



Generation or transfer: this is still the question

Nianwen Xue

Brandeis University

(Joint work with Dun Deng)

2014-07-21

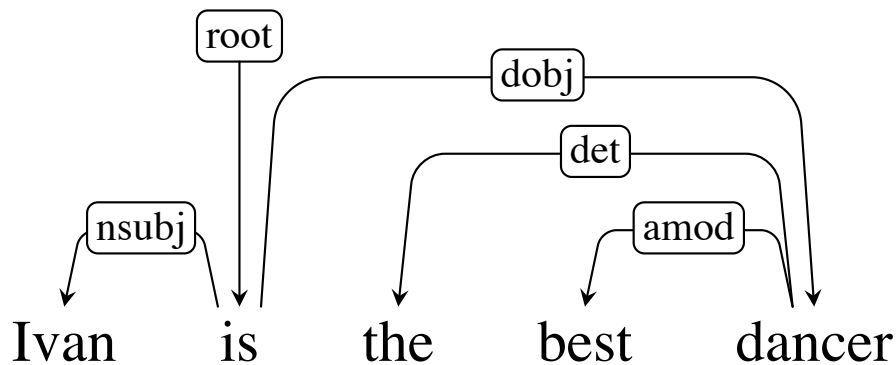
PIRE Workshop



Outline

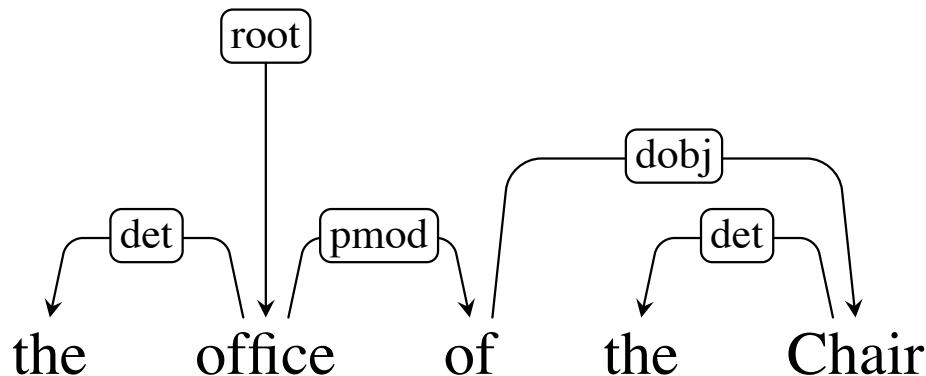
- The trend to go deeper: dependency tree to Abstract Meaning Representation (AMR)
- A shallower alternative for MT: a hierarchically aligned Chinese-English parallel treebank

Dependency representation revolves around the notion of “head”



Root is 'is'
'is' is the head of 'Ivan'
'is' is the head of 'dancer'
'dancer' is the head of 'best'
'dancer' is the head of 'the'

Dependency representation revolves around the notion of “head”



Root is “office”

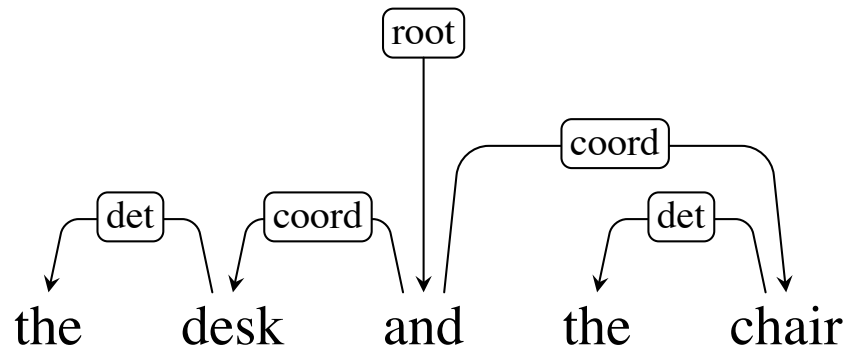
“office” is the head of “of”

“office” is the head of “the”

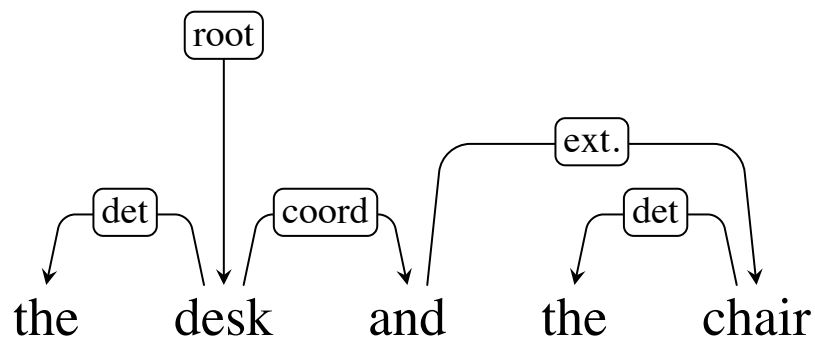
“of” is the head of “Chair”

“Chair” is the head of “the”

It's not always this clean

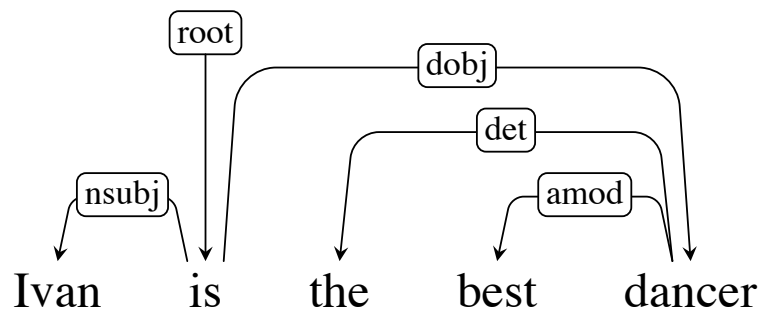


Prague dependency

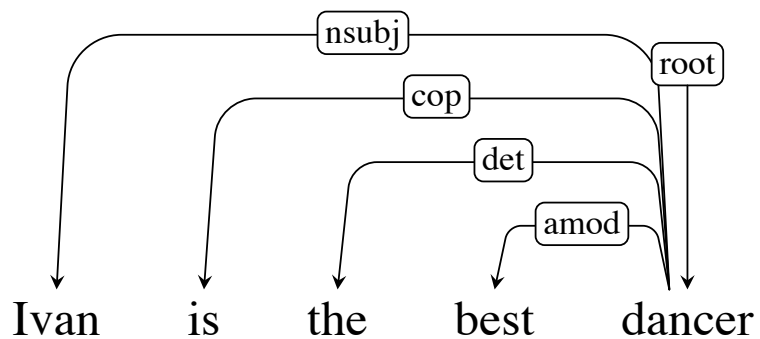


Moscow dependency

Universal Stanford dependency (USD)

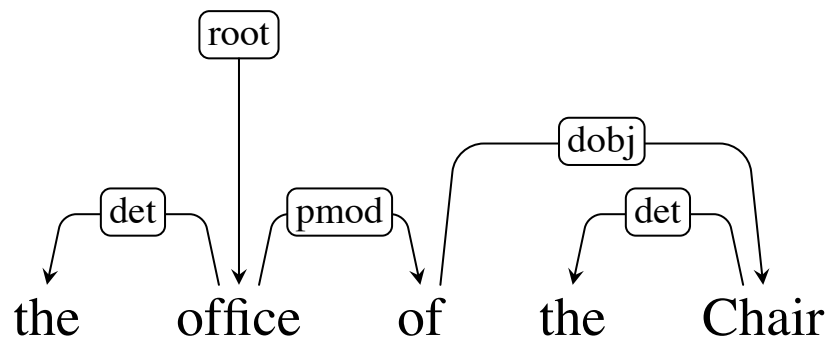


Relation between content words
mediated by function words

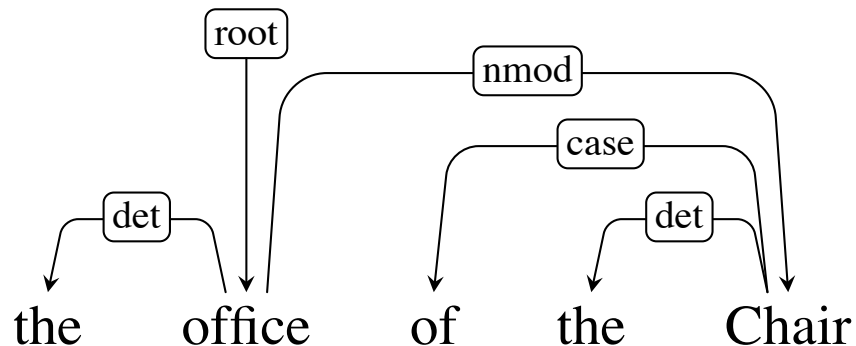


Direct links between content words

Universal Stanford dependency (USD)

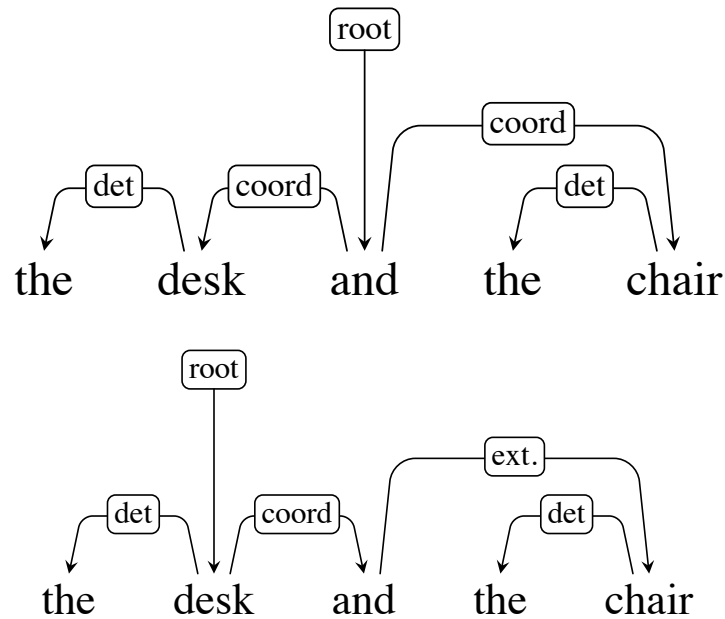


Relation between content words
mediated by function words

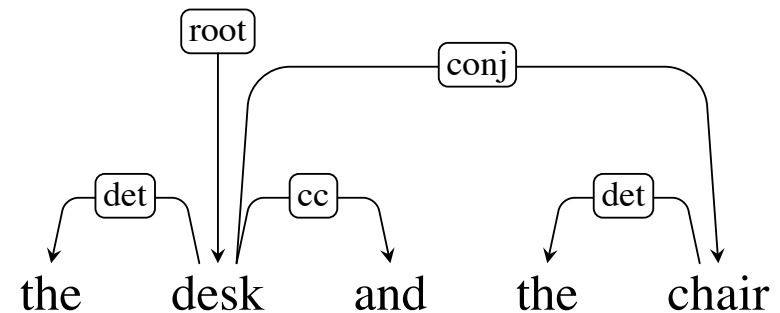


Direct links between content words

Universal Stanford dependency (USD)

