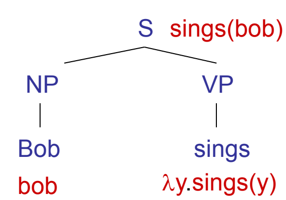
Statistical NLP
Spring 2010

Lecture 20: Compositional Semantics

Dan Klein – UC Berkeley

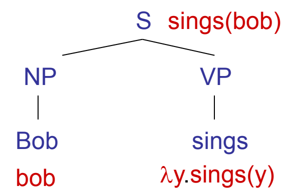
Truth-Conditional Semantics

- Linguistic expressions:
  - “Bob sings”
- Logical translations:
  - sings(bob)
  - Could be p_1218(e_397)
- Denotation:
  - [[bob]] = some specific person (in some context)
  - [[sings(bob)]] = ???
- Types on translations:
  - bob : e (for entity)
  - sings(bob) : t (for truth-value)



Truth-Conditional Semantics

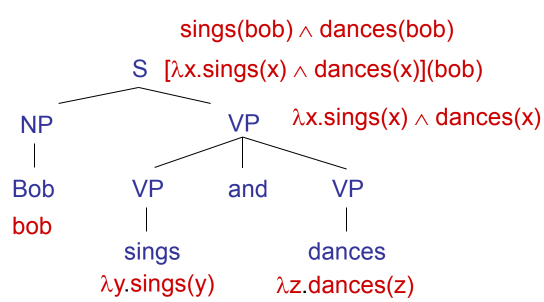
- Proper names:
  - Refer directly to some entity in the world
  - Bob : bob [[bob]]^{W} → ???
- Sentences:
  - Are either true or false (given how the world actually is)
  - Bob sings : sings(bob)



- So what about verbs (and verb phrases)?
  - sings must combine with bob to produce sings(bob)
  - The λ-calculus is a notation for functions whose arguments are not yet filled.
  - sings : λx.sings(x)
  - This is *predicate* – a function which takes an entity (type e) and produces a truth value (type t). We can write its type as e→t.
  - Adjectives?

Compositional Semantics

- So now we have meanings for the words
- How do we know how to combine words?
- Associate a combination rule with each grammar rule:
  - S : β(α) → NP : α VP : β (function application)
  - VP : λx . α(x) ∧ β(x) → VP : α and : ∅ VP : β (intersection)
- Example:

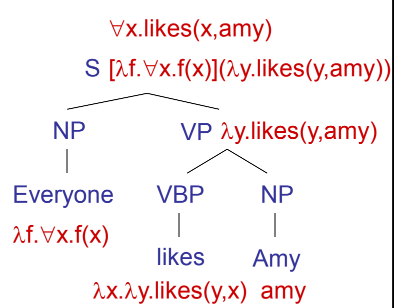


Denotation

- What do we do with logical translations?
  - Translation language (logical form) has fewer ambiguities
  - Can check truth value against a database
    - Denotation (“evaluation”) calculated using the database
  - More usefully: assert truth and modify a database
  - Questions: check whether a statement in a corpus entails the (question, answer) pair:
    - “Bob sings and dances” $\rightarrow$ “Who sings?” + “Bob”
  - Chain together facts and use them for comprehension

Other Cases

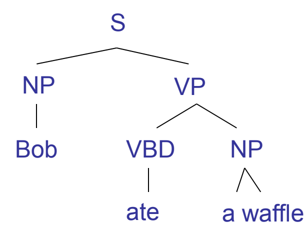
- Transitive verbs:
  - likes : $\lambda x.\lambda y.likes(y,x)$
  - Two-place predicates of type $e \rightarrow (e \rightarrow t)$.
  - likes Amy : $\lambda y.likes(y,Amy)$ is just like a one-place predicate.
- Quantifiers:
  - What does “Everyone” mean here?
  - Everyone : $\lambda f.\forall x.f(x)$
  - Mostly works, but some problems
    - Have to change our NP/VP rule.
    - Won’t work for “Amy likes everyone.”
  - “Everyone likes someone.”
  - This gets tricky quickly!



Indefinites

- First try
  - "Bob ate a waffle" : ate(bob,waffle)
  - "Amy ate a waffle" : ate(amy,waffle)

- Can't be right!
  - $\exists$ x : waffle(x) $\wedge$ ate(bob,x)
  - What does the translation of "a" have to be?
  - What about "the"?
  - What about "every"?



Grounding

- Grounding
  - So why does the translation likes : λx.λy.likes(y,x) have anything to do with actual liking?
  - It doesn't (unless the denotation model says so)
  - Sometimes that's enough: wire up bought to the appropriate entry in a database

- Meaning postulates
  - Insist, e.g $\forall$x,y.likes(y,x) $\rightarrow$ knows(y,x)
  - This gets into lexical semantics issues

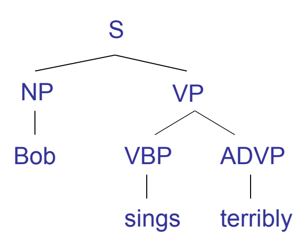
- Statistical version?

Tense and Events

- In general, you don't get far with verbs as predicates
- Better to have event variables e
  - "Alice danced" : danced(alice)
  - ∃ e : dance(e) ∧ agent(e,alice) ∧ (time(e) < now)
- Event variables let you talk about non-trivial tense / aspect structures
  - "Alice had been dancing when Bob sneezed"
  - ∃ e, e' : dance(e) ∧ agent(e,alice) ∧
    sneeze(e') ∧ agent(e',bob) ∧
    (start(e) < start(e') ∧ end(e) = end(e')) ∧
    (time(e') < now)

Adverbs

- What about adverbs?
  - "Bob sings terribly"
  - terribly(sings(bob))?
  - (terribly(sings))(bob)?
  - ∃e present(e) ∧
    type(e, singing) ∧
    agent(e,bob) ∧
    manner(e, terrible) ?
  - It's really not this simple..



