Modeling
Morphologically Rich Languages

Chris Dyer

25 July 2013 – LxMLS

ML MACHINE LEARNING DEPARTMENT lti Carnegie Mellon clab

Two kinds of language processing

- Natural language as **input**
  - **Output space**
    - Primarily determined by task: *language identification, parsing, part-of-speech tagging, topic modeling, authorship identification, sentiment analysis, information extraction*
    - Can be relatively low dimensional
      is this email important or not?
  - **Input space**
    - Words, **sentences**, documents, or entire corpora

Two kinds of language processing

- Natural language as **output**
  - **Output space**
    - Sentences (rarely entire documents or corpora)
    - Always relatively high dimensional
      How many grammatical sentences are there?
      How many English/Russian/Portuguese words are there?
  - **Input space**
    - Determined by task: *speech recognition, summarization, **translation**, "generation"*

Two kinds of language processing

- Natural language as **output**
  - **Output space**
    - Sentences (rarely entire documents or corpora)
    - Always relatively high dimensional
      How many grammatical sentences are there?
      How many English/Russian/Portuguese words are there?
  - **Input space**
    - Determined by task: *speech recognition, summarization,* ***translation****, "generation"*

Translation: a statistical perspective

$$\hat{y} = \arg \max_{y \in \text{English}} p(y \mid \text{português})$$

Translation: a statistical perspective

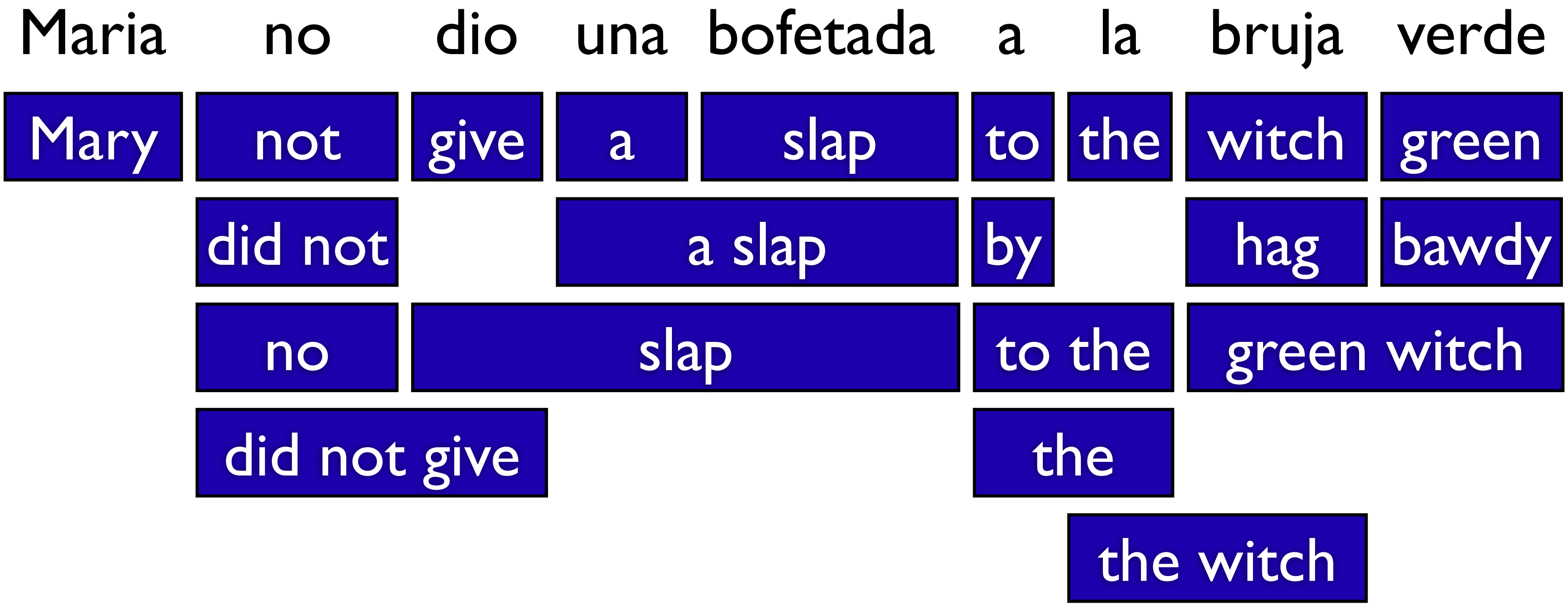
$$\hat{y} = \arg \max_{y \in \text{English}} p(y \mid \text{português})$$

Maria no dio una bofetada a la bruja verde

Adapted from Koehn (2006)

Translation: a statistical perspective

$$\hat{y} = \arg\max_{y \in \text{English}} p(y \mid \text{português})$$



Maria no dio una bofetada a la bruja verde

Mary | not | give | a | slap | to | the | witch | green

did not | a slap | by | hag | bawdy

no | slap | to the | green witch

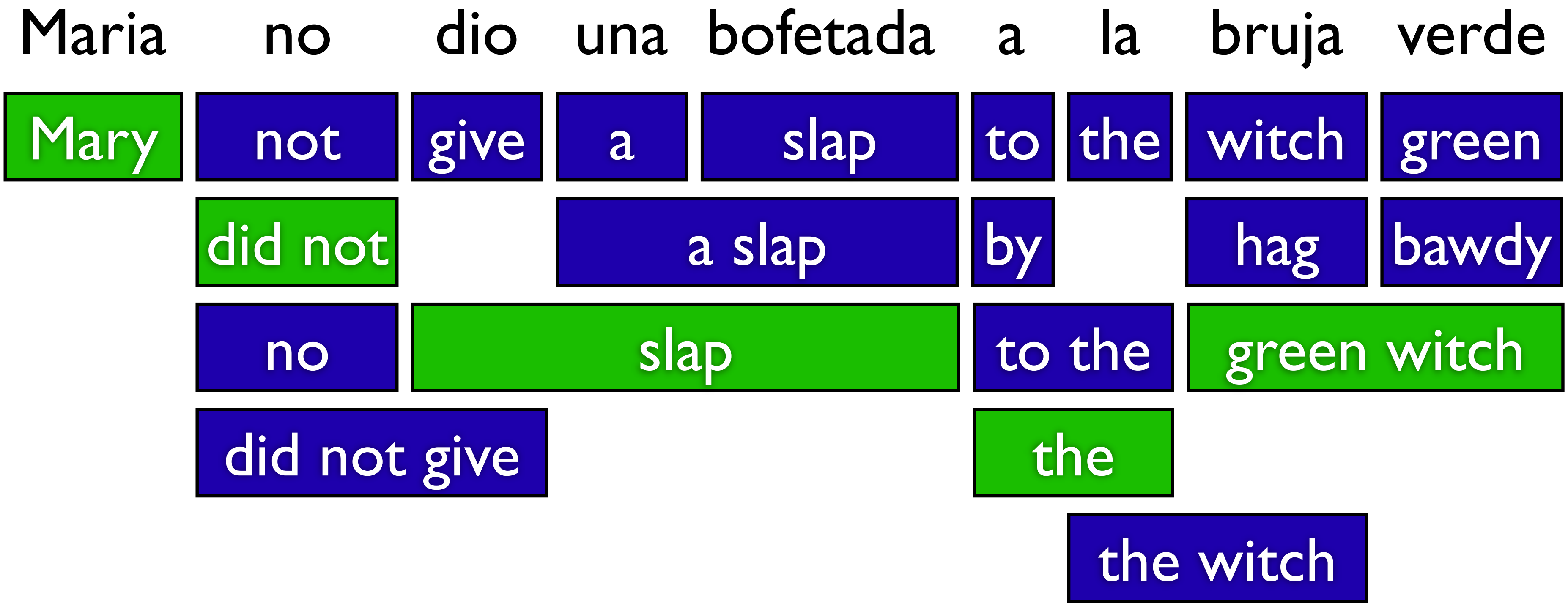
did not give | the

the witch

Adapted from Koehn (2006)

Translation: a statistical perspective

$$\hat{y} = \arg \max_{y \in \text{English}} p(y \mid \text{português})$$



Maria no dio una bofetada a la bruja verde

Mary | not | give | a | slap | to | the | witch | green

did not | a slap | by | hag | bawdy

no | slap | to the | green witch

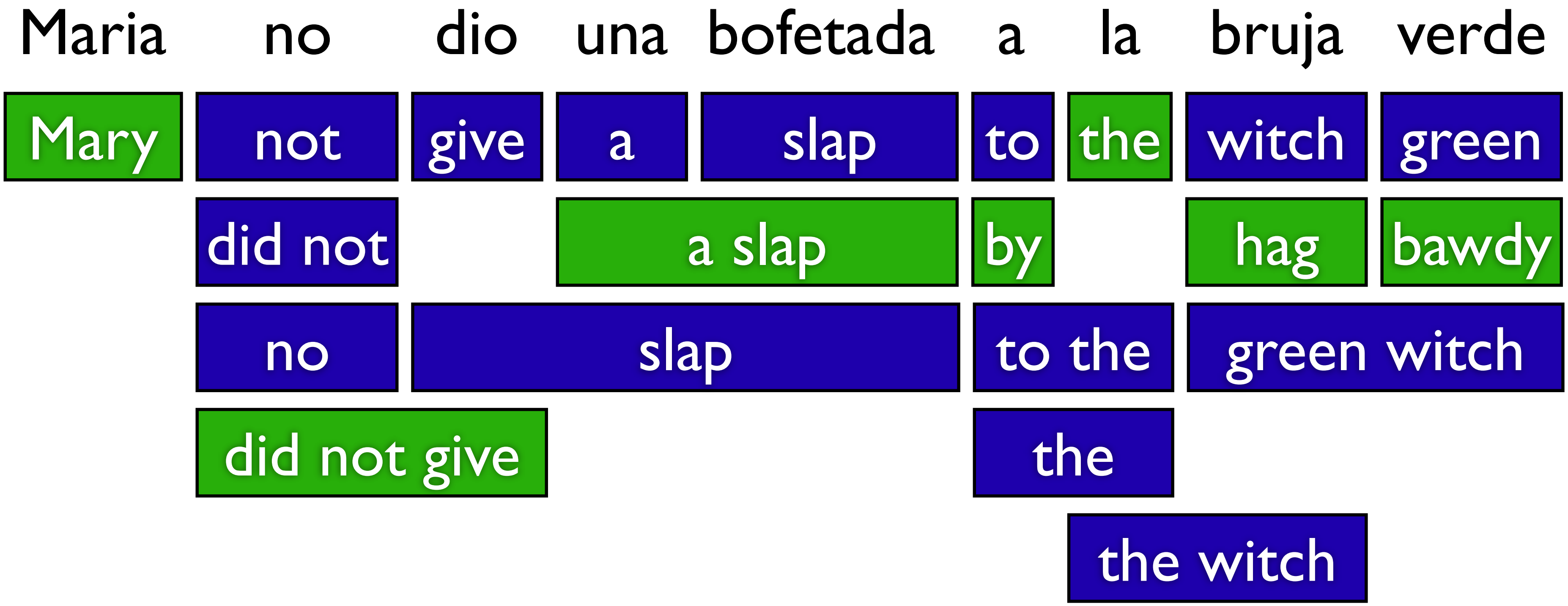
did not give | the

the witch

Adapted from Koehn (2006)