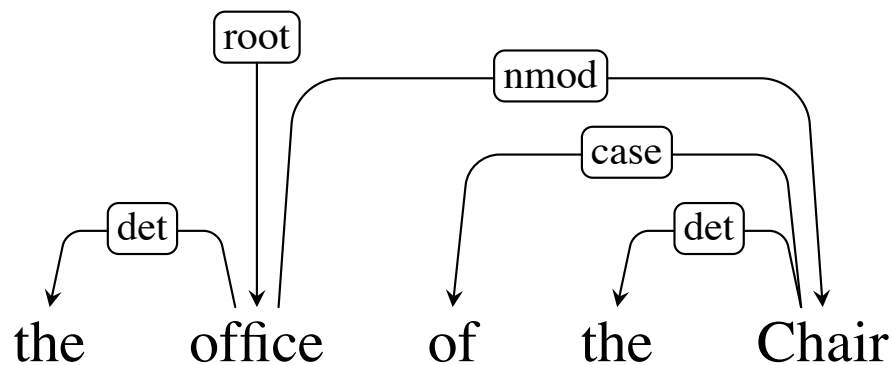


Relation between content words
mediated by function words

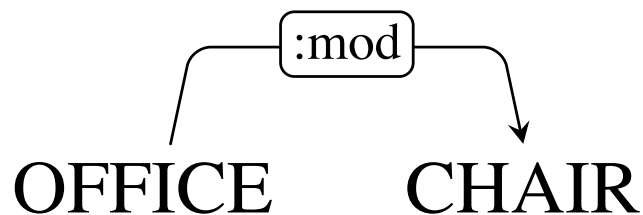


Direct links between content words

Abstract Meaning Representation (AMR)

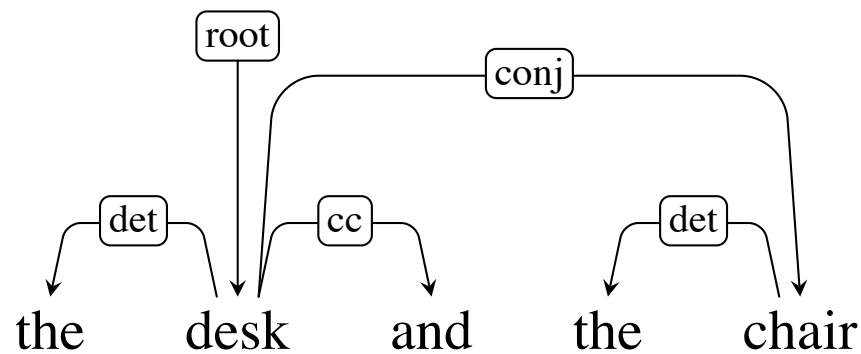


Function words kept

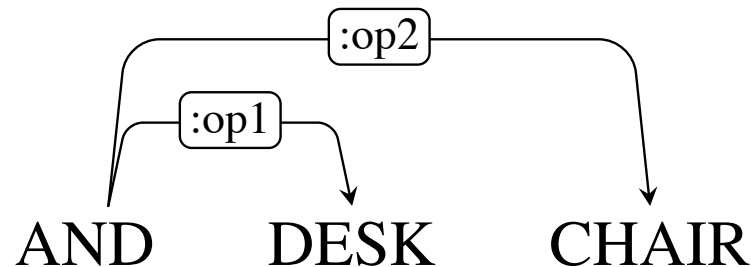


Function words thrown away
(or mapped to abstract concepts
or relations, or attributes)

Abstract Meaning Representation (AMR)

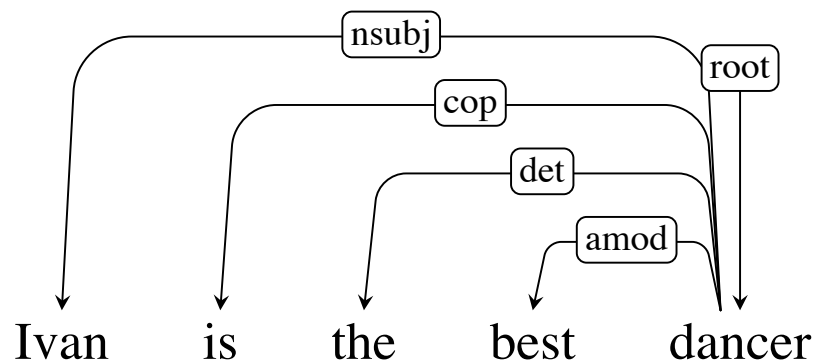


Function words kept



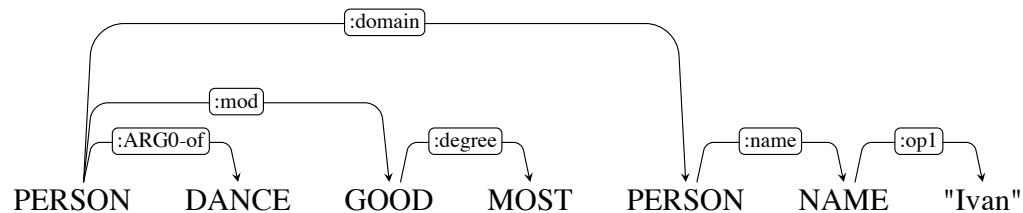
Function words thrown away
(or mapped to abstract concepts,
attributes, or relations)

Abstract Meaning Representation (AMR)



Function words kept

Lexical Integrity respected



Function words thrown away
(or mapped to abstract concepts,
attributes, or relations)

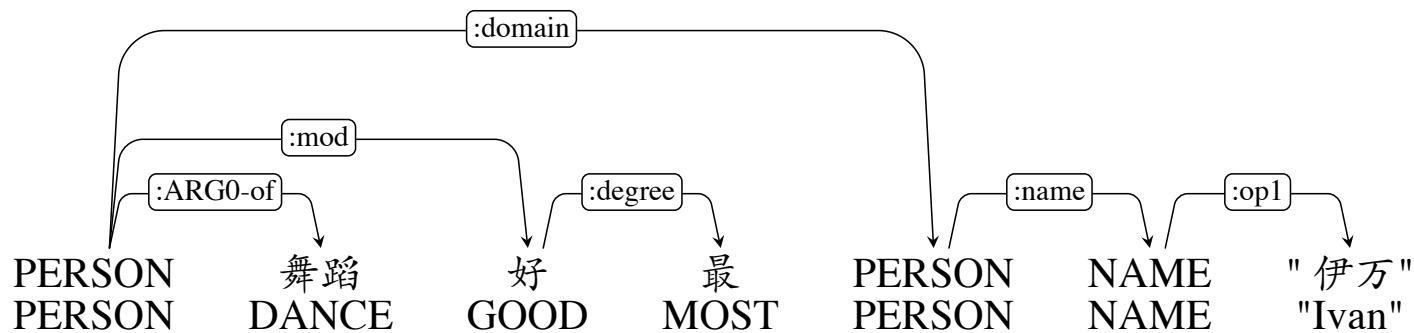
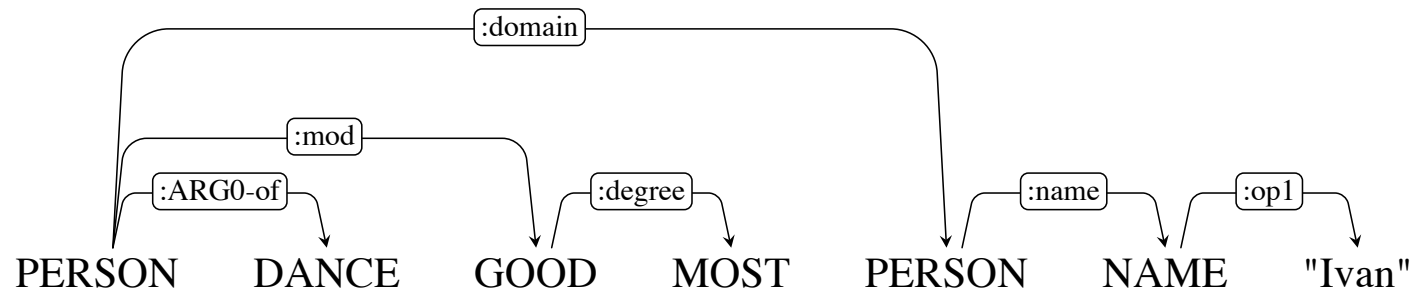
Lexical integrity violated



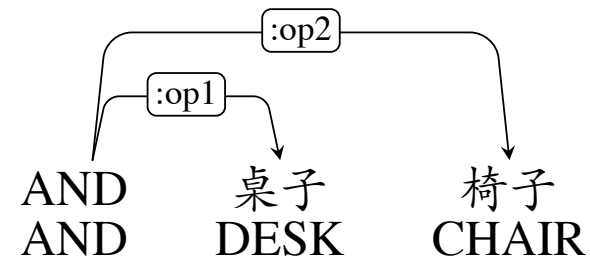
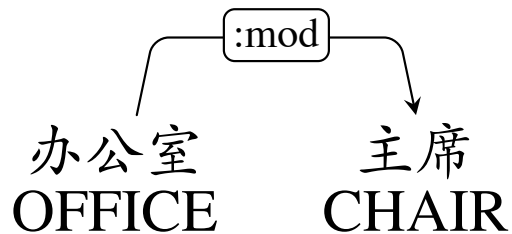
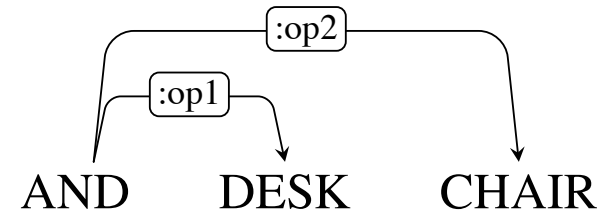
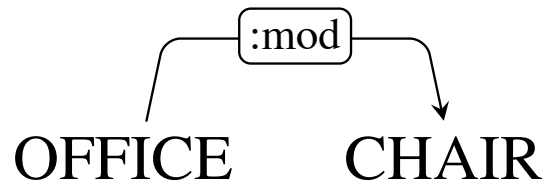
The trend

