switch (9)}}$ 3: $\frac{X_{0,3} \text{ por f.}}{\text{please, (6)}}$ 4: $\frac{X_{0,3} \text{ por f.}}{\text{please, (7)}}$	1,5 off		switch the	light off	12
		e_1	$x_3: \frac{X_{0,3} \text{ por for please, } (6)}{\text{please, } (6)}$	avor 3)		please ,	light off	13
		e_1	$A: \frac{X_{0,3} \text{ por formula}}{\text{please, } (7)}$	avor 7)		please ,	off light	14
								15
								16
								17
								18

	apague	а	luz	por	favor	Left	RIGHT	Node
$X_{1,2}$		$e_1: \frac{a}{the}$				the	the	1
$X_{2,3}$			$e_2: rac{luz}{light}$			light	light	2
$X_{1,3}$			$\frac{X_{1,2}X_{2,3}}{(1)(2)}$			the light	the light	3
$X_{0,2}$	$e_4:rac{apa}{swi}$	ague $X_{1,2}$ tch (1) off				switch the	the off	4
$X_{2,5}$			e_5	$: \frac{X_{2,3} \text{ por f}}{\text{please, (}}$	favor 2)	please ,	, light	5
$X_{0,3}$	e	$e_6: rac{ ext{apague } X}{ ext{switch } (3)} \ e_7: rac{X_{0,2}X_2}{(4)} \ (2)$	1,3) off			switch the	light off	6
110,3		$e_7: \frac{X_{0,2}X_2}{(4)(2)}$)			switch the	off light	7
$X_{1,5}$			$e_8:rac{X_{1,3}}{ple}$ $e_9:rac{X_1}{(}$	ase, (3)		please ,	the light	8
211,5			$e_9: \frac{X_1}{(}$	$\frac{1,2X_{2,5}}{1)(5)}$		the please	, light	9
			$e_{10}: \frac{X_{0,2}X_{2,}}{(4)(5)}$. <u>5</u>		switch the	, light	10
$X_{0,5}$		e:	$11 : \frac{\text{apague } X_1}{\text{switch } (8)}$	0ff		switch please	light off	11
		$e_{:}$	$12: \frac{\text{apague } X_1}{\text{switch } (9)}$	0ff		switch the	light off	12
		e_1	$_3: \frac{X_{0,3} \text{ por fa}}{\text{please, } (6)}$	avor 3)		please ,	light off	13
		e_1	$4: \frac{X_{0,3} \text{ por fa}}{\text{please, } (7)}$	avor 7)		please ,	off light	14
			3: $\frac{X_{0,3} \text{ por fo}}{\text{please, (6)}}$ 4: $\frac{X_{0,3} \text{ por fo}}{\text{please, (7)}}$ e_{15} : $\frac{\vdash X_{0,5} \vdash}{\vdash (10) \vdash}$	1		⊢ switch	light ⊣	15
$S_{0,5}$								16
,,,								17
								18

