Hierarchy of Tai mapping for rooted labeled trees

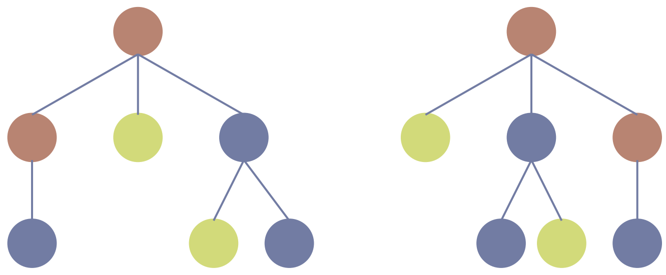
平田耕一(Kouichi Hirata)
(九州工業大学 (Kyushu Institute of Technology))

Contents

- Tree edit distance
- Tai mapping
- Variations of Tai mapping
- Hierarchy of Tai mapping

Ordered trees vs. unordered trees

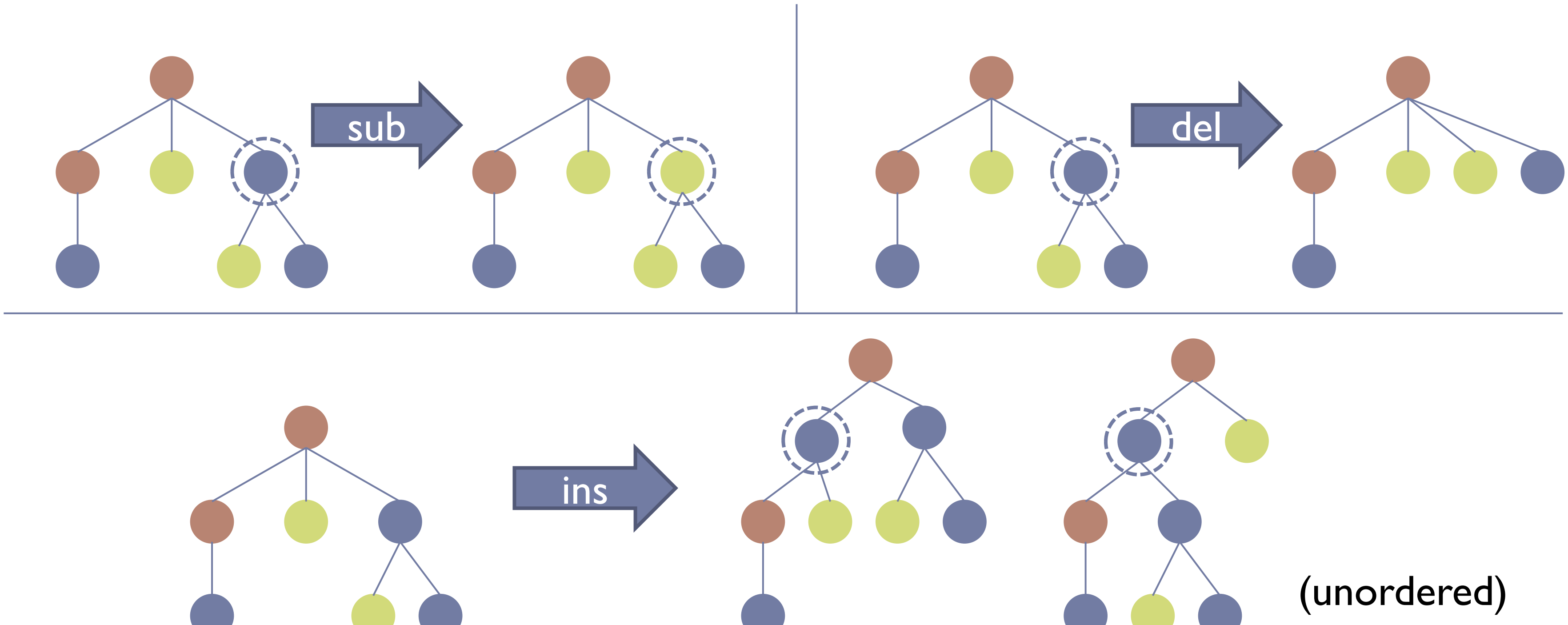
- Rooted labeled trees
  - Rooted tree : a tree with root
  - Labeled tree : each node is assigned to a label
- Ordered trees
  - Left-to-right order among siblings is given
- Unordered trees
  - Left-to-right order among siblings is not given



Non-isomorphic for ordered trees
Isomorphic for unordered trees

Tree edit distance [Tai 79 (J. ACM 26)]

- Edit operation
  - Substitution (relabeling) (sub)
  - Deletion (del)
  - Insertion (ins)

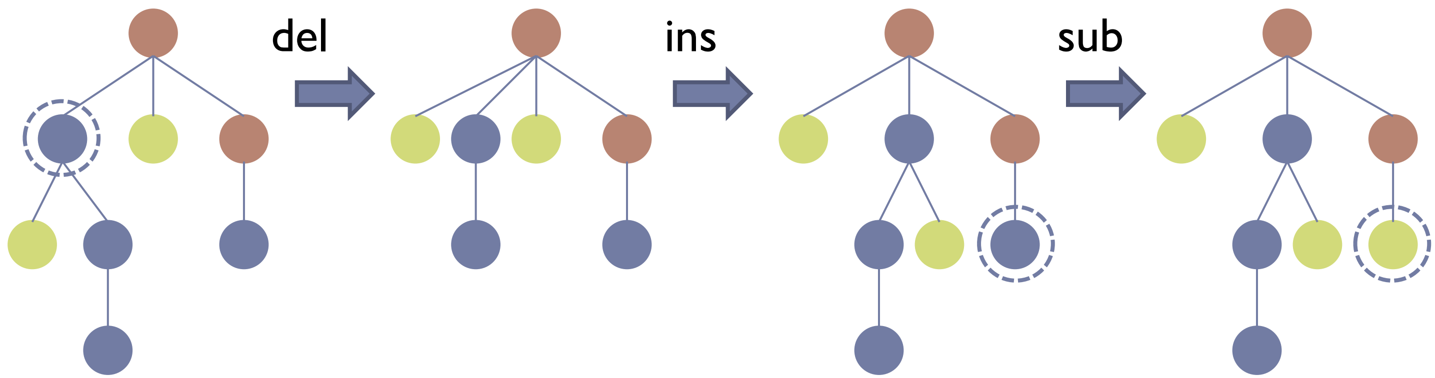


Tree edit distance

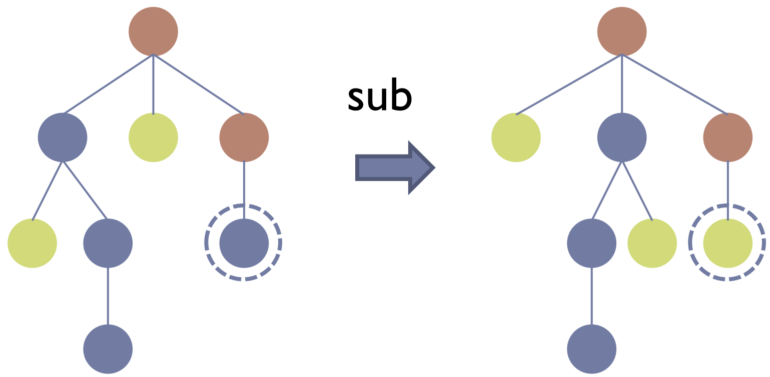
- Edit operation
  - Substitution (relabeling) (sub)
  - Deletion (del)
  - Insertion (ins)
  - Each edit operation is assigned to a cost
- Cost
  - Unit cost : sub=del=ins=1
  - Indel cost : sub=2, del=ins=1
- Tree edit distance
  - The minimum cost to transform from a tree to another tree by applying edit operations

Tree edit distance

- Unit cost
- Ordered tree edit distance = 3



- Unordered tree edit distance = 1

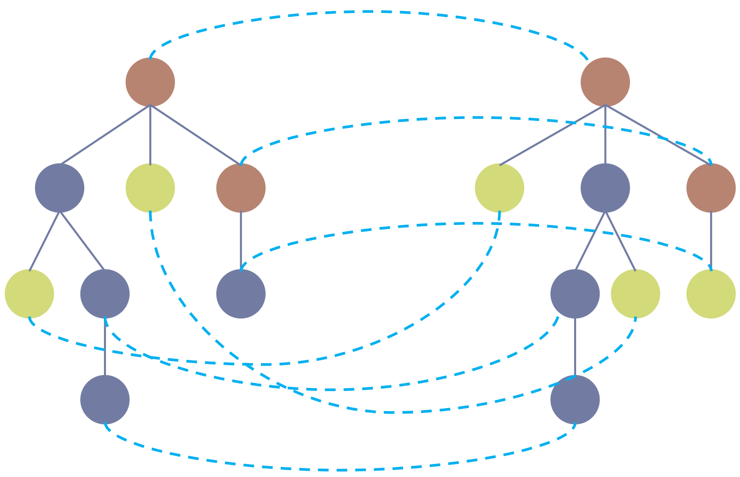


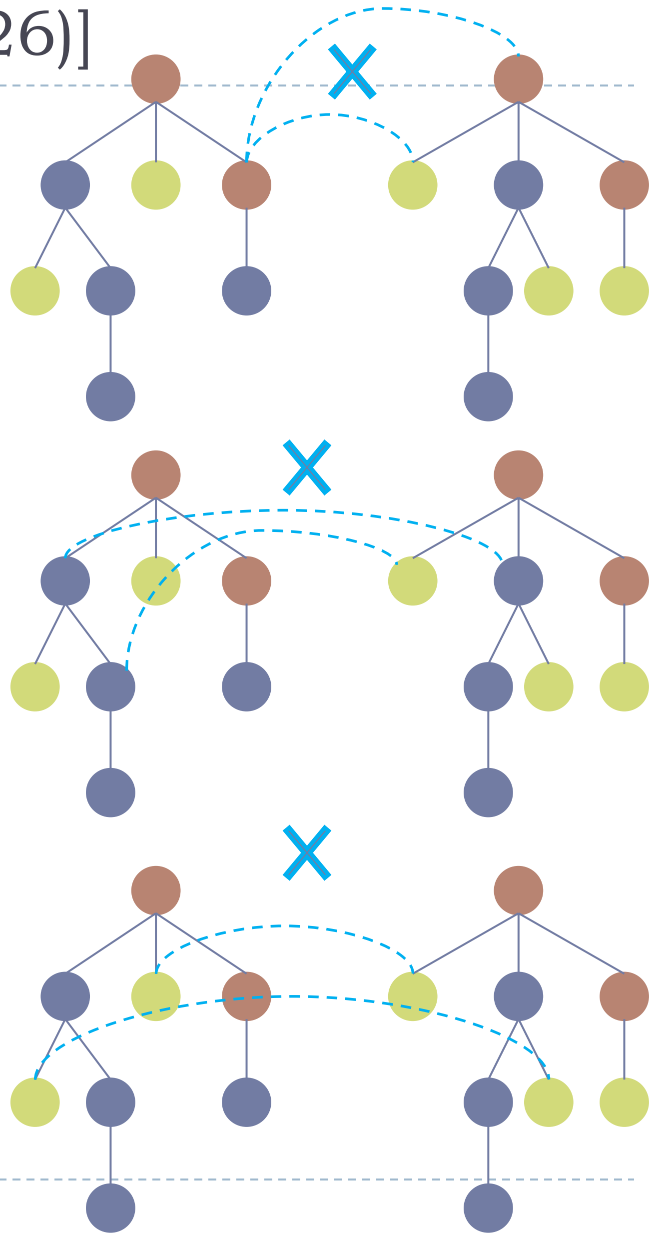
Time complexity

- Ordered tree edit distance
  - $O(n^6)$ [Tai 79 (J. ACM 26)]
  - $O(n^4)$ [Zhang & Shasha 89 (SIAM J. Comput. 18)]
  - $O(n^3 \log n)$ [Klien 98 (Proc. ESA'98)]
  - $O(n^3)$ [Demaine et al. 09 (ACM Trans. Algorithms 6)]
  - n : maximum number of nodes in given two trees
- Unordered tree edit distance
  - NP-hard [Zhang et al. 92 (Inform. Process. Let. 42)]
    - Unit cost, degree 2 or height 2
  - MAX SNP-hard [Zhang & Jiang 94 (Inform. Process. Let. 49)]
    - Indel cost, height 7
    - Unit cost, degree 2 or height 3 [CPM 12]