- From “standard” dependency to Stanford dependency to AMR, the community is moving towards increasingly “deeper”, more abstract representations

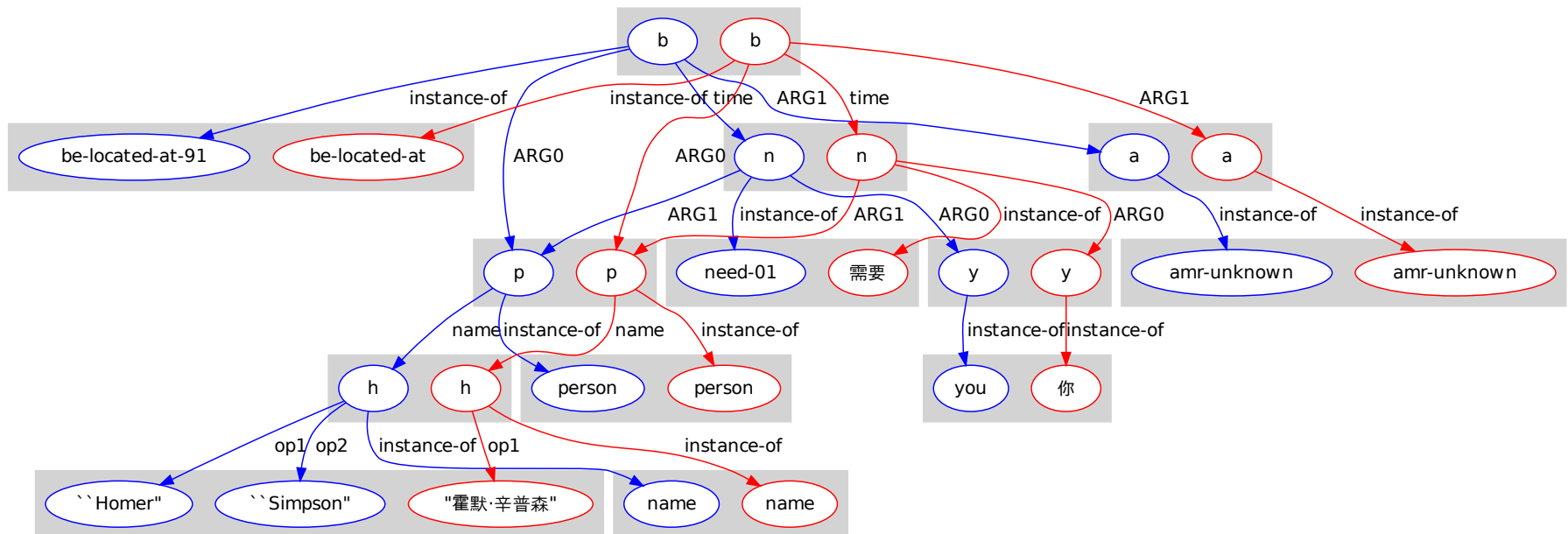
The case for AMR as a representation for MT



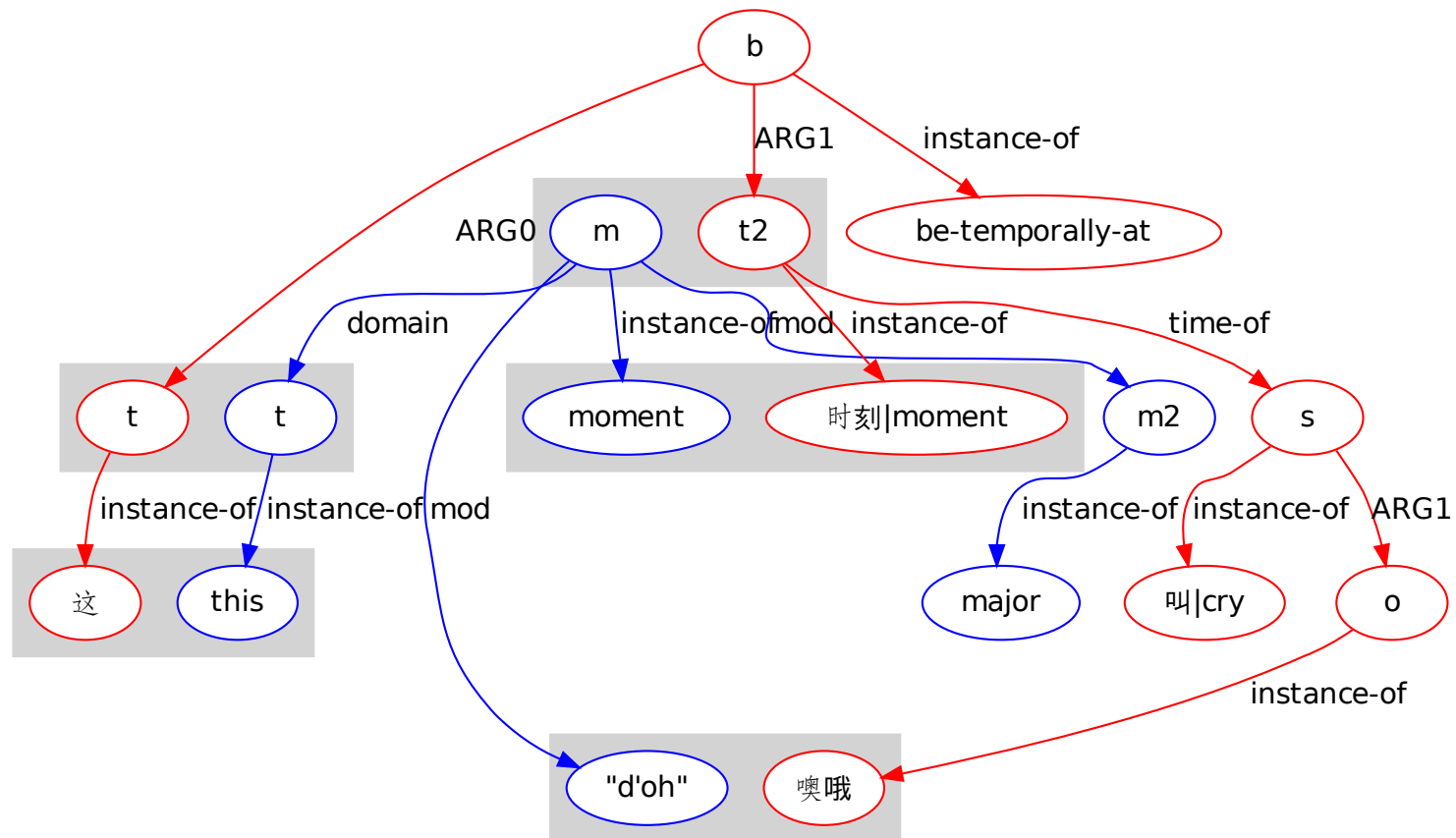
The case for AMR as a representation for MT



当你需要他时, 霍默辛普森在哪里?
Where is Homer Simpson when you need him?



这是一个大叫“噢哦！”的时刻。
This is a major ``D'oh!'' moment.





The representation calculus

- Deeper, more abstract representations lead to better alignment
- But they come at a cost for MT purposes:
 - Deeper analyzers tend to be less accurate
 - For MT, AMR necessitates a generation step: you have to get back the function words you've thrown away or mapped to abstract concepts, attributes, or relations
 - Source string \rightarrow (source AMR) \rightarrow target AMR \rightarrow target string
- Less of a problem for applications such as IE, QA, etc. where the generation step is less involved

Hierarchical alignment between constituent parse trees

Let's not throw away the function words.
Just don't align them (at the word level).

Assume neural networks do not solve all the problems
in MT and representation is still relevant

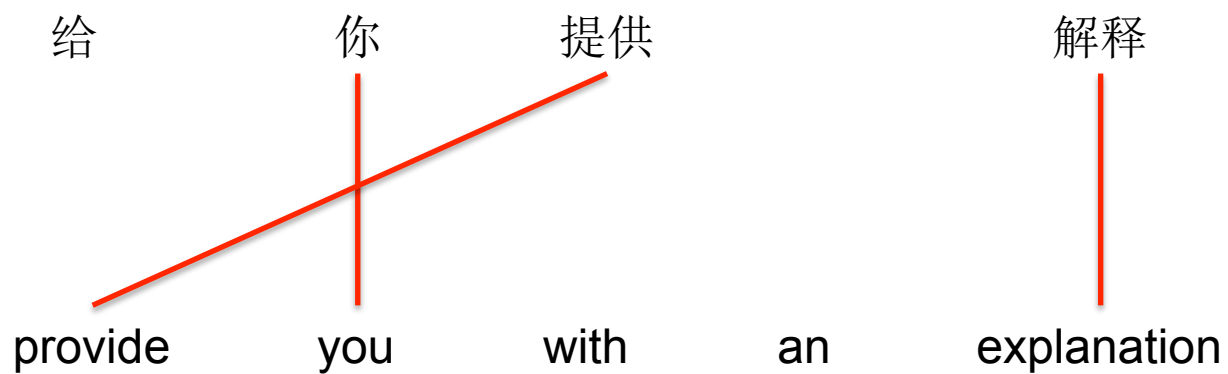
Dun Deng and Nianwen Xue. 2014 (To appear). Building a Hierarchically Aligned Chinese-English Parallel Treebank. In Proceedings of COLING. Dublin, Ireland

Dun Deng and Nianwen Xue. 2014 (To appear). Aligning Chinese English Parallel Parse Trees: Is it Feasible? In Proceedings of Linguistic Annotation Workshop VIII. Dublin, Ireland.





