CAUSE cross Chronology of Chinese:
A corpus-based analysis of Chinese mono-morphemic causatives

Yuxin He & Dick Spearsman
Research Unit QMUL
Department of Linguistics, Faculty of Arts
MQI Warren

Lexical causatives

Morphological causatives

Chinese analytic mono-morphemic causatives
CAUSE cross Chronology of Chinese: A corpus-based analysis of Chinese mono-morphemic causatives

Yanan Hu & Dirk Speelman

Research Unit QLVL
Department of Linguistics, Faculty of Arts
KU Leuven
Chinese analytic mono-morphemic causatives
Chinese Analytic Causative Construction

I asked the guests to sit around the table.

NP1 + VP1 + NP2 + VP2

Causer + CAUSE + Causee + Caused event
Research Target

7 forms of monomorphemic realization:
- 使 shǐ
- 令 lìng
- 让 ràng
- 叫 jiào1
- 教 jiào2
- 给 gěi
- 要 yào

Auxiliary verbs
- How (dis)similar are they?
- What distinguishes them?
- Who, when and where prefer which of them?
Forefathers' Treasure

Theoretical Background:


Methodological Tools:

Statistics for Corpus Linguistics (Speelman 1997), distinctive collexeme analysis (Gries & Stefanowitsch 2004), multinominal logistic regression analysis, motion chart (Hilpert 2011), etc.
What to Excavate?

Research Questions:

- Is there change?
- Which is relatively stable?
- How do the so-called Chinese "doen" ("shi") and "laten" ("rang") develop, especially along the continuum of (in)direct causation?

Geeraerts 2010

<table>
<thead>
<tr>
<th>Semasiology</th>
<th>Synchrony</th>
<th>Diachrony</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polysemous senses/ usages of one causal auxiliary</td>
<td>↑ polysemy study</td>
<td>meaning change over time (↑ semantic change</td>
</tr>
<tr>
<td>Onomasiology</td>
<td>nitrogen between near-synonyms</td>
<td>↑ synonymy study, variation study</td>
</tr>
</tbody>
</table>
Operation

Materials:
- Sheffield Corpus of Chinese
- The UCLA Chinese Corpus (1st ed)

Predictors: 18

Procedures:
- motion chart (exploratory)
- multinominal logistic regression analysis (confirmatory)
CrSem & CeSem

semantic class of Causer & Causee
- Anim = animate like human, organization, animal, body part
- Act = physical and mental activities
- Evt = event
- Inanim = material, abstract entity
Coref

coreferentiality of Causer & Causee

- Y = yes
- N = no
- Undef
CsedCstr

grammatical construction of the effected predicate (V2)

- Trans = transitive/ditransitive verbs
- Intrans = intransitive verbs
- SVC = serial verb construction
- Copula = copula, adjective, past participle
- Idiom = idiom, noun phrase
SVC:
朵颐命令似地叫他过去陪她聊天。
Duoyi imperatively CAUSE him go there accompany her chat
Duoyi ordered him over to chat with her.

Copula:
快乐会让人晕眩。
Kuài lè hui ràng rén yūn xuàn
Happiness will CAUSE people dizzy
Happiness makes one dizzy.

Idiom:
她害怕的是这场天降的爱情，
Tā hái pà de shì zhè chǎng tiān jiàng de ěr qíng,
She fear (genitive) is this (quantifier) heaven fall (adjectival marker) love,
Her fear is that this heavenly love will make her obsessed.
CsedSem

semantic classification of the caused event predicate

- Action
- State
- Change (Process) in Teng (2009)
- Motion
- Percept = perception
- Emotion = feelings derived from D'Andrade (1987)
Time -> Period

time strata of the corpus

<table>
<thead>
<tr>
<th>Time</th>
<th>Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>1100BC-206BC</td>
<td>1</td>
</tr>
<tr>
<td>206BC-220AD</td>
<td>2</td>
</tr>
<tr>
<td>220AD-581AD</td>
<td>3</td>
</tr>
<tr>
<td>581AD-979AD</td>
<td>4</td>
</tr>
<tr>
<td>860AD-1368AD</td>
<td>5</td>
</tr>
<tr>
<td>1368AD-1644AD</td>
<td>6</td>
</tr>
<tr>
<td>1644AD-1911AD</td>
<td>7</td>
</tr>
<tr>
<td>2000AD-2005AD</td>
<td>8</td>
</tr>
</tbody>
</table>
CrChi/CeChi/CsedChi.colloc.sig

significant collocations of CAUSE and Causer, Causee & V2

- TRUE
- FALSE (see Speelman & Geeraerts 2009)

distinctive collexeme analysis
(Gries & Stefanowitsch 2004)
Operation

Materials:
- Sheffield Corpus of Chinese
- The UCLA Chinese Corpus (1st ed)