	apague	a	luz	por	favor	Left	RIGHT	Node
	apague		IUZ	Poi	lavoi	TABL 1	Tugiii	
$X_{1,2}$		$e_1: rac{a}{the}$				the	the	1
$X_{2,3}$			$e_2: rac{luz}{light}$			light	light	2
$X_{1,3}$			$\frac{X_{1,2}X_{2,3}}{(1)(2)}$		·	the light	the light	3
$X_{0,2}$	$e_4: \frac{apa}{swi}$	ague $X_{1,2}$ tch (1) off				switch the	the off	4
$X_{2,5}$			e_5	$: \frac{X_{2,3} \text{ por f}}{\text{please, (}}$	2)	please ,	, light	5
$X_{0,3}$	e	$r_6: \frac{\text{apague } X}{\text{switch } (3)}$	(1,3) off			switch the	light off	6
0,3		$e_7: \frac{X_{0,2}X_2}{(4)}$	2 <u>,3</u>			switch the	off light	7
$X_{1,5}$			$e_8:rac{X_{1,3}}{ple}$ $e_9:rac{X_1}{(}$	ase, (3)		please ,	the light	8
1,3			$e_9: \frac{X_1}{(}$	$\frac{1,2}{1},\frac{X_{2,5}}{(5)}$		the please	, light	9
			$e_{10}: \frac{X_{0,2}X_{2,1}}{(4)}$	5		switch the	, light	10
$X_{0,5}$		e	11: $\frac{\text{apague } X_1}{\text{switch (8)}}$ 12: $\frac{\text{apague } X_1}{\text{switch (8)}}$ 13: $\frac{X_{0,3} \text{ por fi}}{\text{please, (7)}}$ 14: $\frac{X_{0,3} \text{ por fi}}{\text{please, (7)}}$ 15: $\frac{X_{0,3} \text{ por fi}}{\text{please, (7)}}$ 16: $\frac{X_{0,3} \text{ por fi}}{\text{please, (7)}}$ 17: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 18: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 19: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 10: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 10: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 10: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 11: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 12: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 13: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 14: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 15: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 16: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 16: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 17: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 18: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 19: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 10: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 11: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 12: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 13: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 14: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 15: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 16: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 17: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 18: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 18: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 19: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 10: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 10: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 10: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 11: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 12: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 13: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 14: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 15: $\frac{X_{0,5} \text{ por fi}}{\text{please, (7)}}$ 16: $\frac{X_{0,5} \text{ por fi}}{pl$	0ff		switch please	light off	11
		e	$_{12}: \frac{\text{apague } X_1}{\text{switch } (9)}$	off		switch the	light off	12
		e_1	$X_{0,3}$ por far please, (6)	avor 3)		please ,	light off	13
		e_1	$14: \frac{X_{0,3} \text{ por for please, } (7)}{\text{please, } (7)}$	avor 7)		please ,	off light	14
			$e_{15}: \frac{\vdash X_{0,5}}{\vdash (10)}$	1		⊢ switch	light ⊣	15
$S_{0,5}$			$e_{16}: \frac{\vdash X_{0,5} \dashv}{\vdash (11) \dashv}$			⊢ switch	off ⊣	16
								17
								18

								т
	apague	a	luz	por	favor	Left	Right	Node
$X_{1,2}$		$e_1: \frac{a}{the}$				the	the	1
$X_{2,3}$			$e_2:rac{luz}{light}$			light	light	2
$X_{1,3}$		$e_3:\frac{7}{2}$	$\frac{X_{1,2}X_{2,3}}{(1)(2)}$			the light	the light	3
$X_{0,2}$	$e_4: \frac{apt}{swi}$	ague $X_{1,2}$ itch (1) off				switch the	the off	4
$X_{2,5}$			e_5	$: \frac{X_{2,3} \text{ por f}}{\text{please, ()}}$	avor 2)	please ,	, light	5
$X_{0,3}$	ϵ	$r_6: \frac{\text{apague } X}{\text{switch } (3)}$	1,3) off			switch the	light off	6
0,3		$e_7: \frac{X_{0,2}X_2}{(4)}$)			switch the	off light	7
$X_{1,5}$			$e_8: \frac{X_{1,3}}{ple}$	ase, (3)		please ,	the light	8
111,5			$e_9: \frac{X_1}{(}$	$\frac{1,2X_{2,5}}{1)(5)}$		the please	, light	9
			$e_{10}: \frac{X_{0,2}X_{2,1}}{(4)}$	<u>,5</u>)		switch the	, light	10
$X_{0,5}$		e	$_{11}: \frac{apague\ X_1}{switch\ (8)}$	0ff		switch please	light off	11
		e	$_{12}: \frac{\text{apague } X_1}{\text{switch } (9)}$	off		switch the	light off	12
		e_1	$3: \frac{X_{0,3} \text{ por f}}{\text{please, } (6)}$	avor 3)		please ,	light off	13
		e_1	$A: \frac{X_{0,3} \text{ por formula}}{\text{please, } (7)}$	avor 7)		please ,	off light	14
			11: $\frac{\text{apague } X_1}{\text{switch } (8)}$ 12: $\frac{\text{apague } X_1}{\text{switch } (8)}$ 13: $\frac{X_{0,3} \text{ por fi}}{\text{please, } (6)}$ 4: $\frac{X_{0,3} \text{ por fi}}{\text{please, } (6)}$ 16: $\frac{X_{0,3} \text{ por fi}}{\text{please, } (6)}$	1		⊢ switch	light ⊣	15
$S_{0,5}$		$e_{16}: \frac{\vdash X}{\vdash (}$	$c_{0,5}$ or e_{17}	$\therefore \frac{\vdash X_{0,5} \dashv}{\vdash (12) \dashv}$		⊢ switch	off ⊣	16
								17
								18