Word alignment



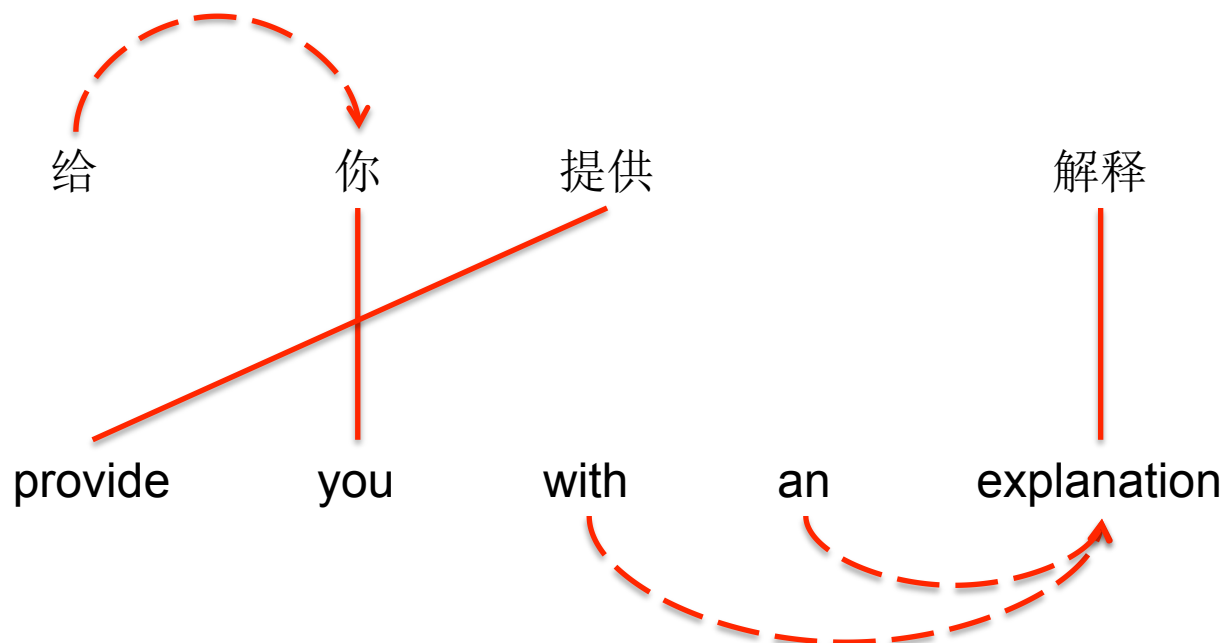


Existing word alignment practice

- Function words are attached to a “head”
 - Determiners (e.g., “the”, “a”) and Chinese classifiers (e.g., 个) are attached to a noun before they are aligned
 - Auxiliary verbs, tense, aspect markers are attached to their main verbs
 - Prepositions are attached to their NP object
 - Complementizers attached to matrix verbs,
 - Relative pronouns attached to head of relative clauses
- Function words (or punctuation) that anchor larger patterns are aligned at the word level as well
 - Chinese comma aligned with “and”
 - Chinese sentence-final markers aligned with punctuation marks
 - Chinese 的, a modifier marker, attached to and aligned with various things
- Pronouns are linked to their coreferent before they are aligned

Linguistic problems

- Cram too much into word-level alignment

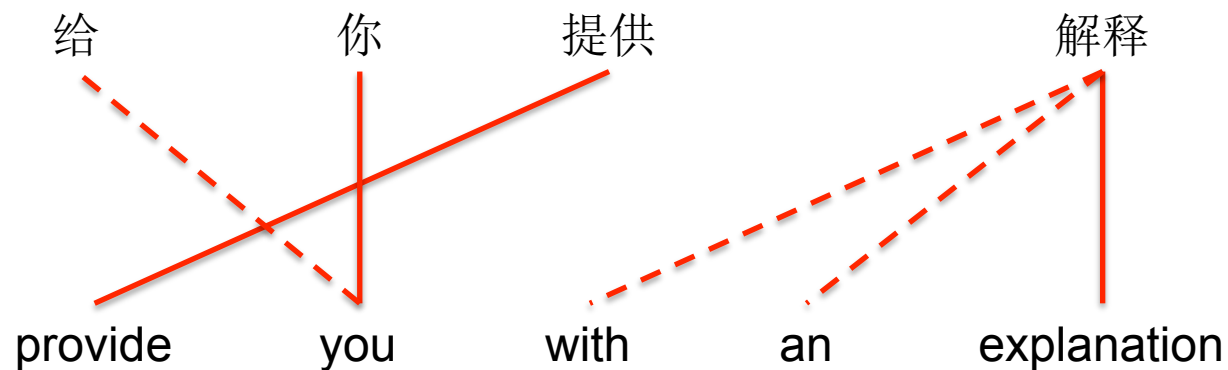




Computational problems

- Creates pervasive *one-to-many* or *many-to-many* alignments, where the “many” are often non-consecutive, making the automatic alignment difficult.
- Severs crucial dependencies and creates many haphazard alignments, increases ambiguity

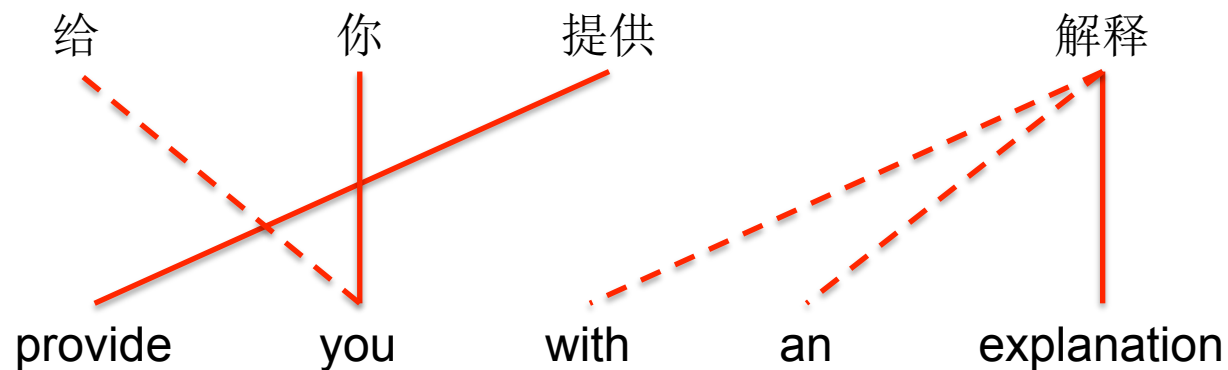
Attaching function words to their “heads”



你 -- you
给你 -- you

解释 -- explanation
解释 -- an explanation
解释 -- with an explanation

Attaching function words to their “heads”