Tai mapping [Tai 79 (J. ACM 26)]

- $M \subseteq T_1 \times T_2$: Tai mapping
  - One-to-one
  - Ancestor condition
  - Sibling condition (ordered trees)



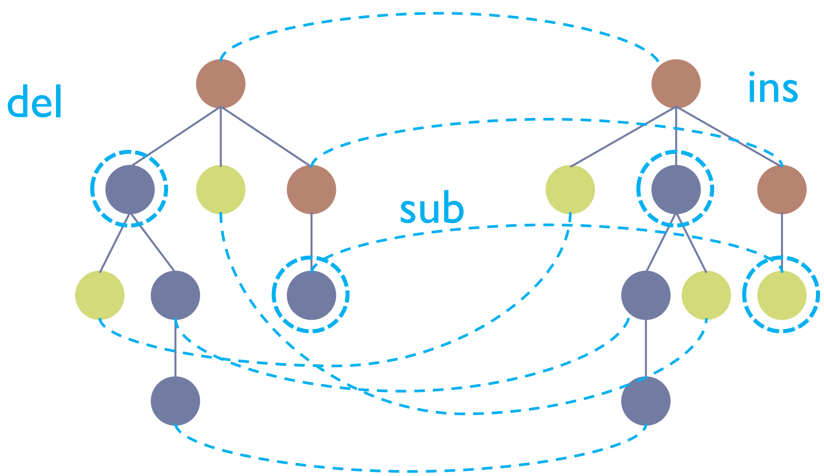


Tai mapping

- The minimum cost of possible Tai mappings

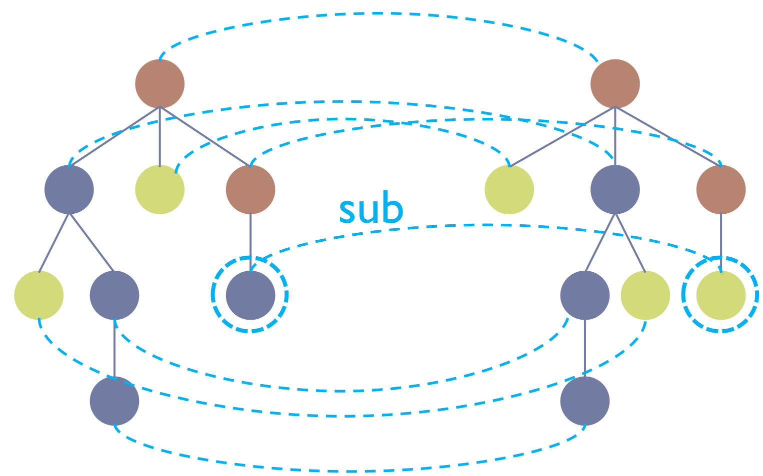
  ＝tree edit distance

Minimum cost Tai mapping (ordered)



Cost of the Tai mapping
＝3
＝ordered tree edit distance

Minimum cost Tai mapping (Unordered)



Cost of the Tai mapping
＝1
＝unordered tree edit distance

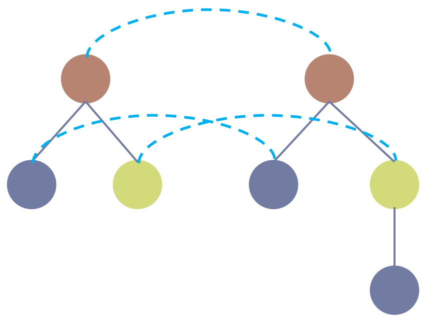
Variation of Tai mapping

- $M \subseteq T_1 \times T_2$: Tai mapping, $M^- = M - \{r(T_1), r(T_2)\}$
- Top-down (degree-1) mapping
  [Selkow 77 (IPL 6), Chawathe 99 (VLDB)]
  - $\forall (v, w) \in M^-$ $(p(v), p(w)) \in M$

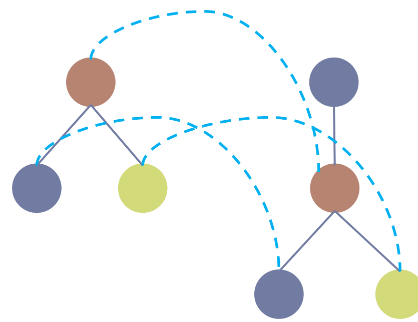
$r(T)$: root of T
$p(v)$: parent of v
$v_1 \sqcup v_2$: LCA of v_1 and v_2

- LCA-preserving (degree-2) mapping [Zhang 95 (IJFCS 7)]
  - $\forall (v_1, w_1), (v_2, w_2) \in M$ $(v_1 \sqcup v_2, w_1 \sqcup w_2) \in M$
- Accordant (Lu's) mapping [Lu 79 (IEEE PAMI 1), Kuboyama 07]
  - $\forall (v_1, w_1), (v_2, w_2), (v_3, w_3) \in M$
    $v_1 \sqcup v_2 = v_1 \sqcup v_3 \Leftrightarrow w_1 \sqcup w_2 = w_1 \sqcup w_3$

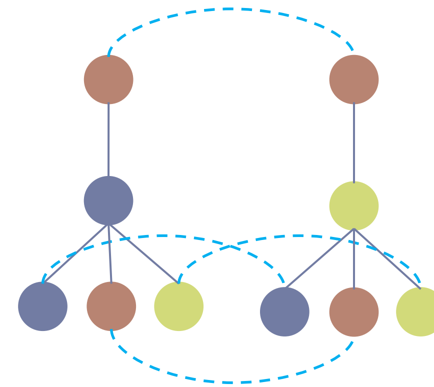
Variation of Tai mapping



Top-down



LCA-preserving
Not top-down



Accordant
Not LCA-preserving

M : top-down mapping $\Rightarrow$ M : LCA-preserving mapping $\Rightarrow$ M : accordant mapping

Variation of Tai mapping

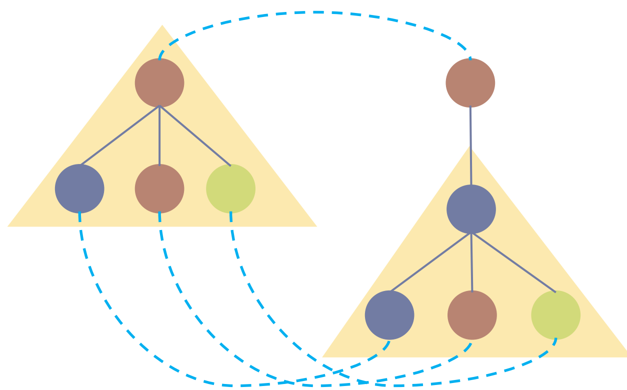
- Constrained (isolated-subtree) mapping

  [Zhang 95 (Patten Recog. 28), Zhang 96 (Algorithmica 15)]
  - $\forall (v_1, w_1), (v_2, w_2), (v_3, w_3) \in M$

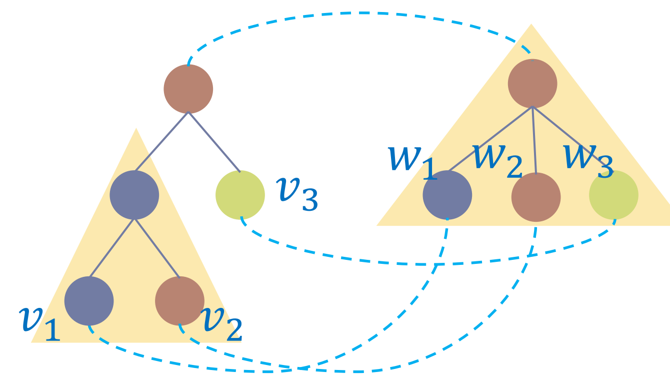
    $v_3 < v_1 \sqcup v_2 \Leftrightarrow w_3 < w_1 \sqcup w_2$

$v_1 \sqcup v_2$: LCA of v_1 and v_2

- A pair in a constrained mapping is "in a pair of subtrees"



Constrained
Not accordant



Not constrained

Variation of Tai mapping

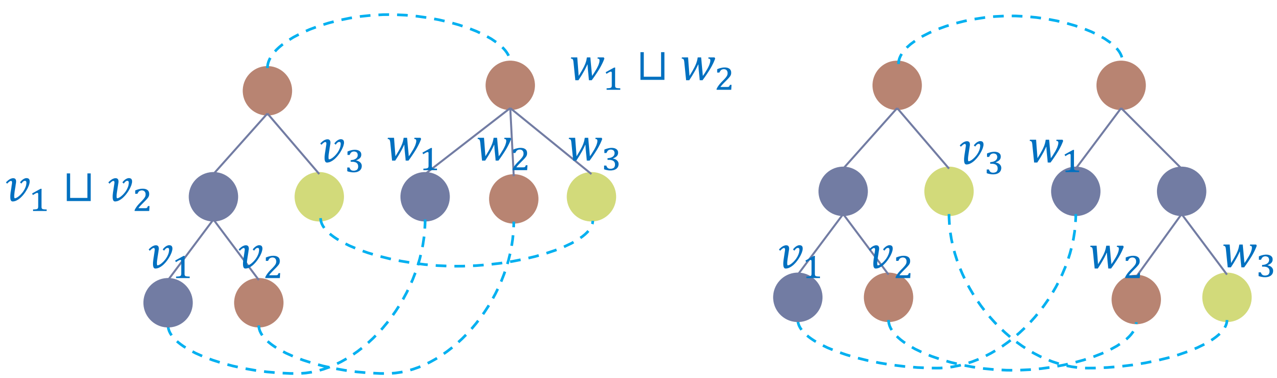
- Less-constrained (alignable) mapping
  [Lu 01 (COCOON), Kuboyama 07]
  - $\forall (v_1, w_1), (v_2, w_2), (v_3, w_3) \in M$