Propositional Attitudes

- “Bob thinks that I am a gummi bear”
  - thinks(bob, gummi(me)) ?
  - thinks(bob, “I am a gummi bear”) ?
  - thinks(bob, ^gummi(me)) ?

- Usual solution involves intensions (^X) which are, roughly, the set of possible worlds (or conditions) in which X is true

- Hard to deal with computationally
  - Modeling other agents models, etc
  - Can come up in simple dialog scenarios, e.g., if you want to talk about what your bill claims you bought vs. what you actually bought

Trickier Stuff

- Non-Intersective Adjectives
  - green ball : $\lambda x.[green(x) \wedge ball(x)]$
  - fake diamond : $\lambda x.[fake(x) \wedge diamond(x)]$? $\longrightarrow$ $\lambda x.[fake(diamond(x))$
- Generalized Quantifiers
  - the : $\lambda f.[\textit{unique-member}(f)]$
  - all : $\lambda f. \lambda g\ [\forall x.f(x) \rightarrow g(x)]$
  - most?
  - Could do with more general second order predicates, too (why worse?)
    - the(cat, meows), all(cat, meows)
- Generics
  - “Cats like naps”
  - “The players scored a goal”
- Pronouns (and bound anaphora)
  - “If you have a dime, put <u>it</u> in the meter.”

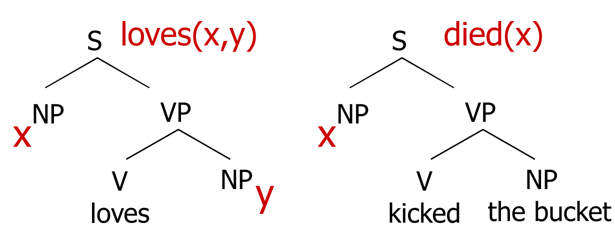
- … the list goes on and on!

Multiple Quantifiers

- Quantifier scope
  - Groucho Marx celebrates quantifier order ambiguity:
    "In this country a woman gives birth every 15 min.
    Our job is to find that woman and stop her."
- Deciding between readings
  - "Bob bought a pumpkin every Halloween"
  - "Bob put a warning in every window"
  - Multiple ways to work this out
    - Make it syntactic (movement)
    - Make it lexical (type-shifting)

Implementation, TAG, Idioms

- Add a "sem" feature to each context-free rule
  - S → NP loves NP
  - S[sem=loves(x,y)] → NP[sem=x] loves NP[sem=y]
  - Meaning of S depends on meaning of NPs
- TAG version:



- Template filling: S[sem=showflights(x,y)] →
  I want a flight from NP[sem=x] to NP[sem=y]

Modeling Uncertainty

- Gaping hole warning!
- Big difference between statistical disambiguation and statistical reasoning.

The scout saw the enemy soldiers with night goggles.

  - With probabilistic parsers, can say things like “72% belief that the PP attaches to the NP.”
  - That means that *probably* the enemy has night vision goggles.
  - However, you can’t throw a logical assertion into a theorem prover with 72% confidence.
  - Not clear humans really extract and process logical statements symbolically anyway.
  - Use this to decide the expected utility of calling reinforcements?

- In short, we need probabilistic reasoning, not just probabilistic disambiguation followed by symbolic reasoning!

CCG Parsing

- Combinatory Categorial Grammar
  - Fully (mono-) lexicalized grammar
  - Categories encode argument sequences
  - Very closely related to the lambda calculus
  - Can have spurious ambiguities (why?)

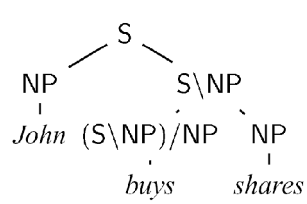
$John \vdash \mathsf{NP} : john'$

$shares \vdash \mathsf{NP} : shares'$

$buys \vdash (\mathsf{S}\backslash\mathsf{NP})/\mathsf{NP} : \lambda x.\lambda y.buys'xy$

$sleeps \vdash \mathsf{S}\backslash\mathsf{NP} : \lambda x.sleeps'x$

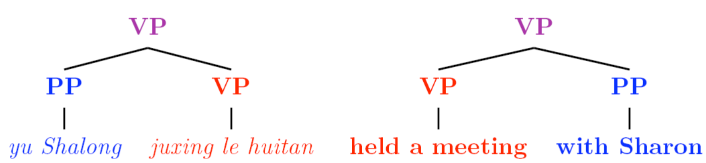
$well \vdash (\mathsf{S}\backslash\mathsf{NP})\backslash(\mathsf{S}\backslash\mathsf{NP}) : \lambda f.\lambda x.well'(fx)$



Syntax-Based MT

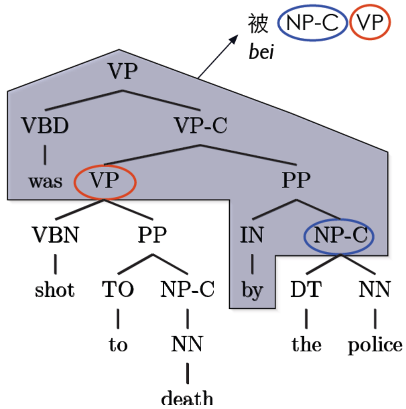
- synchronous context-free grammars (SCFGs)
  - context-free grammar in two dimensions
  - generating pairs of strings/trees simultaneously
  - co-indexed nonterminal further rewritten as a unit

$$
\begin{array}{lcll}
\mathbf{VP} & \rightarrow & \mathbf{PP}^{(1)}\ \mathbf{VP}^{(2)}, & \mathbf{VP}^{(2)}\ \mathbf{PP}^{(1)} \\
\mathbf{VP} & \rightarrow & \textit{juxing le huitan}, & \textbf{held a meeting} \\
\mathbf{PP} & \rightarrow & \textit{yu Shalong}, & \textbf{with Sharon}
\end{array}
$$



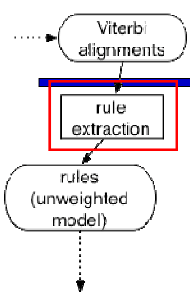
Learning MT Grammars

- syntax-directed, English to Chinese (Huang, Knight, Joshi, 2006)
- first parse input, and then recursively transfer