$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$		apague	a	luz	por	favor	Left	RIGHT	Node
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$		apague		iuz	рог	Tavoi	LEFI	ппсп	NODE
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	$X_{1,2}$		$e_1: \frac{a}{the}$				the	the	1
	$X_{2,3}$			$e_2: rac{luz}{light}$			light	light	2
$X_{2,5} = \begin{array}{c} e_5 : \frac{X_{2,3} \text{ por favor}}{\text{please, (2)}} & \text{please }, & \text{, light} & 5 \\ X_{0,3} = \begin{array}{c} e_6 : \frac{\text{apague } X_{1,3}}{\text{switch (3) off}} & \text{switch the} & \text{light off} & 6 \\ e_7 : \frac{X_{0,2} X_{2,3}}{4 (4 (2)} & \text{switch the} & \text{off light} & 7 \\ \end{array}$	$X_{1,3}$		$e_3:\frac{\lambda}{2}$	$\frac{X_{1,2}X_{2,3}}{(1)(2)}$			the light	the light	3
$X_{0,3} = \begin{array}{ c c c c }\hline e_6: & \frac{\text{apague } X_{1,3}}{\text{switch } (3) \text{ off}} \\ e_7: & \frac{X_{0,2} X_{2,3}}{4(4)(2)} \\ \hline \\ X_{1,5} = \\ \hline \\ & e_8: & \frac{X_{1,3} \text{ por favor}}{\text{please, } (3)} \\ \hline \\ & e_9: & \frac{X_{1,2} X_{2,5}}{(1)(5)} \\ \hline \\ & e_9: & \frac{X_{1,2} X_{2,5}}{(1)(5)} \\ \hline \\ & e_{10}: & \frac{X_{0,2} X_{2,5}}{(4)(5)} \\ \hline \\ & & e_{10}: & \frac{X_{0,2} X_{2,5}}{(4)(5)} \\ \hline \\ & & e_{11}: & \frac{\text{apague } X_{1,5}}{\text{switch } (8) \text{ off}} \\ \hline \\ & & e_{12}: & \frac{\text{apague } X_{1,5}}{\text{switch } (9) \text{ off}} \\ \hline \\ & & e_{13}: & \frac{X_{0,3} \text{ por favor}}{\text{please, } (6)} \\ \hline \\ & & e_{14}: & \frac{X_{0,3} \text{ por favor}}{\text{please, } (7)} \\ \hline \\ & & e_{15}: & \frac{-X_{0,5} + 1}{-(10) + 1} \\ \hline \\ & & e_{15}: & \frac{-X_{0,5} + 1}{-(10) + 1} \\ \hline \\ & & e_{16}: & \frac{-X_{0,5} + 1}{-(10) + 1} \\ \hline \\ & & e_{18}: & \frac{-X_{0,5} + 1}{-(12) + 1} \\ \hline \\ & & & e_{18}: & \frac{-X_{0,5} + 1}{-(12) + 1} \\ \hline \\ & & & & e_{18}: & \frac{-X_{0,5} + 1}{-(12) + 1} \\ \hline \\ & & & & & & e_{18}: & \frac{-X_{0,5} + 1}{-(12) + 1} \\ \hline \\ & & & & & & & & & & & & & & & & &$	$X_{0,2}$	$e_4: \frac{apa}{swi}$	$\frac{\text{ague } X_{1,2}}{\text{tch } (1) \text{ off}}$				switch the	the off	4
$X_{1,5} = \begin{bmatrix} e_8 : \frac{X_{1,3} \text{ por favor}}{\text{please. } (3)} & \text{please }, & \text{the light} & 8 \\ e_9 : \frac{X_{1,2} X_{2,5}}{(1) (5)} & \text{the please} & , \text{light} & 9 \\ \end{bmatrix}$ $E_{10} : \frac{X_{0,2} X_{2,5}}{(4) (5)} & \text{switch the} & , \text{light} & 10 \\ \end{bmatrix}$ $X_{0,5} = \begin{bmatrix} e_{11} : \frac{X_{0,2} X_{2,5}}{(4) (5)} & \text{switch the} & , \text{light} & 10 \\ \end{bmatrix}$ $E_{11} : \frac{\text{apague } X_{1,5}}{\text{switch } (8) \text{ off}} & \text{switch please} & \text{light off} & 11 \\ \end{bmatrix}$ $E_{12} : \frac{X_{0,3} \text{ por favor}}{\text{switch } (9) \text{ off}} & \text{switch the} & \text{light off} & 12 \\ \end{bmatrix}$ $E_{13} : \frac{X_{0,3} \text{ por favor}}{\text{please, } (6)} & \text{please }, & \text{off light} & 14 \\ \end{bmatrix}$ $E_{13} : \frac{X_{0,3} \text{ por favor}}{\text{please, } (7)} & \text{please }, & \text{off light} & 14 \\ \end{bmatrix}$ $E_{15} : \frac{ X_{0,5} }{ -(10) } & \text{or } e_{17} : \frac{ X_{0,5} }{ -(12) } & \text{please} & \text{off} & 15 \\ \end{bmatrix}$ $E_{16} : \frac{ X_{0,5} }{ -(11) } & \text{or } e_{17} : \frac{ X_{0,5} }{ -(12) } & \text{please} & \text{off} & 16 \\ \end{bmatrix}$ $E_{18} : \frac{ X_{0,2} X_{2,5} }{ -(13) } & \text{please} & \text{off} & 17 \\ \end{bmatrix}$	$X_{2,5}$			e_5	$\frac{X_{2,3} \text{ por }}{\text{please, (}}$	favor 2)	please ,	, light	5
