

语料库语言学国际研讨会  
上海理工大学

# Chinese Causative Constructions with *shi*, *ling* and *rang*: A Cross-variety Perspective

Xiaoyu Tian<sup>12</sup> Weiwei Zhang<sup>2</sup>

<sup>1</sup>Shanghai International Studies University

<sup>2</sup>KU Leuven



1. Introduction
2. Research Questions
3. Method
4. Result
5. Conclusion

# Introduction

## Causation & Causative Constructions(cxs)

- **Causation** is a force-dynamic pattern that involves two main participants: the antagonist (the “**causer**”) and the agonist (the “**causee**”). The causer instigates a causing event or state, which affects the causee, who brings about the caused event. (Talmy 2000: Ch. 7)
- **Causative cxs** are linguistic means to express the concept of causation

# Introduction

## Three types of causative cxs

- Lexical causative cxs  
e.g. *He **killed** the girl.*  
*The girl **died**.*
- Morphological causative cxs  
e.g. Mongolian: *xajl-ax* ‘melt’(intr.) ; *xajl-**uul**-ax* ‘melt’(tr.)
- **Analytical causative cxs**  
e.g. *I **made** John go.*

# Introduction

## Chinese analytic causative construction

火灾

使/让/令

多处建筑

倒塌。

**CAUSER**  
致使者

**AUX**  
致使标记

**CAUSEE**  
被使者

**EFFECTED  
PREDICATE**  
结果谓词

# Introduction

## Cognitive Sociolinguistics

- convergence of cognitive linguistics and sociolinguistics
- examine **lectal** variation of language
  - **lect**: dialects, regiolects, or other language varieties that are specific to a certain region, genre, etc. (Kristiansen & Dirven 2008)
- use multivariate analysis as the main research method

# Research Questions

- ◆ What are the factors that constrain the choice of *shi*, *ling* and *rang*?
- ◆ Do **Mainland**, **Taiwan** and **Singapore** Chinese differ in the choice of causative auxiliaries?

# Method

## Data

- Tagged Chinese Gigaword Corpus Version 2.0 (Huang 2009)
  - News in Mainland, Taiwan and Singapore Chinese (Huang 2009)
  - Around 2 million words from each variety are selected
- 3069 observations with *shi*, *rang* and *ling*
  - Observations are automatically retrieved using Antconc.
  - Then manually checked to avoid spurious hits.
  - Randomly selected 30% observations for annotation.



# Method

## Annotation

- 27 independent variables
  - Variables explored in the literature
  - All manually annotated except for `Variety` .

CAUSER	Causer's syntactic form, semantics, person, definiteness...
CAUSEE	Causee's syntactic form, semantics, person, definiteness...
EFFECTED PREDICATE	syntactic form, semantics, prosody...