    $v_1 \sqcup v_2 < v_1 \sqcup v_3 \Rightarrow w_2 \sqcup w_3 = w_1 \sqcup w_3$

$r(T)$: root of T
$v_1 \sqcup v_2$: LCA of v_1 and v_2

- A less-constrained mapping contains no "twist"
- A less-constrained mapping

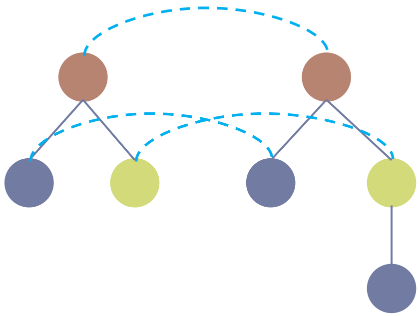
= a mapping for alignment (alignable mapping) [Kuboyama 07]



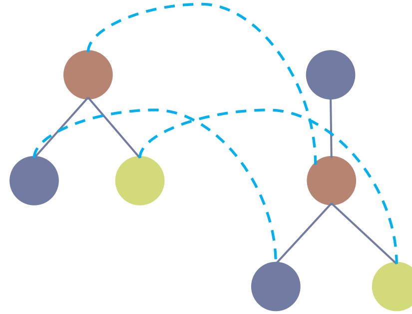
Less-constrained
Not constrained

Not less-constrained

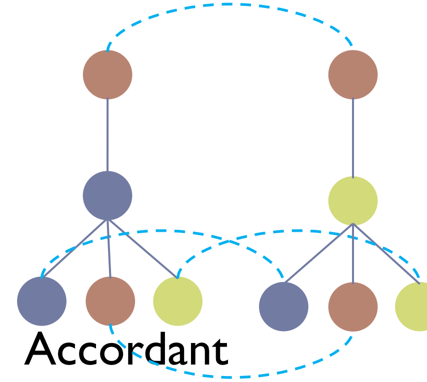
Variation of Tai mapping



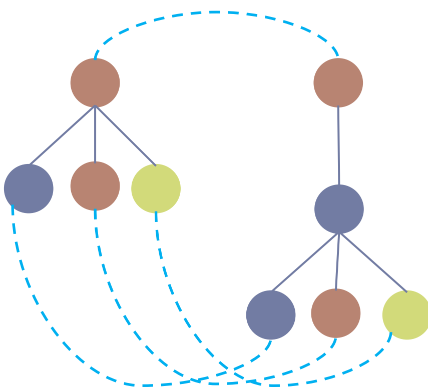
Top-down



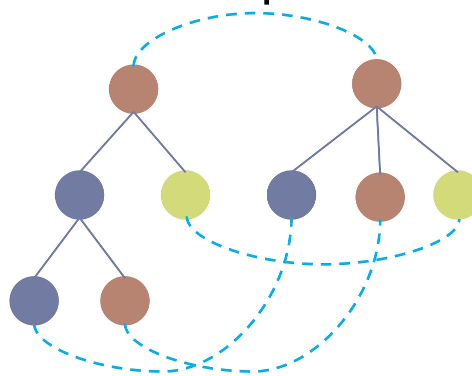
LCA-preserving
Not top-down



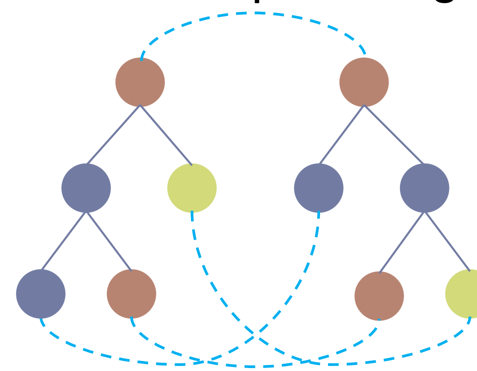
Accordant
Not LCA-preserving



Constrained
Not accordant



Less-constrained
Not constrained



Tai
Not less-constrained

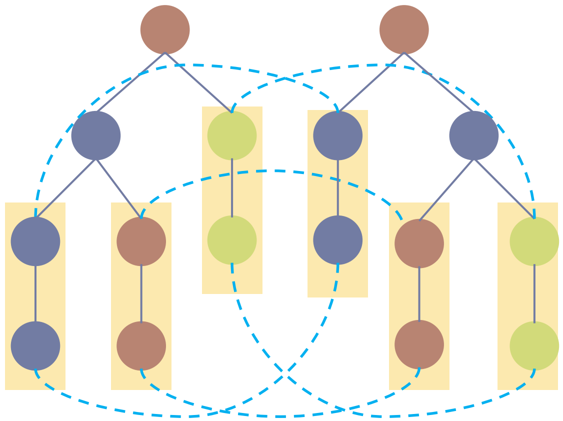
M : top-down mapping $\Rightarrow$ M : LCA-preserving mapping $\Rightarrow$ M : accordant mapping $\Rightarrow$ M : constrained mapping $\Rightarrow$ M : less-constrained mapping $\Rightarrow$ M : Tai mapping

Variation of Tai mapping

- Bottom-up mapping [Valiente 01 (SPIRE)]
  - $\forall (v, w) \in M$ $\qquad T[v]$: complete subtree of T rooted by v

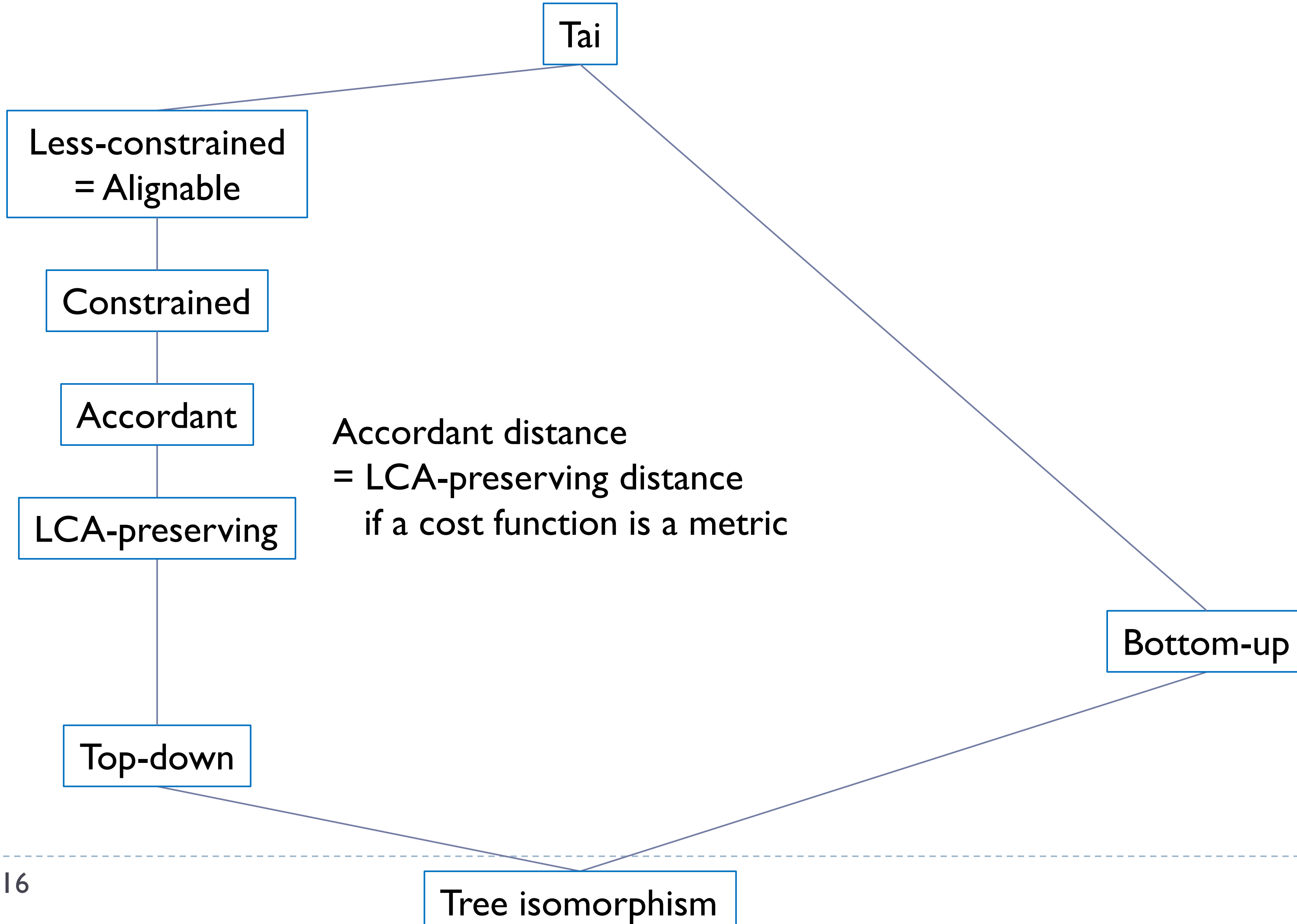
    $\forall v' \in T_1[v] \exists w' \in T_2[w]\ (v', w') \in M$

    $\land \forall w' \in T_2[w]\ \exists v' \in T_1[v]\ (v', w') \in M$



Bottom-up
Not less-constrained

Hierarchy of Tai mapping [cf. Kuboyama 07]



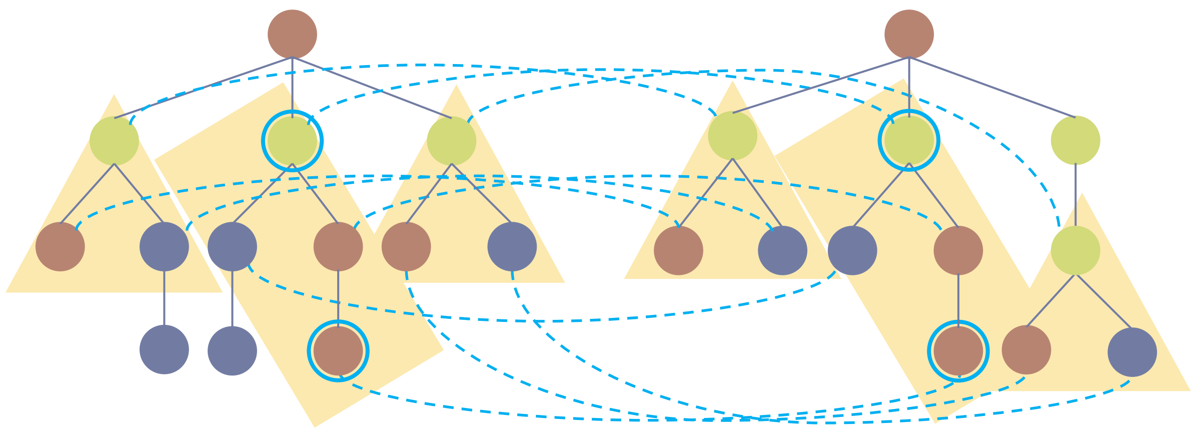
Accordant distance
= LCA-preserving distance
if a cost function is a metric