Translation: a statistical perspective

$$\hat{y} = \arg\max_{y \in \text{English}} p(y \mid \text{português})$$



Maria no dio una bofetada a la bruja verde

Mary | not | give | a | slap | to | the | witch | green

did not | a slap | by | hag | bawdy

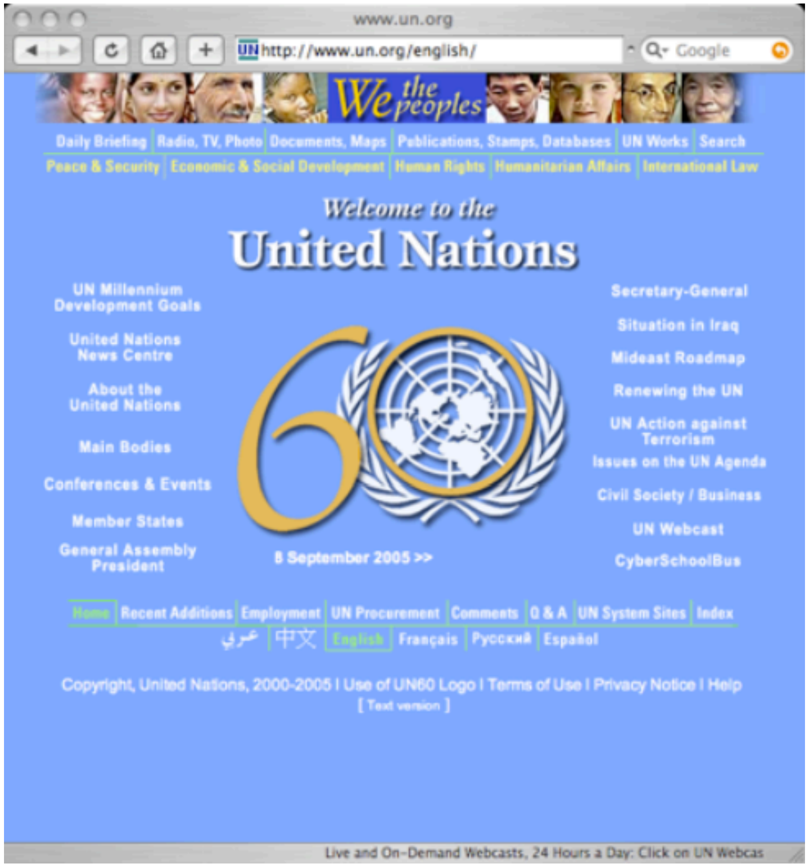
no | slap | to the | green witch

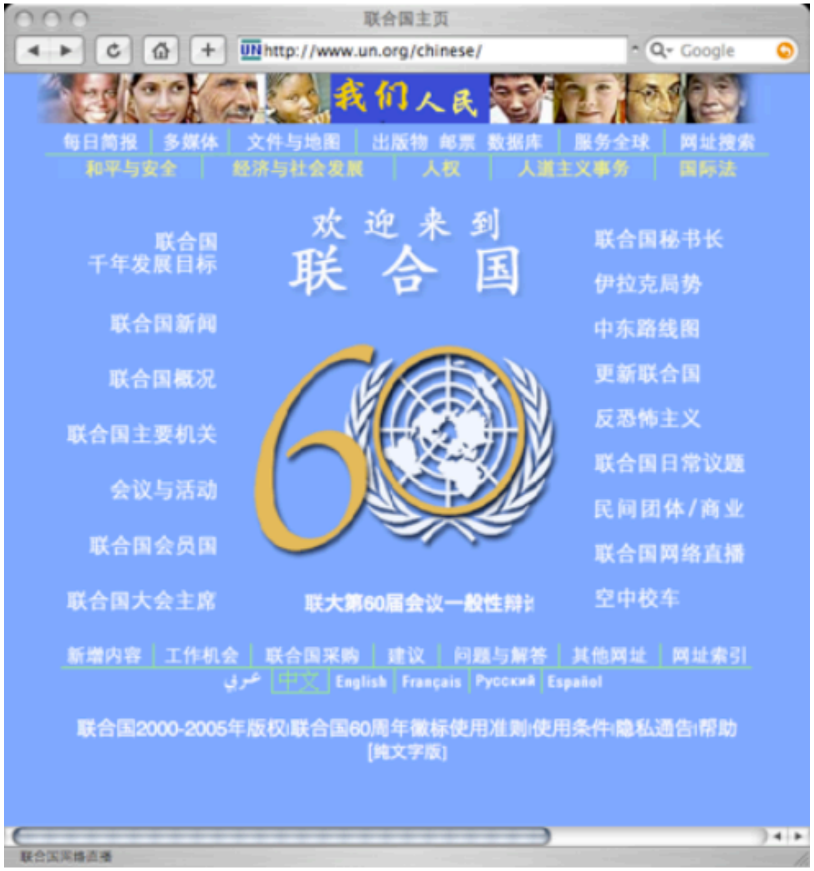
did not give | the

the witch

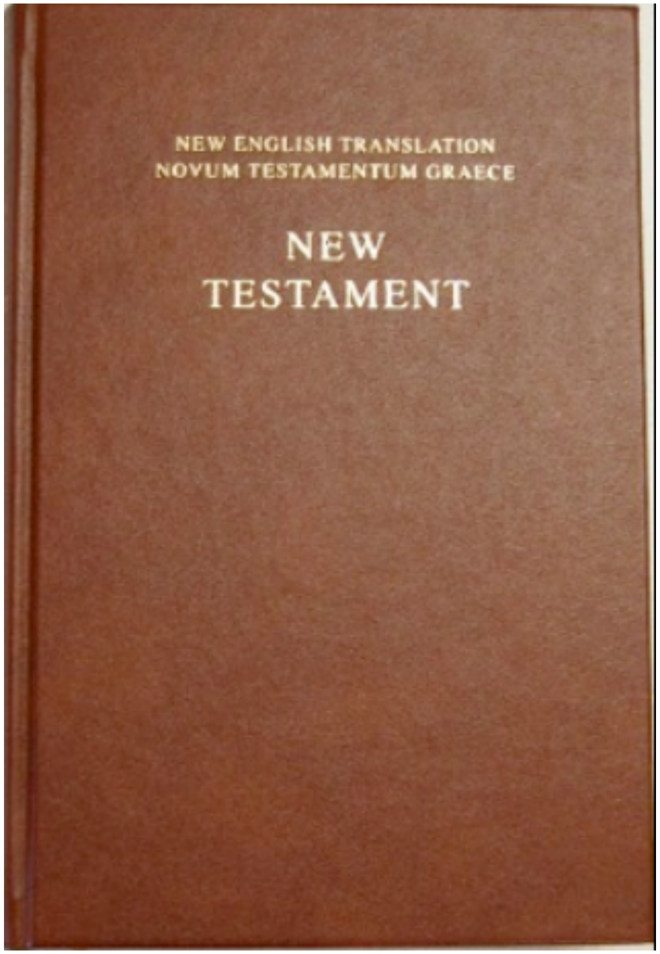
Adapted from Koehn (2006)

Translation: learning distributions

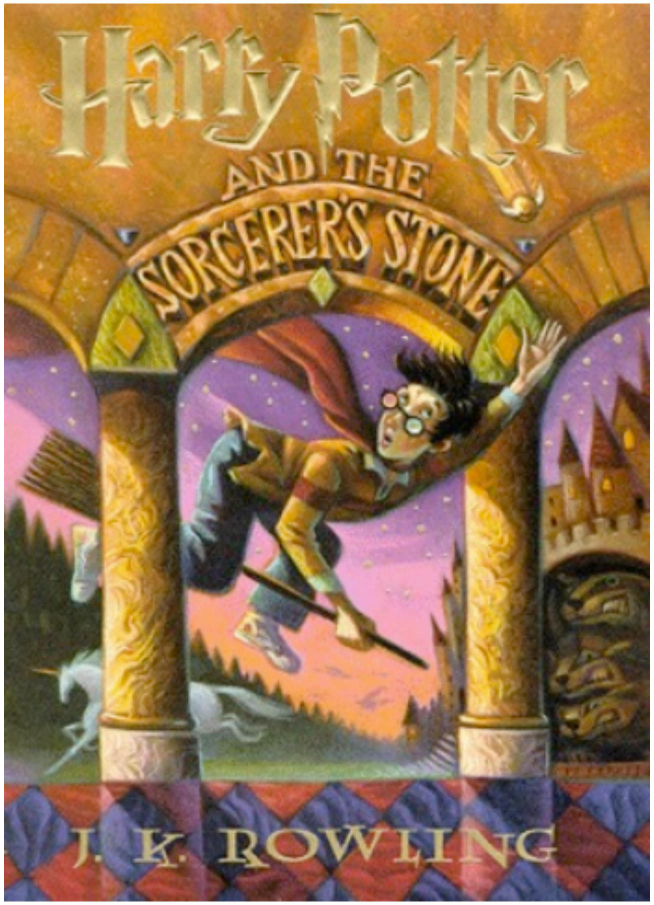

www.un.org
http://www.un.org/english/
Google
We the peoples
Daily Briefing | Radio, TV, Photo | Documents, Maps | Publications, Stamps, Databases | UN Works | Search
Peace & Security | Economic & Social Development | Human Rights | Humanitarian Affairs | International Law
Welcome to the
United Nations
UN Millennium Development Goals
United Nations News Centre
About the United Nations
Main Bodies
Conferences & Events
Member States
General Assembly President
Secretary-General
Situation in Iraq
Mideast Roadmap
Renewing the UN
UN Action against Terrorism
Issues on the UN Agenda
Civil Society / Business
UN Webcast
CyberSchoolBus
8 September 2005 >>
Home | Recent Additions | Employment | UN Procurement | Comments | Q & A | UN System Sites | Index
عربي | 中文 | English | Français | Русский | Español
Copyright, United Nations, 2000-2005 | Use of UN60 Logo | Terms of Use | Privacy Notice | Help
[Text version]
Live and On-Demand Webcasts, 24 Hours a Day: Click on UN Webcas