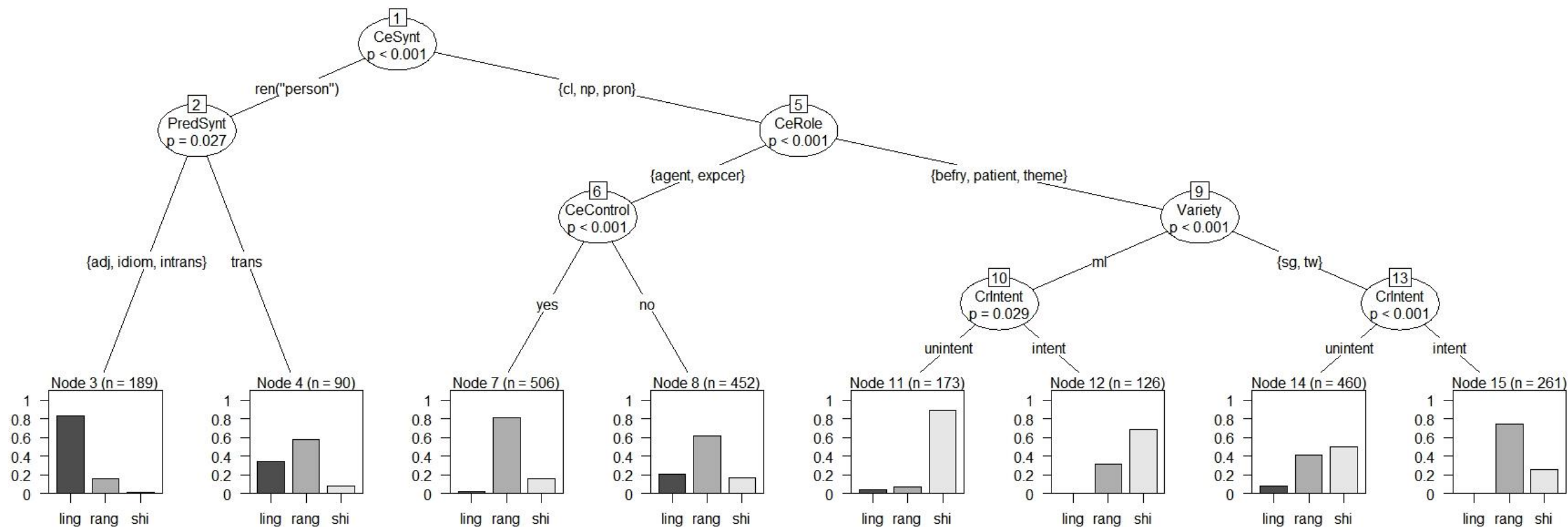
# Method

## Analysis

- Conditional inference tree
  - a **non-parametric** alternative to regression models
  - can handle **complex predictor interactions** well
  - can **visualize** the interactions in an intuitive way
- Random forest
  - an ensemble of multiple individual trees
  - different trees are grown on **randomly bootstrapped data**
  - each split is decided based on a **randomly selected subset of variables**

# Result: ctree

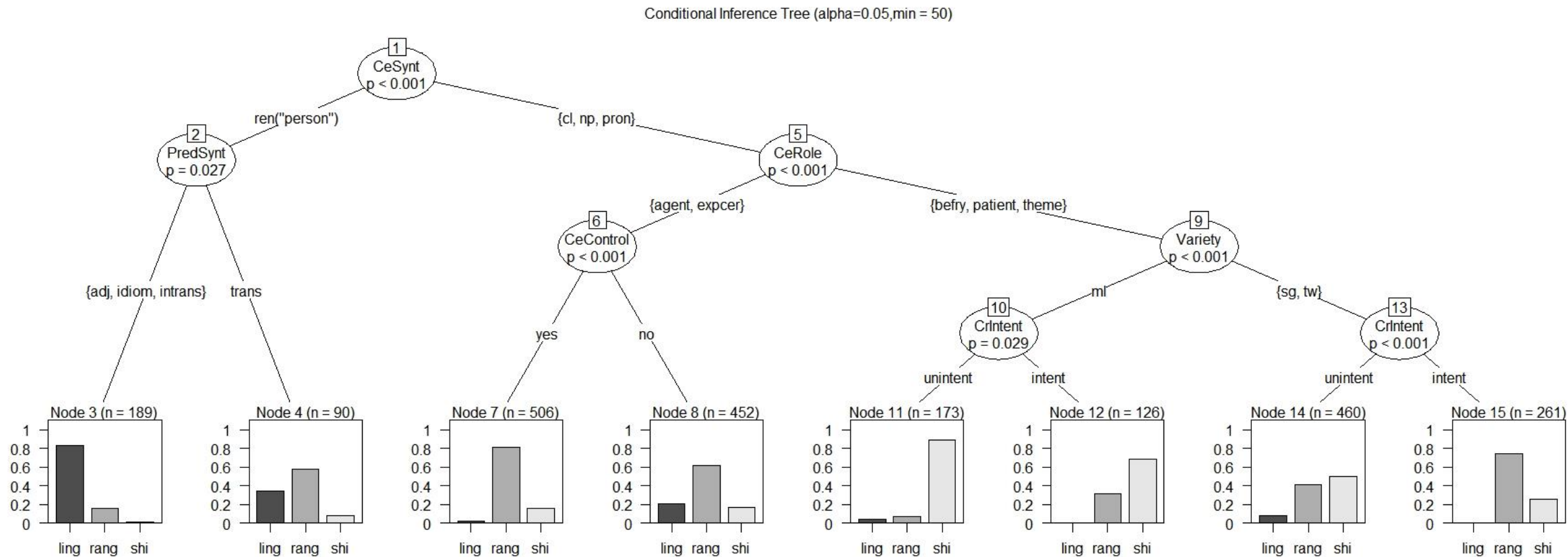
Conditional Inference Tree (alpha=0.05,min = 50)



当 causee 句法形式为“人”时,

- 结果谓词如果不是及物动词 → “令” 比例最高. (Node 3) e.g. 狂犬病问题令人忧心。
- 结果谓词如果是及物动词 → “让” 比例最高. (Node 4) e.g. 此時山風輕拂,亦讓人有暢快淋漓之感。