$X_{1,5} = \begin{bmatrix} e_8 : \frac{X_{1,3} \text{ por favor}}{\text{please. (3)}} & \text{please }, & \text{the light} & 8 \\ e_9 : \frac{X_{1,2}X_{2,5}}{(1)(5)} & \text{the please} & , \text{light} & 9 \\ \end{bmatrix}$ $E_{10} : \frac{X_{0,2}X_{2,5}}{(4)(5)} & \text{switch the} & , \text{light} & 10 \\ \end{bmatrix}$ $X_{0,5} = \begin{bmatrix} e_{11} : \frac{X_{0,2}X_{2,5}}{(4)(5)} & \text{switch the} & , \text{light} & 10 \\ \vdots & \text{switch (8) off} & \text{switch please} & \text{light off} & 11 \\ \end{bmatrix}$ $E_{12} : \frac{X_{0,3} \text{ por favor}}{\text{switch (9) off}} & \text{switch the} & \text{light off} & 12 \\ \end{bmatrix}$ $E_{13} : \frac{X_{0,3} \text{ por favor}}{\text{please, (6)}} & \text{please, } & \text{light off} & 13 \\ \end{bmatrix}$ $E_{14} : \frac{X_{0,3} \text{ por favor}}{\text{please, (7)}} & \text{please, } & \text{off light} & 14 \\ \end{bmatrix}$ $E_{15} : \frac{ -X_{0,5} }{ -(10) } & \text{or } e_{17} : \frac{ -X_{0,5} }{ -(12) } & \text{please, } & \text{off } -15 \\ \end{bmatrix}$ $E_{16} : \frac{ -X_{0,5} }{ -(13) } & \text{or } e_{17} : \frac{ -X_{0,5} }{ -(12) } & \text{please} & \text{off } -17 \\ \end{bmatrix}$ $E_{18} : \frac{ -X_{0,5} }{ -(13) } & \text{please, } & \text{off } -17 \\ \end{bmatrix}$	X _{0.3}	e	$_{6}: \frac{\text{apague } X}{\text{switch } (3)}$	1,3) off			switch the	light off	6
$X_{0,5} = \begin{bmatrix} & c_{10} : \frac{X_{0,2}X_{2,5}}{(4)(5)} & \text{switch the} & \text{, light} & 10 \\ & & \text{apague } X_{1,5} & \text{switch (8) off} \\ & & \text{switch (8) off} & \text{switch please} & \text{light off} & 11 \\ & & & \text{apague } X_{1,5} & \text{switch (9) off} & \text{switch the} & \text{light off} & 12 \\ & & & & \text{20} & \text{20} & \text{20} & \text{20} \\ & & & & \text{20} & \text{20} & \text{20} & \text{20} \\ & & & & & \text{20} & \text{20} & \text{20} \\ & & & & & & \text{20} & \text{20} & \text{20} \\ & & & & & & \text{20} & \text{20} & \text{20} \\ & & & & & & & \text{20} & \text{20} \\ & & & & & & & \text{20} & \text{20} & \text{20} \\ & & & & & & & & \text{20} & \text{20} \\ & & & & & & & & \text{20} & \text{20} \\ & & & & & & & & \text{20} & \text{20} \\ & & & & & & & & \text{20} & \text{20} \\ & & & & & & & & & \text{20} \\ & & & & & & & & & \text{20} \\ & & & & & & & & & \text{20} \\ & & & & & & & & & & \text{20} \\ & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & & & & & $	0,3		$e_7: \frac{X_{0,2}X_2}{(4)(2)}$)			switch the	off light	7
$X_{0,5} = \begin{bmatrix} & c_{10} : \frac{X_{0,2}X_{2,5}}{(4)(5)} & \text{switch the} & \text{, light} & 10 \\ & & \text{apague } X_{1,5} & \text{switch (8) off} \\ & & \text{switch (8) off} & \text{switch please} & \text{light off} & 11 \\ & & & \text{apague } X_{1,5} & \text{switch (9) off} & \text{switch the} & \text{light off} & 12 \\ & & & & \text{20} & \text{20} & \text{20} & \text{20} \\ & & & & \text{20} & \text{20} & \text{20} & \text{20} \\ & & & & & \text{20} & \text{20} & \text{20} \\ & & & & & & \text{20} & \text{20} & \text{20} \\ & & & & & & \text{20} & \text{20} & \text{20} \\ & & & & & & & \text{20} & \text{20} \\ & & & & & & & \text{20} & \text{20} & \text{20} \\ & & & & & & & & \text{20} & \text{20} \\ & & & & & & & & \text{20} & \text{20} \\ & & & & & & & & \text{20} & \text{20} \\ & & & & & & & & \text{20} & \text{20} \\ & & & & & & & & & \text{20} \\ & & & & & & & & & \text{20} \\ & & & & & & & & & \text{20} \\ & & & & & & & & & & \text{20} \\ & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & & & & & $	X _{1.5}			$e_8: \frac{X_{1,3}}{ple}$	ase, (3)		please ,	the light	8
$X_{0,5} = \begin{bmatrix} & c_{10} : \frac{X_{0,2}X_{2,5}}{(4)(5)} & \text{switch the} & \text{, light} & 10 \\ & & \text{apague } X_{1,5} & \text{switch (8) off} \\ & & \text{switch (8) off} & \text{switch please} & \text{light off} & 11 \\ & & & \text{apague } X_{1,5} & \text{switch (9) off} & \text{switch the} & \text{light off} & 12 \\ & & & & \text{20} & \text{20} & \text{20} & \text{20} \\ & & & & \text{20} & \text{20} & \text{20} & \text{20} \\ & & & & & \text{20} & \text{20} & \text{20} \\ & & & & & & \text{20} & \text{20} & \text{20} \\ & & & & & & \text{20} & \text{20} & \text{20} \\ & & & & & & & \text{20} & \text{20} \\ & & & & & & & \text{20} & \text{20} & \text{20} \\ & & & & & & & & \text{20} & \text{20} \\ & & & & & & & & \text{20} & \text{20} \\ & & & & & & & & \text{20} & \text{20} \\ & & & & & & & & \text{20} & \text{20} \\ & & & & & & & & & \text{20} \\ & & & & & & & & & \text{20} \\ & & & & & & & & & \text{20} \\ & & & & & & & & & & \text{20} \\ & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & & & & \text{20} \\ & & & & & & & & & & & & & & & & & & $	11,5			$e_9: \frac{X_1}{(}$	$\frac{1,2X_{2,5}}{1)(5)}$		the please	, light	9
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$			($e_{10}: \frac{X_{0,2}X_{2,}}{(4)(5)}$	<u>5</u>		switch the	, light	10
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$X_{0,5}$		e_1	$_{11}: \frac{\text{apague } X_1}{\text{switch } (8)}$.,5 off		switch please	light off	11
$S_{0,5}$ $e_{16}: \frac{ X_{0,5} }{\vdash (11)\dashv}$ or $e_{17}: \frac{ X_{0,5} }{\vdash (12)\dashv}$ \vdash switch off \dashv 16 $e_{18}: \frac{ X_{0,5} }{\vdash (13)\dashv}$ \vdash please off \dashv 17			e_1	$12 : \frac{\text{apague } \Lambda_1}{\text{apague } (0)}$.,5		switch the	light off	12
$S_{0,5}$ $e_{16}: \frac{ X_{0,5} }{\vdash (11)\dashv}$ or $e_{17}: \frac{ X_{0,5} }{\vdash (12)\dashv}$ \vdash switch off \dashv 16 $e_{18}: \frac{ X_{0,5} }{\vdash (13)\dashv}$ \vdash please off \dashv 17			e_1	$3: \frac{X_{0,3} \text{ por fa}}{\text{please, } (6)}$	avor 3)		please ,	light off	13
$S_{0,5}$ $e_{16}: \frac{ X_{0,5} }{\vdash (11)\dashv}$ or $e_{17}: \frac{ X_{0,5} }{\vdash (12)\dashv}$ \vdash switch off \dashv 16 $e_{18}: \frac{ X_{0,5} }{\vdash (13)\dashv}$ \vdash please off \dashv 17			e_1	$_4: \frac{X_{0,3} \text{ por fa}}{\text{please, } (7)}$	avor 7)		please ,	off light	14
$S_{0,5}$ $e_{16}: \frac{ X_{0,5} }{\vdash (11)\dashv}$ or $e_{17}: \frac{ X_{0,5} }{\vdash (12)\dashv}$ \vdash switch off \dashv 16 $e_{18}: \frac{ X_{0,5} }{\vdash (13)\dashv}$ \vdash please off \dashv 17				$e_{15}: \frac{\vdash X_{0,5}}{\vdash (10)}$	 		⊢ switch	light ⊣	15
. (20)	$S_{0,5}$		$e_{16}: \frac{1}{1}$	$\frac{0.5}{11)\dashv}$ or e_{17}	$: \frac{A_{0,5}}{\vdash (12)} \dashv$		⊢ switch	off -	16
18				$e_{18}: \frac{\vdash X_{0,5}}{\vdash (13)}$	 		⊢ please	off ⊣	17
									18