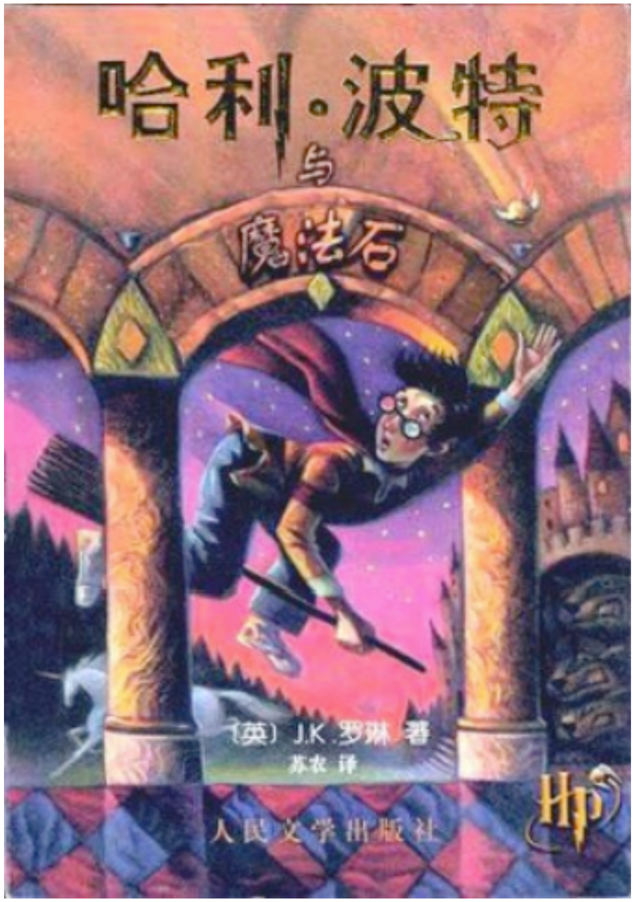

联合国主页
http://www.un.org/chinese/
Google
我们人民
每日简报 | 多媒体 | 文件与地图 | 出版物 邮票 数据库 | 服务全球 | 网址搜索
和平与安全 | 经济与社会发展 | 人权 | 人道主义事务 | 国际法
欢迎来到
联合国
联合国千年发展目标
联合国新闻
联合国概况
联合国主要机关
会议与活动
联合国会员国
联合国大会主席
联合国秘书长
伊拉克局势
中东路线图
更新联合国
反恐怖主义
联合国日常议题
民间团体/商业
联合国网络直播
空中校车
联大第60届会议一般性辩论
新增内容 | 工作机会 | 联合国采购 | 建议 | 问题与解答 | 其他网址 | 网址索引
عربي | 中文 | English | Français | Русский | Español
联合国2000-2005年版权|联合国60周年徽标使用准则|使用条件|隐私通告|帮助
[纯文字版]
联合国网络直播

Translation: learning distributions

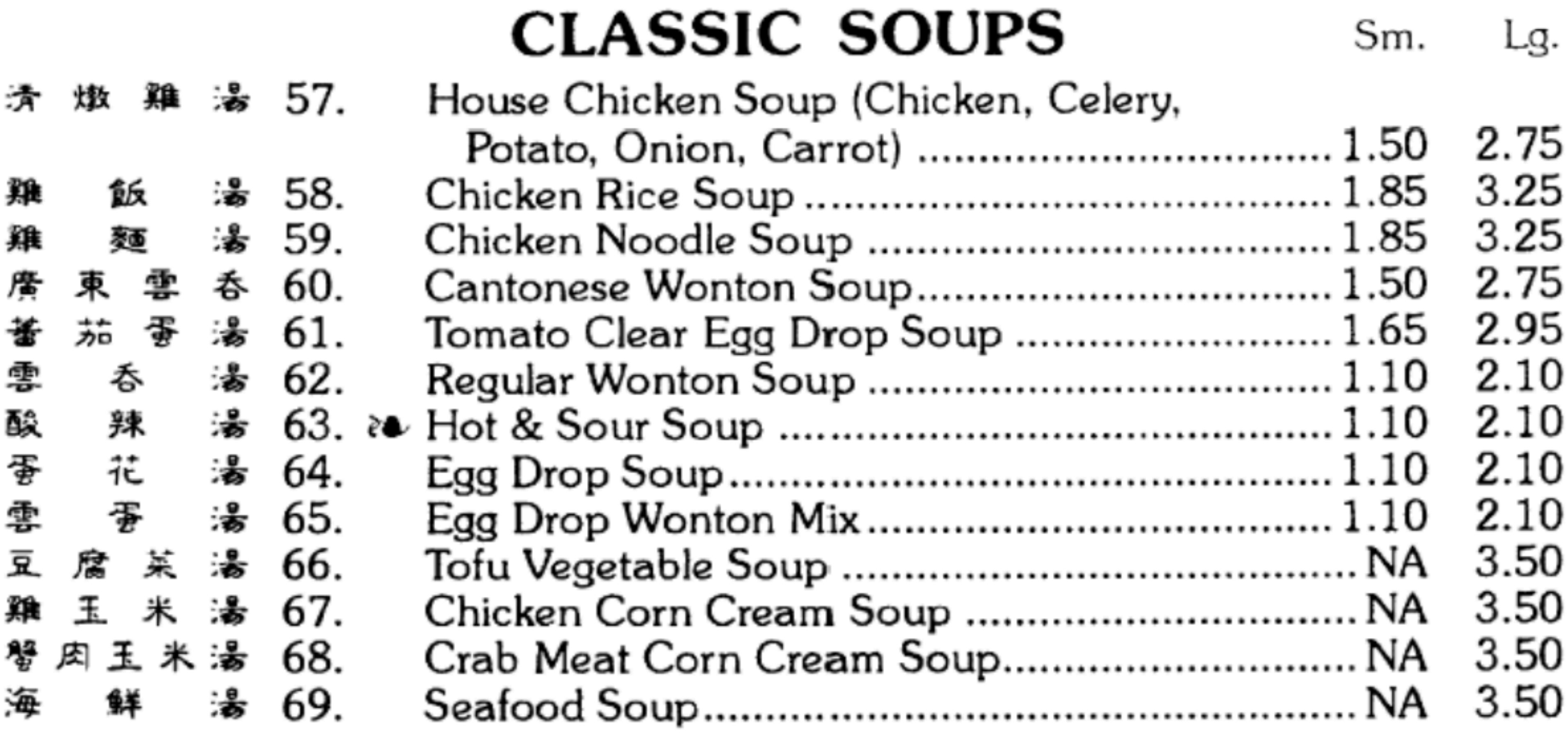


Translation: learning distributions





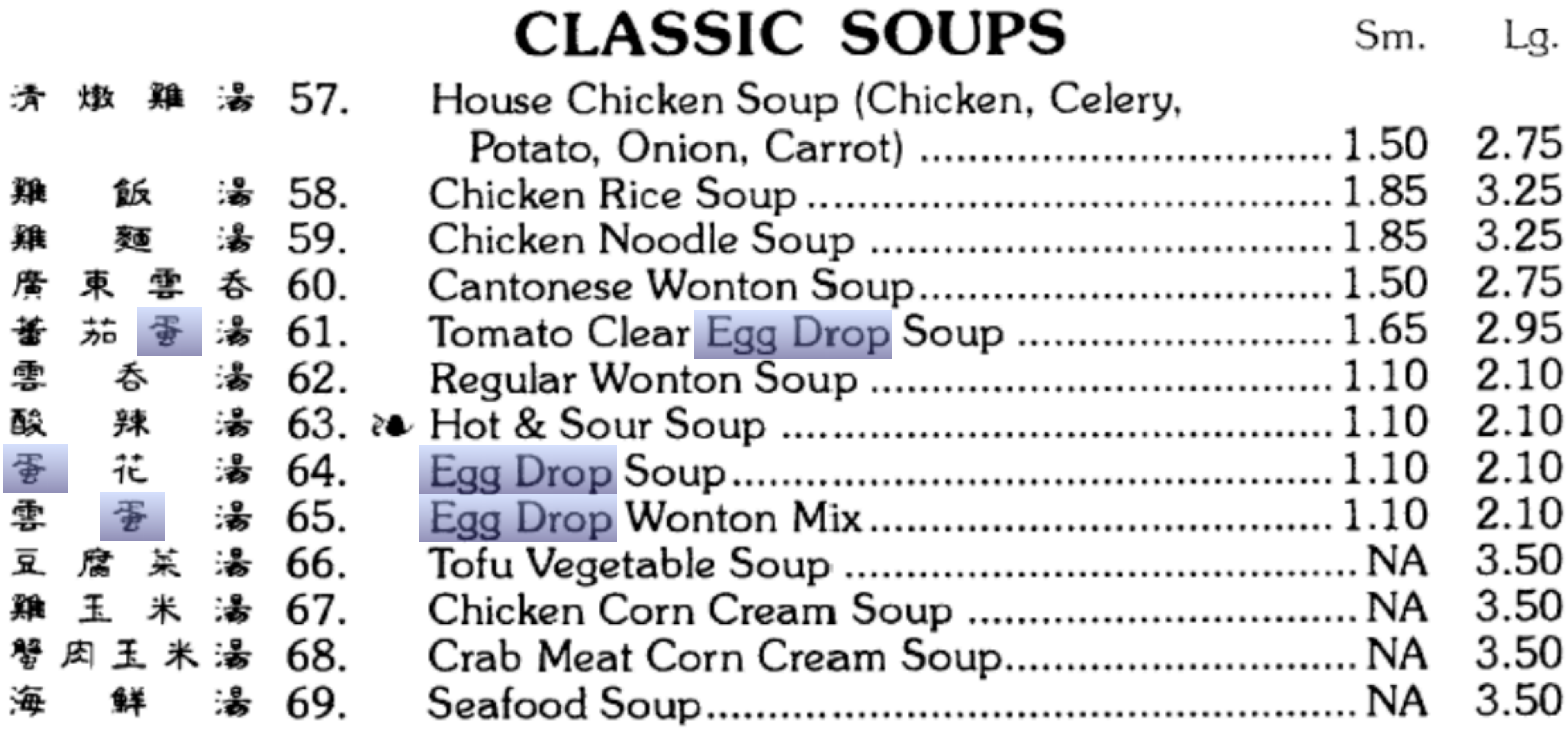
Translation: learning distributions



CLASSIC SOUPS

			Sm.	Lg.
清燉雞湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
雞飯湯	58.	Chicken Rice Soup	1.85	3.25
雞麵湯	59.	Chicken Noodle Soup	1.85	3.25
廣東雲吞	60.	Cantonese Wonton Soup	1.50	2.75
蕃茄蛋湯	61.	Tomato Clear Egg Drop Soup	1.65	2.95
雲吞湯	62.	Regular Wonton Soup	1.10	2.10
酸辣湯	63.	🌶 Hot & Sour Soup	1.10	2.10
蛋花湯	64.	Egg Drop Soup	1.10	2.10
雲蛋湯	65.	Egg Drop Wonton Mix	1.10	2.10
豆腐菜湯	66.	Tofu Vegetable Soup	NA	3.50
雞玉米湯	67.	Chicken Corn Cream Soup	NA	3.50
蟹肉玉米湯	68.	Crab Meat Corn Cream Soup	NA	3.50
海鮮湯	69.	Seafood Soup	NA	3.50