Variation of Tai mapping

- Segmental mapping [Kan 12 (ISAAC)]
  - $\forall (v, w) \in M^{-}$

$anc(v)$: ancestors of v

$$(v', w') \in M \land (v' \in anc(v)) \land (w' \in anc(w))$$
$$\Rightarrow \big(p(v), p(w)\big) \in M$$



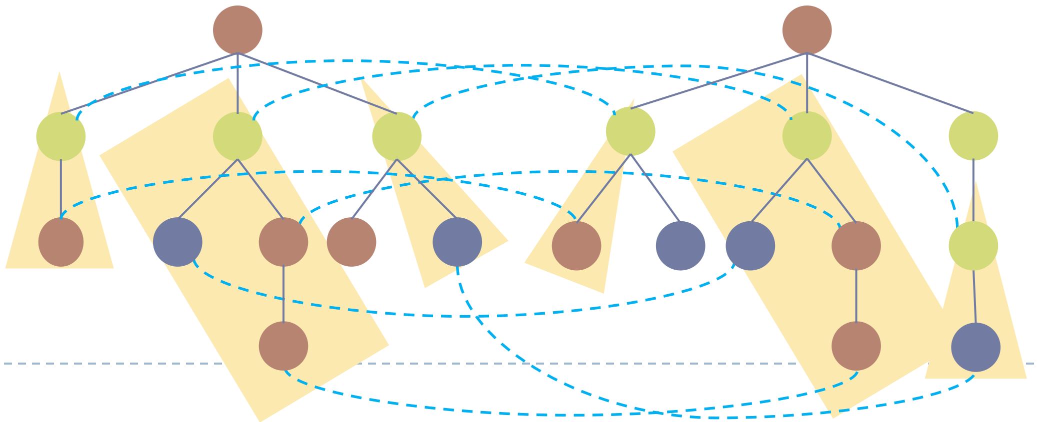
Variation of Tai mapping

- Top-down segmental mapping [Kan 12 (ISAAC)]
  - Segmental and $\left(r(T_1), r(T_2)\right) \in M$
- Bottom-up segmental mapping [Kan 12 (ISAAC)]
  - $\forall (v, w) \in M$

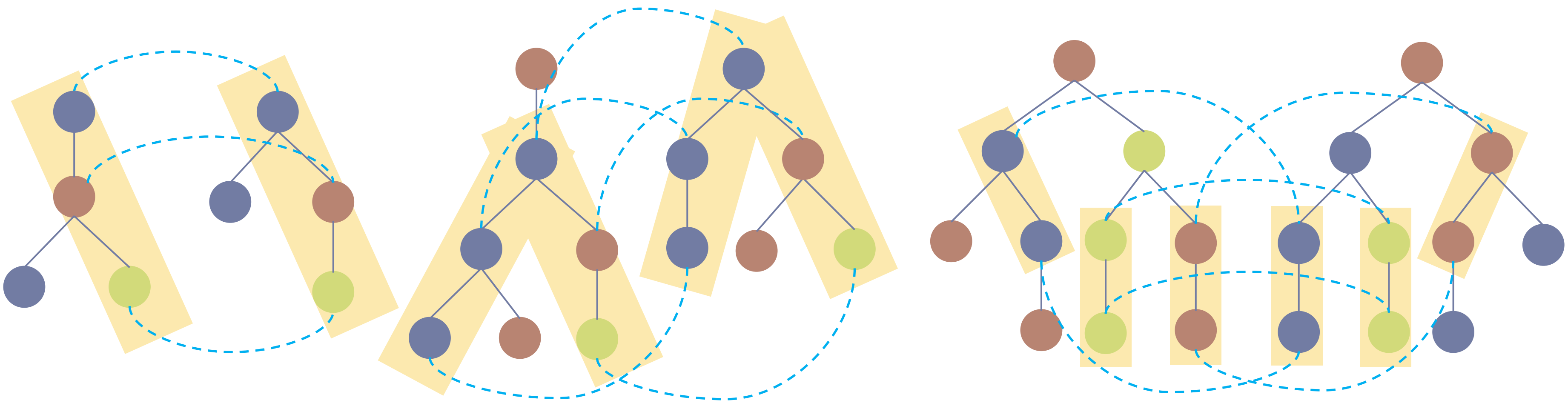
$$\exists (v', w') \in M \; (v \in anc(v') \wedge w \in anc(w') \wedge v' \in lv(T_1) \wedge w' \in lv(T_2))$$

$$\vee \; (v \in lv(T_1) \wedge w \in lv(T_2))$$

$lv(T)$: set of leaves in T
$anc(v)$: ancestors of v



Variation of Tai mapping

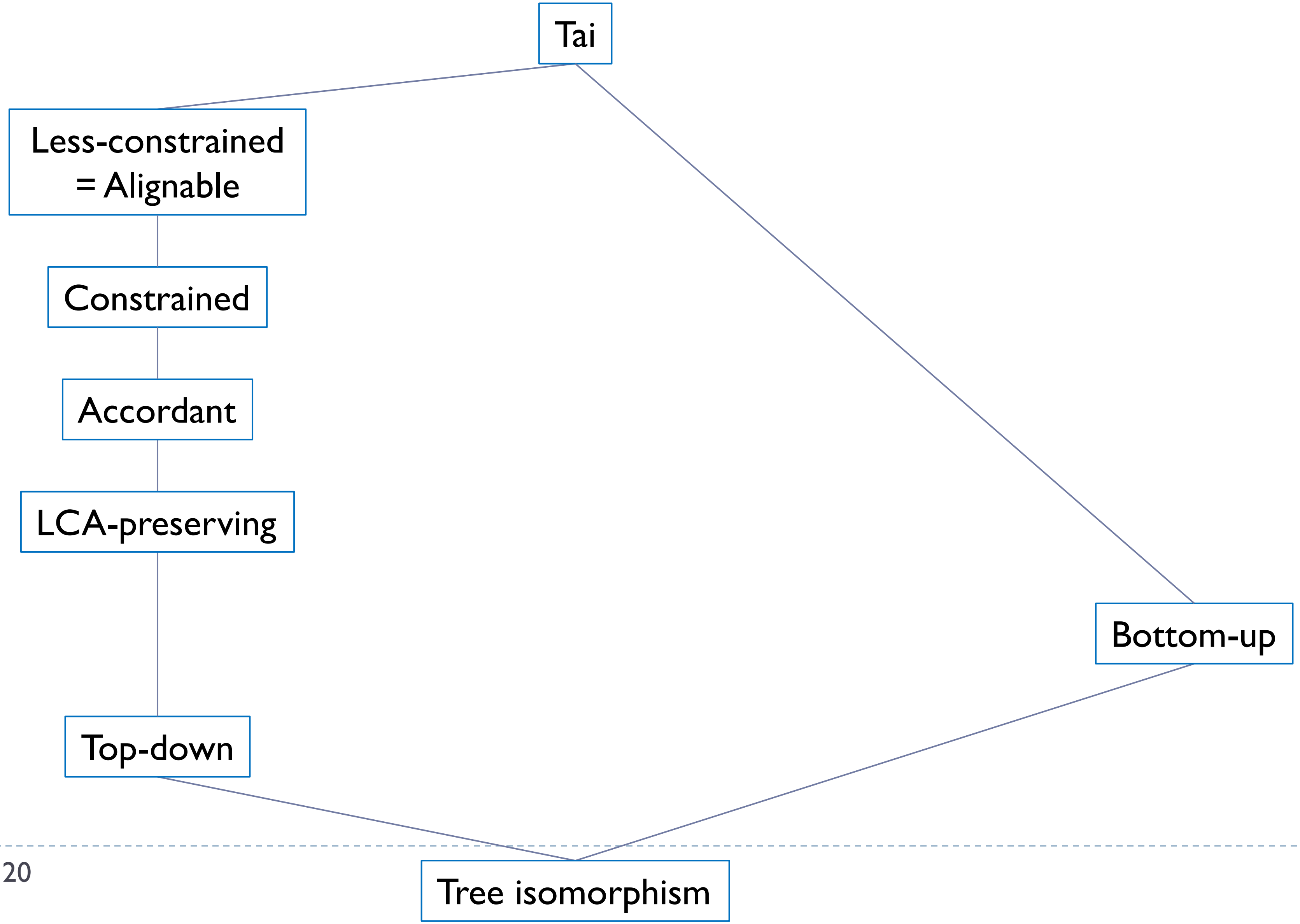


Bottom-up segmental
Top-down segmental
Not bottom-up

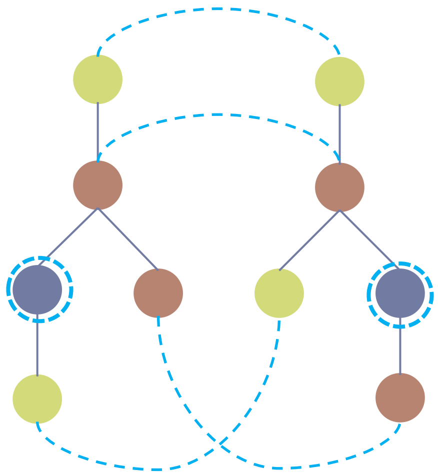
Bottom-up segmental
Not top-down segmental
Not bottom-up

Segmental
Not bottom-up segmental
Not top-down segmental

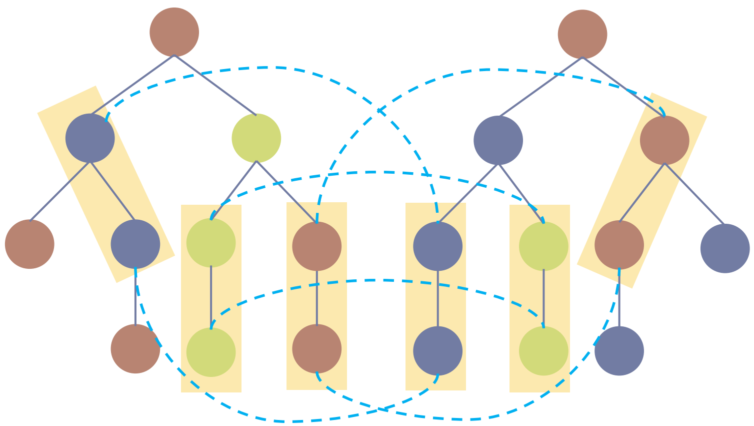
Hierarchy of Tai mapping [cf. Kuboyama 07]