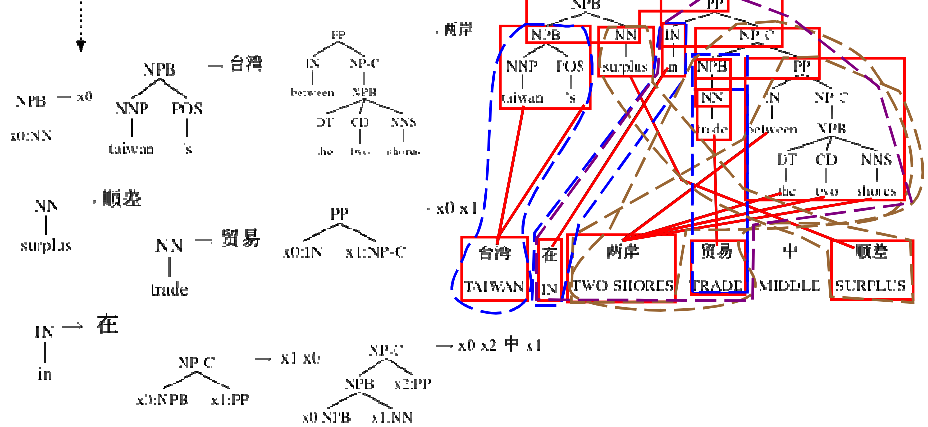
synchronous tree-substitution grammars (STSG)

(Galley et al., 2004; Eisner, 2003)

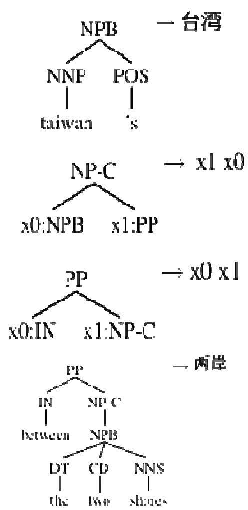


Extracting syntactic rules

Extract rules (Galley et. al. '04, '06)

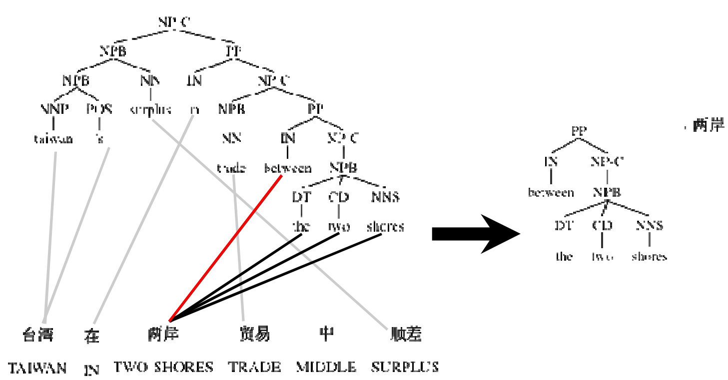


Rules can...



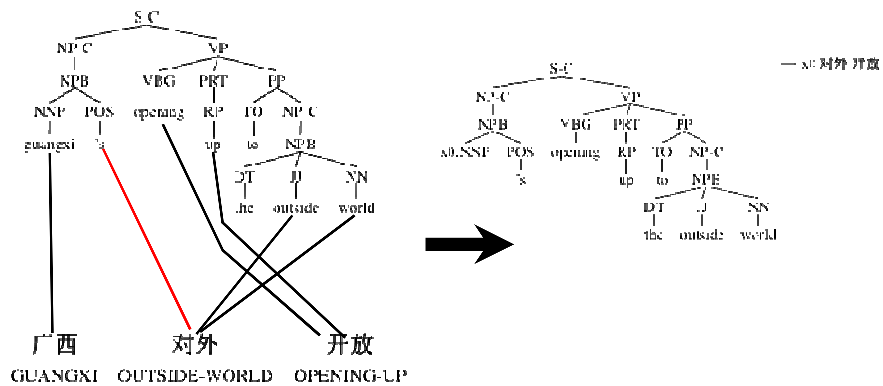
- capture phrasal translation
- reorder parts of the tree
- traverse the tree without reordering
- insert (and delete) words

Bad alignments make bad rules



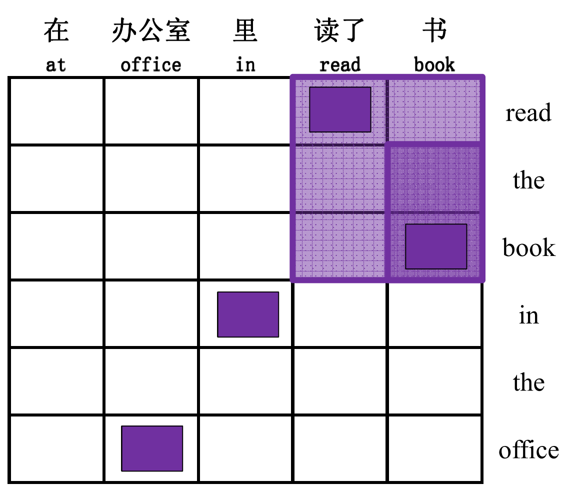
This isn't very good, but let's look at a worse example...

Sometimes they're really bad

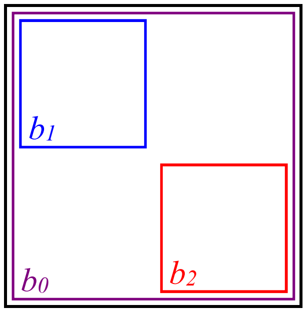


One bad link makes a totally unusable rule!