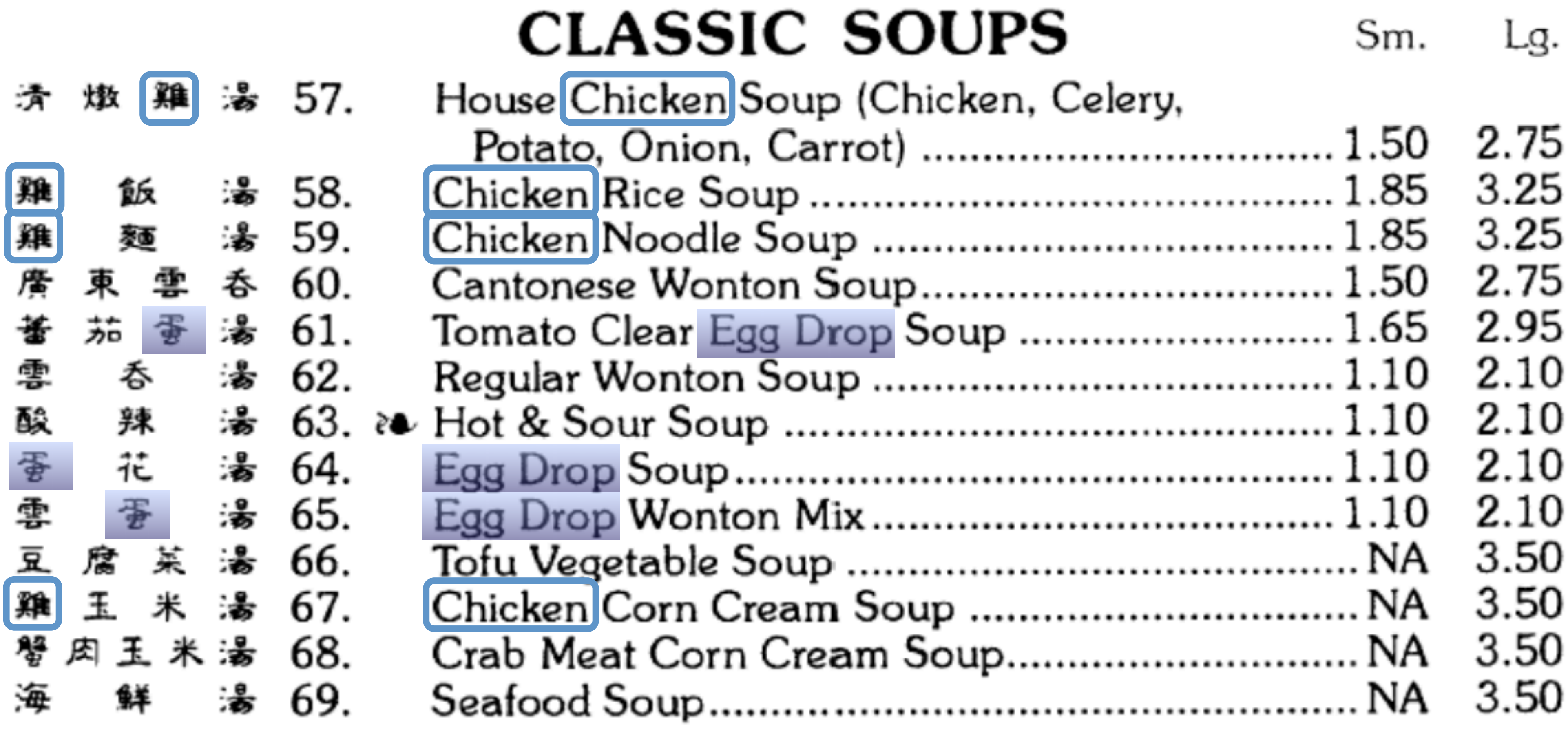
Translation: learning distributions



CLASSIC SOUPS

			Sm.	Lg.
清燉雞湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
雞飯湯	58.	Chicken Rice Soup	1.85	3.25
雞麵湯	59.	Chicken Noodle Soup	1.85	3.25
廣東雲吞	60.	Cantonese Wonton Soup	1.50	2.75
蕃茄蛋湯	61.	Tomato Clear Egg Drop Soup	1.65	2.95
雲吞湯	62.	Regular Wonton Soup	1.10	2.10
酸辣湯	63.	🌶 Hot & Sour Soup	1.10	2.10
蛋花湯	64.	Egg Drop Soup	1.10	2.10
雲蛋湯	65.	Egg Drop Wonton Mix	1.10	2.10
豆腐菜湯	66.	Tofu Vegetable Soup	NA	3.50
雞玉米湯	67.	Chicken Corn Cream Soup	NA	3.50
蟹肉玉米湯	68.	Crab Meat Corn Cream Soup	NA	3.50
海鮮湯	69.	Seafood Soup	NA	3.50

Translation: learning distributions



CLASSIC SOUPS

			Sm.	Lg.
清燉雞湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
雞飯湯	58.	Chicken Rice Soup	1.85	3.25
雞麵湯	59.	Chicken Noodle Soup	1.85	3.25
廣東雲吞	60.	Cantonese Wonton Soup	1.50	2.75
蕃茄蛋湯	61.	Tomato Clear Egg Drop Soup	1.65	2.95
雲吞湯	62.	Regular Wonton Soup	1.10	2.10
酸辣湯	63.	🌶 Hot & Sour Soup	1.10	2.10
蛋花湯	64.	Egg Drop Soup	1.10	2.10
雲蛋湯	65.	Egg Drop Wonton Mix	1.10	2.10
豆腐菜湯	66.	Tofu Vegetable Soup	NA	3.50
雞玉米湯	67.	Chicken Corn Cream Soup	NA	3.50
蟹肉玉米湯	68.	Crab Meat Corn Cream Soup	NA	3.50
海鮮湯	69.	Seafood Soup	NA	3.50

Model form: naïve multinomials

$p(\cdot \mid c\tilde{a}o)$

e	*p*
the	0.0001
and	0.0001
a	0.0001
dog	**0.8**
dogs	**0.18**
canine	**0.01**
cat	0.0001
cats	0.0001
walk	0.0001
walks	0.0001
walked	0.0001
…	

$p(\cdot \mid gato)$

e	*p*
the	0.0001
and	0.0001
a	0.0001
dog	0.0001
dogs	0.0001
canine	0.0001
cat	**0.75**
cats	**0.24**
walk	0.0001
walks	0.0001
walked	0.0001
…	

$p(\cdot \mid andar)$

e	*p*
the	0.0001
and	0.0001
a	0.0001
dog	0.0001
dogs	0.0001
canine	0.0001
cat	0.0001
cats	0.0001
walk	**0.33**
walks	**0.33**
walked	**0.33**
…	

Naïve multinomials: problem?

$$p(\cdot \mid andar)$$

e	*p*
the	0.0001
and	0.0001
a	0.0001
dog	0.0001
dogs	0.0001
canine	0.0001
cat	0.0001
cats	0.0001
walk	**0.33**
walks	**0.33**
walked	**0.33**
…	

Naïve multinomials: problem?

- The vocabularies of languages have **regularities**
  - (English doesn't have many)
  - Russian, Finnish, Turkish have LOTS more regularities
- Can our models exploit such regularities? **YES.**
- Do we need this in the world of big data? **YES.**

$p(\cdot \mid andar)$

e	*p*
the	0.0001
and	0.0001
a	0.0001
dog	0.0001
dogs	0.0001
canine	0.0001
cat	0.0001
cats	0.0001
walk	**0.33**
walks	**0.33**
walked	**0.33**
…	

Outline

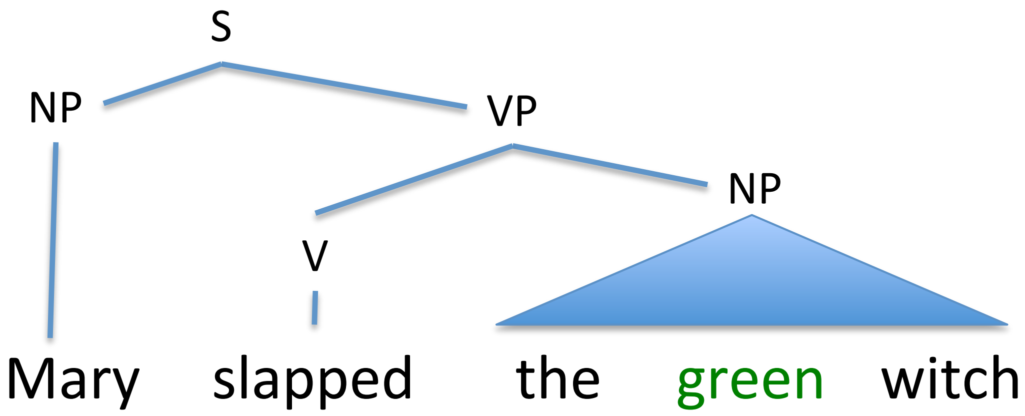
- Introduction to statistical translation
- Introduction to morphology
- Modeling morphologically rich translation
- Aside: Unsupervised morphology
- Experiments

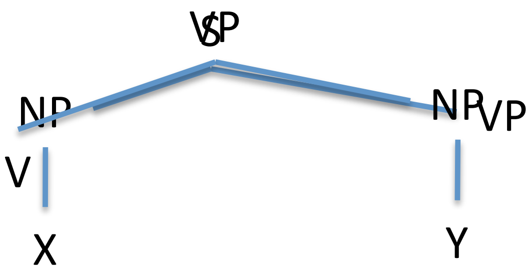
Outline

- Introduction to statistical translation
- Introduction to morphology
- Modeling morphologically rich translation
- Aside: Unsupervised morphology
- Experiments