Hierarchy of Tai mapping

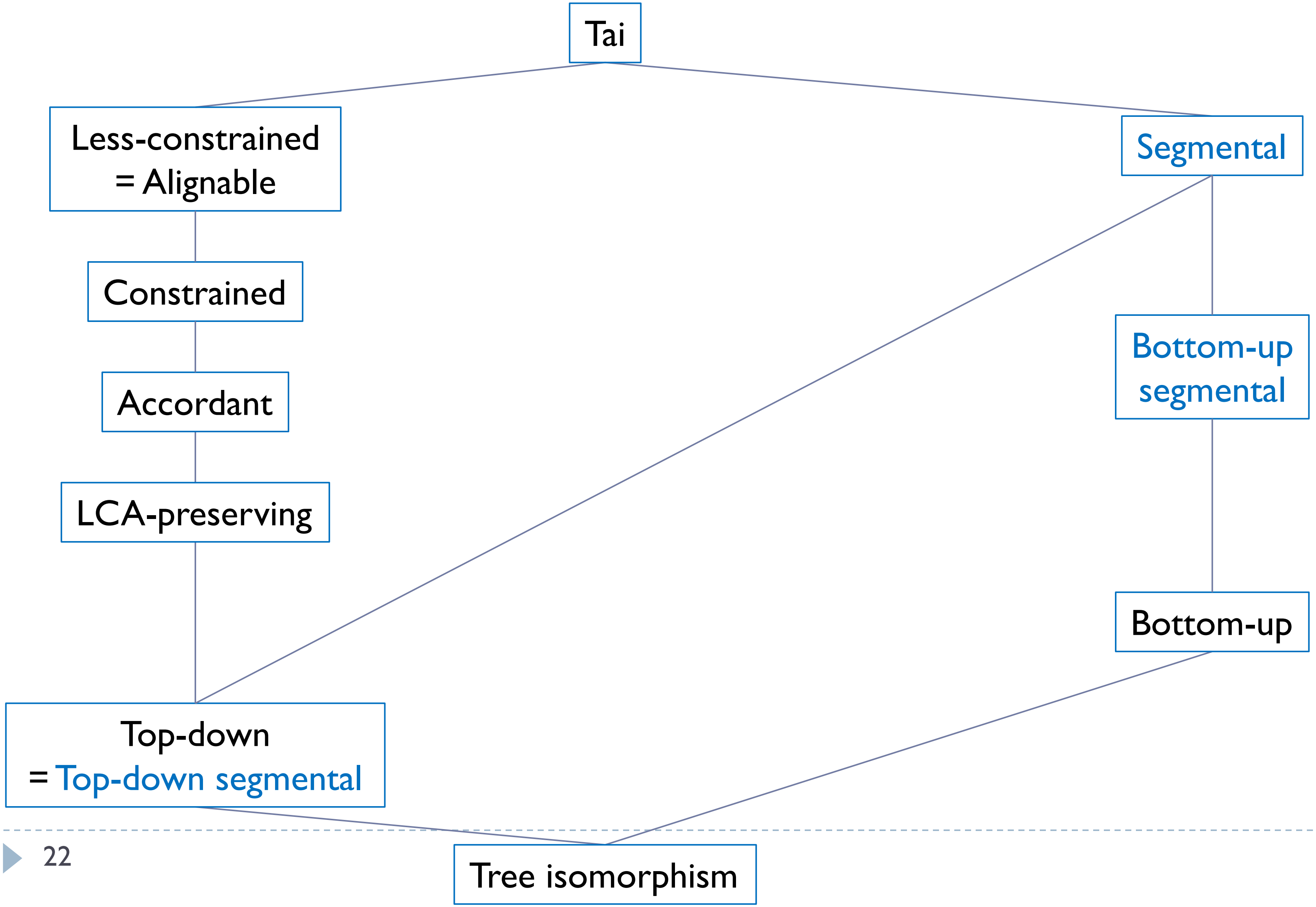


LCA-preserving
Not segmental
Not bottom-up segmental
Not bottom-up

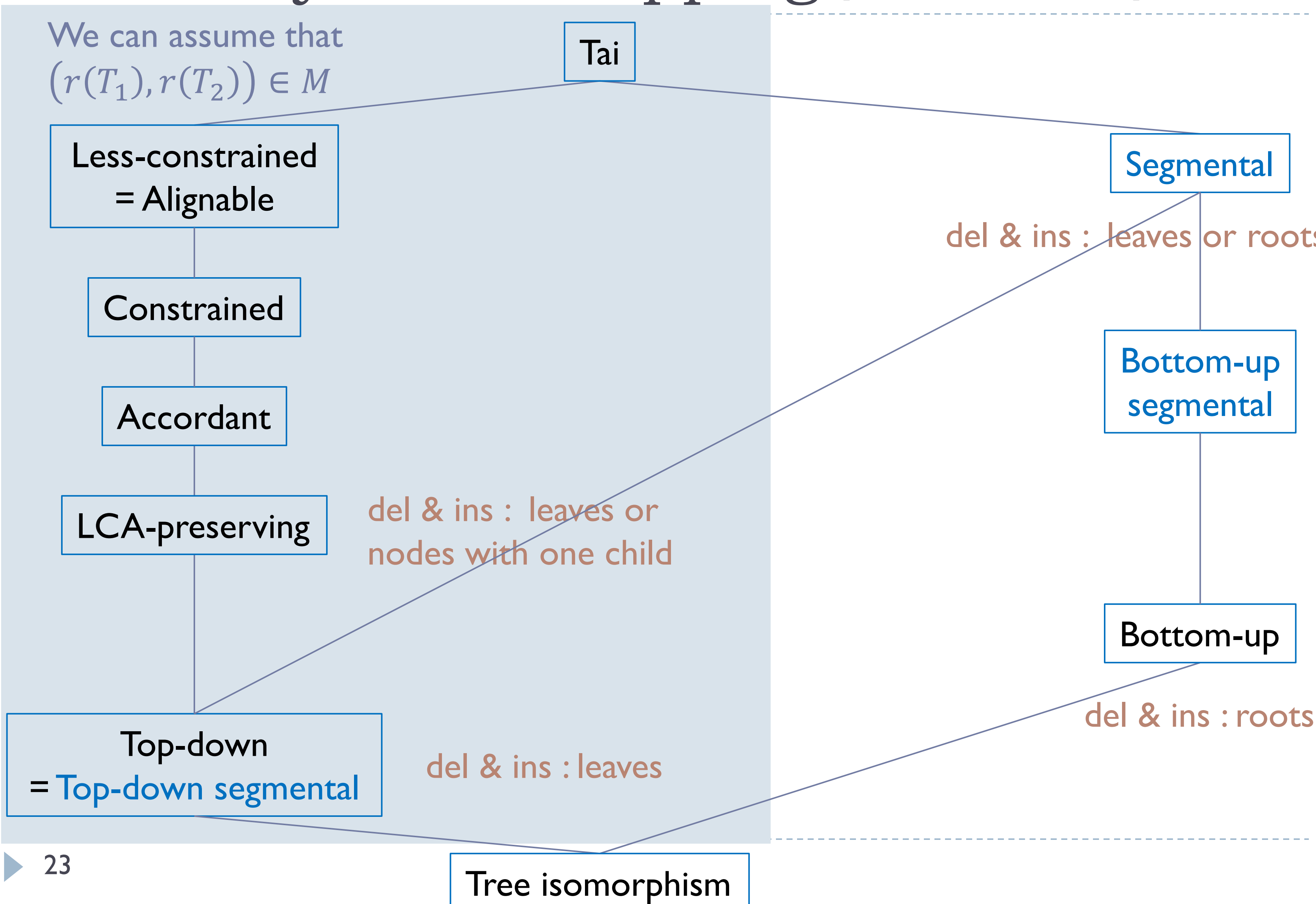


Segmental
Not LCA-preserving
Not accordant
Not constrained
Not less-constrained

Hierarchy of Tai mapping [Kan 12 ISAAC]



Hierarchy of Tai mapping [Kan 12 ISAAC]



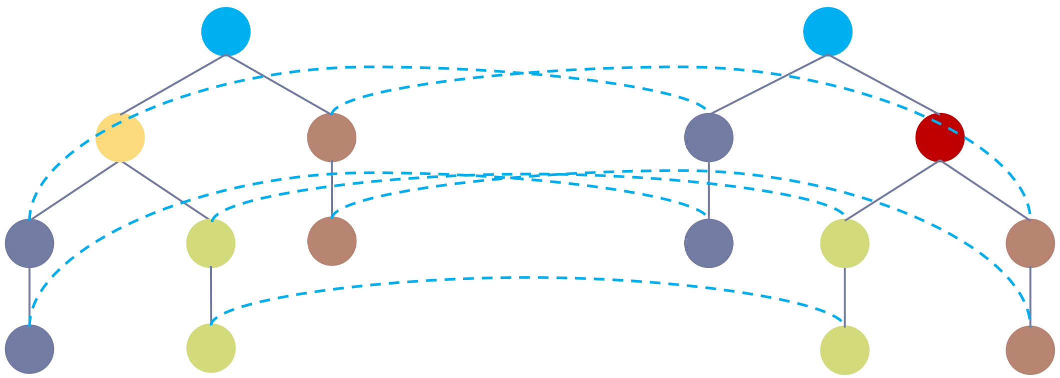
Hierarchy of Tai mapping

- Intersection to alignable mapping

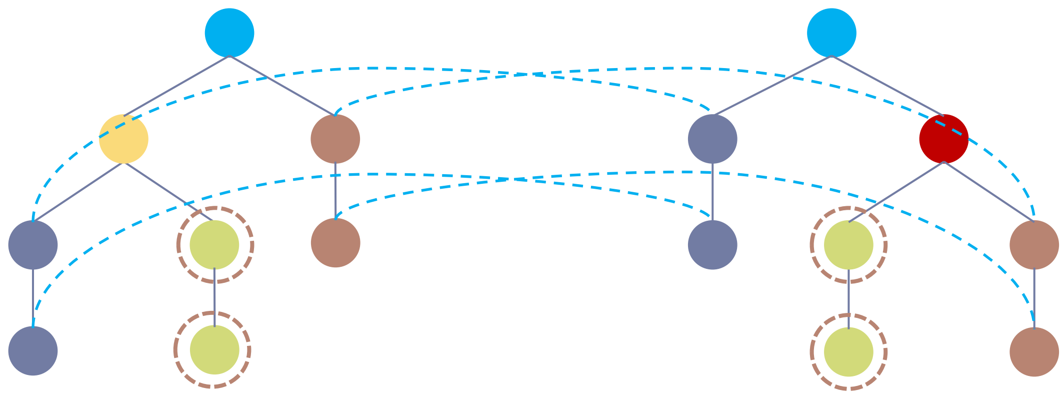
- Segmental alignable mapping
- Bottom-up segmental alignable mapping
- Bottom-up alignable mapping

