

SSST-8, Doha, Qatar 2014.10.24

# Ternary Segmentation

for improving search in  
top-down induction of segmental ITGs

**Markus SAERS Dekai WU**  
{ masaers | dekai }@cs.ust.hk



**HKUST**

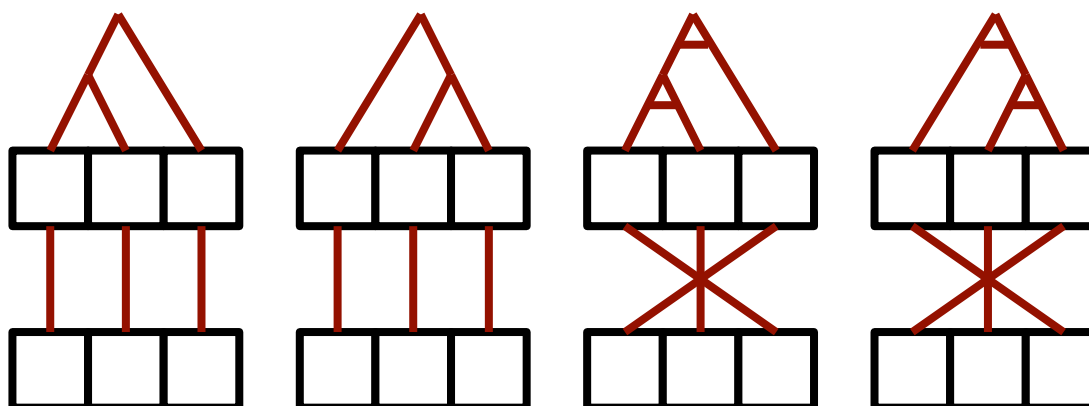
Human Language Technology Center  
Department of Computer Science and Engineering  
University of Science and Technology, Hong Kong



# What is ternary segmentation?

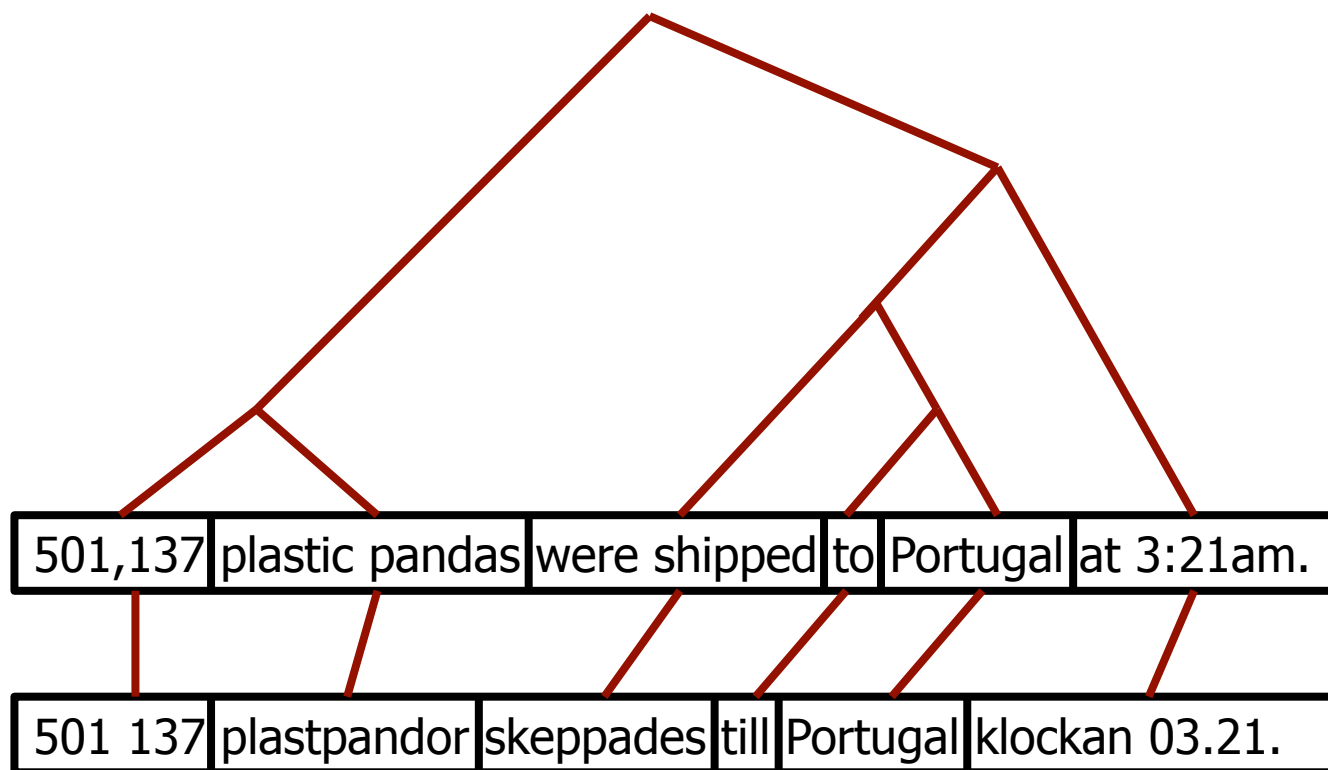
501,137 plastic pandas were shipped to Portugal at 3:21am.