Expressing Grammatical Relations

English uses **syntactic structure** to express grammatical relations like ***argumentation*** and ***modification***



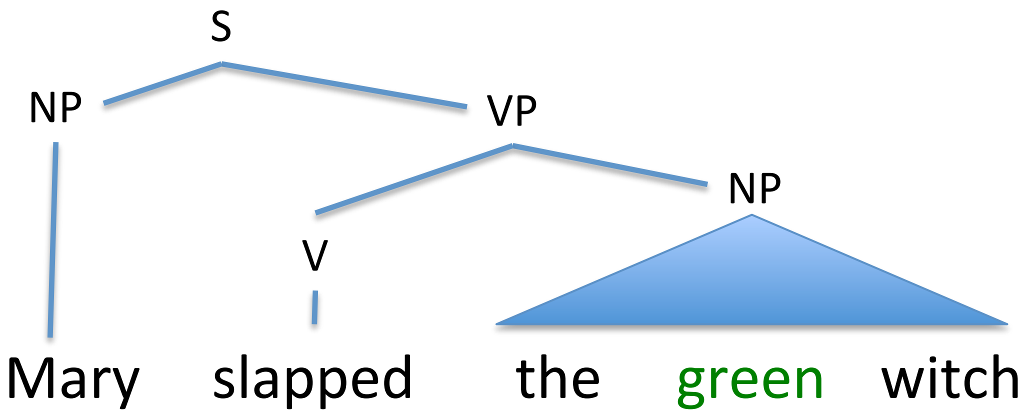


“X is the subject”

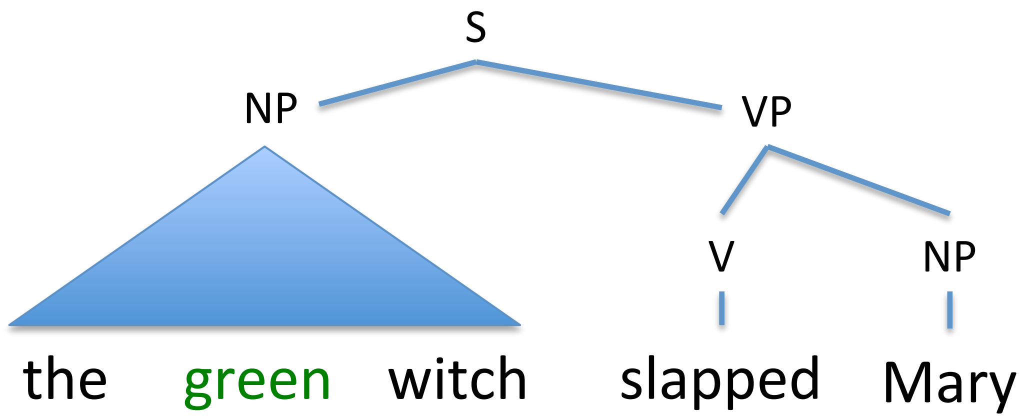
“X is the direct object”

Expressing Grammatical Relations

English uses **syntactic structure** to express grammatical relations like ***argumentation*** and ***modification***



S
NP VP
V NP
Mary slapped the green witch



S
NP VP
V NP
the green witch slapped Mary

Some Russian Data

Мери	ударила	зеленую	ведьму
mary	*udarila*	*zelenuyu*	*ved’mu*
MARY	SLAPPED	GREEN	WITCH

зеленую	ведьму	ударила	Мери
zelenuyu	*ved’mu*	*udarila*	*mary*
GREEN	WITCH	SLAPPED	MARY

ударила	зеленую	ведьму	Мери
udarila	*zelenuyu*	*ved’mu*	*mary*
SLAPPED	GREEN	WITCH	MARY

Some Russian Data

Мери	ударила	зеленую	ведьму
mary	*udarila*	*zelenuyu*	*ved’mu*
MARY	SLAPPED	GREEN	WITCH

зеленую	ведьму	ударила	Мери
zelenuyu	*ved’mu*	*udarila*	*mary*
GREEN	WITCH	SLAPPED	MARY

ударила	зеленую	ведьму	Мери
udarila	*zelenuyu*	*ved’mu*	*mary*
SLAPPED	GREEN	WITCH	MARY

Morphology instead of syntax

Russian uses **morphological inflection** to express the **same grammatical relations**.

Morphology instead of syntax

Russian uses **morphological inflection** to express the **same grammatical relations**.

Here are a few things that different languages use inflectional morphology for:

- Tense
- Mood
- Aspect
- Negation
- Voice
- Ability
- Applicativity
- Factivity
- Definiteness
- Agreement
- Gender
- Spatial relations
- Person
- Number

Inflectional Morphology

- The part-of-speech of the **stem** determines the required/possible inflections
  - English nouns express number (singual *vs*. plural) cat/cats
  - Portuguese adjectives express number *and* gender louco/louca/loucos/loucas

Inflectional Morphology

- Inflection can express **multiple grammatical features**

  {+ACC,+DAT,+NOM,+ERG} x {+FUT,+PAST} x ...

  - With a **single morpheme** (**fusional** languages)
    Indo-European [Russian, Portuguese, Hindi, Greek]

  - With ~**one morpheme per feature** (**agglutinative** languages)
    Turkish, Finnish, Hungarian, Basque, Japanese

Inflectional Morphology

- **Underlying forms**
  - Example: walk +PROG
  - Example: sing +PAST
  - Example: k-t-b +FUT+1P+DUAL+IND

Inflectional Morphology

- **Underlying forms**
  - Example: walk +PROG
  - Example: sing +PAST
  - Example: k-t-b +FUT+1P+DUAL+IND
- **Surface realization ("exponence")**
  - **Concatenation**
    Prefixes, suffixes, circumfixes, infixes
    Add *–ing* to a verb to express +PROG
  - **Ablaut**
    Change vowel (usually) template of stem
    Change /i/ to /a/ to express +PAST
  - **Reduplication**
    Repeat the first syllable of the word to express +PLURAL

Morphological Analysis

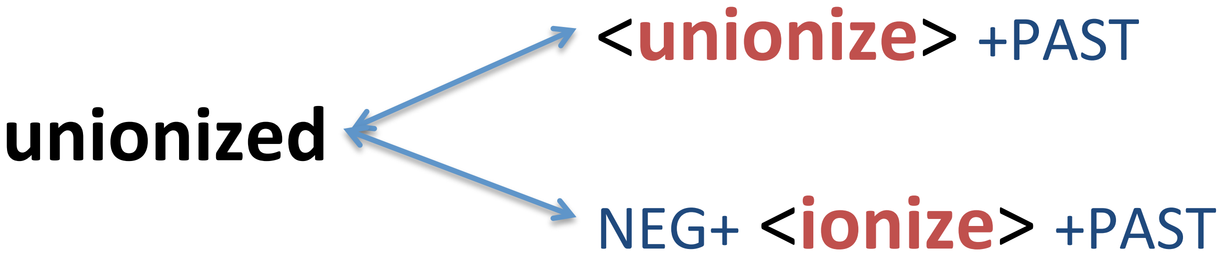
- Decompose an inflected word into its **stem**(s) and **inflectional morphemes**

walking → <**walk**> +PROG

пыталась → <**пытаться**> +IND+PAST+SING+FEM+MED+PERF

Morphological Analysis

- Decompose an inflected word into its **stem**(s) and **inflectional morphemes**
- Two approaches
  - **Rule-based morphological analyzer**
    - Computationally tractable with **finite-state transducers**
    - In general: one word-to-many analyses mapping
    - Use statistical model to **disambiguate** analyses **in context**



unionized

<**unionize**> +PAST

NEG+ <**ionize**> +PAST

Morphological Analysis

- Decompose an inflected word into its **stem**(s) and **inflectional morphemes**
- Two approaches
  - **Rule-based morphological analyzer**
    - Computationally tractable with **finite-state transducers**
    - In general: one word-to-many analyses mapping
    - Use statistical model to **disambiguate** analyses **in context**
  - **Morphology light: segment word into morphemes**
    - Challenges: *allomorphy*, *nonconcatenative morphology*
      analyzed = <**analyze**>**+d** or <**analyz**>**+ed**?
      sang = ???
    - Good unsupervised algorithms (we give one later)

Outline

- Introduction to statistical translation
- Introduction to morphology
- Modeling morphologically rich translation
- Aside: Unsupervised morphology
- Experiments

Task: Translate into a MRL

- Given English, generate {Russian, Swahili, Hebrew, …}
- **This is an important problem!**
  - Lots of information published in English
  - Lots of people who would prefer to read it in other languages

Model desiderata

- Words with **common stems** should share statistical strength
- Source syntactic context should be used to predict inflection
- Inflection should be modeled using features (+MASC+PL is more similar to +MASC+SING than to +FEM+SING)

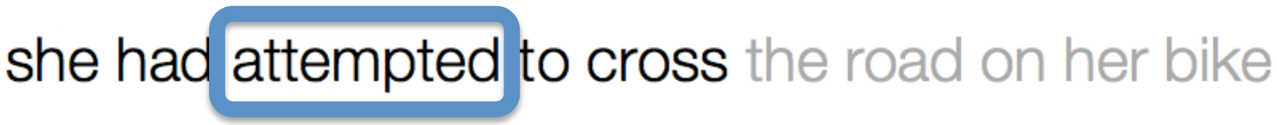
Model desiderata

- Words with **common stems** should share statistical strength
- Source syntactic context should be used to predict inflection
- Inflection should be modeled using features (+MASC+PL is more similar to +MASC+SING than to +FEM+SING)

$$\underset{\text{Stem}}{\sigma} \star \underset{\text{Inflection}}{\mu} = \underset{\textbf{Inflected word}}{f}$$