Hierarchy of Tai mapping

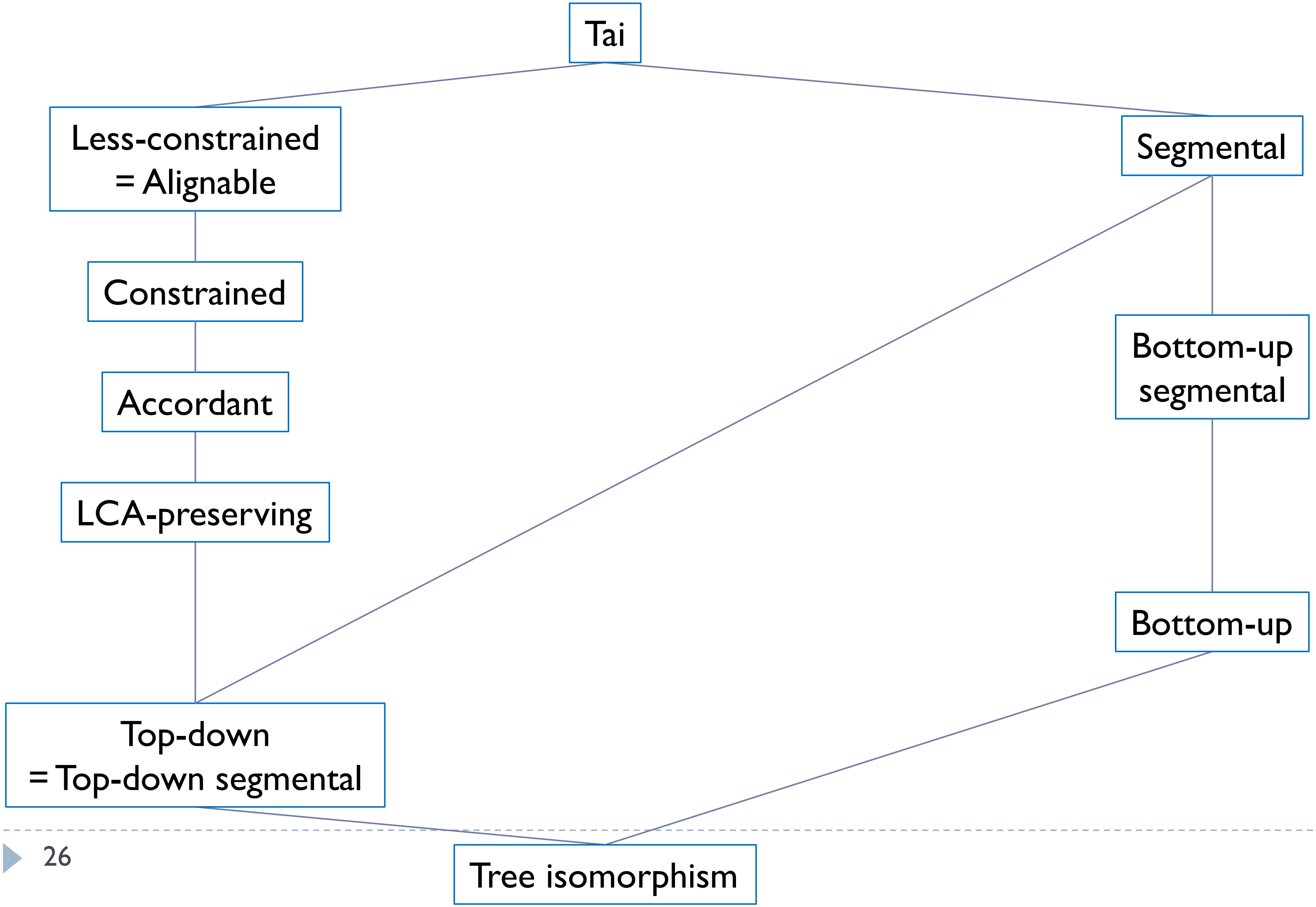
bottom-up mapping



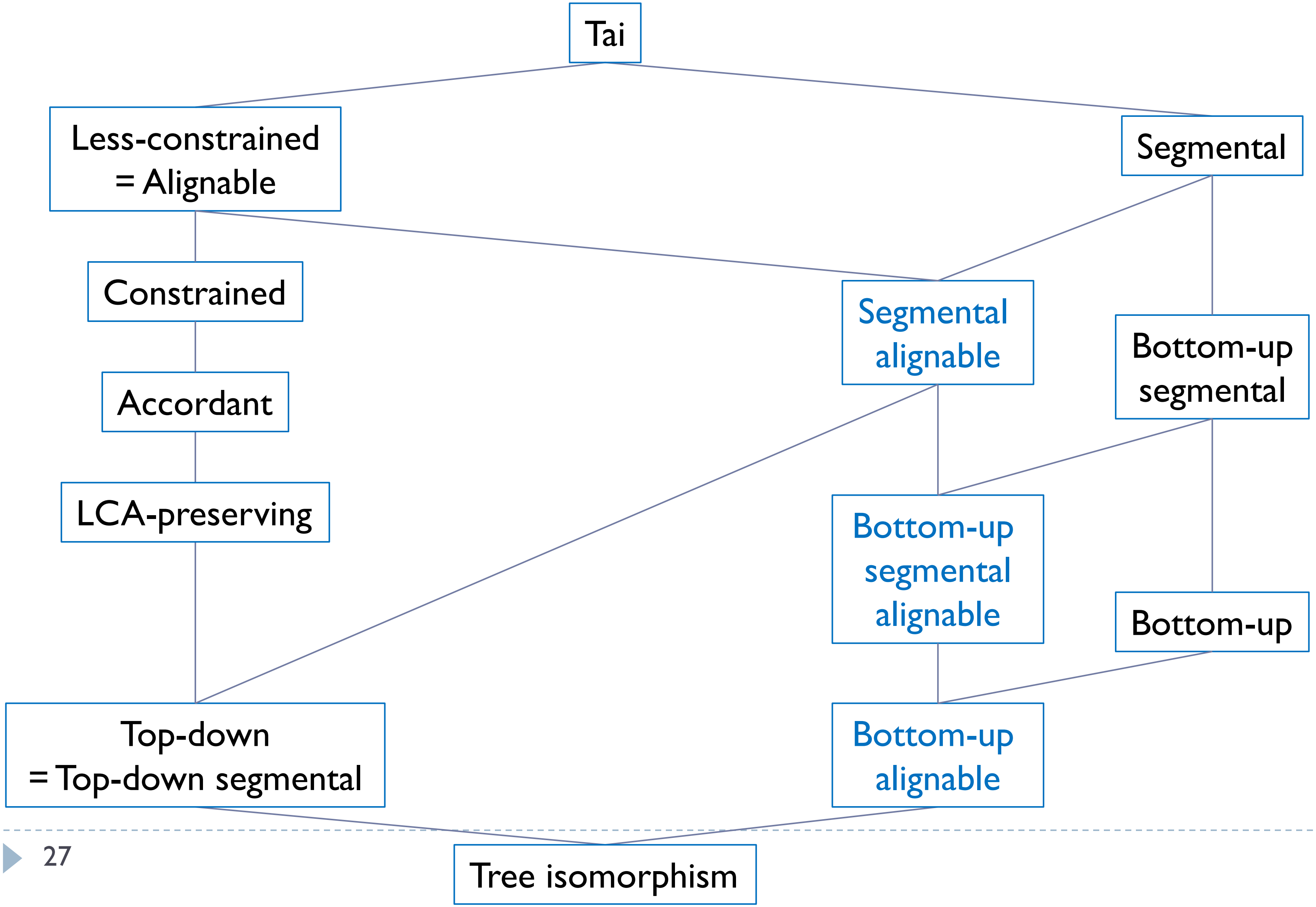
bottom-up alignable mapping



Hierarchy of Tai mapping [Kan 12 ISAAC]



Hierarchy of Tai mapping [Hirata 13 SIGFPAI88]

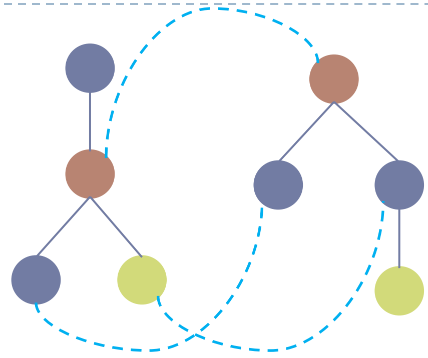


Hierarchy of Tai mapping

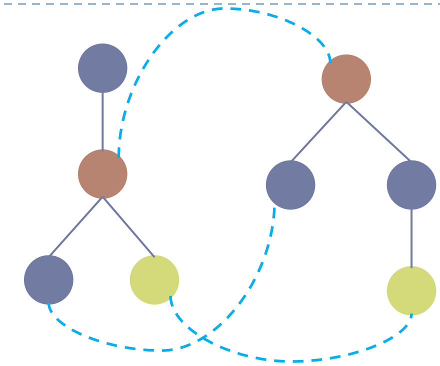
- Intersection to segmental mapping

- Constrained segmental mapping
- Accordant segmental mapping
- LCA-preserving segmental mapping

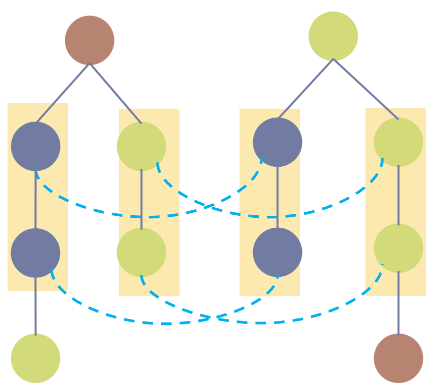
Hierarchy of Tai mapping



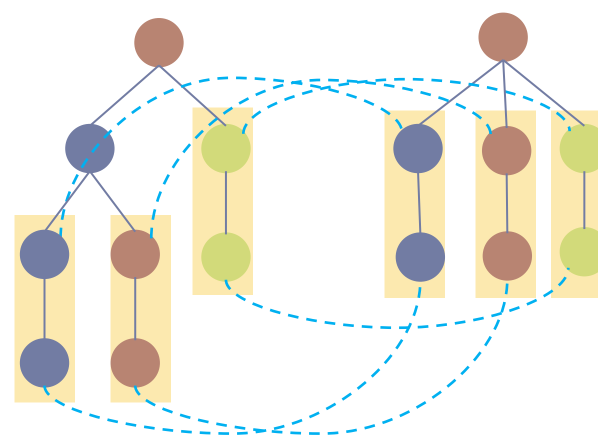
LCA-preserving
LCA-preserving segmental
Not top-down