Predictors: 18

Procedures:
- motion chart (exploratory)
- multinomial logistic regression analysis (confirmatory)

2,531 observations
1100BC-2005AD
Motion Chart

Time is...
Change

Onomasiological Perspective

Semasiological Perspective
Onomasiological Perspective

RQ 1
Is there expression change?
- auxiliary position; e.g., "is" in "I am" vs. "he"
- which can serve to modify, e.g., "it"
- which can serve to be changed, e.g., "be" next

RQ 2
Which concept is relatively weak?
- Expressive concepts, which serve to make language more rich or clear
  - e.g., "beautiful" in "beautiful flowers"

RQ 3
Which concept is relatively strong?
- Procedural concepts, which serve to make language more rich or clear
  - e.g., "easy" in "easy to do"
Is there expression change?

- auxiliary preference, e.g. "shi" vs "ling"
- which ones come into being, e.g. "gei"
- which ones are boosted, e.g. "rang"
- which ones are given up, e.g. "jiao2"
Semasiological Perspective

RQ 1
Is there meaning change?
- Effect of time configuration
- Semantic change over time
- Semantic similarity analysis

RQ 2
Which sequences are relatively stable?
- Same change over time
- "Yes" and "No"
- Semantic similarity analysis

RQ 3
Is there meaning change?

different plot configuration

diachronic change of their usage
  e.g. CrSem_Anim, CeSem_Anim
Onomasiological Perspective

RQ 1
- Is there expression change?
  - auxiliary permissive: e.g., "may" vs "can"
  - which means the enabling, e.g., "may"

RQ 2
- Which meaning is relatively static?
  - Verbs: means "to do" vs "are doing"
  - "be" vs "are"

RQ 3
- Histories of "be" & "are"
  - Stable relationship: "to do" and "are doing"
  - "be" vs "are" dialectal terms: "be" as "am" or "are"
Which concept is relatively stable?

To express this concept, which causative do language users tend to choose all the time?

e.g. CrSem_Evt*CeSem_Act
"shi"
Semasiological Perspective

RQ 1
Is there meaning change?
- Cultural change of their usage
- Change in the environment

RQ 2
Which causes is relatively stable?
- Word order
- “And”/“So”
- “Should” usually has definite

RQ 3
Which causative is relatively stable?

less change over time

"ling" & "shi"
"shi" usually take detours
Onomasiological Perspective

RQ 1

Is there expression change?
- auxiliary present, e.g., "is" vs. "are"
- which means is lacking, e.g., "is" vs.
- which means are presenting, e.g., "are"

RQ 2

Which concept is relatively weak?

- To approximate concepts, which means are significant to overall the concept
  - e.g., "high" vs. "lower"

RQ 3

- Which concept is relatively weak?

- Which concept is relatively weak?
Histories of "shi" & "rang"

**Similarity:**
"shi" = archaic form = "doen"
"rang" = modern default unmarked form = "laten"

**Dissimilarity:**
Colloc. sig gradually favors "rang" over "shi".
"doen" over "laten"

? reasoning on historical relic
(Speelman & Geeraerts 2009)
Semasiological Perspective

RQ 1
Is there meaning change?
- Effective plus configuration
- Ambiguity change of their usage
- Other factors (time, context)

RQ 2
Which connotes is relatively stable?
- Even change over time
- “As” vs. “is”
-fify in the sentence

RQ 3
Moving along (in)direct causation continuum

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Prediction</th>
<th>Developing trace</th>
<th>Endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>CrSem</td>
<td>Inanimate causer = direct causation</td>
<td>“shi” anim → inanim</td>
<td>“shi”, more direct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“rang” too, but lags behind</td>
<td></td>
</tr>
<tr>
<td>Coref</td>
<td>Coreference = direct causation</td>
<td>both non-coref → more coref</td>
<td>“shi”, more direct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“shi” coref &gt; “rang”</td>
<td></td>
</tr>
<tr>
<td>CsedCitr</td>
<td>Intransitive (compared to transitive) = direct causation</td>
<td>“shi”, smaller percentage to take both trans and intrans</td>
<td>“shi”, more direct;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“rang” intransitive → transitive</td>
<td>“rang”, indirect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a bit more transitive than “shi” in modern times</td>
<td></td>
</tr>
<tr>
<td>Copula (compared to SVC) = direct causation</td>
<td>“shi”, more copulas, fewer SVC</td>
<td>“rang”, more copulas, more SVC</td>
<td>“rang”, a bit more direct (towards different direction)</td>
</tr>
<tr>
<td>CsedSem</td>
<td>Perception verb (compared to emotion verb) = direct causation</td>
<td>“shi” percept → emotion (2 times)</td>
<td>“rang”, only percept → emotion (not up to 2 times)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“rang” only percept → emotion (not up to 2 times)</td>
<td></td>
</tr>
</tbody>
</table>