Alignment: Words, Blocks, Phrases



Discriminative Block ITG



Features

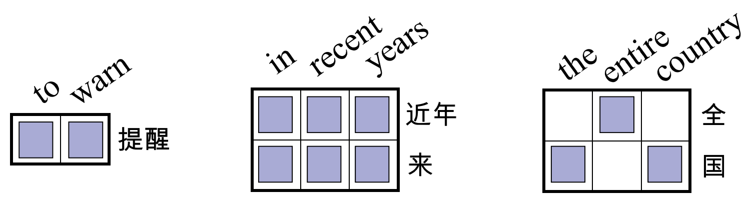
$\varphi(b_0, s, s')$

$\varphi(b_1, s, s')$

$\varphi(b_2, s, s')$

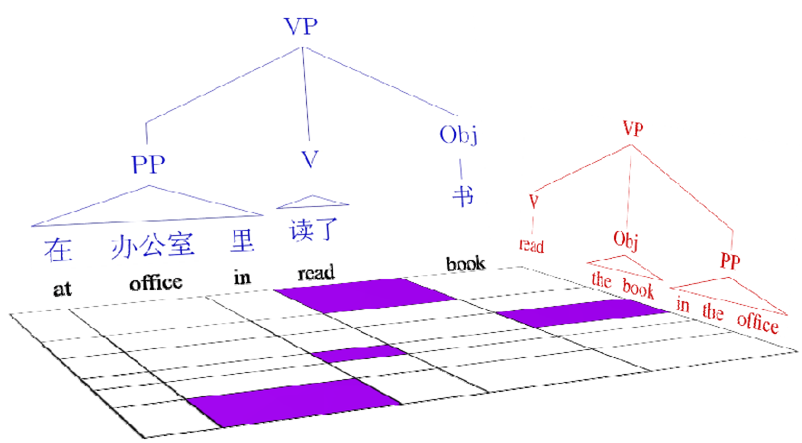
$$\phi(\mathcal{A}) = \sum_{b \in \mathcal{A}} \phi(b, s, s')$$

$$P(\mathcal{A}) \propto \exp\langle \theta, \phi(\mathcal{A}) \rangle$$



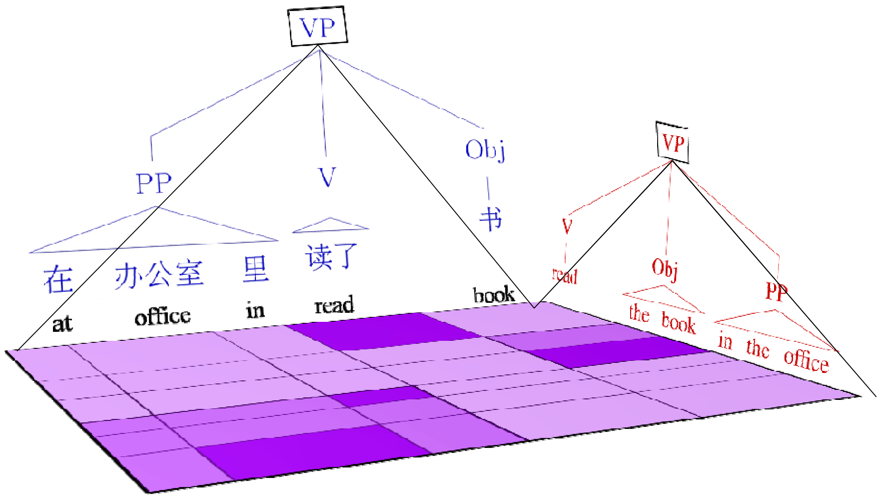
[Haghighi, Blitzer, Denero,and Klein, ACL 09]

Syntactic Correspondence

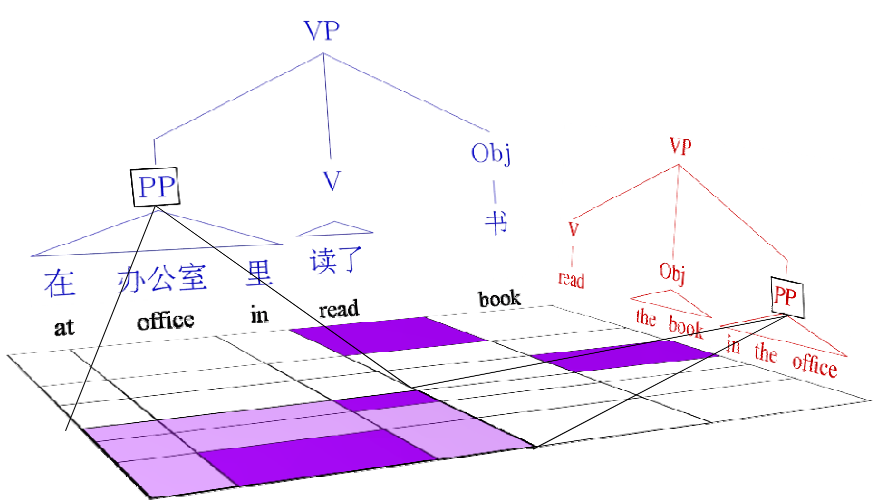


Build a model p_θ (, , 中文, EN)

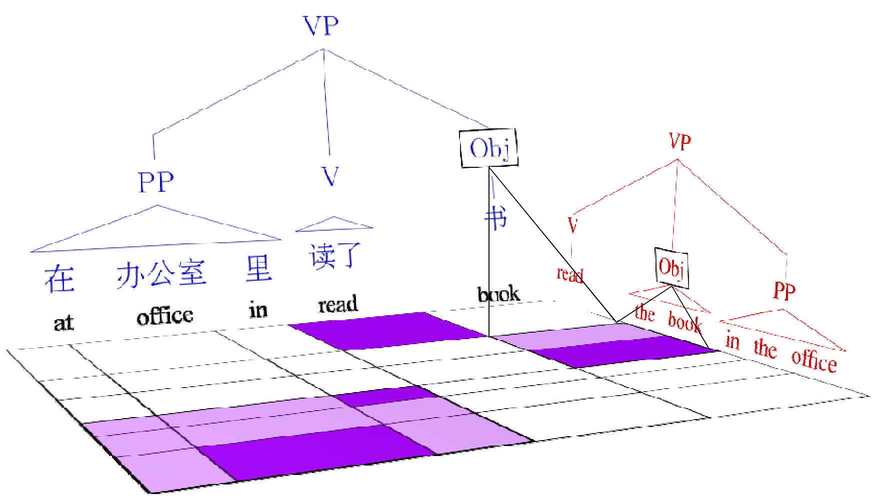
Synchronous Grammars?