LCA-preserving
Not LCA-preserving segmental
Not top-down

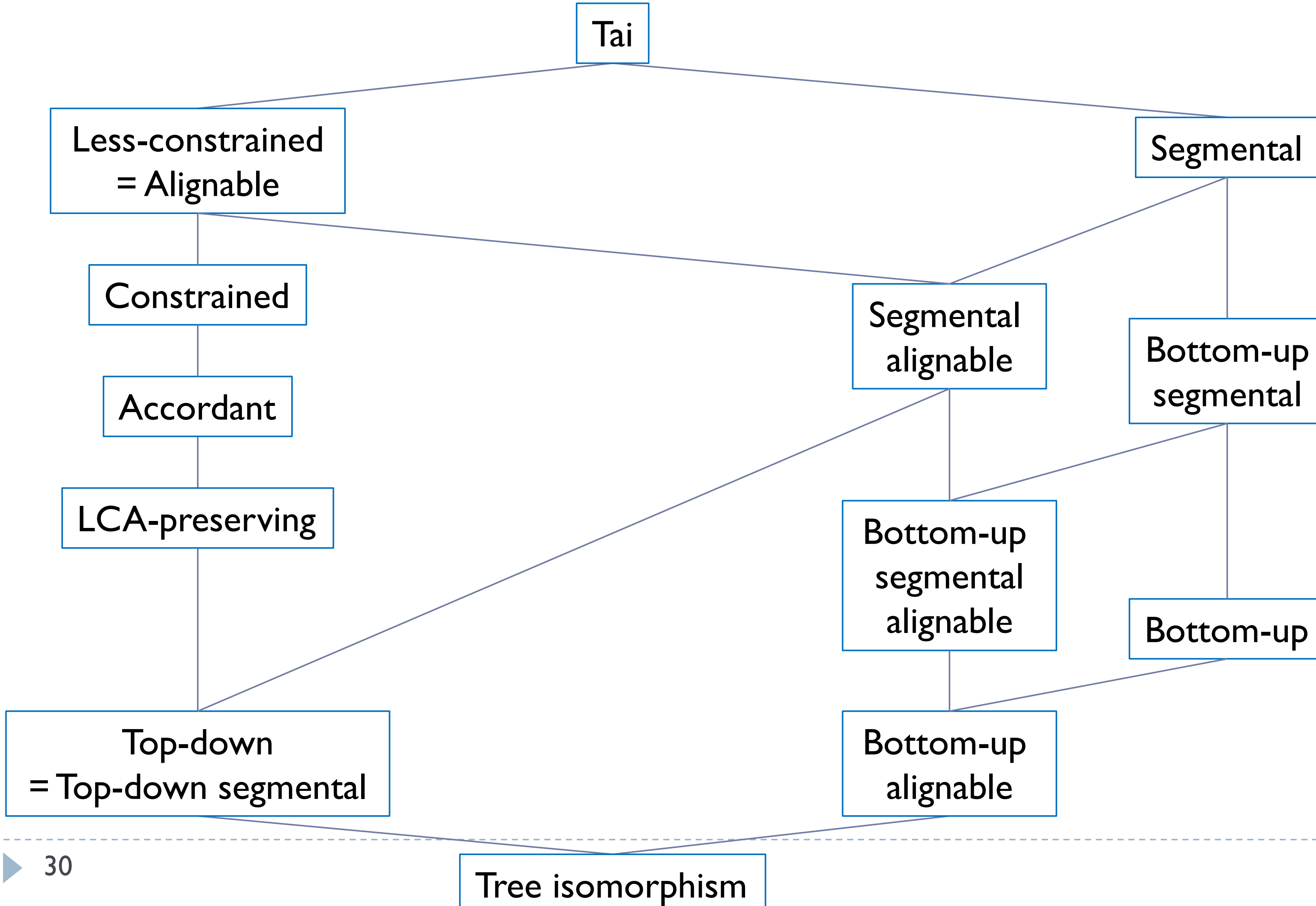


Constrained segmental
Accordant segmental
Not LCA-preserving segmental

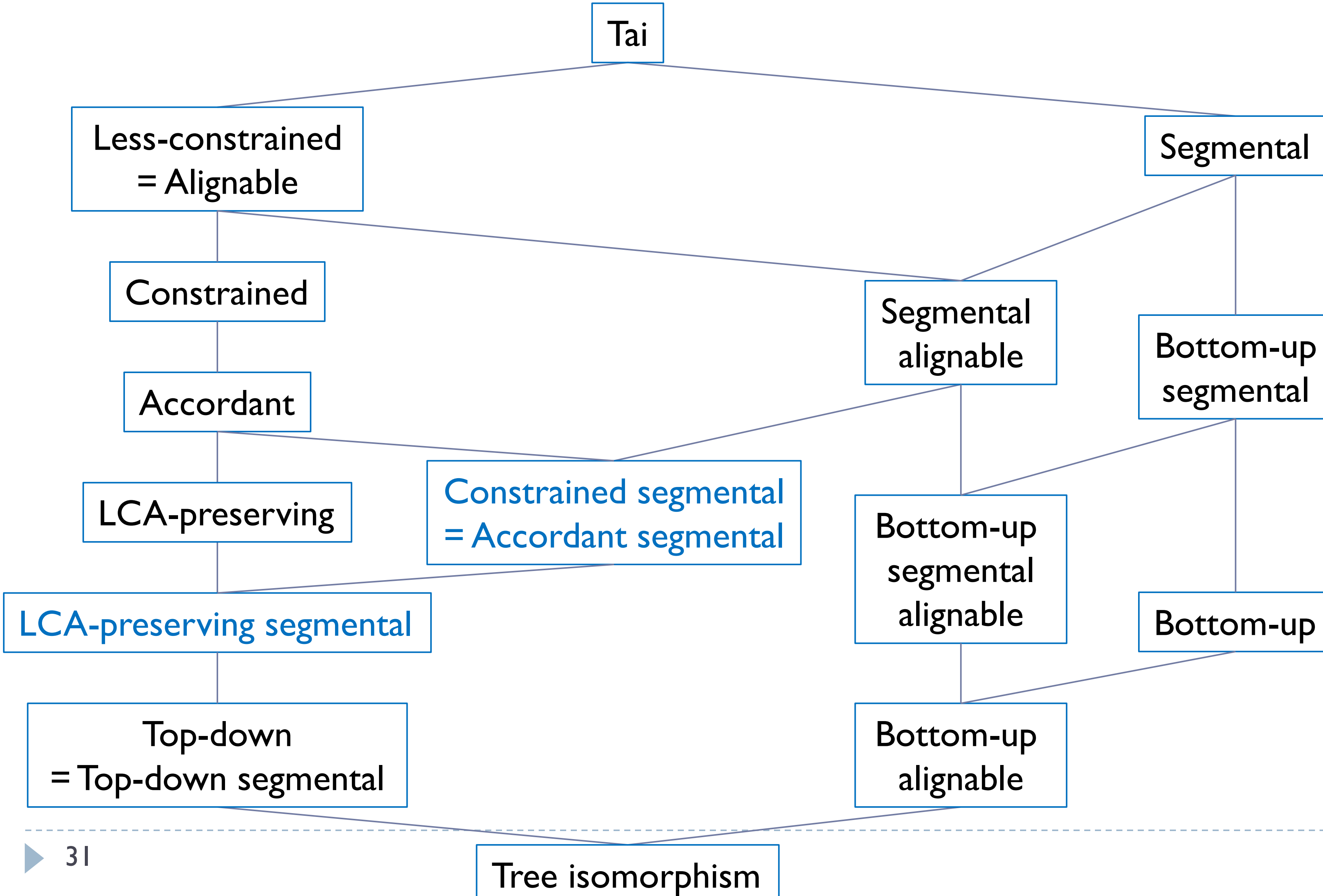


Segmental alignable
Not constrained segmental
Not accordant segmental

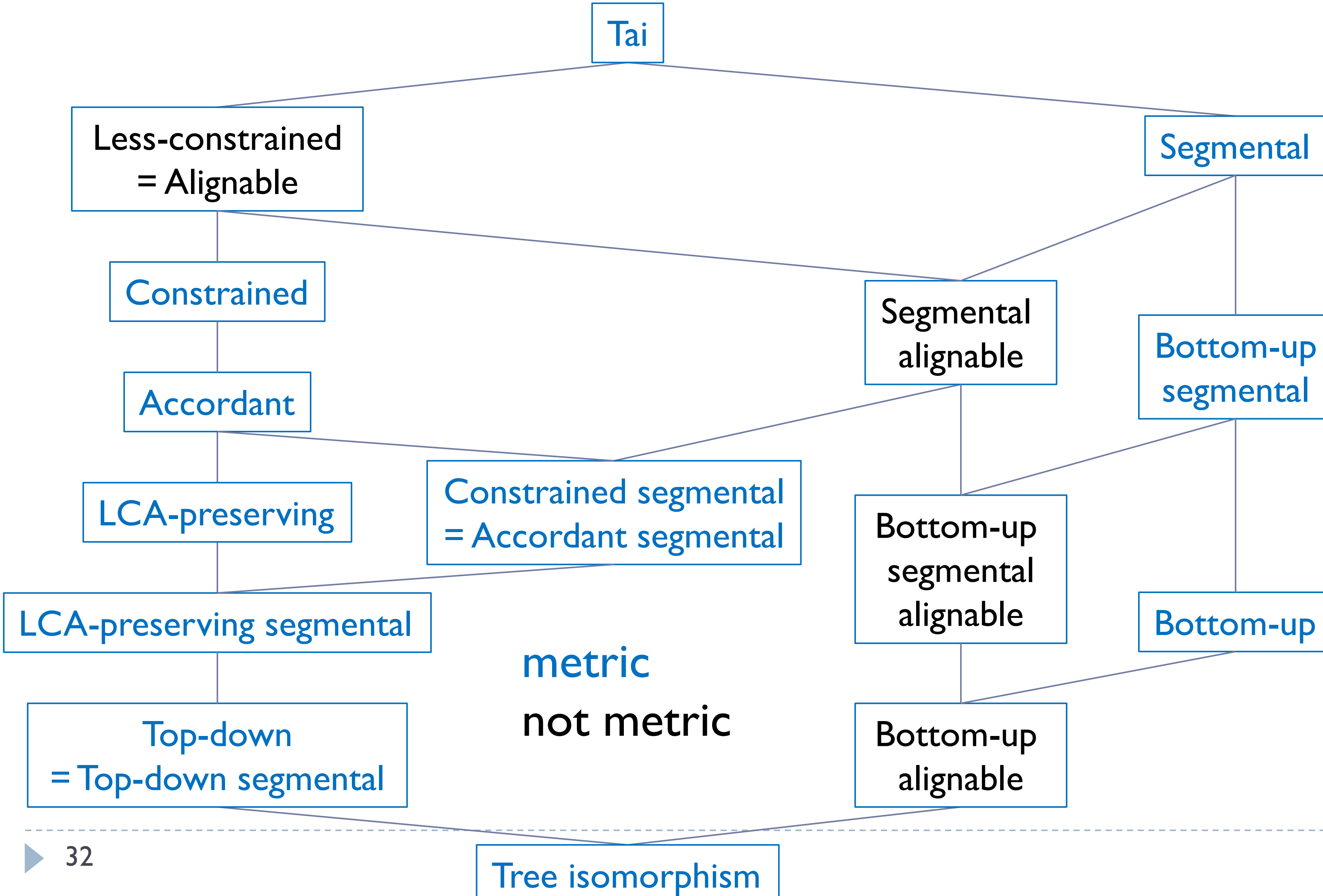
Hierarchy of Tai mapping [Hirata 13 SIGFPAI88]