你 -- you
给你 -- you

provide sb **with** sth

解释 -- explanation
解释 -- an explanation
解释 -- with an explanation

Spurious ambiguity magnified

Eat apples <> 吃 苹果

Eat an apple <> 吃 苹果

Eat the apple <> 吃 苹果

Fond of apples <> 喜欢 苹果

Talk about apples <> 谈论 苹果

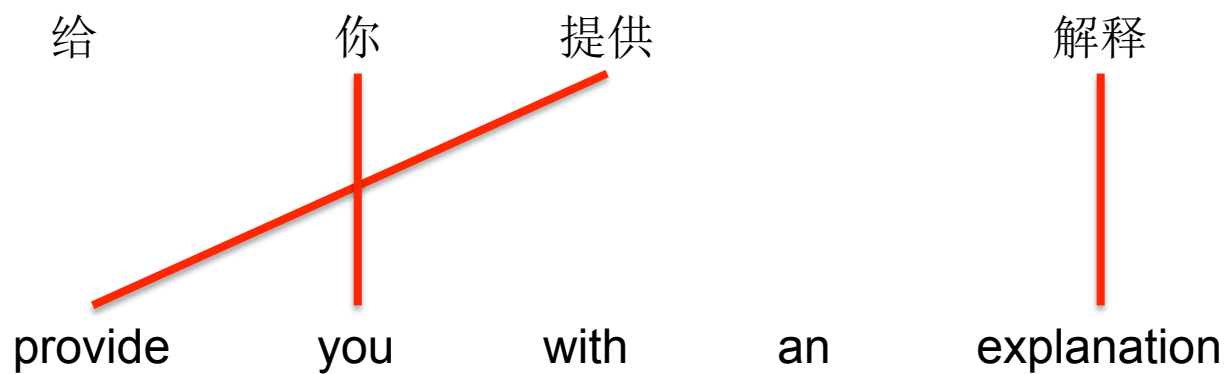
Provide them with apples <> 给 他们 提供 苹果



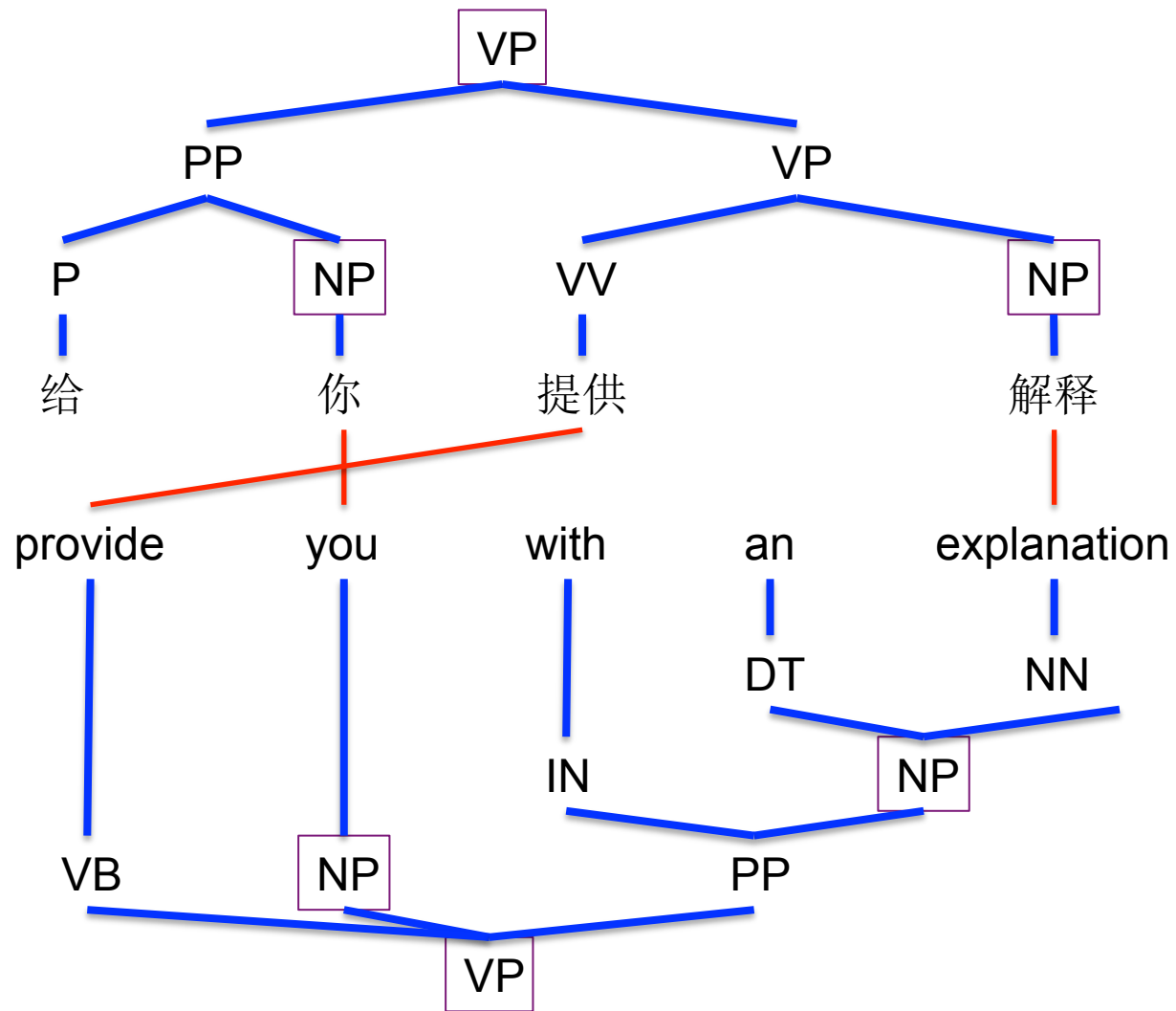
Hierarchical alignment

- Aligning function words in their syntactic context
 - Leaving function words (“glue words”) that signal syntactic patterns (的, “so that”) or semantic attributes (“a”, “the”) unaligned at the word level, and migrating their alignment to the phrase level
- What is the proper syntactic context?

Word alignment



Hierarchical alignment

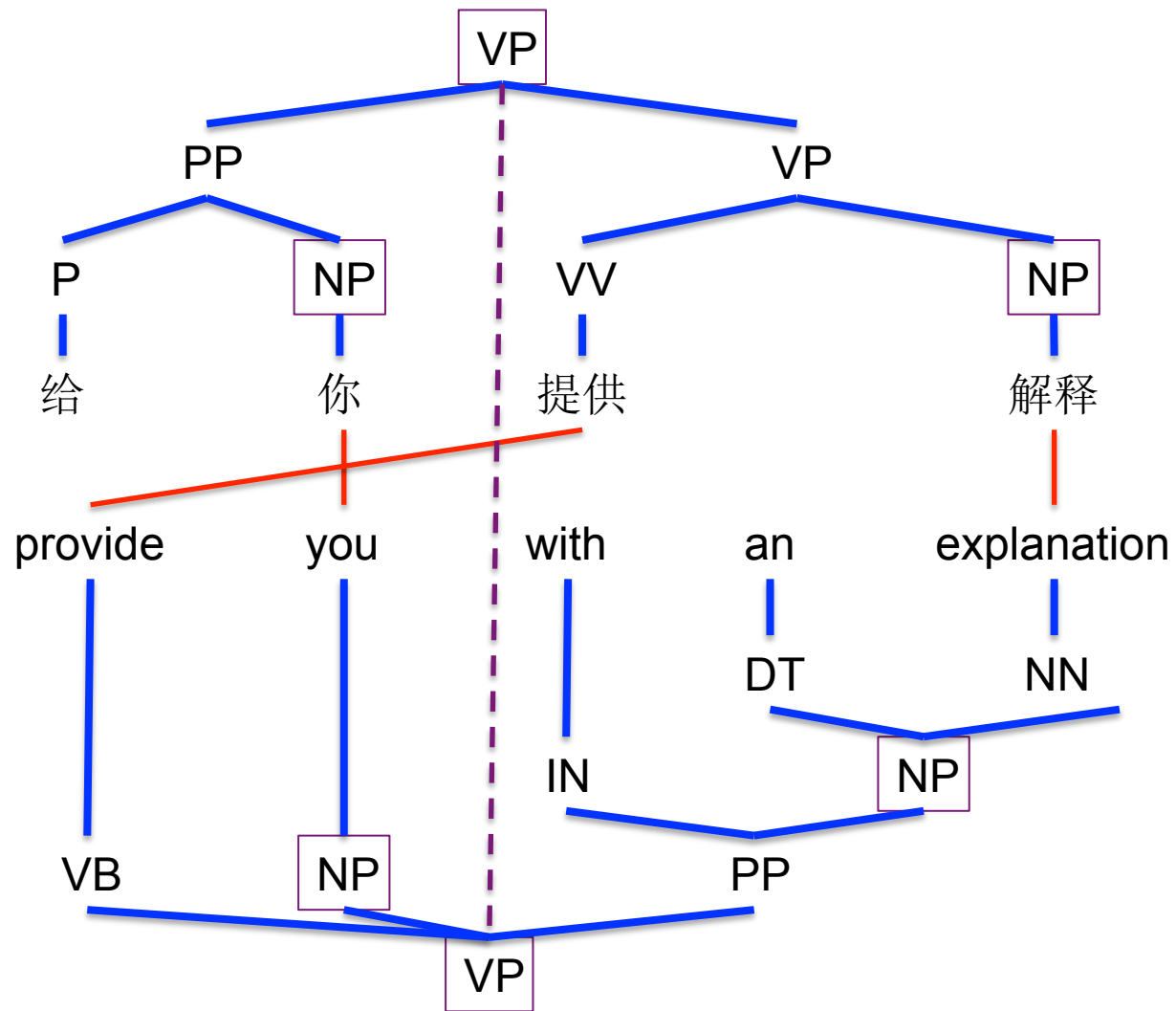




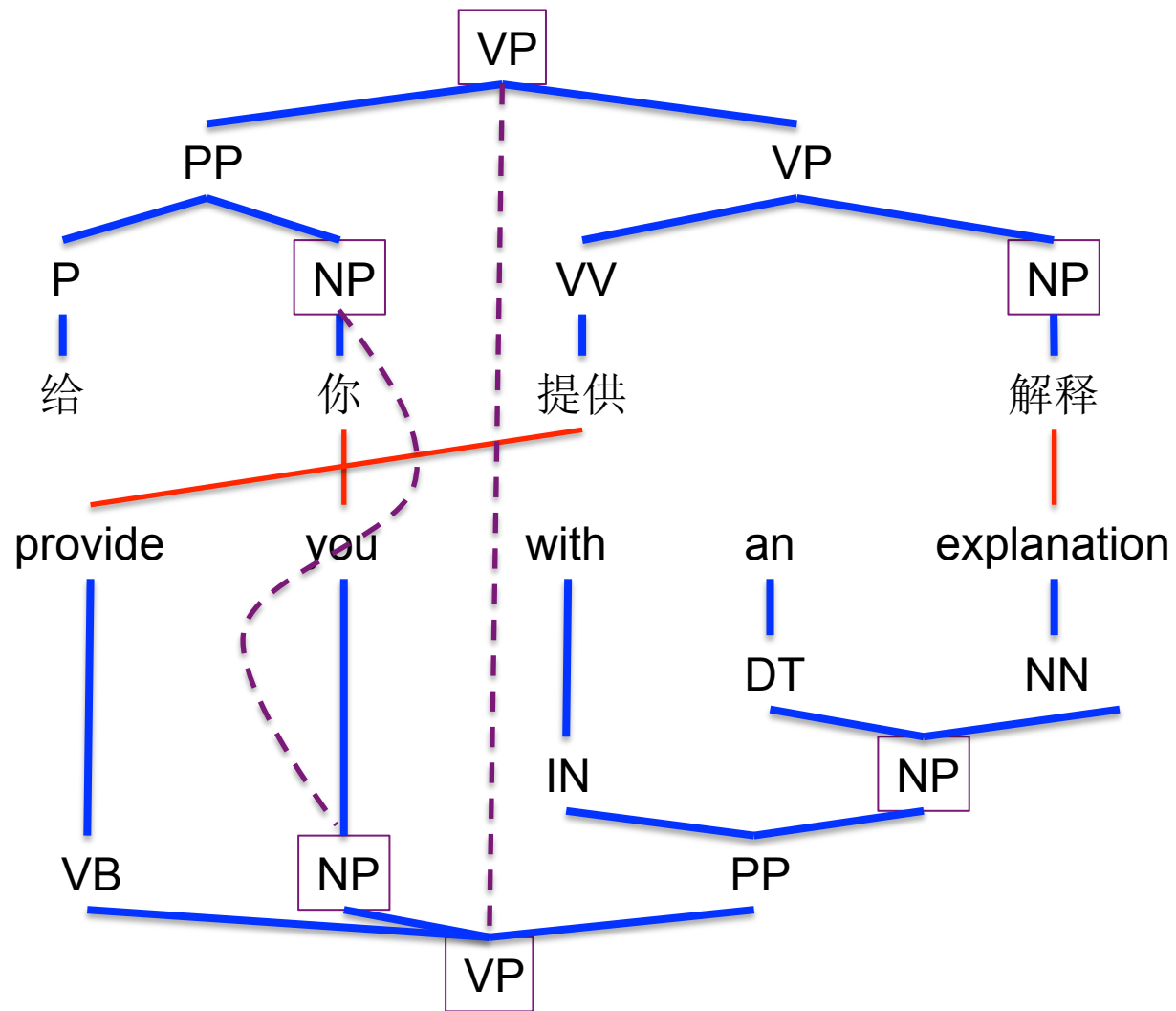
Hierarchical alignment constraints

- If Node n_c is aligned to Node n_e , then the descendants of n_c can only be aligned to the descendants of n_e
- If Node n_c is aligned to Node n_e , then the ancestors of n_c can only be aligned to the ancestors of n_e
- A non-terminal node can only be aligned once
(Tinsley et al, 2007)