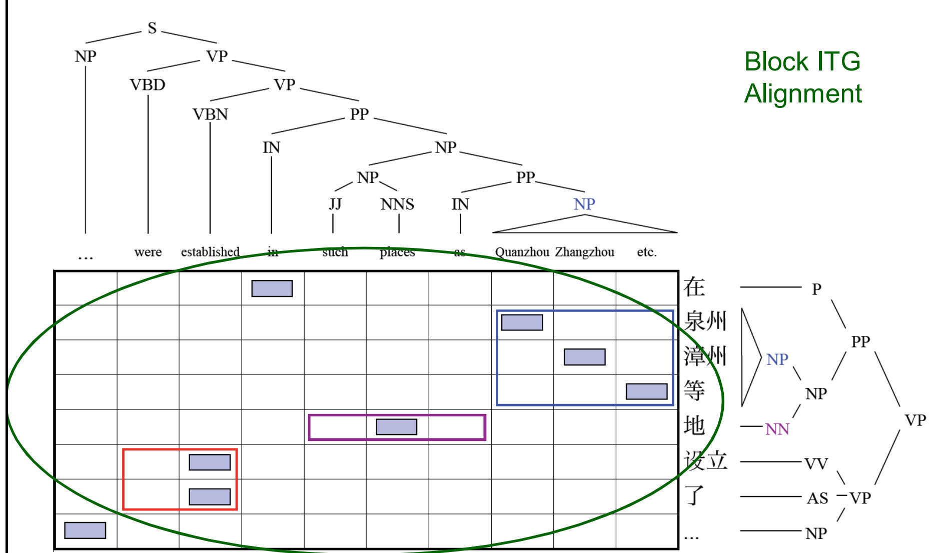
Synchronous Grammars?



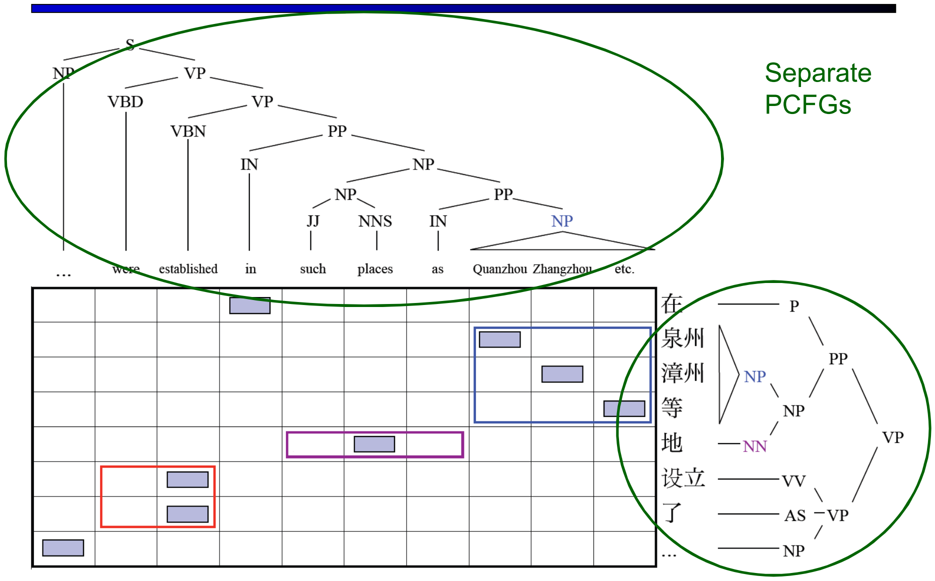
Synchronous Grammars?