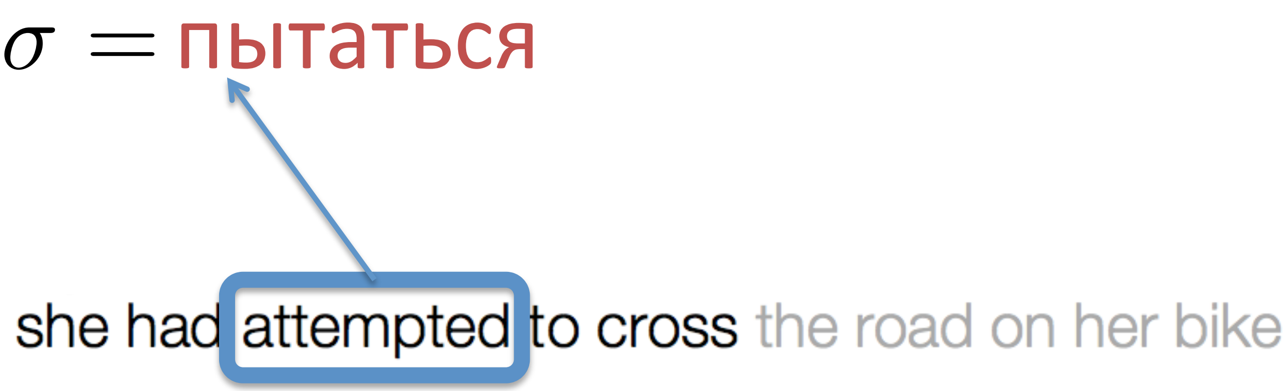
$$p(\sigma,\mu \mid \text{context}) = p(\sigma \mid \text{context}) \times \boxed{p(\mu \mid \sigma, \text{context})}$$

Predicting Inflection in Translation



she had attempted to cross the road on her bike

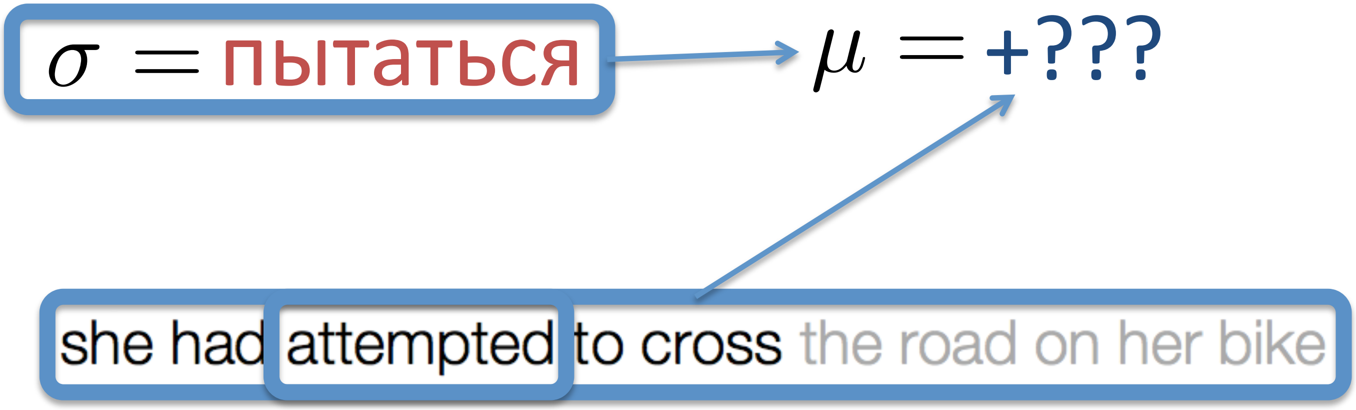
Predicting Inflection in Translation



$\sigma =$ пытаться

she had attempted to cross the road on her bike

Predicting Inflection in Translation

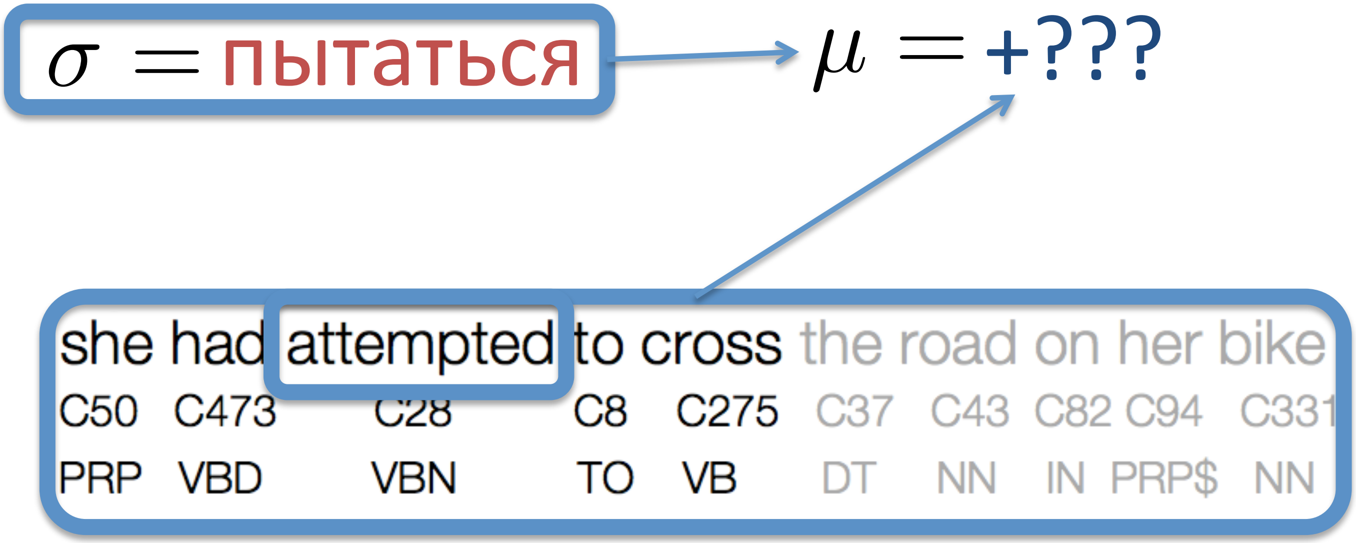


$\sigma =$ пытаться

$\mu = +???$

she had attempted to cross the road on her bike

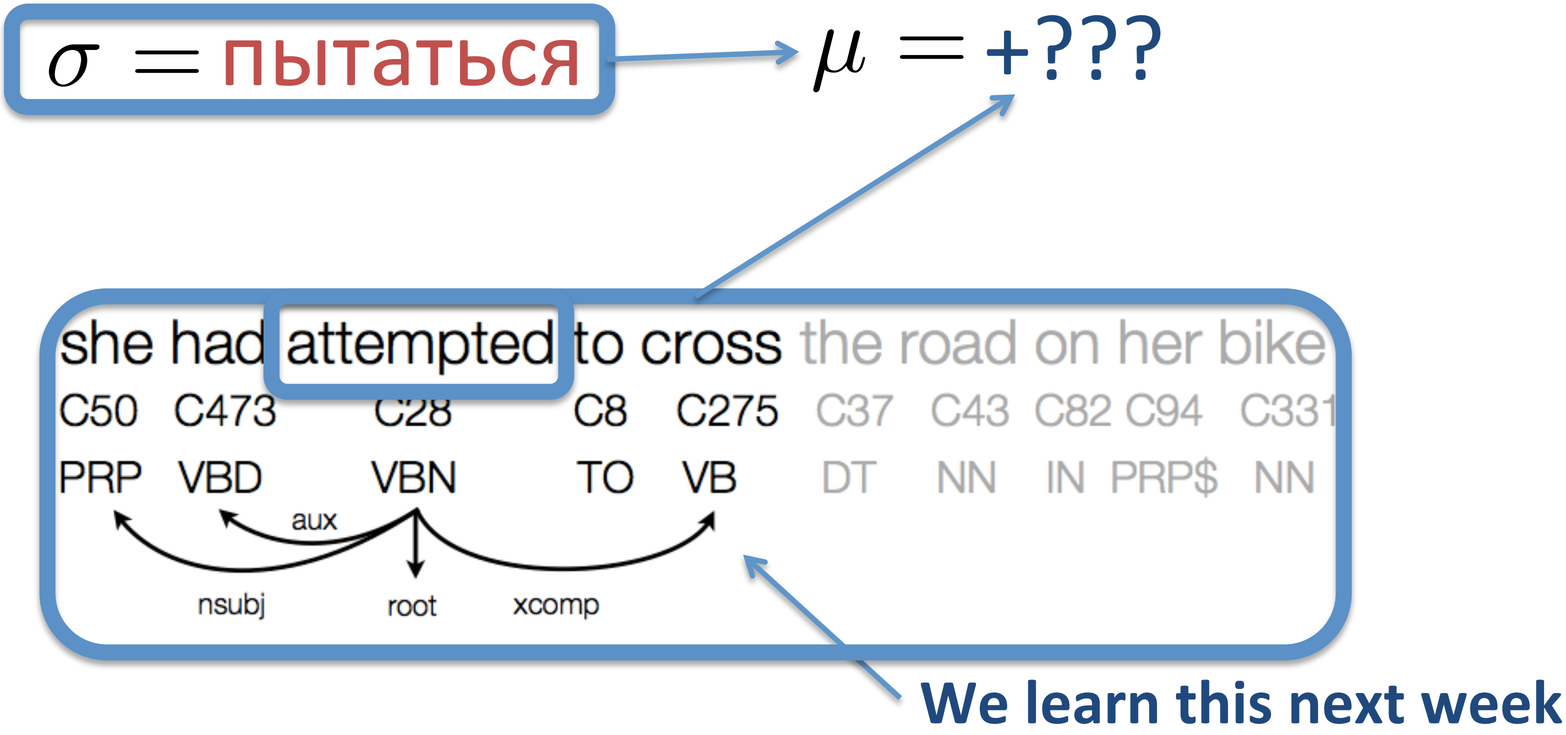
Predicting Inflection in Translation



$\sigma =$ пытаться → $\mu =$ +???

she	had	attempted	to	cross	the	road	on	her	bike
C50	C473	C28	C8	C275	C37	C43	C82	C94	C331
PRP	VBD	VBN	TO	VB	DT	NN	IN	PRP$	NN

Predicting Inflection in Translation



$\sigma =$ пытаться → $\mu = +???$

she had attempted to cross the road on her bike
C50 C473 C28 C8 C275 C37 C43 C82 C94 C331
PRP VBD VBN TO VB DT NN IN PRP$ NN
aux
nsubj
root
xcomp

We learn this next week

Inflection Model: Logistic Regression

$$p(\mu \mid \mathbf{x}) = \frac{\exp \boldsymbol{w}^\top \boldsymbol{f}(\mu, \mathbf{x})}{\sum_{\mu'} \exp \boldsymbol{w}^\top \boldsymbol{f}(\mu', \mathbf{x})}$$

Features of $\mathbf{x}$

Parent of the source is NNS

Source word is VBD

Source word has 3 dependents

Source word is attempted

Source word is the object of a verb

Source word -1 is would

Inflection Model: Logistic Regression

$$p(\mu \mid \mathbf{x}) = \frac{\exp \boldsymbol{w}^\top \boldsymbol{f}(\mu, \mathbf{x})}{\sum_{\mu'} \exp \boldsymbol{w}^\top \boldsymbol{f}(\mu', \mathbf{x})}$$