Hierarchical alignment



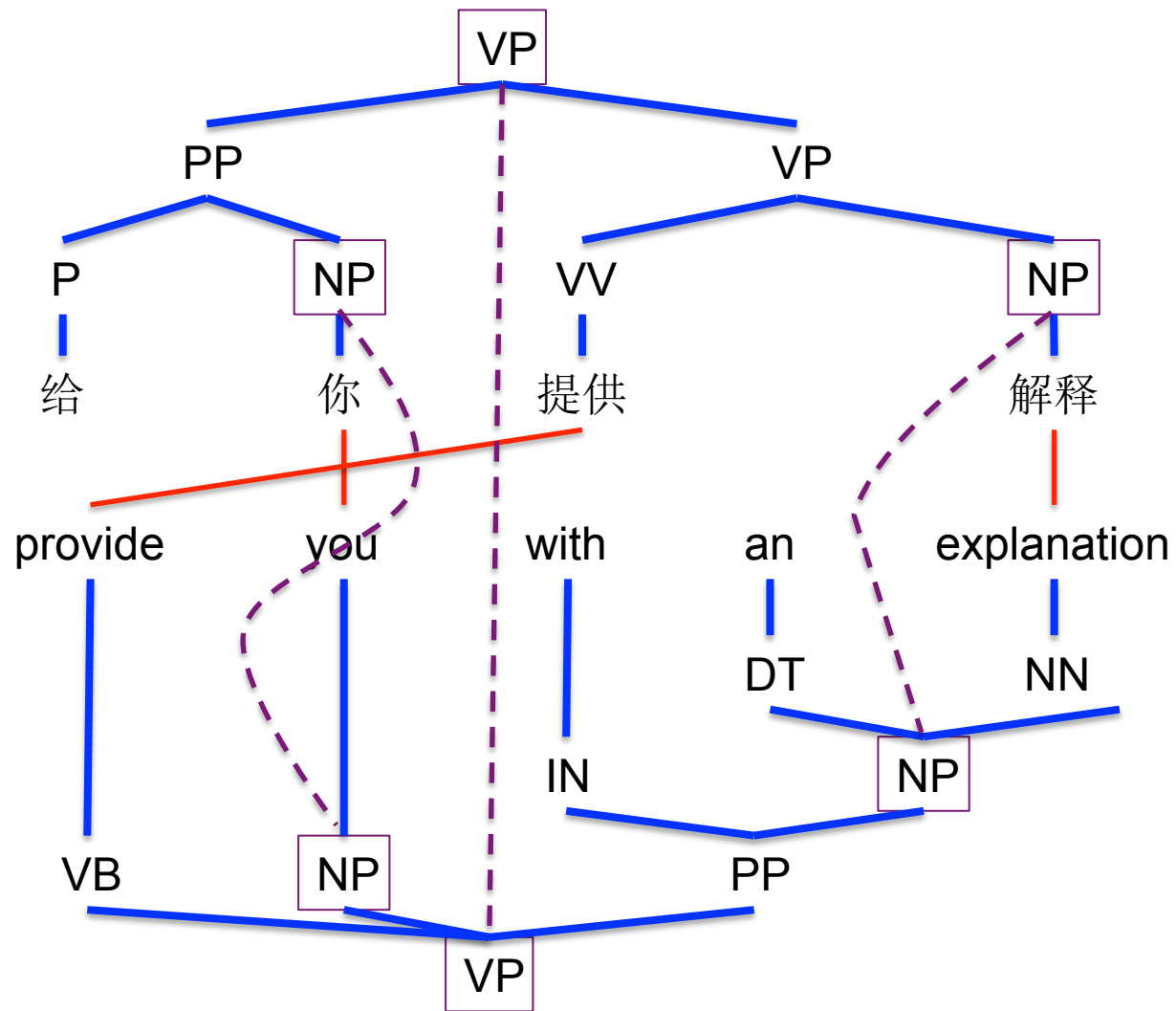
Hierarchical alignment



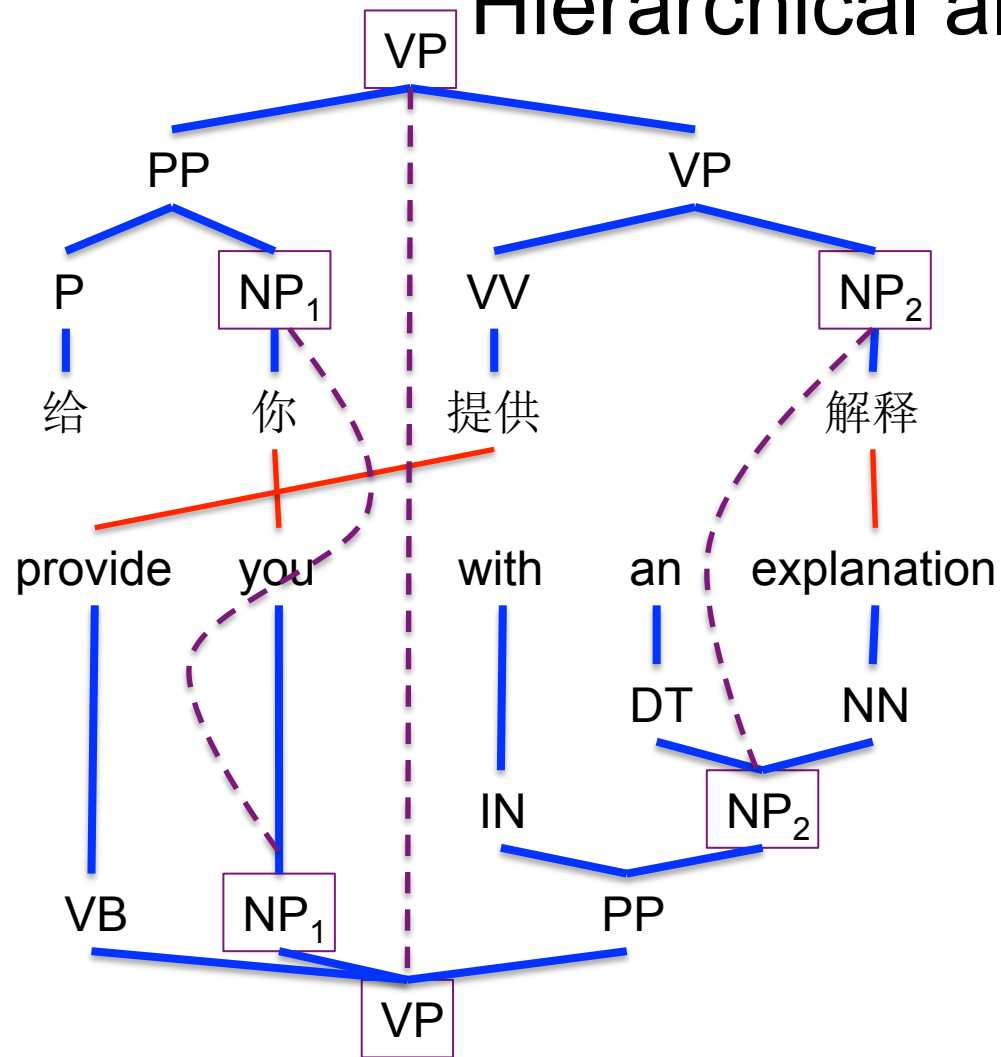
Guiding principles for determining the alignment of non-terminal nodes

- Provide enough context for the alignment
 - Alignment should not sever crucial dependencies or patterns
 - Reduces the number of haphazard alignments to its minimum
- Minimum alignment
 - Aligned (non-terminal) nodes dominate only crucial dependencies, nothing more

Hierarchical alignment



Hierarchical alignment



[你 <--> you]

[解释 <--> an explanation]

[给X₁提供X₂ <--> provide X₁ with X₂]

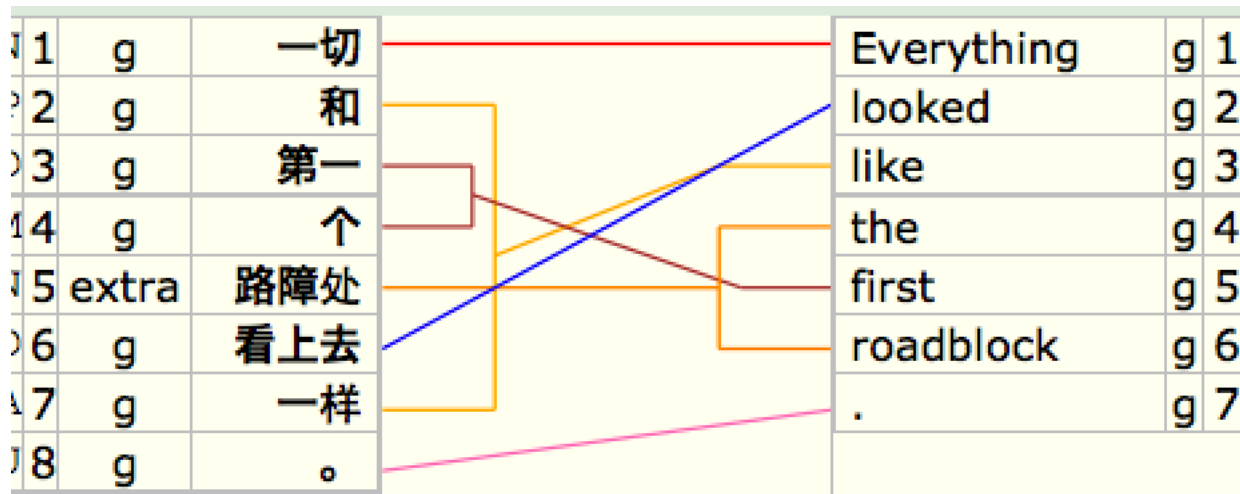


Alignment procedure

- Take a portion of the Penn Chinese TreeBank (CTB) that has English translations that have also been manually parsed (PTB) style and word aligned
- Revising the word alignments, and then align the non-terminal nodes, using a web-based tool developed by IBM

Attaching a determiner to its head

S3878



Hierarchically aligned

S3878

TOP—IP—NP_SBJ—PN	1	g	一切		Everything	g	1	NN—NP_SBJ—S—TOP
—VP—PP_ADV—P	2	extra	和		looked	g	2	VBD—VP/REU—
— —NP—QP—OD	3	g	第一		like	extra	3	IN—PP_CLR—
— — —CLP—M	4	extra	个		the	extra	4	DT—NP/UFW—
— —NP—NN	5	extra	路障处		first	g	5	JJ—
—ADVP—AD	6	g	看上去		roadblock	g	6	NN—
—VP—VA	7	extra	一样		.	g	7	.
—PU	8	g	。					

Attaching 把 to its NP “object”