Adding Syntax: Weak Synchronization

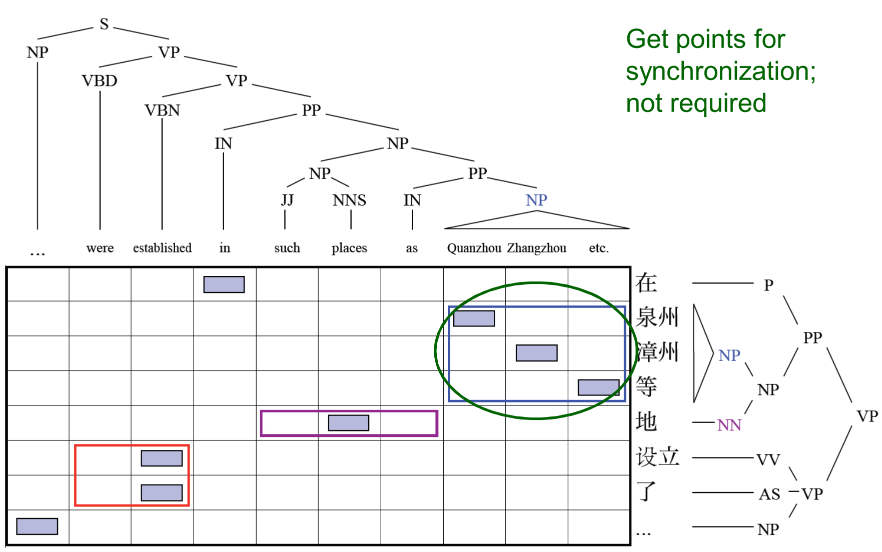


Adding Syntax: Weak Synchronization



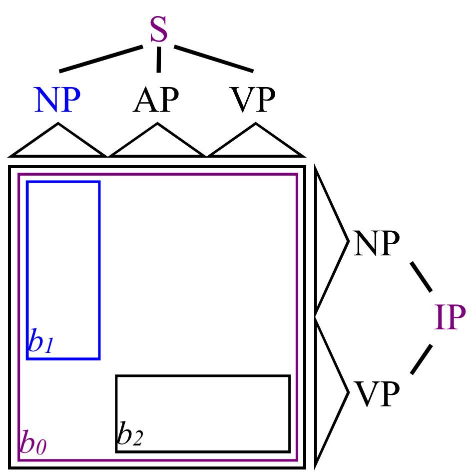
Separate PCFGs

Adding Syntax: Weak Synchronization



Get points for synchronization; not required

Weakly Synchronous Features



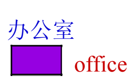
Parsing	Alignment
$\phi_{\mathcal{F}}(\mathrm{IP}, s)$	$\phi_{\mathcal{A}}(b_0, s, s')$
$\phi_{\mathcal{F}}(\mathrm{NP}, s)$	$\phi_{\mathcal{A}}(b_1, s, s')$
$\phi_{\mathcal{F}}(\mathrm{VP}, s)$	$\phi_{\mathcal{A}}(b_2, s, s')$
	Agreement
$\phi_{\mathcal{E}}(\mathrm{S}, s')$	$\phi_{\triangleright}(\mathrm{IP}, b_0)$
$\phi_{\mathcal{E}}(\mathrm{NP}, s')$	$\phi_{\triangleleft}(b_0, \mathrm{S})$
$\phi_{\mathcal{E}}(\mathrm{AP}, s')$	$\phi_{\triangleleft}(b_1, \mathrm{NP})$
$\phi_{\mathcal{E}}(\mathrm{VP}, s')$	$\phi_{\bowtie}(\mathrm{IP}, b_0, \mathrm{S})$

Weakly Synchronous Model

p_θ (, , | EN,中文)

Feature Type 1: Word Alignment

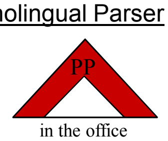
$\phi($, EN , 中文 $)$



[HBDK09]

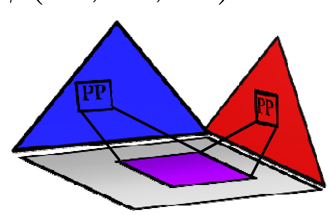
Feature Type 2: Monolingual Parser

$\phi($, EN $)$



Feature Type 3: Agreement

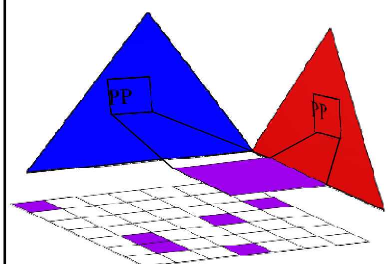
$\phi($, , $)$



Inference: Structured Mean Field

- Problem: Summing over weakly aligned hypotheses is intractable
- Factored approximation: $p_\theta(\blacktriangle, \blacktriangle, \boxplus \mid \text{EN}, \text{中文}) \approx q(\blacktriangle)q(\blacktriangle)q(\boxplus)$
- Set q to minimize $KL\left(\, q(\blacktriangle)q(\blacktriangle)q(\boxplus),\;\; p_\theta(\blacktriangle, \blacktriangle, \boxplus \mid \text{EN}, \text{中文})\,\right)$

<u>Algorithm</u>

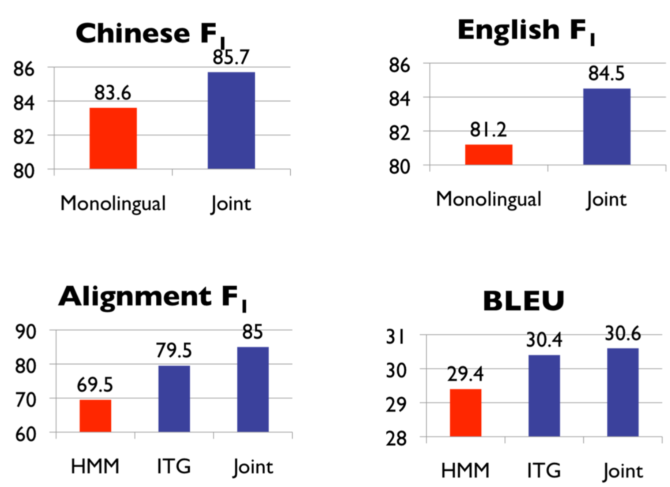


1) Initialize: $q(\blacktriangle)\; q(\blacktriangle)\; q(\boxplus)$

2) Iterate:

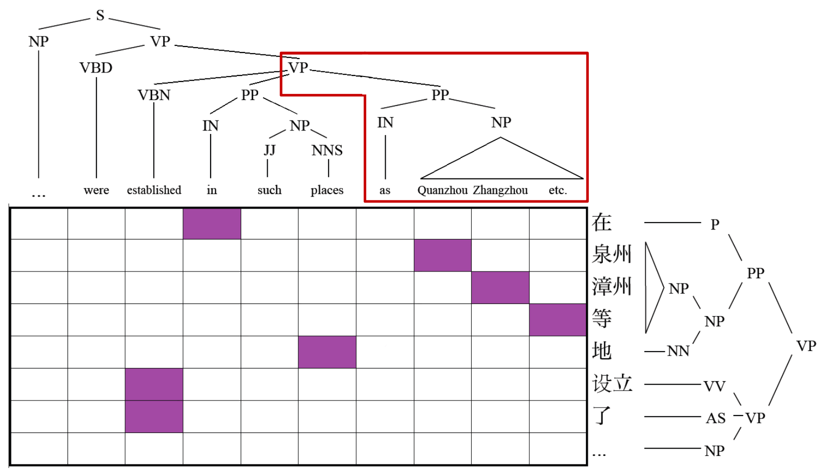
$$q(\blacktriangle) \propto \exp\{\langle \theta, \phi(\blacktriangle, E_q(\blacktriangle), E_q(\boxplus))\rangle\}$$
$$q(\blacktriangle) \propto \exp\{\langle \theta, \phi(E_q(\blacktriangle), \blacktriangle, E_q(\boxplus))\rangle\}$$
$$q(\boxplus) \propto \exp\{\langle \theta, \phi(E_q(\blacktriangle), E_q(\blacktriangle), \boxplus)\rangle\}$$