→ Do these really matter?
## Moving along (in)direct causation continuum

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Prediction</th>
<th>Developing trace</th>
<th>Endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>CrSem</td>
<td>Inanimate causer = direct causation</td>
<td>“shi” anim $\rightarrow$ inanim</td>
<td>“shi”, more direct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“rang” too, but lags behind</td>
<td></td>
</tr>
<tr>
<td>Coref</td>
<td>Coreference = direct causation</td>
<td>both non-coref $\rightarrow$ more coref</td>
<td>“shi”, more direct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“shi” coref $\geq$ “rang”</td>
<td></td>
</tr>
<tr>
<td>CsedCstr</td>
<td>Intransitive (compared to transitive) = direct causation</td>
<td>“shi”, smaller percentage to take both trans and intrans</td>
<td>“shi”, more direct;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“rang” intransitive $\rightarrow$ transitive</td>
<td>“rang”, indirect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a bit more transitive than “shi” in modern times</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Copula (compared to SVC) = direct causation</td>
<td>“shi”, more copulas, fewer SVC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>“rang”, more copulas, more SVC</td>
<td></td>
</tr>
<tr>
<td>CsedSem</td>
<td>Perception verb (compared to emotion verb) = direct causation</td>
<td>“shi” percept $\rightarrow$ emotion (2 times)</td>
<td>“rang”, a bit more direct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“rang” only percept $\rightarrow$ emotion (not up to 2 times)</td>
<td>(towards different direction)</td>
</tr>
</tbody>
</table>

→ Do these really matter?
Multinomial Logistic Regression Analysis

Multinomial Regression Model (ref="rang")

\[
\text{dS} = \text{Causatives-ref}(\text{dS} \mid \text{Causatives-ref} = \text{"rang"})
\]

\[
\text{fit} = \text{multinom}(\text{Causatives-ref} \mid \text{CrSem} + \text{CrDef} + \text{CrPers} + \text{CrSem} + \text{CrDef} + \text{CrPers} + \text{Cldef} + \text{Manner} + \text{CrdNeg} + \text{Clcolloc} + \text{Clcolloc} + \text{CrdNeg} + \text{Implicit} + \text{SyntFun} + \text{Time} + \text{CrdColloc} + \text{Clcolloc} + \text{Clcolloc}.
\]

\[
data = 1
\]

\[
\text{maxit} = 1000
\]
Multinomial Regression Model (ref="rang")

d$Causatives=relevel(d$Causatives,ref="rang")

fit1=multinom(Causatives~CrSem + CrDef + CrPers + CeSem + CeDef + CePers + Coref + Manner + CseNeg + CsedCstr + CsedSem + CsedNeg + Implicit + SyntFun + Time + CrChi.colloc.sig + CeChi.colloc.sig + CsedChi.colloc.sig, data =d, maxit=1000)
Model: Good enough?

pseudo R square

```r
> fit1.pR2
      l1h      l1hNull G2 McFadden   r2ML   r2CU
-1867.1820104 -3895.2062641  4056.0485074  0.5206462  0.7986172  0.8371700
```
**Predictors: Important?**