S3818

1	g	UPGA		UPGA	g	1
2	n/a	完全		held	g	2
3	g	把		me	g	3
4	g	我		in	g	4
5	g	控制		its	g	5
6	g	在		sway	g	6
7	g	它的		.	g	7
8	g	的				
9	g	掌心				
10	g	。				

Hierarchically aligned

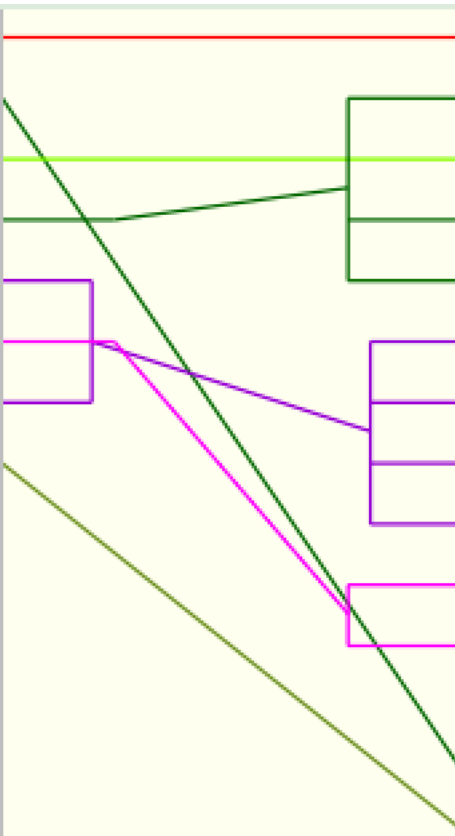
S3818

TOP	IP	NP_PN_SBJ	NR	1	g	UPGA		UPGA	g	1	NNP	NP_SBJ	S	TOP
	VP	ADVP	AD	2	n/a	完全		held	g	2	VBD	VP/REU		
	LVP		BA	3	extra	把		me	g	3	PRP	NP		
	LIP_OBJ	NP_SBJ	PN	4	g	我		in	g	4	IN	PP_CLR		
	LVP		VV	5	g	控制		its	g	5	PRP\$	NP/UFW		
	LPP		P	6	g	在		sway	g	6	NN			
	LNP	DNP	NP	7	g	它的		.	g	7	.			
			DEG	8	extra	的								
	LNP		NN	9	g	掌心								
			PU	10	g	。								

Verb prepositions/particles

S3877

1	g	他		he	g	1
2	g	也		had	g	2
3	g	没		n't	g	3
4	g	能		been	g	4
5	g	拿出		able	g	5
6	extra	会费		to	extra	6
7	g	来		come	g	7
8	g	。		up	g	8
				with	g	9
				the	g	10
				dues	g	11
				,	extra	12
				either	g	13
				.	g	14



Hierarchically aligned

S3877

TOP	IP	NP_SBJ	PN	1	g	他		he	g	1	PRP	NP_SBJ_1	S	TOP
		VP	ADVP	AD	2	g	也	had	extra	2	VBD	VP/UFW		
			ADVP	AD	3	g	没	n't	g	3	RB			
			LVP	VV	4	g	能	been	extra	4	VCN	VP		
			LVP	VP	VV	5	g	able	g	5	JJ	ADJP_PRD		
				NP_OBJ	NN	6	extra	to	extra	6	TO	VP-S		
				LVP	VV	7	extra	come	g	7	VB	vp/UFW	VP	
					PU	8	g	up	g	8	RP	PRT		
								with	g	9	IN	PP_CLR		
								the	extra	10	DT	NP/UFW		
								dues	g	11	NNS			
								,	extra	12	,			
								either	g	13	RB	ADVP		
								.	g	14	.			

Auxiliary verbs

S3705

1	g	那		how	g	1
2	extra	是		did	TEN	2
3	g	怎么		that	g	3
4	g	发生		happen	g	4
5	extra	的		?	g	5
6	g	?				

Hierarchically aligned

S3705

TOP-CP_Q-IP-NP_SBJ	PN	1	g	那	how	g	1WRB-WHADVP_1-SBARQ-TOP
	LVP	VC	2	extra	是	extra	2VBD
	LVP-ADVP_WH-AD	AD	3	g	怎么	g	3DT-NP_SBJ
	LVP	VV	4	g	发生	g	4VB-VP
		SP	5	extra	的	g	5.
		PU	6	g	?		



Are the PTB and CTB trees compatible?

- Not always. Some of the incompatibilities are well-documented in MT literature
- Three types of incompatibilities:
 - Incompatibilities between lexico-semantic differences between the two languages
 - Incompatibilities caused by translation-related reasons
 - 5 cases out of 500 sentence pairs
 - Incompatibilities caused by bracketing annotation
 - 20 cases out of 500 sentence pairs
- The last one we can do something about

Idiomatic translations

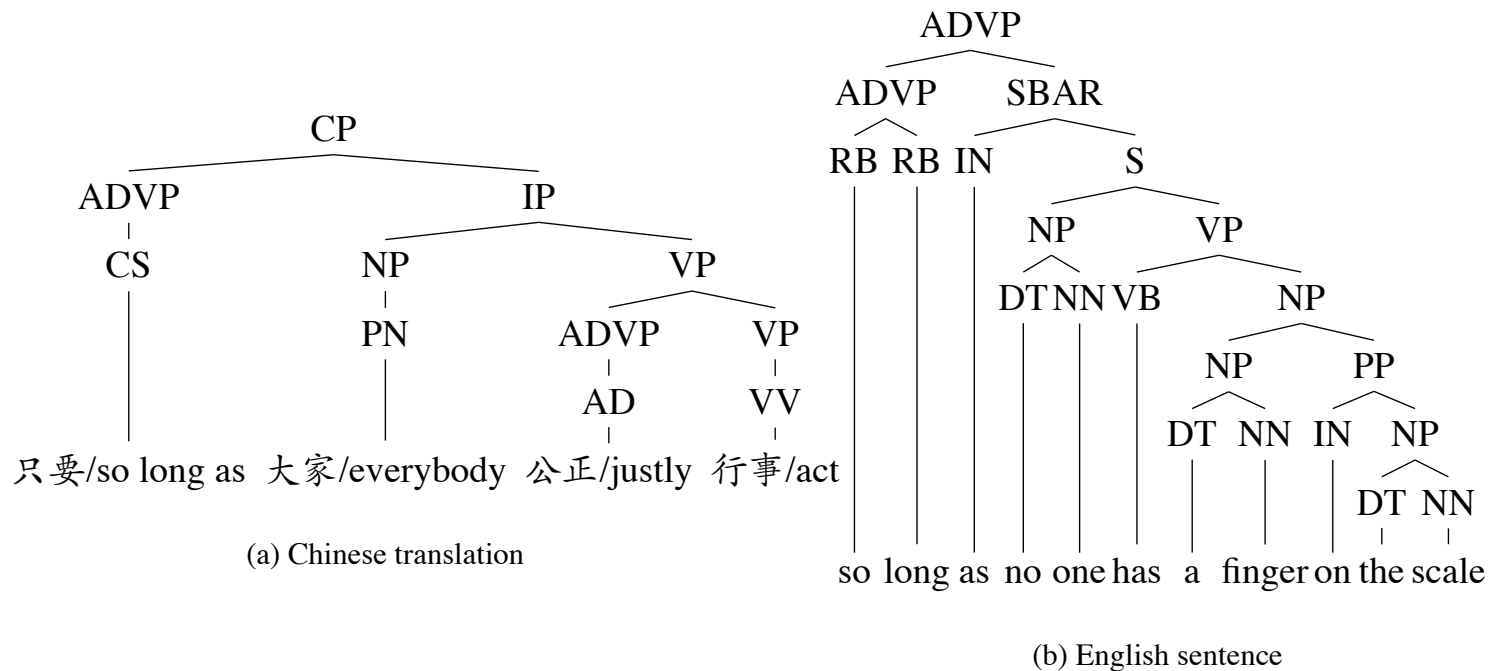
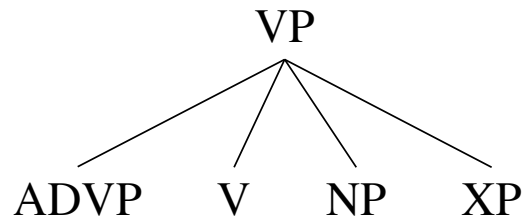
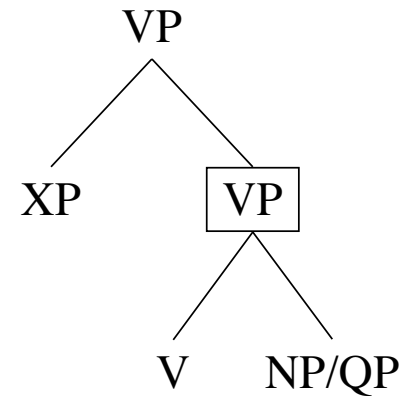


Figure 2: Translation of idiomatic expressions.

Divergent annotation standards



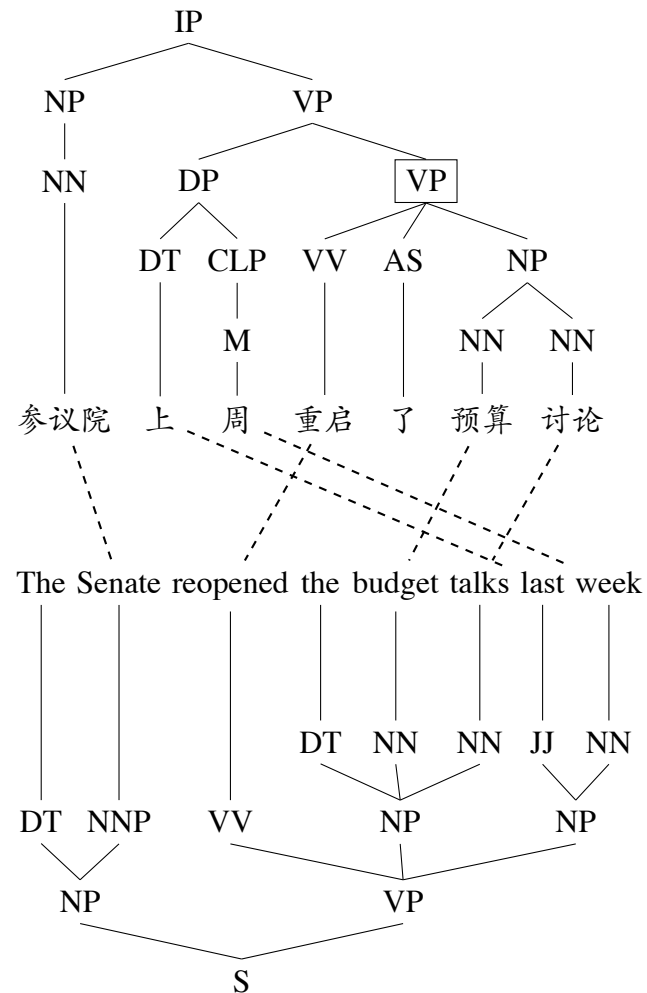
(a) Penn English TreeBank structure



(b) Chinese TreeBank structure

Figure 4: Differences in treebank annotation standards. $XP = \{NP, PP, ADVP, S\}$

Divergent annotation standards





Can this be done quickly and consistently?