Results

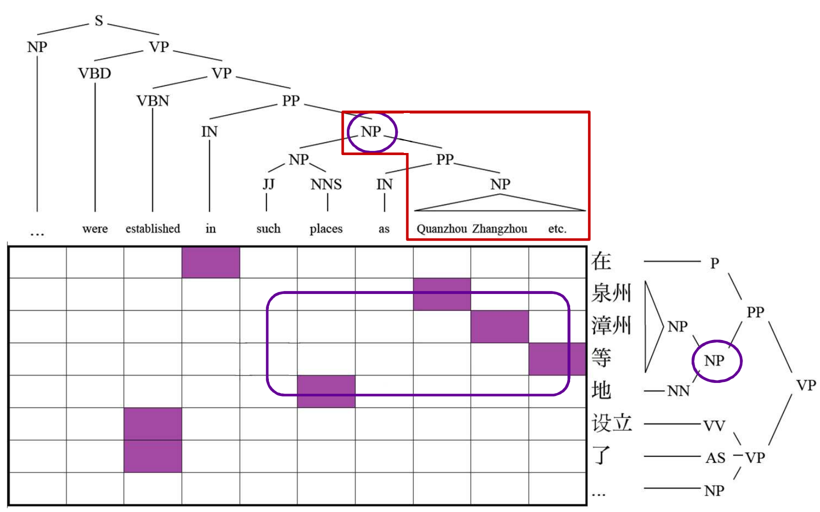


[Burkett, Blitzer, and Klein, NAACL 10]

Incorrect English PP Attachment



Corrected English PP Attachment



Improved Translations

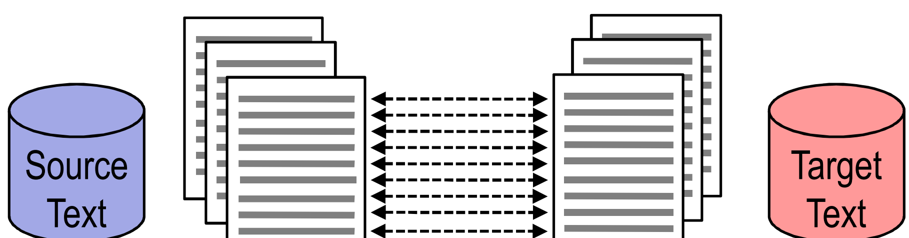
目前	导致	飞机	相撞	的	原因	尚	不	清楚,	当地	民航	部门	将	对此	展开	调查
Cur- rently	cause	plane	crash	DE	reason	yet	not	clear,	local	civil aero- nautics	bureau	will	toward	open	investi- gations

Reference
At this point the cause of the plane collision is still unclear. The local caa will launch an investigation into this .

Baseline (GIZA++)
The cause of planes is still not clear yet, local civil aviation department will investigate this .

Bilingual Adaptation Model
The cause of plane collision remained unclear, local civil aviation departments will launch an investigation .

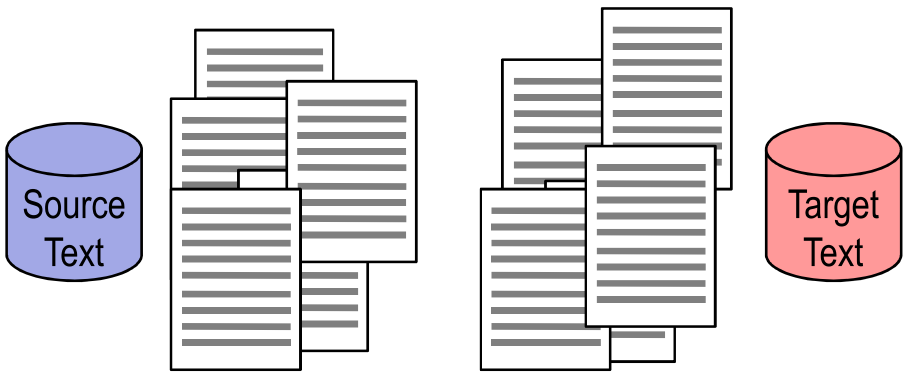
Machine Translation Approach



nous acceptons votre opinion .

we accept your view .

Translations from Monotexts

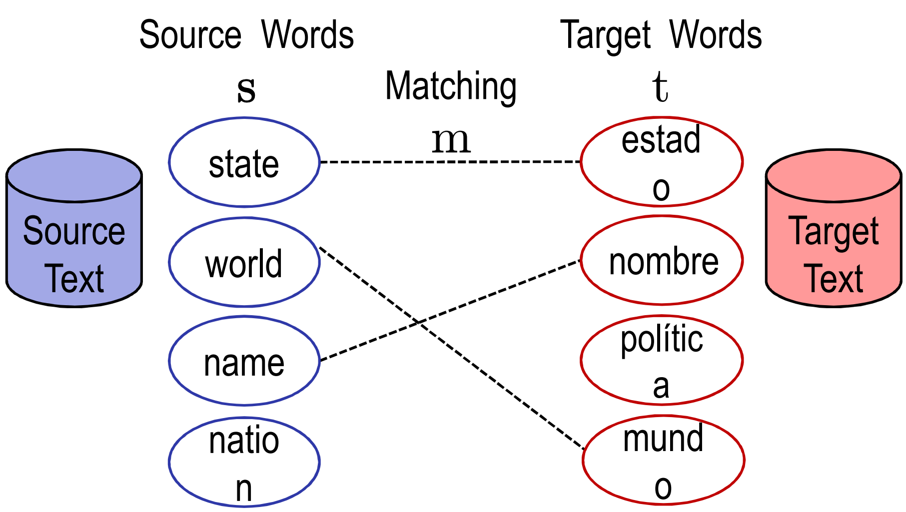


- Translation without parallel text?