Features of $\mathbf{x}$

Parent of the source is **NNS**

Source word is **VBD**

Source word has **3** dependents

Source word is **attempted**

Source word is the object of a verb

Source word **-1** is **would**

$\Omega = \{$

+IND+PAST+SING+FEM+MED+PERF,

+IND+FUT+SING+FEM+MED,

+IND+PAST+PL+FEM+MED,

+IND+PAST+SING+MASC+MED,

+IND+PAST+PL+MASC+MED,

$\}$

Inflection Model: Logistic Regression

$$p(\mu \mid \mathbf{x}) = \frac{\exp \boldsymbol{w}^\top \boldsymbol{f}(\mu, \mathbf{x})}{\sum_{\mu'} \exp \boldsymbol{w}^\top \boldsymbol{f}(\mu', \mathbf{x})}$$

Features of $\mathbf{x}$

Parent of the source is **NNS**

Source word is **VBD**

Source word has **3** dependents

Source word is **attempted**

Source word is the object of a verb

Source word **-1** is **would**

Features of μ $=$

+IND+PAST+SING+FEM+MED+PERF

+IND
+PAST
+SING
+FEM
+MED
+PERF

Inflection Model

$$p(\mu \mid \mathbf{x}) = \frac{\exp\left[\overset{\text{input-output correlations}}{\boldsymbol{f}(\mathbf{x})^\top \mathbf{W} \boldsymbol{g}(\mu)} + \overset{\text{output correlations}}{\boldsymbol{g}(\mu)^\top \mathbf{V} \boldsymbol{g}(\mu)}\right]}{Z(\mathbf{x})}$$

$$\boldsymbol{f}(\mathbf{x})$$

Parent of the source is NNS

Source word is VBD

Source word has 3 dependents

Source word is attempted

Source word is the object of a verb

Source word -1 is would

$$\boldsymbol{g}(\mu)$$

+IND

+PAST

+SING

+FEM

+MED

+PERF

Inflection Model – Feature Space

Linear in $\boldsymbol{f}'(\mu, \mathbf{x}) = \boldsymbol{f}(\mathbf{x})\boldsymbol{g}(\mu)^\top$

	+ACC	+NOM	+DAT	+SG	+PL	+MASC	…
Parent_NN	x	x	x	x	x	x	…
Parent_NNS	x	x	x	x	x	x	…
Parent_VBD	x	x	x	x	x	x	…
Parent_VBG	x	x	x	x	x	x	…
Left_NN	x	x	x	x	x	x	…
Left_NNS	x	x	x	x	x	x	…
Left_VBD	x	x	x	x	x	x	…
…	…	…	…	…	…	…	…

Infection Model: Training

- Training data extracted from parallel corpus
  - Morphologically analyze and disambiguate target side of parallel corpus
  - Syntactic analysis of English source
  - Align words
  - **Every word pair in the parallel corpus becomes a training instance for the inflection model**
- Stochastic gradient descent, LBFGS, etc.

Outline

- Introduction to statistical translation
- Introduction to morphology
- Modeling morphologically rich translation
- Aside: Unsupervised morphology
- Experiments

Aside: Unsupervised Morphology

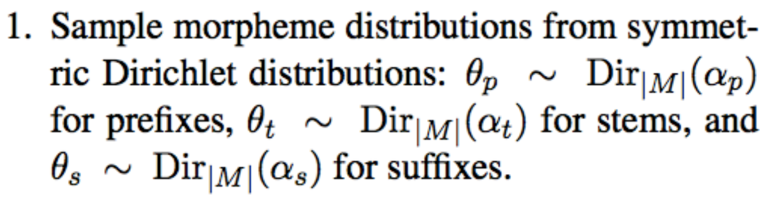
- Morphological analyzers may not exist for a language we want to translate into
- We would like to be able to use **unsupervised morphological analysis**
  - We assume words **decompose concatenatively**
  - We require the model to distinguish between the **stem** and **non-stem** parts of the word

Unsupervised Morphology

- Bayesian methods are effective
  - there are very nice nonparametric solutions to the problem (Goldwater & Griffiths, Johnson et al)
  - Nonparametrics can be slow, so we are going to introduce a slightly simpler parametric model

Grammar: M^*MM^*

Unsupervised Morphology



1. Sample morpheme distributions from symmetric Dirichlet distributions: $\theta_p \sim \mathrm{Dir}_{|M|}(\alpha_p)$ for prefixes, $\theta_t \sim \mathrm{Dir}_{|M|}(\alpha_t)$ for stems, and $\theta_s \sim \mathrm{Dir}_{|M|}(\alpha_s)$ for suffixes.

Hyperparameters: $\alpha_p, \alpha_t, \alpha_s$

By setting $0 \ll \alpha_p, \alpha_t \ll \alpha_s \ll 1$ we find we learn the high-entropy stem part of the word reliably.

Sampling representation:

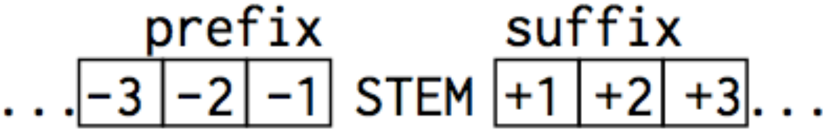
<walk>+ing

<sing>+ing

<fasten>+ing

Unsupervised Morphology: Features

- For defining output features $\boldsymbol{g}(\mu)$ we use:



wa+ki+wa+<**piga**>

Prefix[-1][wa]

Prefix[-2][ki]

Prefix[-3][wa]

Outline

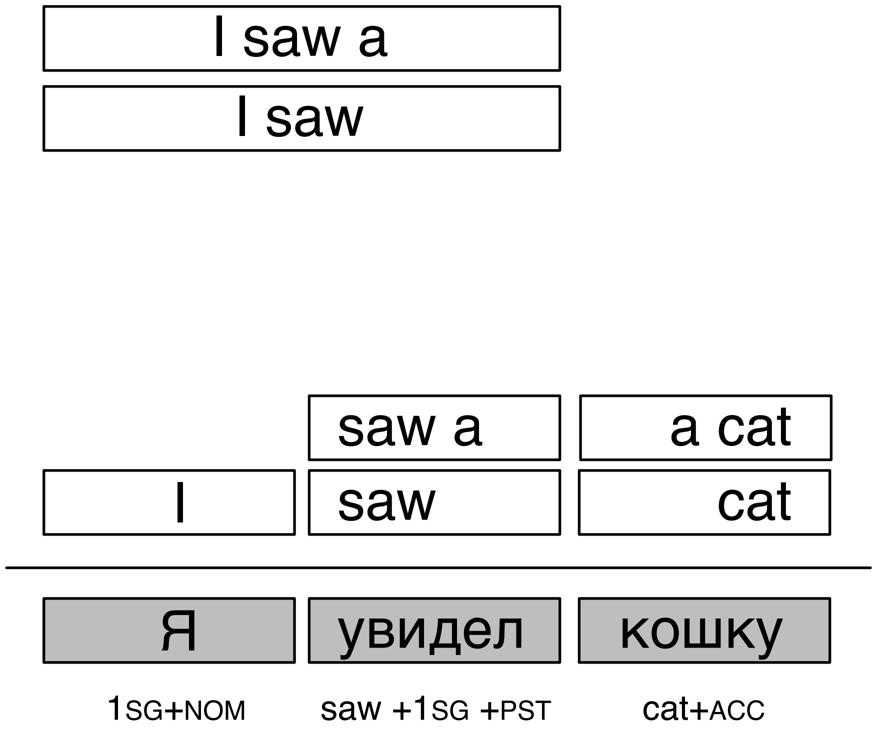
- Introduction to statistical translation
- Introduction to morphology
- Modeling morphologically rich translation
- Aside: Unsupervised morphology
- Experiments

Back to translation

How might this sentence be translated?

Я	увидел	кошку
1SG+NOM	saw +1SG +PST	cat+ACC

Back to translation



I saw a

I saw

saw a

a cat

I

saw

cat

Я

увидел

кошку

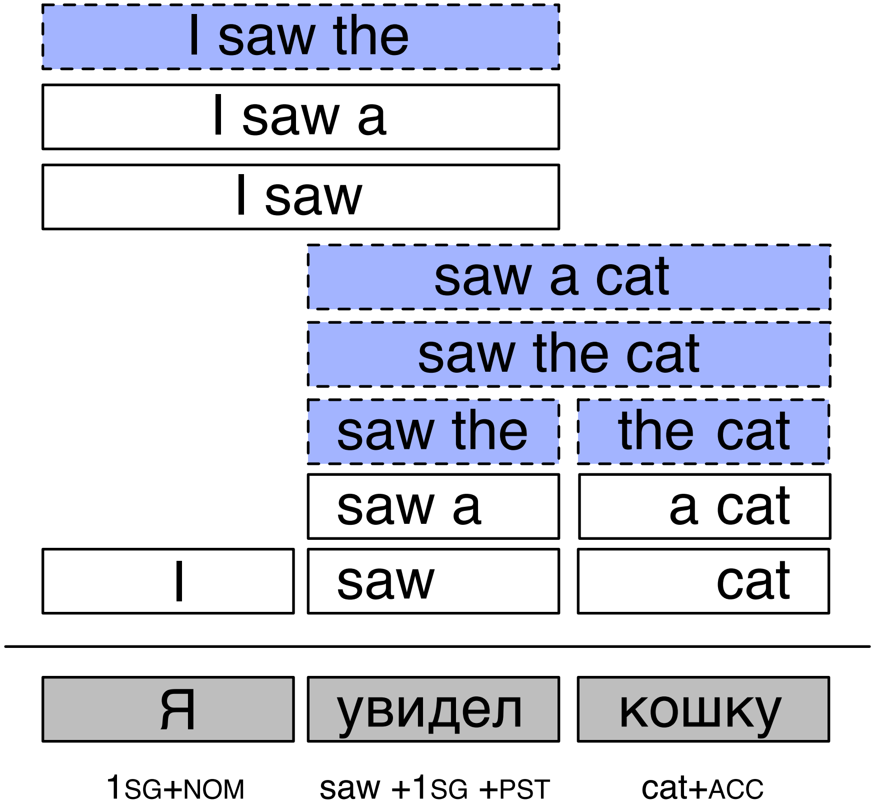
1SG+NOM

saw +1SG +PST

cat+ACC

What about ***I saw the cat***?