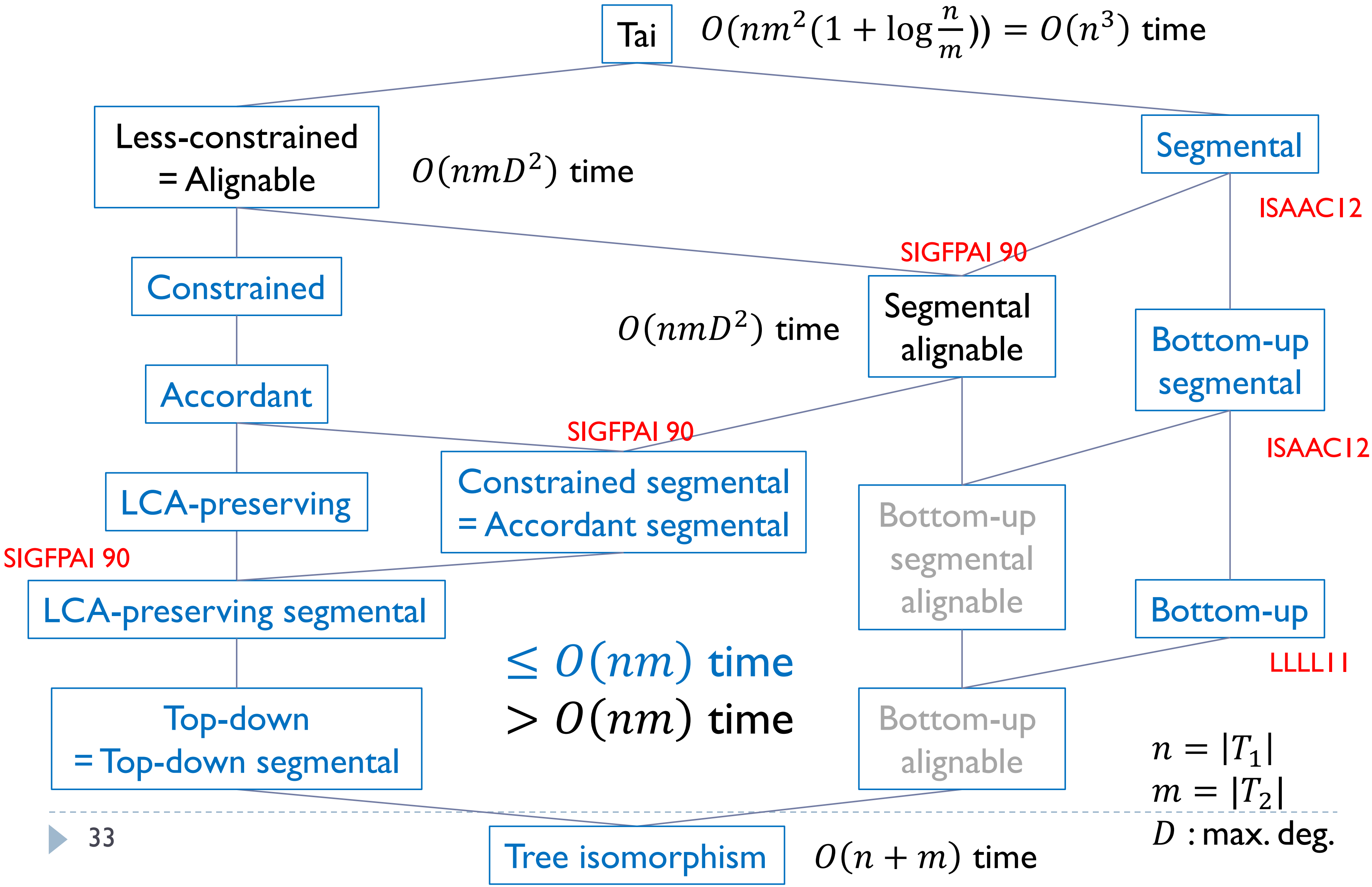
Hierarchy of Tai mapping [Hirata 13 SIGFPAI90]



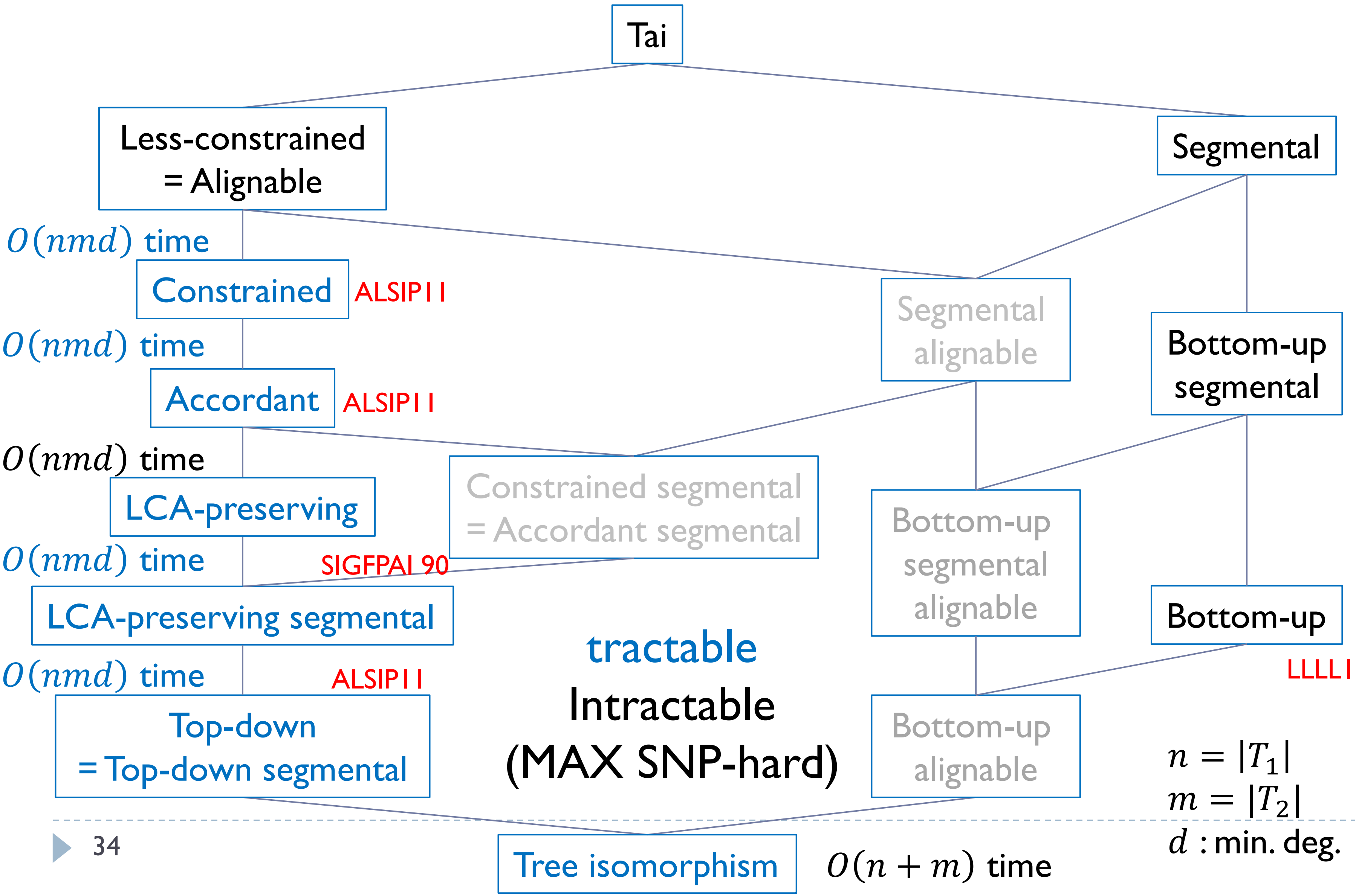
Hierarchy of Tai mapping : Metricity



Hierarchy of Tai mapping : Time complexity for ordered trees



Hierarchy of Tai mapping : Time complexity for unordered trees



Edit distance, indel distance: ordered trees

Mapping	Edit distance	Indel distance
Tai	$O(nm^2(1+\log\frac{n}{m}))$	$O(\Delta H^2 \log\log m)$, $O(L\Delta \log \Delta \log\log m)$
less-constrained (alignable)	$O(nmD^2)$	[Mozes 09 (TCS 410)]
constrained (isolated-subtree)	$O(nm)$	
accordant (Lu)	$O(nm)$	[Mozes 09 (TCS 410)]
LCA-preserving (degree-2)	$O(nm)$	$O(\Delta H \log\log m)$
LCA-preserving segmental	$O(nm)$ [SIGPPAI]	
top-down (degree-1)	$O(nm)$	$O(n+m)$
segmental	$O(nm)$ [ISAAC 12]	
bottom-up segmental	$O(nm)$ [ISAAC 12]	[Valiente 01 (SPIRE)]
bottom-up	$O(nm)$ [LLLL 11]	$O(n+m)$
segmental alignable	$O(nmD^2)$ [SIGPPAI]	
accordant segmental	$O(nm)$ [SIGPPAI]	

Δ : sum of pairs of nodes in T_1 and T_2 whose labels are different

Edit distance, indel distance, inclusion: unordered trees

Mapping	Edit distance	Indel distance	Inclusion
Tai	MAX SNP-hard	MAX SNP-hard	NP-complete
less-constrained (alignable)	MAX SNP-hard	MAX SNP-hard	
constrained (isolated-subtree)	$O(nmd)$		$O(nm^{1.5})$
accordant (Lu) [ALSIP 12]	$O(nmd)$	[Kao 01 (J. Algo. 40)]	(same as LCA)
LCA-preserving (degree-2)	$O(nmd)$	$O(\sqrt{D}\Delta \log \frac{2n}{D})$	$O(nm^{1.5})$ [Valiente 05 (J. Disc. Algo. 3)]
LCA-preserving segmental	$O(nmd)$ [SIGPPAI]		$O(nm^{1.5}/\log m)$
top-down (degree-1)	$O(nmd)$	$O(nm^{1.5}/\log m)$	$O(nm^{1.5}/\log m)$
segmental	MAX SNP-hard	[Shamir & Tsur 99 (J. Algo. 33)]	$O(nm^{1.5}/\log m)$
bottom-up segmental	MAX SNP-hard		[Shamir & Tsur 99 (J. Algo. 33)]
bottom-up [LLLL 11]	MAX SNP-hard	$O(n+m)$	$O(n+m)$
accordant segmental		[Valiente 01 (SPIRE)]	$O(nm^{1.5}/\log m)$
segmental alignable			

Δ : sum of pairs of nodes in T_1 and T_2 whose labels are different

Future works

- Improving time complexity computing the following distances
  - Segmental alignable distance (= segmental alignment)
  - Alignable distance (= alignment)
    - Alignment : $O(nmD^2) = O(n^4)$
    - Edit distance : $O(nm^2(1 + \log \frac{n}{m})) = O(n^3)$
- Analyzing whether computing the following distance is tractable or intractable
  - Accordant segmental distance
    - Accordant distance : tractable, segmental distance : intractable
  - Segmental alignable distance
    - Segmental distance, alginable distance : intractable