[Fung 95, Koehn and Knight 02, Haghighi and Klein 08]

Task: Lexicon Matching

[Haghighi and Klein 08]



Data Representation

state

Source Text

Orthographic Features

#st	1.0
tat	1.0
te#	1.0

Context Features

world	20.0
politics	5.0
society	10.0

Data Representation

state

Source Text

Orthographic Features

#st	1.0
tat	1.0
te#	1.0

Context Features

world	20.0
politics	5.0
society	10.0

estado

Target Text

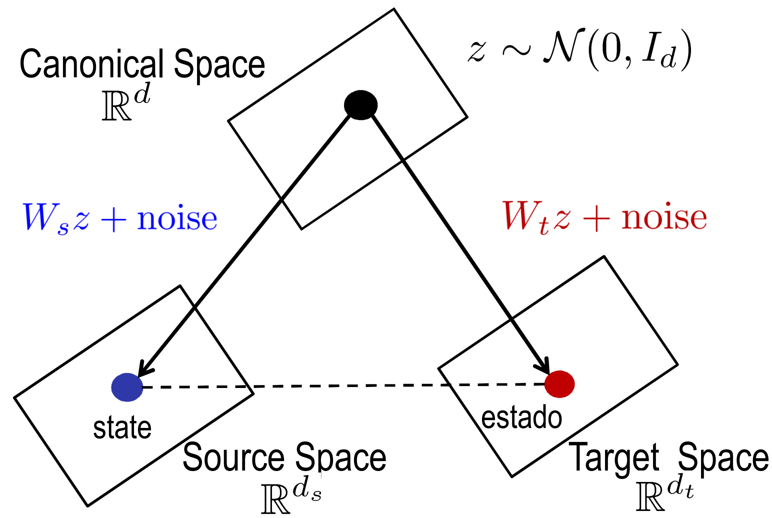
Orthographic Features

#es	1.0
sta	1.0
do#	1.0

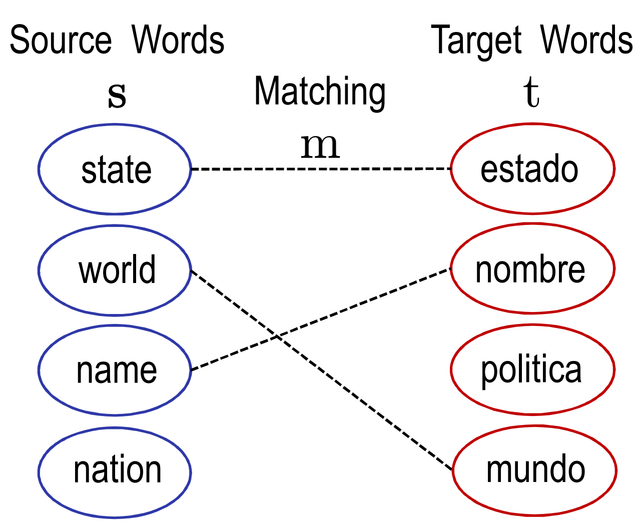
Context Features

mundo	17.0
politica	10.0
sociedad	6.0

Generative Model (CCA)



Generative Model (Matching)



Inference: Hard EM

E-Step: Find best matching

$$w_{ij} = \log p(s_i, t_j | \mathbf{m}; W_s, W_t) - \log \mathrm{NULL}_S(s_i) - \log \mathrm{NULL}_T(t_j)$$

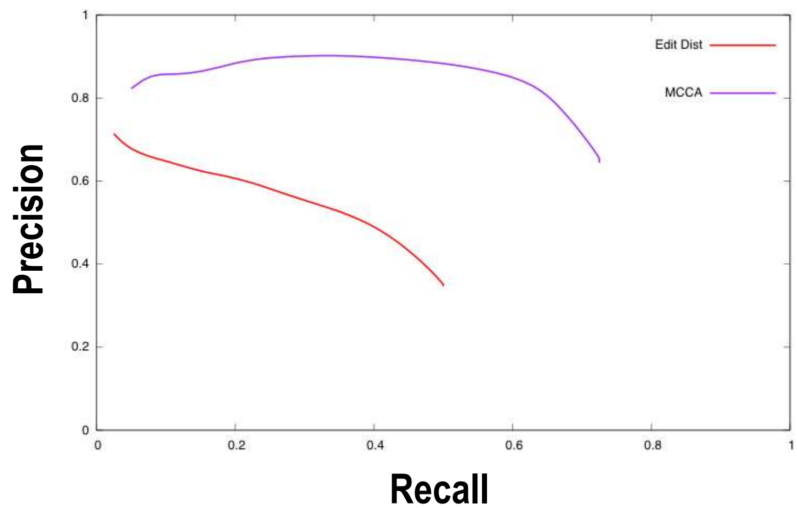
M-Step: Solve a CCA problem

$$\max_{(W_s, W_t)} \left[\sum_{(i,j) \in \mathbf{m}} \log p(s_i, t_j | \mathbf{m}; W_s, W_t) \right]$$

Experimental Setup

- Data: 2K most frequent nouns, texts from Wikipedia
- Seed: 100 translation pairs
- Evaluation: Precision and Recall against lexicon obtained from Wiktionary
  - Report $p_{0.33}$, precision at recall 0.33

Lexicon Quality (EN-ES)



Analysis

English-Spanish

Source	Target	Correct
education	educación	Y
pacto	pact	Y
stability	estabilidad	Y
corruption	corrupción	Y
tourism	turismo	Y
organisation	organización	Y
convenience	conveniencia	Y
syria	siria	Y
cooperation	cooperación	Y
culture	cultura	Y
protocol	protocolo	Y
north	norte	Y
health	salud	Y
action	reacción	N

Analysis

Interesting Matches

health	salud
traceability	rastreabilidad
youth	juventud
report	informe
advantages	ventajas

Interesting Mistakes

liberal	partido
Kirkhope	Gorsel
action	reacción
Albanians	Bosnia
a.m.	horas

Language Variation

English-Chinese

Source	Target	Correct
prices	价格	Y
network	网络	Y
population	人口	Y
reporter	孙	N
oil	石油	Y