	apague	a	luz	por	favor	LEFT	RIGHT	Node
$X_{1,2}$	-1-0	$e_1: \frac{a}{the}$		1, 7,		the	the	1
$X_{2,3}$		the .	$e_2: \frac{luz}{light}$			light	light	2
$X_{1,3}$		$e_3:$	$\frac{X_{1,2}X_{2,3}}{(1)(2)}$			the light	the light	3
$X_{0,2}$	$e_4: \frac{apa}{swi}$	ague $X_{1,2}$ tch (1) off	(1) (2)			switch the	the off	4
$X_{2,5}$			e_5	$\frac{X_{2,3} \text{ por }}{\text{please, (}}$	2)	please ,	, light	5
$X_{0,3}$	e	$r_6: \frac{apague\ X}{switch\ (3)}$	(1,3) off			switch the	light off	6
210,3		$e_7: \frac{X_{0,2}X_2}{(4)}$	2,3			switch the	off light	7
$X_{1,5}$			$e_8:rac{X_{1,3}}{ple}$ $e_9:rac{X_1}{(}$	por favor ease, (3)		please ,	the light	8
21,5			$e_9: \frac{X_1}{(}$	$\frac{1,2X_{2,5}}{1)(5)}$		the please	, light	9
			$e_{10}: \frac{X_{0,2}X_{2,}}{(4)}$ $11: \frac{X_{0,2}X_{2,}}{(4)}$ switch (8)	.5		switch the	, light	10
$X_{0,5}$		e	$11 : \frac{\text{apague } X_1}{\text{switch } (8)}$	off		switch please	light off	11
,		e	12: $\frac{\text{apague } X_1}{\text{switch } (9)}$	0ff		switch the	light off	12
		e_1	$X_{0,3} = \frac{X_{0,3} \text{ por for } x_{0,3}}{x_{0,3}}$	avor		please ,	light off	13
		e_1	$14: \frac{X_{0,3} \text{ por fa}}{\text{please, } (7)}$	avor 7)		please ,	off light	14
			$X_{0,3}$ por for please, (7) $e_{15}: \frac{X_{0,5}}{\vdash (10)}$	 		⊢ switch	light ⊣	15
$S_{0,5}$		$e_{16}: \frac{\vdash x}{\vdash (}$	$\frac{\alpha_{0,5}}{11)}$ or e_{17}	$: \frac{\vdash X_{0,5} \dashv}{\vdash (12) \dashv}$		⊢ switch	off ⊣	16
			$e_{18}: \frac{\vdash X_{0,5}}{\vdash (13)}=$	 		⊢ please	off ⊣	17
			$e_{19}: \frac{ X_{0,5} }{ (14) }$	1		⊢ please	light ⊣	18