# Result: ctree

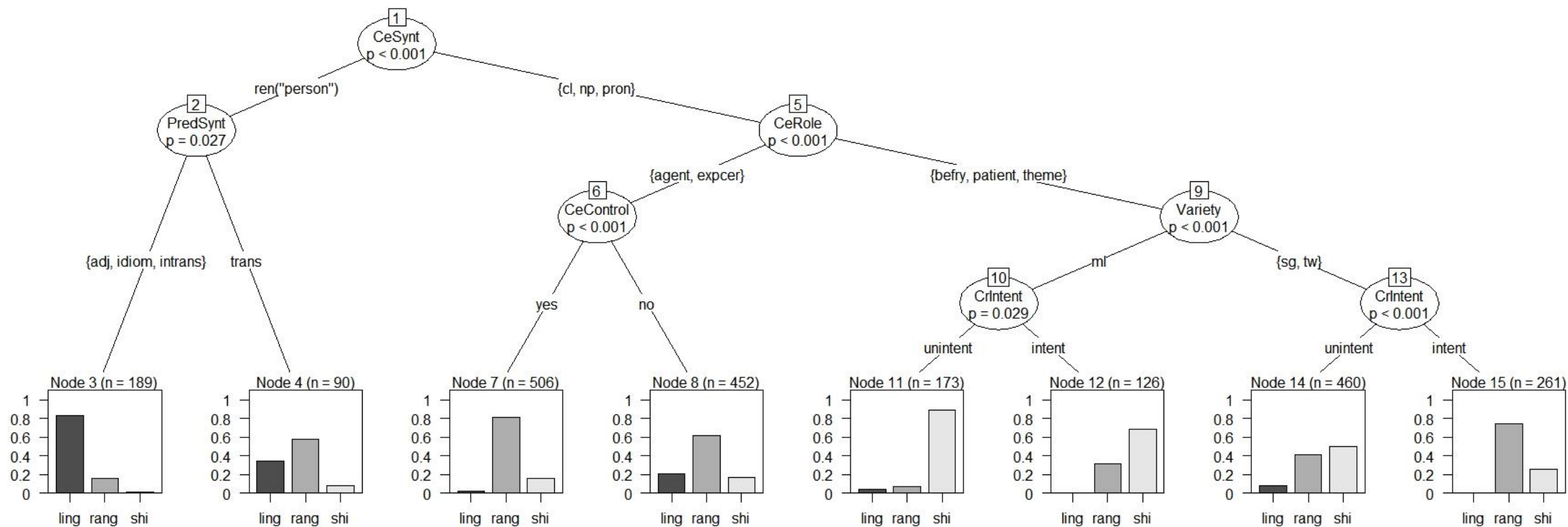


当causee是除“人”以外的句法形式，且题元角色为施事或感事时：

- Causee能够控制被使事件 → “让”的比例最高，其次是“使”，“令”的比例极低。e. g. …要让她做好打后卫的准备。
- Causee无法控制被使事件 → “让”的优势减少，“令”比例增大超过了“使”。e. g. 上海话常让外地人头疼不已。

# Result: ctree

Conditional Inference Tree (alpha=0.05,min = 50)



当causee是除“人”以外的句法形式，且题元角色为受益（害）者，受事或主体时：

- 大陆汉语 → “使”的比例最高，尤其当致使者为无意时。 e. g. 政府的“弃渔从耕”政策使他们的生活发生了很大变化。
- 台湾/新加坡 → 致使者为有意时“让”比例最高；致使者为无意时“使”略高于“让”。 e. g. 連日暴雨已使班瓦水壩注滿。

# Result

Prediction of conditional inference tree on 30% test data:

<div>Pred</div> <div>Obs</div>	<i>ling</i>	<i>rang</i>	<i>shi</i>
<i>ling</i>	64	73	9
<i>rang</i>	16	409	91
<i>shi</i>	2	109	192

- Prediction accuracy: 68.91%
- Slightly better than always picking the most frequent item (53.47%)
- Much better than random assignment (33.33%)

# Result: random forest

Pred \ Obs	<i>ling</i>	<i>rang</i>	<i>shi</i>
<i>ling</i>	77	59	10
<i>rang</i>	16	445	55
<i>shi</i>	1	126	176

- Prediction accuracy: 72.33%
- Slightly better than always picking the most frequent item (53.47%)
- Much better than random assignment (33.33%)

# Conclusion

- Significant lectal variation
  - *shi* is more favoured in Mainland Chinese; Taiwan and Singapore prefer *rang*.
- Lexical fixation effect of *ling* and *ren*
  - in accord with previous Chinese literature (e.g. Niu, 2007).
- Hyponymy relation between *rang* and *shi*, *ling*
  - predominant in general, *rang* can be seen as a hypernym of *shi* and *ling*.
- Causee's features more important than causer's
  - different from the findings of research on Dutch (e.g. Levshina 2011)



If you have any questions or comments, please contact:

[xiaoyu.tian@shisu.edu.cn](mailto:xiaoyu.tian@shisu.edu.cn)

[weiwei.zhang@kuleuven.be](mailto:weiwei.zhang@kuleuven.be)

### **Acknowledgements:**

**CHILECTO**: EU-Horizon 2020 Marie Skłodowska-Curie Action / Individual Fellowship 793920 @ KU Leuven