- Annotators with no prior linguistic training can perform this kind of alignment with good consistency given proper training and guidelines
 - 87% Inter-annotator agreement for (non-terminal) node alignment
- Tree synchronization requires linguistic training
 - We currently just mark places of tree divergence
- Status:
 - Double annotation of about 10,000 sentence pairs



Use cases for hierarchical alignment

- Extracting Hierarchical rules
 - From an unaligned function word, find its alignment domain by identifying its closest ancestor node N that is aligned.
 - From N, all descendant nodes that are not node-aligned themselves are visible as potential **triggers** of a hierarchical rule
- Alignment-driven flattening of syntactic trees
 - We can assume the unaligned nodes are unimportant to MT and can be automatically eliminated to create a simplified tree for MT purposes

Passives

S66

TOP-IP-ADVP-AD	1	g	最终	Eventually	g	1	RB-ADVP_TMP-S-TOP
PU	2	extra	,	we	g	2	PRP-NP_SBJ
NP_SBJ-PN	3	g	我们	will	g	3	MD-VP
VP-VP-ADVP-AD	4	g	将	outlaw	g	4	VB-VP/REU-VP
LVP-BA	5	extra	把	gravity	g	5	NN-NP
LIP_OBJ-NP_SBJ-NN	6	g	重力	so	extra	6	IN-SBAR_PRP
LVP-VV	7	g	定为	that	extra	7	IN
LNP_OBJ-ADJP-JJ	8	g	非法	sludge	g	8	NN-NP_SBJ_1-S/REO
LNP-NN	9	g	因素	is	extra	9	VB-VP
PU	10	extra	,	prohibited	g	10	VRN-VP
LVP-ADVP-AD	11	extra	这样	from	extra	11	IN-PP_CLR
LVP-VV	12	extra	可	running	g	12	VBG-VP-S_NOM
LVP-VV	13	g	禁止	downhill	g	13	RB-ADVP_DIR
NP_OBJ-NN	14	g	淤泥	.	g	14	.
LIP-VP-VV	15	g	流到				
LNP_OBJ-NN	16	g	山下				
PU	17	g	。				

[可 禁止 X_1 X_2 \leftrightarrow X_1 is prohibited from X_2]

把 construction

S66

TOP	IP	ADVP	AD	1	g	最终		Eventually	g	1	RB	ADVP_TMP	S	TOP
			PU	2	extra	,		we	g	2	PRP	NP_SBJ		
	NP_SBJ		PN	3	g	我们		will	g	3	MD	VP		
	VP	VP	ADVP	4	g	将		outlaw	g	4	VB	VP/REU	VP	
		LVP		5	extra	把		gravity	g	5	NN	NP		
			LIP_OBJ-NP_SBJ-NN	6	g	重力		so	extra	6	IN	SBAR_PRP		
			LVP	7	g	定为		that	extra	7	IN			
			LNP_OBJ-ADJP-JJ	8	g	非法		sludge	g	8	NN	NP_SBJ_1-S/REO		
			LNP	9	g	因素		is	extra	9	VB	VP		
			PU	10	extra	,		prohibited	g	10	VRN	VP		
	LVP	ADVP	AD	11	extra	这样		from	extra	11	IN	PP_CLR		
		LVP		12	extra	可		running	g	12	VBG	VP-S_NOM		
		LVP		13	g	禁止		downhill	g	13	RB	ADVP_DIR		
		NP_OBJ	NN	14	g	淤泥		.	g	14	.			
		LIP	VP	15	g	流到								
			LNP_OBJ	NN	16	g	山下							
			PU	17	g	。								

[把 X 定为非法因素 <--> outlaw X]

Connectives

S66

TOP-IP-ADVP-AD	1	g	最终	Eventually	g	1	RB-ADVP_TMP-S-TOP
PU	2	extra	,	we	g	2	PRP-NP_SBJ
NP_SBJ-PN	3	g	我们	will	g	3	MD-VP
VP-VP-ADVP-AD	4	g	将	outlaw	g	4	VB-VP/REU-VP
LVP-BA	5	extra	把	gravity	g	5	NN-NP
LIP_OBJ-NP_SBJ-NN	6	g	重力	so	extra	6	IN-SBAR_PRP
LVP-VV	7	g	定为	that	extra	7	IN
LNP_OBJ-ADJP-JJ	8	g	非法	sludge	g	8	NN-NP_SBJ_1-S/REO
LNP-NN	9	g	因素	is	extra	9	VB-VP
PU	10	extra	,	prohibited	g	10	VRN-VP
LVP-ADVP-AD	11	extra	这样	from	extra	11	IN-PP_CLR
LVP-VV	12	extra	可	running	g	12	VBG-VP-S_NOM
LVP-VV	13	g	禁止	downhill	g	13	RB-ADVP_DIR
NP_OBJ-NN	14	g	淤泥	.	g	14	.
LIP-VP-VV	15	g	流到				
LNP_OBJ-NN	16	g	山下				
PU	17	g	。				

[将 X_1 , 这样 $X_2 \leftrightarrow$ will X_1 so that X_2]

Conjunction

S3850

TOP-IP-NP_SBJ	PN	1	g	我	I	g	1	PRP	NP_SBJ-S-TOP
VP	VV	2	g	明白	understood	g	2	VBD	VP/UFW
L-IP-OBJ-IP-NP_SBJ	NN	3	g	讨论	that	extra	3	IN	SBAR
VP-ADVP	AD	4	extra	已	the	extra	4	DT	NP_SBJ/UFW-S-S/UFW
LVP	VV	5	g	结束	discussion	g	5	NN	
PU		6	extra	,	had	extra	6	VBD	VP/UFW
CP-IP-NP_SBJ	PN	7	g	我	ended	g	7	VBN	VP
LVP	VE	8	g	没有	and	extra	8	CC	
NP-OBJ	NN	9	g	出路	I	g	9	PRP	NP_SBJ-S/UFW
SP		10	extra	了	had	g	10	VBD	VP
PU		11	g	。	no	g	11	DT	NP-NP
					way	g	12	NN	
					out	g	13	RB	ADVP
					.	g	14	.	

[X , Y <> X and Y]

Questions

S1550

TOP-IP-Q	PU	1	extra	+		Why	g	1	WRB-WHADVP_2-SBARQ/REU-TOP
NP-SBJ	PN	2	g	他们		do	extra	2	VBP-SQ
VP-ADVP_WH-AD	AD	3	g	为什么		n't	g	3	RB
ADVP	AD	4	g	不		the	g	4	PRP-NP-SBJ-1
VP-VSB	VV	5	g	出去		go	g	5	VB-VP/UFW
	VV	6	g	打仗		out	g	6	RB-ADVP_DIR
	PU	7	g	?		to	extra	7	TO-VP-S_PRP
						fight	g	8	VB-VP
						?	g	9	.

[X₁ 为什么 不 X₂ <--> Why do n't X₁ X₂]

Questions

S3250

TOP-CP-Q-IP-ADVP-AD	1	g	那		Then	g	1RB-ADVP-SBARQ/REU-TOP
					why	g	2WRB-WHADVP_1-
					do	extra	3VB-SQ-
					I	g	4PRP-NP_SBJ-
					still	g	5RB-ADVP_TMP-
					like	g	6VB-VP-
					Inter	g	7NNP-NP-
					Milan	g	8NNP-
					?	g	9.

那X为什么Y呢? <> Then why do X Y?

Negation

S1400

TOP	IP-NP_SBJ	PN	1	g	我		I	g	1	PRP	NP_SBJ-S-TOP
	VP	VV	2	g	认为		do	extra	2	VBP	VP/REU
	LIP_OBJ-NP_SBJ-PN	PN	3	g	我们		n't	extra	3	RB	
	LVP-ADVP	AD	4	extra	不必		think	g	4	VB	VP
	PP_LOC	P	5	extra	在		we	g	5	PRP	NP_SBJ_1-S-SBAR
	NP-DP	DT	6	g	每一		need	extra	6	VBP	VP
	LQP	CD	7	extra	一个		see	g	7	VB	VP-S
	LCLP	M	8	extra	个		a	extra	8	DT	NP/UFW-NP
	NP	NN	9	g	地方		plot	g	9	NN	
	ADVP	AD	10	extra	都		in	extra	10	IN	PP_LOC
	LVP	VV	11	g	看到		every	g	11	DT	NP/UFW
	NP_OBJ	NN	12	g	阴谋		wood	g	12	NN	
	PU		13	g	。		pile	g	13	NN	
							.	g	14	.	

[认为 我们 不必 X 都 看到 Y <> do n't think we need see Y X]



Summary

- Strategically select non-terminal node pairs to align
 - Effectively creates an hierarchical partition for each sentence
- Harmonize word alignment and (non-terminal) node alignment
 - Content words are (generally) aligned the word level
 - Function words are (generally) aligned via the alignment of non-terminal nodes
- Synchronize parallel parse trees
 - Address divergences in PTB and CTB annotation styles



Cases for and against hierarchical alignment for MT

- Supports transfer-based approach to MT, no need for a separate generation step. More steps in the pipeline mean more chances for errors
- Can be readily exploited by existing statistical techniques (supervised synchronized parsing?)
- But need to be done for each language pair. May not work for some language pairs.



Conclusions

- Familiar tradeoffs between depth of representation (therefore better alignment) and processing difficulty (analysis, generation)
- A hierarchically aligned parallel corpus might be a viable alternative as a representation for MT



Acknowledgement

- Libin Shen for inspirations
- Tool support by the IBM Multilingual NLP Technologies Group: Niyu Ge, Abe Ittycheriah, Salim Roukos
- Funded by DARPA BOLT Program via IBM