501 137 plastpandor skeppades till Portugal klockan 03.21.





# Desired parse





# Desired ITG

---

A→ 501,137 | 501 137  
A→ plastic pandas | plastpandor  
A→ were shipped | skeppades  
A→ to | till |  
A→ Portugal | Portugal  
A→ at 3:21am | klockan 03.21



# Inducing with binary segmentation

501,137 plastic pandas were shipped to Portugal at 3:21am.

501 137 plastpandor skeppades till Portugal klockan 03.21.

Not a useful  
segment

Not a useful  
segment



# Inducing with binary segmentation

501,137 plastic pandas were shipped to Portugal at 3:21am.

501 137 plastpandor skeppades till Portugal klockan 03.21.

Not a useful  
segment

Not a useful  
segment



# Inducing with binary segmentation

501,137 plastic pandas were shipped to Portugal at 3:21am.

501 137 plastpandor skeppades till Portugal klockan 03.21.

Not a useful  
segment

Not a useful  
segment



# Inducing with binary segmentation

501,137 plastic pandas were shipped to Portugal at 3:21am.

501 137 plastpandor skeppades till Portugal klockan 03.21.

Not a useful  
segment

Not a useful  
segment





# Inducing with binary segmentation

501,137 plastic pandas were shipped to Portugal at 3:21am.

501 137 plastpandor skeppades till Portugal klockan 03.21.

Not a useful  
segment

Induction stops

Not a useful  
segment



# Induction with ternary segmentation

501,137 plastic pandas | were shipped | to Portugal at 3:21am.

501 137 plastpandor | skeppades | till Portugal klockan 03.21.

Not a useful  
segment

Very useful  
segment

Not a useful  
segment

A→ 501,137 plastic pandas | 501 137 plastpandor

A→ were shipped | skeppades

A→ to Portugal at 3:21am | till Portugal klockan 03.21



# Induction with ternary segmentation

501,137 plastic pandas were shipped to Portugal at 3:21am.

501 137 plastpandor skeppades till Portugal klockan 03.21.

A→ 501,137 | 501 137

A→ plastic pandas | plastpandor

A→ were shipped | skeppades

A→ to | till

A→ Portugal at 3:21am | Portugal klockan 03.21



# Induction with ternary segmentation

501,137 plastic pandas were shipped to Portugal at 3:21am.

501 137 plastpandor skeppades till Portugal klockan 03.21.

A→ 501,137 | 501 137  
A→ plastic pandas | plastpandor  
A→ were shipped | skeppades  
A→ to | till  
A→ Portugal | Portugal  
A→ at 3:21am | klockan 03.21



# How did we do?

## Desired ITG

A→ 501,137 | 501 137  
A→ plastic pandas | plastpandor  
A→ were shipped | skeppades  
A→ to | till  
A→ Portugal | Portugal  
A→ at 3:21am | klockan 03.21

## Induced ITG

A→ 501,137 | 501 137  
A→ plastic pandas | plastpandor  
A→ were shipped | skeppades  
A→ to | till  
A→ Portugal | Portugal  
A→ at 3:21am | klockan 03.21



# Induction with ternary segmentation

---

501,137 plastic pandas were shipped to Portugal at 3:21am.

501 137 plastpandor skeppades till Portugal klockan 03.21.



# Why ternary segmentation?

---

- Equivalent to two binary segmentations
- But...
  - search algorithm may decide against the first step...
  - when it leads to a temporarily worse model...
  - that would allow it to find the desired model
- Ternary segmentation allows the search algorithm...
  - to **skip** the intermediate – **bad** – model...
  - and **find** the desired – **good** – model



# Why segmental ITGs?

---

- Inversion Transduction Grammars (ITGs)
  - model bilingual composition
  - model abstractions over categories
  - can be biparsed in polynomial time
  
- Segmental ITGs
  - also model bilingual lexicalized expressions
    - has proved to be very important (see PBSMT)