**Anova**

```
> Anova(fit1, type = 2)
Analysis of Deviance Table (Type II tests)

Response: CausalRes

<table>
<thead>
<tr>
<th>Predictor</th>
<th>LR Chi sq</th>
<th>DF</th>
<th>Pr(&gt;Chi sq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CausalRes</td>
<td>62.42</td>
<td>18</td>
<td>&lt;2.2e-16</td>
</tr>
<tr>
<td>Cervix</td>
<td>11.42</td>
<td>6</td>
<td>0.074546</td>
</tr>
<tr>
<td>Cephas</td>
<td>70.32</td>
<td>36</td>
<td>0.0000366</td>
</tr>
<tr>
<td>Cedef</td>
<td>39.49</td>
<td>18</td>
<td>0.0024468</td>
</tr>
<tr>
<td>Cefrat</td>
<td>10.97</td>
<td>6</td>
<td>0.1020970</td>
</tr>
<tr>
<td>CefRes</td>
<td>103.79</td>
<td>36</td>
<td>1.71e-38</td>
</tr>
<tr>
<td>Cenrat</td>
<td>20.18</td>
<td>12</td>
<td>0.0460327</td>
</tr>
<tr>
<td>Cenres</td>
<td>417.45</td>
<td>6</td>
<td>&lt;2.2e-16</td>
</tr>
<tr>
<td>Cenres</td>
<td>14.11</td>
<td>6</td>
<td>0.0281492</td>
</tr>
<tr>
<td>Cenres</td>
<td>39.48</td>
<td>12</td>
<td>0.0282212</td>
</tr>
<tr>
<td>Cenres</td>
<td>156.07</td>
<td>30</td>
<td>&lt;2.2e-16</td>
</tr>
<tr>
<td>Cenres</td>
<td>20.00</td>
<td>6</td>
<td>0.0003337</td>
</tr>
<tr>
<td>Cenres</td>
<td>22.55</td>
<td>6</td>
<td>0.009951</td>
</tr>
<tr>
<td>Cenres</td>
<td>59.10</td>
<td>12</td>
<td>0.0000472</td>
</tr>
<tr>
<td>Cenres</td>
<td>150.53</td>
<td>42</td>
<td>&lt;2.2e-16</td>
</tr>
<tr>
<td>Cenres</td>
<td>94.92</td>
<td>6</td>
<td>&lt;2.2e-16</td>
</tr>
<tr>
<td>Cenres</td>
<td>206.99</td>
<td>6</td>
<td>&lt;2.2e-16</td>
</tr>
<tr>
<td>Cenres</td>
<td>242.70</td>
<td>6</td>
<td>&lt;2.2e-16</td>
</tr>
</tbody>
</table>

Signif. codes: 0 **** 0.001 *** 0.01 ** 0.05 * 0.1 . 1
```
> Anova(fit1)
Analysis of Deviance Table (Type II tests)

Response: Causatives

<table>
<thead>
<tr>
<th></th>
<th>LR</th>
<th>Chisq</th>
<th>Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CrSem</td>
<td>62.42</td>
<td>18</td>
<td>18</td>
<td>8.258e-07  ***</td>
</tr>
<tr>
<td>CrDef</td>
<td>11.43</td>
<td>6</td>
<td>6</td>
<td>0.0759486  .</td>
</tr>
<tr>
<td>CrPers</td>
<td>70.32</td>
<td>36</td>
<td>36</td>
<td>0.0005369  ***</td>
</tr>
<tr>
<td>CeSem</td>
<td>39.49</td>
<td>18</td>
<td>18</td>
<td>0.0024445  **</td>
</tr>
<tr>
<td>CeDef</td>
<td>10.57</td>
<td>6</td>
<td>6</td>
<td>0.1025079  .</td>
</tr>
<tr>
<td>CePers</td>
<td>103.79</td>
<td>36</td>
<td>36</td>
<td>1.717e-08  ***</td>
</tr>
<tr>
<td>Coref</td>
<td>20.15</td>
<td>12</td>
<td>12</td>
<td>0.0643229  .</td>
</tr>
<tr>
<td>Manner</td>
<td>417.45</td>
<td>6</td>
<td>6</td>
<td>&lt; 2.2e-16  ***</td>
</tr>
<tr>
<td>CseNeg</td>
<td>14.11</td>
<td>6</td>
<td>6</td>
<td>0.0284492  *</td>
</tr>
<tr>
<td>CsedCstr</td>
<td>39.46</td>
<td>24</td>
<td>24</td>
<td>0.0244060  *</td>
</tr>
<tr>
<td>CsedSem</td>
<td>186.07</td>
<td>30</td>
<td>30</td>
<td>&lt; 2.2e-16  ***</td>
</tr>
<tr>
<td>CsedNeg</td>
<td>25.05</td>
<td>6</td>
<td>6</td>
<td>0.0003337  ***</td>
</tr>
<tr>
<td>Implicit</td>
<td>22.58</td>
<td>6</td>
<td>6</td>
<td>0.0009511  ***</td>
</tr>
<tr>
<td>SyntFun</td>
<td>59.10</td>
<td>12</td>
<td>12</td>
<td>3.287e-08  ***</td>
</tr>
<tr>
<td>Time</td>
<td>1503.53</td>
<td>42</td>
<td>42</td>
<td>&lt; 2.2e-16  ***</td>
</tr>
<tr>
<td>CrChi.colloc.sig</td>
<td>94.26</td>
<td>6</td>
<td>&lt; 2.2e-16  ***</td>
<td></td>
</tr>
<tr>
<td>CeChi.colloc.sig</td>
<td>206.99</td>
<td>6</td>
<td>&lt; 2.2e-16  ***</td>
<td></td>
</tr>
<tr>
<td>CsedChi.colloc.sig</td>
<td>242.73</td>
<td>6</td>
<td>&lt; 2.2e-16  ***</td>
<td></td>
</tr>
</tbody>
</table>