“Synthetic Translation Options”



I saw the

I saw a

I saw

saw a cat

saw the cat

saw the

the cat

saw a

a cat

I

saw

cat

Я

увидел

кошку

1SG+NOM

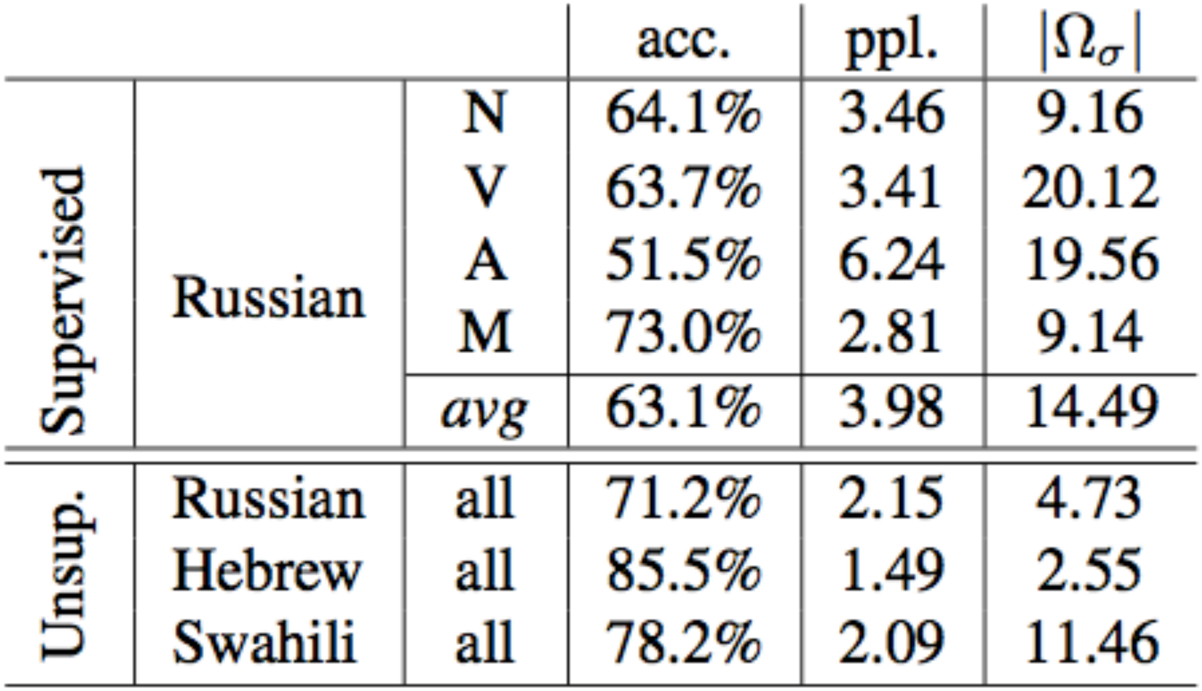
saw +1SG +PST

cat+ACC

Data

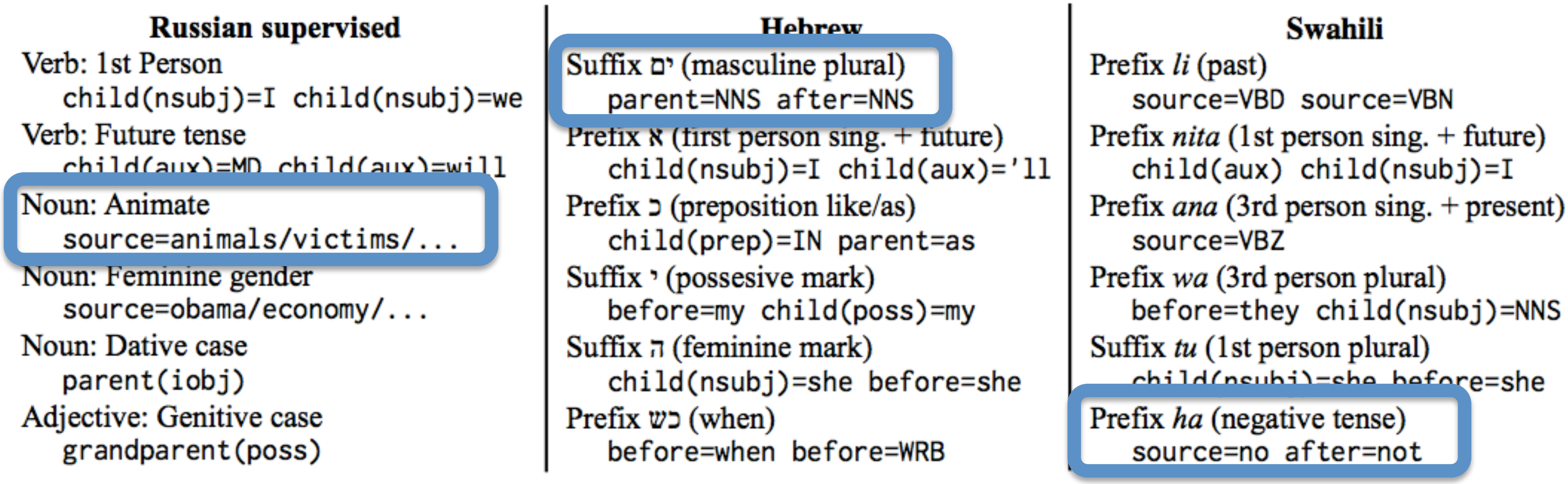
- English—Russian
  - Supervised morphological analyzer
  - Unsupervised morphological analyzer
  - 150k sentence pairs
- English—Hebrew
  - Unsupervised morphological analyzer only
  - 134k sentence pairs
- English—Swahili
  - Unsupervised morphological analyzer only
  - 15k sentence pairs

Intrinsic Evaluation: Quantitative



			acc.	ppl.	$\lvert\Omega_\sigma\rvert$
Supervised	Russian	N	64.1%	3.46	9.16
		V	63.7%	3.41	20.12
		A	51.5%	6.24	19.56
		M	73.0%	2.81	9.14
		avg	63.1%	3.98	14.49
Unsup.	Russian	all	71.2%	2.15	4.73
	Hebrew	all	85.5%	1.49	2.55
	Swahili	all	78.2%	2.09	11.46

Intrinsic Evaluation: Qualitative



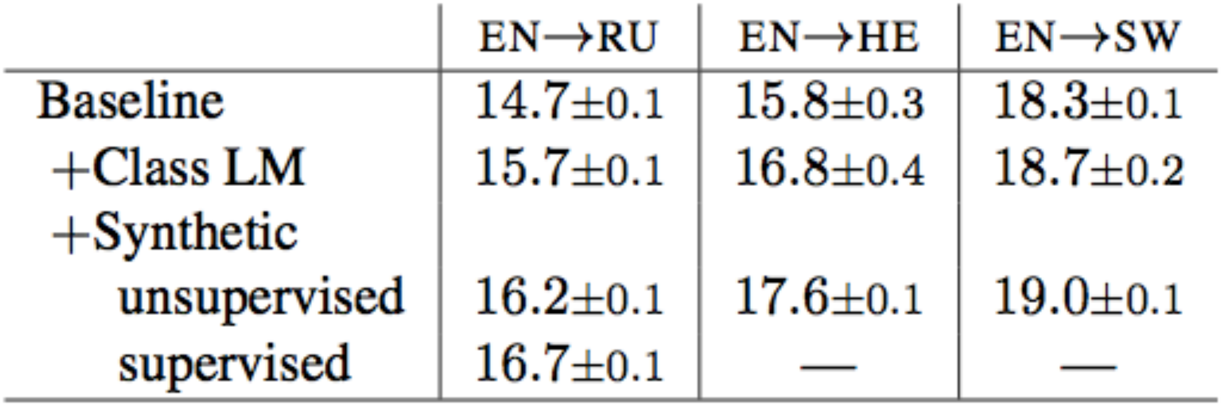
Russian supervised	Hebrew	Swahili
Verb: 1st Person `child(nsubj)=I child(nsubj)=we`	Suffix ים (masculine plural) `parent=NNS after=NNS`	Prefix *li* (past) `source=VBD source=VBN`
Verb: Future tense `child(aux)=MD child(aux)=will`	Prefix א (first person sing. + future) `child(nsubj)=I child(aux)='ll`	Prefix *nita* (1st person sing. + future) `child(aux) child(nsubj)=I`
Noun: Animate `source=animals/victims/...`	Prefix כ (preposition like/as) `child(prep)=IN parent=as`	Prefix *ana* (3rd person sing. + present) `source=VBZ`
Noun: Feminine gender `source=obama/economy/...`	Suffix י (possesive mark) `before=my child(poss)=my`	Prefix *wa* (3rd person plural) `before=they child(nsubj)=NNS`
Noun: Dative case `parent(iobj)`	Suffix ה (feminine mark) `child(nsubj)=she before=she`	Suffix *tu* (1st person plural) `child(nsubj)=she before=she`
Adjective: Genitive case `grandparent(poss)`	Prefix כש (when) `before=when before=WRB`	Prefix *ha* (negative tense) `source=no after=not`

- Highly weighted features learned in training
  - Many highly interpretable features
  - **Semantics for inflection?**

Extrinsic Evaluation: Translation

- **Synthetic translation options**
  - Create default phrase table
  - Create synthetic translation options
    - Create “stemmed” target phrase table
    - For the sentence being translated,
      - For every stem in phrase table, predict MAP inflected form using source context
      - Add resulting phrase (features: stem translation probability, inflection probability, synthetic indicator)
- **Language modeling**
  - N-grams don’t work well in MRLs
  - Add a secondary “Brown Cluster” LM
  - *More interesting approaches, but that’s another talk*

Extrinsic Evaluation: Translation



	EN→RU	EN→HE	EN→SW
Baseline	14.7±0.1	15.8±0.3	18.3±0.1
+Class LM	15.7±0.1	16.8±0.4	18.7±0.2
+Synthetic			
unsupervised	16.2±0.1	17.6±0.1	19.0±0.1
supervised	16.7±0.1	—	—

Summary

- **Morphology matters**
  - Big data is big, but not limitless
  - *English is **not** typologically representative – but most of our models were developed with!*
  - Rule-based morphology is good, but imperfect unsupervised morphology can work well
- The "output feature" formulation of LR is flexible and easy to implement
  - Next stop: unsupervised learning of feature representations (just another partial derivative!)

Obrigado!

Victor Chahuneau
Eva Schlinger
Yulia Tsvetkov
Noah A. Smith

clab