The problem with the solution

The problem with the solution

Computational complexity!

The problem with the solution

Computational complexity!

- 1 it seems like the underlying grammar is growing
- ${\bf 2}$ there are way too many n-grams leading to way too many nonterminals
- 3 the graphical representation (forest) is growing

What is really going on?

We are transferring memory from an automaton to the forest

What is really going on?

We are transferring memory from an automaton to the forest

- n-gram LMs can be thought of as an automaton where each state q uniquely represents a k-gram prefix α_q (k < n)

What is really going on?

We are transferring memory from an automaton to the forest

- $n\text{-}\mathrm{gram}$ LMs can be thought of as an automaton where each state q uniquely represents a $k\text{-}\mathrm{gram}$ prefix α_q (k < n)
- a transition from a state q labelled with word w is weighted by the probability $p(w|\alpha_q)$

What is really going on?

We are transferring memory from an automaton to the forest

- n-gram LMs can be thought of as an automaton where each state q uniquely represents a k-gram prefix α_q (k < n)
- a transition from a state q labelled with word w is weighted by the probability $p(w|\alpha_q)$

A nonterminal in the forest yields a set of strings

What is really going on?

We are transferring memory from an automaton to the forest

- n-gram LMs can be thought of as an automaton where each state q uniquely represents a k-gram prefix α_q (k < n)
- a transition from a state q labelled with word w is weighted by the probability $p(w|\alpha_q)$

A nonterminal in the forest yields a set of strings

strings project onto paths in the LM automaton

What is really going on?

We are transferring memory from an automaton to the forest

- n-gram LMs can be thought of as an automaton where each state q uniquely represents a k-gram prefix α_q (k < n)
- a transition from a state q labelled with word w is weighted by the probability $p(w|\alpha_q)$

A nonterminal in the forest yields a set of strings

- strings project onto paths in the LM automaton
- paths are weighted

What is really going on?

We are transferring memory from an automaton to the forest

- n-gram LMs can be thought of as an automaton where each state q uniquely represents a k-gram prefix α_q (k < n)
- a transition from a state q labelled with word w is weighted by the probability $p(w|\alpha_q)$

A nonterminal in the forest yields a set of strings

- strings project onto paths in the LM automaton
- paths are weighted

Nonterminals must be aware of (parts of) the strings they yield

What is really going on?

We are transferring memory from an automaton to the forest

- n-gram LMs can be thought of as an automaton where each state q uniquely represents a k-gram prefix α_q (k < n)
- a transition from a state q labelled with word w is weighted by the probability $p(w|\alpha_q)$

A nonterminal in the forest yields a set of strings

- strings project onto paths in the LM automaton
- paths are weighted

Nonterminals must be aware of (parts of) the strings they yield

• they must be annotated with states of the automaton

How hard is it?

Weighted intersection between a wCFG and a wFSA

Weighted intersection between a wCFG and a wFSA

Generalisation of parsing for

Weighted intersection between a wCFG and a wFSA

Generalisation of parsing for

arbitrary automata

Weighted intersection between a wCFG and a wFSA

Generalisation of parsing for

- arbitrary automata
- weighted sets

Weighted intersection between a wCFG and a wFSA

Generalisation of parsing for

- arbitrary automata
- weighted sets

Complexity

pace of translations Formal devices Linear models Decision rules **Decoding** Reference

How hard is it?

Weighted intersection between a wCFG and a wFSA

Generalisation of parsing for

- arbitrary automata
- weighted sets

Complexity

• input: $X_0 \to X_1 X_2 \dots X_a$ where $X_i \in N$

Weighted intersection between a wCFG and a wFSA

Generalisation of parsing for

- arbitrary automata
- weighted sets

Complexity

- input: $X_0 \to X_1 X_2 \dots X_a$ where $X_i \in N$
- output: $X_0^{(q_1,q_a)} o X_1^{(q_1,q_2)} X_2^{(q_2,q_3)} \dots X_a^{(q_{a-1},q_a)}$ where $q_i \in Q$

Weighted intersection between a wCFG and a wFSA

Generalisation of parsing for

- arbitrary automata
- weighted sets

Complexity

- input: $X_0 \to X_1 X_2 \dots X_a$ where $X_i \in N$
- output: $X_0^{(q_1,q_a)} o X_1^{(q_1,q_2)} X_2^{(q_2,q_3)} \dots X_a^{(q_{a-1},q_a)}$ where $q_i \in Q$
- complexity: $O(|N||Q|^{a+1})$

Solution

The usual suspect

pruning

Space of translations Formal devices Linear models Decision rules **Decoding** References

Solution

The usual suspect

pruning

Alternatives

- local search (greedy methods)
- relaxation techniques
- sampling

Approximate intersection by budgeting the combination of "comparable" nodes

Approximate intersection by budgeting the combination of "comparable" nodes

nodes that share structure

Approximate intersection by budgeting the combination of "comparable" nodes

- nodes that share structure
 - phrase-based: coverage vector
 - hierarchical: input spans

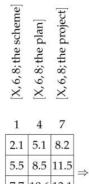
Approximate intersection by budgeting the combination of "comparable" nodes