if. codes: 0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ' ' 1
<table>
<thead>
<tr>
<th></th>
<th>LR</th>
<th>Chisq</th>
<th>Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1503.53</td>
<td>42</td>
<td>&lt;2.2e-16</td>
<td>***</td>
</tr>
<tr>
<td>Manner</td>
<td>417.45</td>
<td>6</td>
<td>&lt;2.2e-16</td>
<td>***</td>
</tr>
<tr>
<td>CsedChi.colloc.sig</td>
<td>242.73</td>
<td>6</td>
<td>&lt;2.2e-16</td>
<td>***</td>
</tr>
<tr>
<td>CeChi.colloc.sig</td>
<td>206.99</td>
<td>6</td>
<td>&lt;2.2e-16</td>
<td>***</td>
</tr>
<tr>
<td>CsedSem</td>
<td>186.07</td>
<td>30</td>
<td>&lt;2.2e-16</td>
<td>***</td>
</tr>
<tr>
<td>CePers</td>
<td>103.79</td>
<td>36</td>
<td>1.72E-08</td>
<td>***</td>
</tr>
<tr>
<td>CrChi.colloc.sig</td>
<td>94.26</td>
<td>6</td>
<td>&lt;2.2e-16</td>
<td>***</td>
</tr>
<tr>
<td>CrPers</td>
<td>70.32</td>
<td>36</td>
<td>0.000537</td>
<td>***</td>
</tr>
<tr>
<td>CrSem</td>
<td>62.42</td>
<td>18</td>
<td>8.26E-07</td>
<td>***</td>
</tr>
<tr>
<td>SyntFun</td>
<td>59.1</td>
<td>12</td>
<td>3.29E-08</td>
<td>***</td>
</tr>
<tr>
<td>CeSem</td>
<td>39.49</td>
<td>18</td>
<td>0.002445</td>
<td>**</td>
</tr>
<tr>
<td>CsedCstr</td>
<td>39.46</td>
<td>24</td>
<td>0.024406</td>
<td>*</td>
</tr>
<tr>
<td>CsedNeg</td>
<td>25.05</td>
<td>6</td>
<td>0.000334</td>
<td>***</td>
</tr>
<tr>
<td>Implicit</td>
<td>22.58</td>
<td>6</td>
<td>0.000951</td>
<td>***</td>
</tr>
<tr>
<td>Coref</td>
<td>20.15</td>
<td>12</td>
<td>0.064323</td>
<td>.</td>
</tr>
<tr>
<td>CseNeg</td>
<td>14.11</td>
<td>6</td>
<td>0.028449</td>
<td>*</td>
</tr>
<tr>
<td>CrDef</td>
<td>11.43</td>
<td>6</td>
<td>0.075949</td>
<td>.</td>
</tr>
<tr>
<td>CeDef</td>
<td>10.57</td>
<td>6</td>
<td>0.102508</td>
<td>.</td>
</tr>
</tbody>
</table>
Levels: Significant?

rang vs shi

<table>
<thead>
<tr>
<th>Term</th>
<th>2.5 %</th>
<th>97.5 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>26.530350317</td>
<td>29.97251283</td>
</tr>
<tr>
<td>CrSemAnim</td>
<td>-0.840626523</td>
<td>0.18438545</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CorefY</td>
<td>-0.009365415</td>
<td>0.70745477</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CsedCstrIntrans</td>
<td>-0.535536219</td>
<td>0.40467330</td>
</tr>
<tr>
<td>CsedCstrSVC</td>
<td>-1.125579358</td>
<td>1.07083698</td>
</tr>
<tr>
<td>CsedCstrTrans</td>
<td>-0.672091046</td>
<td>0.11261454</td>
</tr>
<tr>
<td>CsedSemEmotion</td>
<td>-0.573517155</td>
<td>0.69330695</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(In)direct causation hypothesis

Ni (2012)

- "shi"="doen", "rang"="laten"
- factors → different type of causality
(In)direct causation hypothesis

Ni (2012)

- "shi"="doen", "rang"="laten"
- factors → different type of causality
What do we get already?

- There is CHANGE!
- It's hard to say which is stable.
  - occurrence
  - different complexions
- Witness of development
  - default form
  - (in)direct causation continuum 🎃