- nodes that share structure
 - phrase-based: coverage vector
 - hierarchical: input spans
- heuristic view of interaction with LM
 - phrase-based: approximate future cost
 - local approximation based on limited context

Naive beam search

- enumerate combinations
- $oldsymbol{2}$ sort and prune all but the k best

Naive beam search



 $[X, 5, 8; from the \star the scheme] : 2.1$ $[X, 5, 8; from the \star the plan]: 5.1$ $[X, 5, 8; from the \star the scheme] : 5.5$ $[X, 5, 8; since the \star the scheme] : 7.7$ Space of translations Formal devices Linear models Decision rules **Decoding** Reference

Cube pruning

An agenda for pruning [Chiang, 2007]

- tries to enumerate combinations in best-first order
- stops after k items have been enumerated
- inspiration: product of sorted lists
- heuristic: assumes the LM a monotone function over edges

Cube pruning

(b) [X, 6, 8; the scheme [X, 6, 8; the project [X, 6, 8; the plan] 1 $X \rightarrow \langle \operatorname{cong} X_{[1]}, \operatorname{from} X_{[1]} \rangle$ 2.1 5.1 $X \rightarrow \langle \text{cong } X_{\square}, \text{ from the } X_{\square} \rangle$ 5.5 $X \rightarrow \langle cong X_{\square}, since X_{\square} \rangle$ 6 $X \rightarrow \langle cong X_{\square}, through X_{\square} \rangle$

[X, 6, 8]; the schem	(X, 6, 8; the plan	J [X, 6, 8; the proje	
1	4		
2.1	5.1	8.2	
5.5	8.5		

e

Space of translations Formal devices Linear models Decision rules **Decoding** References

Problem with pruning

- unbounded approximation
- 2 approximating the Viterbi solution
- 3 incompatible with models which have a probabilistic interpretation
- 4 cannot handle arbitrarily nonlocal dependencies

Space of translations Formal devices Linear models Decision rules **Decoding** Reference

Beyond beam search

Local search [Hardmeier et al., 2012]

- computationally cheap
- unbounded approximation
- approximate Viterbi
- can handle arbitrarily nonlocal dependencies
- too local view of the distribution (bad for tuning)

Beyond beam search

Relaxation methods [Chang and Collins, 2011, Rush and Collins, 2011]

- computationally expensive
- bounded approximation
- (approximate) Viterbi
- may handle arbitrarily nonlocal dependencies

Space of translations Formal devices Linear models Decision rules **Decoding** Reference

Beyond beam search

Sampling [Arun et al., 2009, Aziz et al., 2013, Aziz, 2014]

- (bounded) approximation
- (approximate) Viterbi, expectations
- handle arbitrarily nonlocal dependencies
- in principle ideal for tuning (global view of distribution)
- potentially computationally expensive



References I

Abhishek Arun, Chris Dyer, Barry Haddow, Phil Blunsom, Adam Lopez, and Philipp Koehn. Monte Carlo inference and maximization for phrase-based translation. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, CoNLL '09, pages 102–110, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics. ISBN 978-1-932432-29-9. URL http://dl.acm.org/citation.cfm?id=1596374.1596394.

Wilker Aziz, Marc Dymetman, and Sriram Venkatapathy.
Investigations in exact inference for hierarchical translation. In Proceedings of the Eighth Workshop on Statistical Machine Translation, pages 472–483, Sofia, Bulgaria, August 2013.
Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W13-2260.

References II

- Wilker Ferreira Aziz. Exact Sampling and Optimisation in Statistical Machine Translation. PhD thesis, University of Wolverhampton, 2014.
- Yin-Wen Chang and Michael Collins. Exact decoding of phrase-based translation models through Lagrangian relaxation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 26–37, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics. ISBN 978-1-937284-11-4.
- David Chiang. Hierarchical phrase-based translation. Computational Linguistics, 33(2):201–228, June 2007. ISSN 0891-2017. doi: 10.1162/coli.2007.33.2.201. URL http://dx.doi.org/10.1162/coli.2007.33.2.201.

References III

Christian Hardmeier, Joakim Nivre, and Jörg Tiedemann.

Document-wide decoding for phrase-based statistical machine translation. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1179–1190, Jeju Island, Korea, July 2012. Association for Computational Linguistics. URL

http://www.aclweb.org/anthology/D12-1108.

Alexander M. Rush and Michael Collins. Exact decoding of syntactic translation models through Lagrangian relaxation. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, HLT '11, pages 72–82, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics. ISBN 978-1-932432-87-9.